

Lawrence Berkeley National Laboratory

LBL Publications

Title

Draft genome sequence of *Yarrowia lipolytica* NRRL Y-64008, an oleaginous yeast capable of growing on lignocellulosic hydrolysates.

Permalink

<https://escholarship.org/uc/item/9ff281kc>

Journal

Microbiology Resource Announcements, 12(12)

Authors

Jagtap, Sujit

Liu, Jing-Jing

Walukiewicz, Hanna

et al.

Publication Date

2023-12-14

DOI

10.1128/MRA.00435-23

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Draft genome sequence of *Yarrowia lipolytica* NRRL Y-64008, an oleaginous yeast capable of growing on lignocellulosic hydrolysates

Sujit Sadashiv Jagtap,^{1,2} Jing-Jing Liu,^{1,2} Hanna E. Walukiewicz,^{1,2} Robert Riley,³ Steven Ahrendt,³ Maxim Koriabine,³ Kelly Cobaugh,³ Asaf Salamov,³ Yuko Yoshinaga,³ Vivian Ng,³ Chris Daum,³ Igor V. Grigoriev,^{3,4} Patricia J. Slininger,⁵ Bruce S. Dien,⁵ Yong-Su Jin,^{2,6} Christopher V. Rao^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Yarrowia lipolytica* is an oleaginous yeast that produces high titers of fatty acid-derived biofuels and biochemicals. It can grow on hydrophobic carbon sources and lignocellulosic hydrolysates. The genome sequence of *Y. lipolytica* NRRL Y-64008 is reported to aid in its development as a biotechnological chassis for producing biofuels and bioproducts.

KEYWORDS *Yarrowia lipolytica*, genome analysis, oleaginous yeast

Yarrowia lipolytica NRRL Y-64008 is an oleaginous yeast that is generally recognized as a safe organism (1, 2). It is a promising host for high-titer production of sugar alcohols, citric acid, lipases, proteases, and lipid-based chemicals (3–6). It grows in the presence of organic acids at low pH and ionic liquids and is salt-tolerant (7–9). *Y. lipolytica* can grow on sugars and hydrophobic substrates and produce fatty acid-based bioproducts (10, 11). It can also produce lipids from lignocellulosic hydrolysates (12–14). It was originally isolated from the soil (7). This yeast was named *Y. lipolytica* NRRL Y-64008 after a BLAST analysis of its internal transcribed spacer (ITS) sequence (GenBank accession number [ON114154.1](https://ncbi.nlm.nih.gov/nucl/ON114154.1)) against the UNITE database (<https://unite.ut.ee/>) and genome sequencing results. Interesting phenotypes observed for this strain include the ability to utilize plant-based sugars and growth on undetoxified biomass hydrolysates (12–14). We sequenced the genome of *Y. lipolytica* NRRL Y-64008 to allow for further research into its metabolism and enable metabolic engineering to produce biofuels and bioproducts from renewable carbon sources.

The cells were grown for 2 days at 30°C and 250 rpm in a YPG medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose) and harvested at an OD₆₀₀ of 10. Genomic DNA was extracted using the Dr. GenTLE (from yeast) High Recovery Kit (Takara Bio Inc., Shiga, Japan).

Libraries larger than 10 kb for PacBio sequencing were prepared using 5 µg of genomic DNA (15, 16). The sheared DNA was treated with exonuclease, followed by end repair and blunt adapter ligation with the SMRTbell Template Prep Kit 1.0. BluePippin (Sage Science) was used to select the library at 6 kb cutoff, which was purified with AMPure PB beads. Following the annealing of the PacBio sequencing primers to the SMRTbell template library, a Sequel Binding Kit 2.0 was used to bind the sequencing polymerase to the primers. On a Sequel II sequencer from Pacific Biosystems, the SMRTbell template libraries were then sequenced using a v3 sequencing primer, 8M v1 SMRT cells, and version 2.0 sequencing chemistry with 1 × 900 sequencing movie run durations. Reads containing inverted repeats were removed with BBTools version 38.79 (<https://jgi.doe.gov/data-and-tools/bbtools/>), and the default

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

Address correspondence to Sujit Sadashiv Jagtap, jagtap@illinois.edu, or Christopher V. Rao, [cvrho@illinois.edu](mailto:cvrhao@illinois.edu).

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 18 May 2023

Accepted 10 October 2023

Published 20 November 2023

Copyright © 2023 Jagtap 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/).

minimum read length was 40 bp. Filtered circular consensus sequencing (CCS) reads were assembled with Flye version 2.3.6 release (<https://github.com/fenderglass/Flye>) (-g 40M --asm-coverage 50 --pacbio-corr) and polished with arrow version SMRTLink v8.0.0.80529 (<https://www.pacb.com/support/software-downloads>). All CCS reads were used in both assembly and polishing. Assembled scaffolds were screened using BLAST vs the National Center for Biotechnology Information (NCBI) RefSeq and nt databases to ensure that no non-fungal contaminants were assembled.

The 20.5 Mb genome assembly was in six contigs (N_{50} value of 3 Mb), with a 149.31× sequencing read coverage depth and a 49.04% GC content. A combination of gene predictors including transcriptome-based, protein homology-based, and GeneWise seeded by BLASTx alignments of fungal sequences in the NCBI's non-redundant protein database, and *de novo* and GeneMark was applied to a repeat-masked genome assembly (17–21). Functional annotations were inferred from the best BLAST hits against the NCBI non-redundant protein databank and InterProScan (22). The JGI Annotation process was used to predict 6,577 protein-coding genes from the genome (23).

ACKNOWLEDGMENTS

We thank Dr. Kyria Boundy-Mills affiliated with Phaff Yeast Culture Collection at the University of California Davis for useful discussions. Microbial strains used in this work were provided by the USDA-ARS Culture Collection (NRRL).

This material is based upon the work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under Award Number DE-SC0018420. The work (proposal: 10.46936/10.25585/60001206) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

AUTHOR AFFILIATIONS

¹Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

²DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

³Lawrence Berkeley National Laboratory, US Department of Energy Joint Genome Institute, Berkeley, California, USA

⁴Department of Plant and Microbial Biology, University of California, Berkeley, California, USA

⁵Bioenergy Research Unit, National Center for Agricultural Utilization Research, USDA-ARS, Peoria, Illinois, USA

⁶Department of Food Science and Nutrition, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

AUTHOR ORCID*s*

Sujit Sadashiv Jagtap  <http://orcid.org/0000-0001-9402-2259>

Robert Riley  <http://orcid.org/0000-0003-0224-0975>

Yong-Su Jin  <http://orcid.org/0000-0002-4464-9536>

Christopher V. Rao  <http://orcid.org/0000-0001-5606-4204>

FUNDING

Funder	Grant(s)	Author(s)
U.S. Department of Energy (DOE)	DE-SC0018420	Yong-Su Jin Christopher V. Rao Hanna E. Walukiewicz Sujit Sadashiv Jagtap Patricia J. Slininger Bruce S. Dien Jing-Jing Liu
U.S. Department of Energy (DOE)	DE-AC02-05CH11231	Igor V. Grigoriev Robert Riley Asaf Salamov Yuko Yoshinaga Chris Daum Vivian Ng Maxim Koriabine Kelly Cobaugh Steven Ahrendt

AUTHOR CONTRIBUTIONS

Sujit Sadashiv Jagtap, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing | Jing-Jing Liu, Conceptualization, Data curation, Investigation, Methodology | Hanna E. Walukiewicz, Conceptualization, Investigation, Methodology | Robert Riley, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation | Steven Ahrendt, Data curation, Formal analysis, Investigation, Methodology, Software, Validation | Maxim Koriabine, Investigation, Methodology | Kelly Cobaugh, Data curation, Formal analysis, Investigation, Methodology, Software, Validation | Asaf Salamov, Data curation, Formal analysis, Investigation, Methodology, Software, Validation | Yuko Yoshinaga, Data curation, Formal analysis, Investigation, Resources, Software, Validation | Vivian Ng, Data curation, Formal analysis, Investigation, Methodology, Software, Validation | Chris Daum, Data curation, Formal analysis, Investigation, Methodology, Resources, Software | Igor V. Grigoriev, Data curation, Formal analysis, Investigation, Methodology, Resources, Software | Patricia J. Slininger, Conceptualization, Resources | Bruce S. Dien, Conceptualization, Resources | Yong-Su Jin, Conceptualization, Funding acquisition, Investigation, Methodology | Christopher V. Rao, Conceptualization, Funding acquisition, Project administration, Writing – review and editing

DATA AVAILABILITY

MycCosm (https://myccosm.jgi.doe.gov/YarliY64008_1) provides access to the whole-genome assemblies and annotation. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number [JAKETR000000000](#). The version [JAKETR000000000.1](#) is discussed in this paper. The accession numbers for the project and reads are [PRJNA791961](#) and [SRR17752527](#), respectively.

REFERENCES

- Blazek J, Hill A, Liu L, Knight R, Miller J, Pan A, Otoupal P, Alper HS. 2014. Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* 5:3131. <https://doi.org/10.1038/ncomms4131>
- Rakicka M, Lazar Z, Dulermo T, Fickers P, Nicaud JM. 2015. Lipid production by the oleaginous yeast *Yarrowia lipolytica* using industrial by-products under different culture conditions. *Biotechnol Biofuels* 8:104. <https://doi.org/10.1186/s13068-015-0286-z>

3. Barth G, Gaillardin C. 1997. Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. FEMS Microbiol Rev 19:219–237. <https://doi.org/10.1111/j.1574-6976.1997.tb00299.x>
4. Beopoulos A, Cescut J, Haddouche R, Uribelarrea JL, Molina-Jouve C, Nicaud JM. 2009. *Yarrowia lipolytica* as a model for bio-oil production. Prog Lipid Res 48:375–387. <https://doi.org/10.1016/j.plipres.2009.08.005>
5. Finogenova TV, Morgunov IG, Kamzolova SV, Chernyavskaya OG. 2005. Organic acid production by the yeast *Yarrowia lipolytica*: a review of prospects. Appl Biochem Microbiol 41:418–425. <https://doi.org/10.1007/s10438-005-0076-7>
6. Liu HH, Ji XJ, Huang H. 2015. Biotechnological applications of *Yarrowia lipolytica*: past, present and future. Biotechnol Adv 33:1522–1546. <https://doi.org/10.1016/j.biotechadv.2015.07.010>
7. Andreishcheva EN, Isakova EP, Sidorov NN, Abramova NB, Ushakova NA, Shaposhnikov GL, Soares MI, Zvyagilskaya RA. 1999. Adaptation to salt stress in a salt-tolerant strain of the yeast *Yarrowia lipolytica*. Biochemistry (Mosc) 64:1061–1067.
8. Ryu S, Labbé N, Trinh CT. 2015. Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of α -ketoglutaric acid by *Yarrowia lipolytica*. Appl Microbiol Biotechnol 99:4237–4244. <https://doi.org/10.1007/s00253-015-6521-5>
9. Zhang S, Jagtap SS, Deewan A, Rao CV. 2019. pH selectively regulates citric acid and lipid production in *Yarrowia lipolytica* W29 during nitrogen-limited growth on glucose. J Biotechnol 290:10–15. <https://doi.org/10.1016/j.jbiotec.2018.10.012>
10. Ledesma-Amaro R, Nicaud J-M. 2016. *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. Prog Lipid Res 61:40–50. <https://doi.org/10.1016/j.plipres.2015.12.001>
11. Fickers P, Benetti PH, Waché Y, Marty A, Mauersberger S, Smit MS, Nicaud JM. 2005. Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. FEMS Yeast Res 5:527–543. <https://doi.org/10.1016/j.femsyr.2004.09.004>
12. Slininger PJ, Dien BS, Kurtzman CP, Moser BR, Bakota EL, Thompson SR, O'Bryan PJ, Cotta MA, Balan V, Jin M, Sousa L da C, Dale BE. 2016. Comparative lipid production by oleaginous yeasts in hydrolyzates of lignocellulosic biomass and process strategy for high titers. Biotechnol Bioeng 113:1676–1690. <https://doi.org/10.1002/bit.25928>
13. Walker C, Dien B, Giannone RJ, Slininger P, Thompson SR, Trinh CT, Tullman-Ercek D. 2021. Exploring proteomes of robust *Yarrowia lipolytica* isolates cultivated in biomass hydrolysate reveals key processes impacting mixed sugar utilization, lipid accumulation, and degradation. mSystems 6:e0044321. <https://doi.org/10.1128/mSystems.00443-21>
14. Walker C, Ryu S, Na H, Zane M, LaButti K, Lipzen A, Haridas S, Barry K, Grigoriev IV, Quarterman J, Slininger P, Dien B, Trinh CT. 2018. Draft genome assemblies of five robust *Yarrowia lipolytica* strains exhibiting high lipid production, pentose sugar utilization, and sugar alcohol secretion from undetoxified lignocellulosic biomass hydrolysates. Microbiol Resour Announc 7:e01040-18. <https://doi.org/10.1128/MRA.01040-18>
15. Kim KE, Peluso P, Babayan P, Yeadon PJ, Yu C, Fisher WW, Chin C-S, Rapicavoli NA, Rank DR, Li J, Catcheside DEA, Celniker SE, Phillippy AM, Bergman CM, Landolin JM. 2014. Long-read, whole-genome shotgun sequence data for five model organisms. Sci Data 1:140045. <https://doi.org/10.1038/sdata.2014.45>
16. Travers KJ, Chin C-S, Rank DR, Eid JS, Turner SW. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38:e159–e159. <https://doi.org/10.1093/nar/gkq543>
17. Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. Genome Res 14:988–995. <https://doi.org/10.1101/gr.1865504>
18. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezukh Y, Raytselis Y, Sayers EW, Tao T, Ye J, Zaretskaya I. 2013. BLAST: a more efficient report with usability improvements. Nucleic Acids Res 41:W29–W33. <https://doi.org/10.1093/nar/gkt282>
19. Salamov AA, Solovyev VV. 2000. Ab initio gene finding in *Drosophila* genomic DNA. Genome Res 10:516–522. <https://doi.org/10.1101/gr.10.4.516>
20. 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>
21. Zhou K, Salamov A, Kuo A, Aerts AL, Kong X, Grigoriev IV. 2015. Alternative splicing acting as a bridge in evolution. Stem Cell Investig 2:19. <https://doi.org/10.3978/j.issn.2306-9759.2015.10.01>
22. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res 37:D211–D215. <https://doi.org/10.1093/nar/gkn785>
23. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res 42:D699–D704. <https://doi.org/10.1093/nar/gkt1183>