genomeA
Journals.ASM.org

Genome Sequence of *Torulaspora delbrueckii* NRRL Y-50541, Isolated from Mezcal Fermentation

Jorge Gomez-Angulo,[a] Leticia Vega-Alvarado,[b] Zazil Escalante-García,[a] Ricardo Grande,[c] Anne Gschaedler-Mathis,[a] Lorena Amaya-Delgado,[a] Javier Arrizon,[a] Alejandro Sanchez-Flores[c]

Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, AC, Unidad de Biotecnología Industrial, Jalisco, Mexico[a]; Centro de Ciencias Aplicadas y Desarrollo Tecnológico, UNAM, Distrito Federal, Mexico[b]; Unidad de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología de la UNAM, Morelos, Mexico[c]

***Torulaspora delbrueckii* presents metabolic features interesting for biotechnological applications (in the dairy and wine industries). Recently, the *T. delbrueckii* CBS 1146 genome, which has been maintained under laboratory conditions since 1970, was published. Thus, a genome of a new mezcal yeast was sequenced and characterized and showed genetic differences and a higher genome assembly quality, offering a better reference genome.**

Received 27 March 2015 **Accepted** 22 June 2015 **Published** 23 July 2015

Citation Gomez-Angulo J, Vega-Alvarado L, Escalante-García Z, Grande R, Gschaedler-Mathis A, Amaya-Delgado L, Arrizon J, Sanchez-Flores A. 2015. Genome sequence of *Torulaspora delbrueckii* NRRL Y-50541, isolated from mezcal fermentation. Genome Announc 3(4):e00438-15. doi:10.1128/genomeA.00438-15.

Copyright © 2015 Gomez-Angulo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Javier Arrizon, jparrizon@ciatej.mx, or Alejandro Sanchez-Flores, alexsf@ibt.unam.mx.

Torulaspora delbrueckii, an ascomycetous yeast, has high osmotolerance and high freeze tolerance (1–4). These properties make this yeast an interesting organism, with potential biotechnological applications in bakery, wine, and dairy industrial processes. In the last few years, studies about enzyme production in winemaking have been carried out, but only a few genes have been characterized (5). Recently, it has been reported that *T. delbrueckii* yeast strains isolated from the mezcal-fermenting process produced β-fructofuranosidase enzymes with fructosyltransferase activity (6). Therefore, the genome sequencing of *T. delbrueckii* can lead to the discovery of new genes with biotechnological application. Lately, the genome of *T. delbrueckii* CBS 1146 was obtained (using 454 sequencing), with the main purpose of studying the sex chromosome evolution in the *Saccharomycetaceae* family (7). However, the characterized strain has been maintained under laboratory conditions since 1970. Thus, in order to find the differences between the published reference and our mezcal isolate, we sequenced, assembled, and characterized it in order to find genes and variations associated with the fermentation process.

Genomic DNA from *T. delbrueckii* NRRL Y-50540 was isolated and prepared as Illumina sequencing libraries to generate a total of 20,514,013 paired-end reads (estimated coverage, ~328×) with a length of 72 bases, using the Illumina GAIIx platform. The assembly was performed with Velvet version 1.2.10 using a *k*-mer size of 35 (8). An assembly of 11,236,894 bp in 374 contigs with length ≥1,000 bp was obtained, with N_{50} and N_{90} values of 82,617 and 23,849 bp, respectively. The average contig length was 30,012 bp, giving a considerable space to search for genes. Finally, we ordered and scaffolded the assembly using ABACAS (9) against the available *T. delbrueckii* reference genome (7) of a different strain to leave the whole assembly in 8 scaffolds corresponding to 8 chromosomes. The average G+C content was 42%, which is consistent with the reported genome. Gene prediction was performed using AUGUSTUS version 2.7, and using several different yeast species profiles, we predicted 4,714 protein-coding genes by intersecting all predictions (10). Using CEGMA version 2.5, we obtained a 97% genome completeness (11).

In contrast to the genome published by Gordon et al. (7), we found a slightly better value for completeness and fewer open reading frames (ORFs) (4,714 versus 4,972, respectively) in our assembly. Both assemblies presented gaps, but we were able to remove some of them, which led to a more complete genome. Although these differences are not of concern, they are expected, since each genome was assembled using different sequencing technologies and assembly strategies. Currently, we are working on a hybrid assembly strategy using the information from both strains in order to obtain a better assembly, gene prediction, and annotation.

We believe that the *T. delbrueckii* genome sequence presented here can be used as a better reference to perform further analyses, such as differential gene expression of enzymes related to the synthesis and degradation of biotechnological molecules of interest, for example, under different fermentation conditions analyzed using RNA sequencing (RNA-seq) data.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at the NCBI GenBank database under the accession numbers CP011778 to CP011785.

ACKNOWLEDGMENTS

We thank the "Unidad de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología (USMB), UNAM," for DNA sequencing advice and bioinformatics analysis. The USMB is part of the "Laboratorio Nacional de Respuesta a Enfermedades Emergentes," which has been created and funded by the "CONACYT-Programa de Laboratorios Nacionales." We also thank Veronica Jimenez-Jacinto for preparing and submitting the sequencing data to the NCBI SRA and GenBank repositories.

We thank the CONACYT project CB-2012-01-181766 for financial support.

REFERENCES

1. **Hernández-López MJ, Pallotti C, Andreu P, Aguilera J, Prieto JA, Randez-Gil F.** 2007. Characterization of a *Torulaspora delbrueckii* diploid strain with optimized performance in sweet and frozen sweet dough. Int J Food Microbiol **116:**103–110. http://dx.doi.org/10.1016/ j.ijfoodmicro.2006.12.006.
2. **Hérnandez-López MJ, Panadero J, Prieto JA, Randez-Gil F.** 2006. Regulation of salt tolerance by *Torulaspora delbrueckii* calcineurin target Crz1p. Eukaryot Cell **5:**469–479. http://dx.doi.org/10.1128/EC.5.3.469-479.2006.
3. **Tofalo R, Chavez-López C, Di Fabio F, Schirone M, Felis GE, Torriani S, Paparella A, Suzzi G.** 2009. Molecular identification and osmotolerant profile of wine yeasts that ferment a high sugar grape must. Int J Food Microbiol **130:**179–187. http://dx.doi.org/10.1016/ j.ijfoodmicro.2009.01.024.
4. **Warren A, Chasseriaud L, Comte G, Panfili A, Delcamp A, Salin F, Marullo P, Bely M.** 2014. Winemaking and bioprocesses strongly shaped the genetic diversity of the ubiquitous yeast *Torulaspora delbrueckii*. PLoS One **9:**94246. http://dx.doi.org/10.1371/ journal.pone.0094246.
5. **Maturano YP, Rodríguez Assaf LA, Toro ME, Nally MC, Vallejo M, Castellanos de Figueroa LI, Combina M, Vazquez F.** 2012. Multi-enzyme production by pure and mixed cultures of *Saccharomyces* and non-*Saccharomyces* yeasts during wine fermentation. Int J Food Microbiol **155:**43–50. http://dx.doi.org/10.1016/j.ijfoodmicro.2012.01.015.
6. **Arrizon J, Morel S, Gschaedler A, Monsan P.** 2012. Fructanase and fructosyltransferase activity of non-*Saccharomyces* yeasts isolated from fermenting musts of mezcal. Bioresour Technol **110:**560–565. http://dx.doi.org/10.1016/j.biortech.2012.01.112.
7. **Gordon JL, Armisén D, Proux-Wéra E, ÓhÉigeartaigh SS, Byrne KP, Wolfe KH.** 2011. Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. Proc Natl Acad Sci USA **108:**20024–20029. http://dx.doi.org/10.1073/pnas.1112808108.
8. **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.
9. **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.
10. **Stanke M, Schöffmann O, Morgenstern B, Waack S.** 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics **7:**62. http://dx.doi.org/10.1186/1471-2105-7-62.
11. **Parra G, Bradnam K, Ning Z, Keane T, Korf I.** 2009. Assessing the gene space in draft genomes. Nucleic Acids Res **37:**289–297. http://dx.doi.org/10.1093/nar/gkn916.