

# Draft Genome Sequence of the Oleaginous Yeast *Yarrowia lipolytica* PO1f, a Commonly Used Metabolic Engineering Host

Leqian Liu,<sup>a</sup> Hal S. Alper<sup>a,b</sup>

Department of Chemical Engineering, the University of Texas at Austin, Austin, Texas, USA<sup>a</sup>; Institute for Cellular and Molecular Biology, the University of Texas at Austin, Austin, Texas, USA<sup>b</sup>

**The draft genome sequence of the oleaginous yeast *Yarrowia lipolytica* strain PO1f, a commonly used metabolic engineering host, is presented here. The approximately 20.3-Mb genome sequence of PO1f will greatly facilitate research efforts in metabolic engineering of *Yarrowia lipolytica* for value-added chemical production.**

Received 8 June 2014 Accepted 16 June 2014 Published 3 July 2014

**Citation** Liu L, Alper HS. 2014. Draft genome sequence of the oleaginous yeast *Yarrowia lipolytica* PO1f, a commonly used metabolic engineering host. *Genome Announc.* 2(4):e00652-14. doi:10.1128/genomeA.00652-14.

**Copyright** © 2014 Liu and Alper. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Hal S. Alper, halper@che.utexas.edu.

*Yarrowia lipolytica*, a nonconventional oleaginous yeast, has recently emerged as a potential host strain that is recognized both as safe (1) and as a potent producer of value-added chemicals and industrial protein (2, 3). As a result, there has been a growing interest in biotechnological applications in this host strain due to both established biological information and intriguing physiological characteristics. Recent efforts have further expanded the genetic toolbox for *Y. lipolytica* (4, 5) and rewired metabolic networks for high-level production of fatty acid-based value-added chemicals (6–9).

Although a high-quality genome sequence of *Y. lipolytica* strain CLIB122 (E150) has been available (10), this strain is not the most popular for metabolic engineering applications. Specifically, *Y. lipolytica* strain W29 (CLIB89) and its derivatives, such as PO1f, have been more widely used, especially in metabolic engineering studies for value-added chemical production (6, 7, 9, 11–13), therapeutic protein production (14, 15), and fundamental microbiology studies (16–18). As one of the parental strains of the French inbred lines, the wild-type haploid strain, W29, was originally isolated from sewage material (19). A preliminary sequencing effort was conducted with only 4.9 Mb available (20). To gain a better understanding of the strain W29 and its potential for value-added chemical production, we generated the genome sequence for its derivative strain, PO1f.

The genome of *Y. lipolytica* PO1f was sequenced using the Illumina HiSeq DNA sequencing platform (PE2X100). The raw sequence data comprise a total of 8,740,022 reads that together provide very high sampling coverage of the genome (43.7-fold coverage). The reads were assembled using Velvet with a k-mer size of 55 (21). This led to a genome assembly containing 669 contigs (each at a length of  $\geq 500$  bp). The total length of the genome assembly is 20,282,994 bp, with an  $N_{50}$  equal to 58 kbp. The reads were also assembled using the A5 pipeline (22), and gaps were closed with IMAGE (23) to 348 contigs (each at a length of  $\geq 500$  bp) and further scaffolded based on the genome sequence of strain CLIB122 using ABACAS (24). A total of 19,922,824 bp was placed to the final 6 scaffolds.

The final *de novo* assembled genome was analyzed to assign open reading frames (ORFs) with Augustus (25) trained with *Y. lipolytica* CLIB122 data. A total of 6,420 putative ORFs were identified and 4,096 were annotated with Blast2Go (26). The genome sequences of PO1f and strain CLIB122 are very similar in nature. By mapping the Illumina reads to the CLIB122 genome using BWA (27) and analyzing using Samtools (27) and BEDTools (28), a total of 24,675 single nucleotide variations were called in PO1f genome sequences (QUAL  $\geq 30$ ; DP  $\geq 10$ ). Long terminal repeat (LTR)-retrotransposon elements are confirmed to be absent in strain PO1f, matching prior information about this strain (20). There is one large deletion in chromosome A with four ORFs missing. Two of them are weakly similar to the SMC5/6 complex (YAL10A01562p and YAL10A01602p), which are related to double-strand break repairing and homologous recombination (29). These absences may give rise to differences in homologous recombination efficiencies in this strain.

**Nucleotide sequence accession numbers.** This whole-genome shotgun analysis has been deposited at DDBJ/EMBL/GenBank under the accession no. JAFI000000000. The versions described in this paper are versions JAFI01000000 and JAFI02000000.

## ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research Young Investigator Program.

We thank Dhivya Arasappan for her helpful advice and discussions.

## REFERENCES

1. Groenewald M, Boekhout T, Neuvéglise C, Gaillardin C, van Dijck PWM, Wyss M. 2014. *Yarrowia lipolytica*: Safety assessment of an oleaginous yeast with a great industrial potential. *Crit. Rev. Microbiol.* 40: 187–206. [http://dx.doi.org/10.3109/1040841X.2013.770386](https://doi.org/10.3109/1040841X.2013.770386).
2. Bankar AV, Kumar AR, Zinjarde SS. 2009. Environmental and industrial applications of *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol.* 84: 847–865. [http://dx.doi.org/10.1007/s00253-009-2156-8](https://doi.org/10.1007/s00253-009-2156-8).
3. Liu L, Redden H, Alper HS. 2013. Frontiers of yeast metabolic engineering: diversifying beyond ethanol and *Saccharomyces*. *Curr. Opin. Biotechnol.* 24: 1023–1030. [http://dx.doi.org/10.1016/j.copbio.2013.03.005](https://doi.org/10.1016/j.copbio.2013.03.005).
4. Blazek J, Liu L, Redden H, Alper H. 2011. Tuning gene expression in

- Yarrowia lipolytica* by a Hybrid promoter Approach. Appl. Environ. Microbiol. 77:7905–7914. <http://dx.doi.org/10.1128/AEM.05763-11>.
5. Juretzek T, Le Dall M-T, Mauersberger S, Gaillardin C, Barth G, Nicaud J-M. 2001. Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. Yeast 18:97–113. [http://dx.doi.org/10.1002/1097-0061\(20010130\)18:2<97::AID-YEA652>3.0.CO;2-U](http://dx.doi.org/10.1002/1097-0061(20010130)18:2<97::AID-YEA652>3.0.CO;2-U).
  6. Blazek J, Liu L, Knight R, Alper HS. 2013. Heterologous production of pentane in the oleaginous yeast *Yarrowia lipolytica*. J. Biotechnol. 165: 184–194. <http://dx.doi.org/10.1016/j.jbiotec.2013.04.003>.
  7. Tai M, Stephanopoulos G. 2013. Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. Metab. Eng. 15:1–9. <http://dx.doi.org/10.1016/j.ymben.2012.08.007>.
  8. Xue Z, Sharpe PL, Hong S-P, Yadav NS, Xie D, Short DR, Damude HG, Rupert RA, Seip JE, Wang J, Pollak DW, Bostick MW, Bosak MD, Macool DJ, Hollerbach DH, Zhang H, Arcilla DM, Bledsoe SA, Croker K, McCord EF, Tyreus BD, Jackson EN, Zhu Q. 2013. Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. Nat. Biotechnol. 31:734–740. <http://dx.doi.org/10.1038/nbt.2622>.
  9. 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. <http://dx.doi.org/10.1038/ncomms4131>.
  10. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, de Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisramé A, Boyer J, Cattolico L, Confaniolieri F, de Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud J-M, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard G-F, Straub M-L, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P, Souciet J-L. 2004. Genome evolution in yeasts. Nature 430:35–44. <http://dx.doi.org/10.1038/nature02579>.
  11. Dulermo T, Nicaud J-M. 2011. Involvement of the G3P shuttle and  $\beta$ -oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. Metab. Eng. 13:482–491. <http://dx.doi.org/10.1016/j.ymben.2011.05.002>.
  12. Beopoulos A, Verbeke J, Bordes F, Guicherd M, Bressy M, Marty A, Nicaud J-M. 2014. Metabolic engineering for ricinoleic acid production in the oleaginous yeast *Yarrowia lipolytica*. Appl. Microbiol. Biotechnol. 98:251–262. <http://dx.doi.org/10.1007/s00253-013-5295-x>.
  13. Zhang B, Chen H, Li M, Gu Z, Song Y, Ratledge C, Chen YQ, Zhang H, Chen W. 2013. Genetic engineering of *Yarrowia lipolytica* for enhanced production of trans-10, cis-12 conjugated linoleic acid. Microb. Cell Fact. 12:70. <http://dx.doi.org/10.1186/1475-2859-12-70>.
  14. De Pourcq K, Vervecken W, Dewerte I, Valevska A, Van Hecke A, Callewaert N. 2012. Engineering the yeast *Yarrowia lipolytica* for the production of therapeutic proteins homogeneously glycosylated with Man8GlcNAc2 and Man5GlcNAc2. Microb. Cell Fact. 11:53. <http://dx.doi.org/10.1186/1475-2859-11-53>.
  15. Gasmí N, Ayed A, Nicaud J-M, Kallel H. 2011. Design of an efficient medium for heterologous protein production in *Yarrowia lipolytica*: case of human interferon alpha 2b. Microb. Cell Fact. 10:1–13. <http://dx.doi.org/10.1186/1475-2859-10-S1-S1>.
  16. Mlíčková K, Roux E, Athenstaedt K, d'Andrea S, Daum G, Chardot T, Nicaud J-M. 2004. Lipid accumulation, lipid body formation, and acyl coenzyme A oxidases of the yeast *Yarrowia lipolytica*. Appl. Environ. Microbiol. 70:3918–3924. <http://dx.doi.org/10.1128/AEM.70.7.3918-3924.2004>.
  17. Ruiz-Herrera J, Sentandreu R. 2002. Different effectors of dimorphism in *Yarrowia lipolytica*. Arch. Microbiol. 178:477–483. <http://dx.doi.org/10.1007/s00203-002-0478-3>.
  18. Beopoulos A, Mrozova Z, Thevenieau F, Le Dall M-T, Hapala I, Papanikolaou S, Chardot T, Nicaud J-M. 2008. Control of lipid accumulation in the yeast *Yarrowia lipolytica*. Appl. Environ. Microbiol. 74: 7779–7789. <http://dx.doi.org/10.1128/AEM.01412-08>.
  19. Barth G, Gaillardin C. 1996. *Yarrowia lipolytica*, p 313–388. In Wolf K (ed), Nonconventional yeasts in biotechnology. Springer Verlag, Berlin, Germany.
  20. Casaregola S, Neuvéglise C, Lépingle A, Bon E, Feynerol C, Artiguenave F, Wincker P, Gaillardin C. 2000. Genomic exploration of the hemiascomycetous yeasts: 17. *Yarrowia lipolytica*. FEBS. Lett. 487:95–100. [http://dx.doi.org/10.1016/S0014-5793\(00\)02288-2](http://dx.doi.org/10.1016/S0014-5793(00)02288-2).
  21. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
  22. Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012. An integrated pipeline for de novo assembly of microbial genomes. PLoS One 7:e42304. <http://dx.doi.org/10.1371/journal.pone.0042304>.
  23. Tsai JJ, Otto TD, Berriman M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. 11:R41. <http://dx.doi.org/10.1186/gb-2010-11-4-r41>.
  24. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. Bioinformatics 25:1968–1969. <http://dx.doi.org/10.1093/bioinformatics/btp347>.
  25. Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19(Suppl 2):ii215–ii225.
  26. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676. <http://dx.doi.org/10.1093/bioinformatics/bti610>.
  27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>.
  28. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. <http://dx.doi.org/10.1093/bioinformatics/btq033>.
  29. Potts PR. 2009. The Yin and Yang of the MMS21–SMC5/6 Sumo ligase complex in homologous recombination. DNA Repair 8:499–506. <http://dx.doi.org/10.1016/j.dnarep.2009.01.009>.