



CHALMERS
UNIVERSITY OF TECHNOLOGY

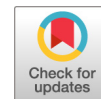
Complete genome sequences of the xylose-fermenting *Candida intermedia* strains CBS 141442 and PYCC 4715

Downloaded from: <https://research.chalmers.se>, 2024-04-27 20:15 UTC

Citation for the original published paper (version of record):

Moreno, D., Tellgren-Roth, C., Soler, L. et al (2017). Complete genome sequences of the xylose-fermenting *Candida intermedia* strains CBS 141442 and PYCC 4715. *Genome Announcements*, 5(14): e0013817-.
<http://dx.doi.org/10.1128/genomeA.00138-17>

N.B. When citing this work, cite the original published paper.



Complete Genome Sequences of the Xylose-Fermenting *Candida intermedia* Strains CBS 141442 and PYCC 4715

Antonio D. Moreno,^{a*} Christian Tellgren-Roth,^b Lucile Soler,^c Jacques Dainat,^c Lisbeth Olsson,^a Cecilia Geijer^a

Department of Biology and Biological Engineering, Industrial Biotechnology, Chalmers University of Technology, Gothenburg, Sweden^a; Department of Immunology, Genetics and Pathology, National Genomics Infrastructure, Science for Life Laboratory, Uppsala University, Uppsala, Sweden^b; Department of Medical Biochemistry and Microbiology, National Bioinformatics Infrastructure Sweden, Uppsala University, Uppsala, Sweden^c

ABSTRACT Sustainable biofuel production from lignocellulosic materials requires efficient and complete use of all abundant sugars in the biomass, including xylose. Here, we report on the *de novo* genome assemblies of two strains of the xylose-fermenting yeast *Candida intermedia*: CBS 141442 and PYCC 4715.

For commercially viable lignocellulose-based ethanol production, the microorganism of choice must be able to ferment all monosaccharides to ethanol, including xylose (1). The yeast *Candida intermedia* is known for its capacity to grow on and ferment xylose (2–4). We have sequenced the genomes of two strains of this yeast: CBS 141442, isolated from the liquid fraction of a steam-pretreated wheat straw hydrolysate in Gothenburg, Sweden, and PYCC 4715, isolated from sewage in Oeiras, Portugal, and which has been characterized previously in terms of xylose growth and transport capacity (2, 5).

DNA was extracted as described elsewhere (6), and samples were sent for single-molecule real-time (SMRT) sequencing (Uppsala Genome Center at the National Genomics Infrastructure, SciLifeLab, Uppsala, Sweden). DNA was sheared into 10-kb fragments using a GeneMachines HydroShear instrument (Digilab, Marlborough, MA, USA). SMRT bells were constructed and sequenced on three SMRT cells on a Pacific Biosciences RSII sequencer according to the manufacturer's instructions (Pacific Biosciences, Menlo Park, CA, USA) with a 4-h movie time.

For *de novo* assembly of the two genomes, reads were assembled using the SMRT Analysis HGAP3 assembly pipeline; 450 Mb of subreads longer than 8.3 kb and 4 kb were used for the preassembly step for the CBS 141442 and PYCC 4715 genomes, respectively, and 369 Mb of corrected reads with an average read length of 8.5 kb (CBS 141442) and 323 Mb of corrected reads with an average read length of 5.5 kb (PYCC 4715) were used to assemble the genomes with the Celera assembler included in SMRT Analysis. The assemblies were polished using Quiver (Pacific Biosciences). To assess the completeness of the assemblies, contig ends were analyzed for repetitive sequence motives. For CBS 141442, contigs were manually joined at unique overlaps to create complete chromosomes. For PYCC 4715, a reference-guided assembly with the CBS 141442 chromosomes as a backbone was used to create full-length chromosomes. Gap-filling was done using Quiver (Pacific Biosciences).

Annotations of the *C. intermedia* strains were computed using the Maker package version 2.31-8 (7). For construction of the gene models, *ab initio* predictions from three sources were combined: a profile model for *Candida guilliermondii* included with Augustus version 2.7 (8), a profile model for the SNAP gene predictor based on the annotation of *Clavispora lusitaniae* (9), and a self-trained GeneMark-ES version 4.3 *ab*

Received 6 February 2017 Accepted 10 February 2017 Published 6 April 2017

Citation Moreno AD, Tellgren-Roth C, Soler L, Dainat J, Olsson L, Geijer C. 2017. Complete genome sequences of the xylose-fermenting *Candida intermedia* strains CBS 141442 and PYCC 4715. *Genome Announc* 5:e00138-17. <https://doi.org/10.1128/genomeA.00138-17>.

Copyright © 2017 Moreno et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Cecilia Geijer, cecilia.geijer@chalmers.se.

* Present address: Antonio D. Moreno, Department of Energy, Biofuels Unit, CIEMAT, Madrid, Spain.

initio model specific for fungi (10). To support gene predictions, a protein data set was provided (manually curated protein sequences from UniProt), and publically available expressed sequence tag data from the genome of *Candida albicans*.

The CBS 141442 and PYCC 4715 genomes each consist of seven chromosomes, totaling 13,162,108 and 13,077,109 nucleotides, respectively. In total, 5,944 (CBS 141442) and 6,082 (PYCC 4715) protein-coding genes were found. The genome sequences and the identified genes provide insights into how *C. intermedia* utilizes xylose, which can be used to improve xylose fermentation in lignocellulosic bioethanol production.

Accession number(s). The sequence data for *C. intermedia* strains CBS 141442 and PYCC 4715 have been deposited in the European Nucleotide Archive (ENA) with the accession numbers [LT635756](#) to [LT635763](#) and [LT635764](#) to [LT635771](#), respectively.

ACKNOWLEDGMENTS

We acknowledge support from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the Swedish Research Council, the National Bioinformatics Infrastructure Sweden, and the Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing, bioinformatics analysis, and access to the UPPMAX computational infrastructure.

This work was financed by the Swedish Energy Agency (project no. 35372-1 and 38779-1).

REFERENCES

- Koppram R, Tomás-Pejó E, Xiros C, Olsson L. 2014. Lignocellulosic ethanol production at high-gravity: challenges and perspectives. *Trends Biotechnol* 32:46–53. <https://doi.org/10.1016/j.tibtech.2013.10.003>.
- Gárdonyi M, Osterberg M, Rodrigues C, Spencer-Martins I, Hahn-Hägerdal B. 2003. High capacity xylose transport in *Candida intermedia* PYCC 4715. *FEMS Yeast Res* 3:45–52. <https://doi.org/10.1111/j.1567-1364.2003.tb00137.x>.
- Langeron MG, Guerra P. 1938. Nouvelles recherches de zymology médicale. *Ann Parasitol Hum Comp* 16:429–476.
- Morikawa Y, Takasawa S, Masunaga I, Takayama K. 1985. Ethanol productions from D-xylose and cellobiose by *Kluyveromyces cellobiovorus*. *Biotechnol Bioeng* 27:509–513. <https://doi.org/10.1002/bit.260270417>.
- Leandro MJ, Gonçalves P, Spencer-Martins I. 2006. Two glucose/xylose transporter genes from the yeast *Candida intermedia*: first molecular characterization of a yeast xylose-H⁺ symporter. *Biochem J* 395: 543–549. <https://doi.org/10.1042/BJ20051465>.
- Rozman D, Komel R. 1994. Isolation of genomic DNA from filamentous fungi with high glucan level. *Biotechniques* 16:382–384.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <https://doi.org/10.1186/1471-2105-12-491>.
- Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 32: W309–W312. <https://doi.org/10.1093/nar/gkh379>.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>.
- Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res* 18:1979–1990. <https://doi.org/10.1101/gr.081612.108>.