

Draft Genome Sequence of the Probiotic Yeast *Saccharomyces cerevisiae* var. *boulardii* Strain ATCC MYA-796

T. M. Batista,^a E. T. A. Marques, Jr.,^c G. R. Franco,^a B. Douradinha^{b,c}

Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil^a; Fondazione RiMED, Palermo, Italy^b; University of Pittsburgh, Center of Vaccine Research, Pittsburgh, Pennsylvania, USA^c

***Saccharomyces boulardii* is the only yeast approved as a probiotic for human consumption. Here, we report the draft genome sequence of the strain ATCC MYA-796, derived from the French Ultra Levure probiotic drug. The genome has a size of 11.6 Mb with 5,305 putative open reading frames predicted.**

Received 12 November 2014 Accepted 14 November 2014 Published 18 December 2014

Citation Batista TM, Marques ETA, Jr., Franco GR, Douradinha B. 2014. Draft genome sequence of the probiotic yeast *Saccharomyces cerevisiae* var. *boulardii* strain ATCC MYA-796. *Genome Announc.* 2(6):e01345-14. doi:10.1128/genomeA.01345-14.

Copyright © 2014 Batista et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to T. M. Batista, thiagomafra@gmail.com, or B. Douradinha, brd49@pitt.edu.

Isolated by the French scientist Henri Boulard in 1920 during a cholera outbreak, the yeast *Saccharomyces cerevisiae* var. *boulardii* is still the only eukaryotic microorganism used as a probiotic in human health for the treatment of gastrointestinal disorders (1). This probiotic yeast is used worldwide and has been tested for clinical efficacy against several diseases, including traveler's diarrhea, antibiotic-associated diarrhea (AAD), acute adult diarrhea, HIV-related diarrhea, *Helicobacter pylori* diseases, *Clostridium difficile* and *Salmonella typhi* infections, and Crohn's disease, among others (2). Recently, a work was published describing the genome of an *S. boulardii* strain commercialized in India (3).

Here, we report the draft genome sequence of *S. cerevisiae* var. *boulardii* strain ATCC MYA-796, derived from the French Ultra Levure probiotic drug. The genomic DNA was sequenced using Illumina HiSeq by Axseq Technologies (<http://www.axseq.com>). A total of 48.3 million paired-end reads of 101 bp with an estimated 403× coverage were produced. Different k-mer values were tested to obtain the ideal value (k 61). The *de novo* assembly was performed using SOAPdenovo version 2.04 (4) with parameters -R -u -F -M 2. The contigs were scaffolded using SSPACE (5) and ordered by CONTIGuator (6) using the *S. cerevisiae* S288c genome as a reference; gaps were closed with GapCloser (4). The resulting assembly has 193 contigs (>400 bp) with a total length of 11,405,855 bp (the largest contig having a length of 459,679 bp), an N_{50} of 203 kb (in 19 contigs), and a G+C content of 38.1% and 100% of completeness, as estimated by mapping the orthologous genes (KOG databases) using CEGMA software (7). Genome annotation performed by MAKER2 (8) revealed 5,305 putative open reading frames (ORFs). An automatic annotation using the BLASTp algorithm (9) revealed 5,321 ORFs with significant similarity (E -value cutoff $\leq 10^{-3}$) to sequences deposited in the non-redundant (nr) protein database from NCBI. Using the tRNAscan-SE version 1.3 software (9), we found 273 tRNA genes scattered across the contigs.

The draft sequence of the yeast *S. cerevisiae* var. *boulardii* strain ATCC MYA-796 complements the genetic information of this probiotic yeast already deposited in GenBank.

Nucleotide sequence accession numbers. Data related to the whole-genome shotgun project of *S. cerevisiae* var. *boulardii* ATCC MYA-796 has been deposited at DDBJ/EMBL/GenBank under the accession number [JRHY00000000](https://www.ncbi.nlm.nih.gov/nuccore/JRHY00000000). The version herein described is under accession number [JRHY01000000](https://www.ncbi.nlm.nih.gov/nuccore/JRHY01000000).

ACKNOWLEDGMENTS

We thank Flaviano S. Martins from the Universidade Federal de Minas Gerais (Belo Horizonte, Brazil) for helpful discussion and suggestions, and EMBRAPA Informática Agropecuária (Campinas, Brazil) for providing access to genome annotation.

This work was partially funded by Fondazione RiMED (Palermo, Italy).

REFERENCES

- Czerucka D, Piche T, Rampal P. 2007. Review article: yeast as probiotics—*Saccharomyces boulardii*. *Aliment. Pharmacol. Ther.* 26:767–78. <http://dx.doi.org/10.1111/j.1365-2036.2007.03442.x>.
- McFarland LV. 2010. Systematic review and meta-analysis of *Saccharomyces boulardii* in adult patients. *World J. Gastroenterol.* 16:2202–2222. <http://dx.doi.org/10.3748/wjg.v16.i18.2202>.
- Khatri I, Akhtar A, Kaur K, Tomar R, Prasad GS, Ramya TNC, Subramanian S. 2013. Gleaning evolutionary insights from the genome sequence of a probiotic yeast *Saccharomyces boulardii*. *Gut Pathog.* 5:30. <http://dx.doi.org/10.1186/1757-4749-5-30>.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience.* 1:18. <http://dx.doi.org/10.1186/2047-217X-1-18>.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffold folding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <http://dx.doi.org/10.1093/bioinformatics/btq683>.
- Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on

- draft genomes. *Source Code Biol. Med.* 6:11. <http://dx.doi.org/10.1186/1751-0473-6-11>.
7. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067. <http://dx.doi.org/10.1093/bioinformatics/btm071>.
 8. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <http://dx.doi.org/10.1186/1471-2105-12-491>.
 9. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <http://dx.doi.org/10.1186/1471-2105-10-421>.