AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Assembly of a Highly Aldehyde-Resistant *Saccharomyces cerevisiae* SA1-Derived Industrial Strain

Sheila Tiemi Nagamatsu,[a] Gleidson Silva Teixeira,[a,b] Fellipe da Silveira Bezerra de Mello,[a] Pedro Augusto Galvão Tizei,[a] Bruna Tatsue Grichoswski Nakagawa,[a] Lucas Miguel de Carvalho,[a] Gonçalo Amarante Guimarães Pereira,[a] Marcelo Falsarella Carazzolle[a]

[a]Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil
[b]Departamento de Engenharia de Alimentos, Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, Campinas, SP, Brazil

ABSTRACT Here, we report the genome assembly of a *Saccharomyces cerevisiae* SA1-derived haploid (FMY097) indigenous strain isolated from a Brazilian ethanol distillery. FMY097 was recently reported to be a highly aldehyde-resistant strain capable of producing bioethanol in the presence of up to 40 mM furfural and 80 mM 5-hydroxymethylfurfural.

The awareness of the environmental consequences resulting from fossil fuel usage leads to an increased interest in the development of renewable energy sources (1). In a biomass-derived ethanol context, different by-products with inhibitory effects on microorganism metabolism are significantly produced during industrial biomass pretreatment (2). In particular, aldehydes—such as furfural and 5-hydroxymethylfurfural (HMF)—induce cell wall and membrane damage and breakdown of DNA and inhibit protein and RNA synthesis (3). To overcome these effects, *Saccharomyces cerevisiae* strains adapted to industrial conditions (industrial strains) have been considered more appropriate platforms for the development of commercial yeasts able to convert C6 and C5 hydrolyzed sugars into ethanol (4, 5). Here, we report the genome assembly of a SA1-derived haploid (FMY097) indigenous strain isolated from a Brazilian ethanol distillery (6). FMY097 was recently reported to be a highly aldehyde-resistant strain, despite its high fermentation efficiency and prolonged persistence in the process, because it is able to produce biomass in the presence of up to 40 mM furfural and 80 mM HMF (7).

Genomic DNA was isolated with a Wizard genomic DNA purification kit (Promega) from haploid cells cultivated on yeast extract-peptone-dextrose (YPD) medium (1% yeast extract, 2% peptone, and 2% glucose). Illumina sequencing libraries were prepared using the Nextera DNA library prep kit and sequenced on a HiSeq 2500 instrument to produce around 8 million paired-end 100-bp reads, representing 134× genome coverage. The quality of sequenced reads was evaluated by the FastQC package (version 3) (8). The reads were assembled with SPAdes (version 3.6.1) (9) configured as the "–careful" parameter in order to decrease erroneous mutations and indels (insertions/deletions) and a k-mer ranging from 21 to 91 with a step of 10. The final assembly generated 228 contigs larger than 1,000 bp that totaled 11.6 Mb, an N_{50} value of 130,414 bp (28 contigs), and a G+C content of 36.8%.

Gene prediction was performed using AUGUSTUS software (version 3.2.3) (10) with default parameters and using the previously available "saccharomyces_cerevisiae_rm11-1a_1" as a training model. A total of 5,330 predicted proteins larger than 50 amino acids were identified, with an average sequence length of 508 amino acids. In order to evaluate the completeness of the predicted proteins, the BUSCO analysis (11) was applied using a set of 1,711 highly conserved orthologous *Saccharomycetales* proteins.

Citation Nagamatsu ST, Teixeira GS, de Mello FDSB, Tizei PAG, Nakagawa BTG, de Carvalho LM, Pereira GAG, Carazzolle MF. 2019. Genome assembly of a highly aldehyde-resistant *Saccharomyces cerevisiae* SA1-derived industrial strain. Microbiol Resour Announc 8:e00071-19. https://doi.org/10.1128/MRA.00071-19.

Editor Christina Cuomo, Broad Institute of MIT and Harvard University

Address correspondence to Gonçalo Amarante Guimarães Pereira, goncalo@unicamp.br.

S.T.N. and G.S.T. are co-first authors with equal contribution.

Received 31 January 2019
Accepted 23 February 2019
Published 28 March 2019

A total of 99.5% of the 1,711 proteins in the BUSCO database were identified in the predicted proteins, and of these, 97.5% were classified as complete and single copy.

The annotation of predicted genes was performed using the well-studied genome of *Saccharomyces cerevisiae* S288c through the analysis of single-copy orthologs using the ORTHOMCL program (version 1.4) (12). A total of 4,747 single-copy genes were identified and annotated by this approach. The remaining 583 genes were annotated by PANNZER2 (version 2.0) (13), which resulted in only 51 unknown genes.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number SDIA00000000 and BioProject number PRJNA515487. The DNA Illumina reads were submitted to the Sequence Read Archive (SRA) at NCBI under SRA accession number SRR8455574.

ACKNOWLEDGMENTS

We thank Piotr Mieczkowski at the University of North Carolina for sequencing these data.

We also thank the following Brazilian funding agencies: the Center for Computational Engineering and Sciences-FAPESP/Cepid (2013/08293-7); the São Paulo Research Foundation (FAPESP) through grants 2014/26905-2, 2015/06263-9, 2015/06677-8, and 2016/02506-7; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)-Finance Code 001.

REFERENCES

1. Tavoni M, Kriegler E, Riahi K, van Vuuren DP, Aboumahboub T, Bowen A, Calvin K, Campiglio E, Kober T, Jewell J, Luderer G, Marangoni G, McCollum D, van Sluisveld M, Zimmer A, van der Zwaan B. 2015. Post-2020 climate agreements in the major economies assessed in the light of global models. Nat Clim Change 5:119–126. https://doi.org/10.1038/nclimate2475.
2. Jayakody LN, Hayashi N, Kitagaki H. 2013. Molecular mechanisms for detoxification of major aldehyde inhibitors for production of bioethanol by *Saccharomyces cerevisiae* from hot-compressed water-treated lignocellulose. *In* Méndez-Vilas A (ed), Materials and processes for energy: communicating current research and technological developments. Formatex Research Center, Badajoz, Spain.
3. Liu ZL. 2011. Molecular mechanisms of yeast tolerance and in situ detoxification of lignocellulose hydrolysates. Appl Microbiol Biotechnol 90:809–825. https://doi.org/10.1007/s00253-011-3167-9.
4. Parreiras LS, Breuer RJ, Narasimhan RA, Higbee AJ, La Reau A, Tremaine M, Qin L, Willis LB, Bice BD, Bonfert BL, Pinhancos RC, Balloon AJ, Uppugundla N, Liu T, Li C, Tanjore D, Ong IM, Li H, Pohlmann EL, Serate J, Withers ST, Simmons BA, Hodge DB, Westphall MS, Coon JJ, Dale BE, Balan V, Keating DH, Zhang Y, Landick R, Gasch AP, Sato TK. 2014. Engineering and two-stage evolution of a lignocellulosic hydrolysate-tolerant *Saccharomyces cerevisiae* strain for anaerobic fermentation of xylose from AFEX pretreated corn stover. PLoS One 9:e107499. https://doi.org/10.1371/journal.pone.0107499.
5. dos Santos LV, Carazzolle MF, Nagamatsu ST, Sampaio NMV, Almeida LD, Pirolla RAS, Borelli G, Corrêa TLR, Argueso JL, Pereira GAG. 2016. Unraveling the genetic basis of xylose consumption in engineered *Saccharomyces cerevisiae* strains. Sci Rep 6:38676. https://doi.org/10.1038/srep38676.
6. Basso LC, de Amorim HV, de Oliveira AJ, Lopes ML. 2008. Yeast selection for fuel ethanol production in Brazil. FEMS Yeast Res 8:1155–1163. https://doi.org/10.1111/j.1567-1364.2008.00428.x.
7. de Mello FDSB, Coradini ALV, Tizei PAG, Carazzolle MF, Pereira GAG, Teixeira GS. 2019. Static microplate fermentation and automated growth analysis approaches identified a highly-aldehyde resistant *Saccharomyces cerevisiae* strain. Biomass Bioenergy 120:49–58. https://doi.org/10.1016/j.biombioe.2018.10.019.
8. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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.
10. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res 34:W435–W439. https://doi.org/10.1093/nar/gkl200.
11. Simão 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.
12. Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert CJ, Jr. 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr Protoc Bioinformatics Chapter 6:Unit 6.12.1–6.12.19. https://doi.org/10.1002/0471250953.bi0612s35.
13. Törönen P, Medlar A, Holm L. 2018. PANNZER2: a rapid functional annotation Web server. Nucleic Acids Res 46:W84–W88. https://doi.org/10.1093/nar/gky350.