

Draft Genome Sequence of the D-Xylose-Fermenting Yeast *Spathaspora arborariae* UFMG-HM19.1A^T

Francisco P. Lobo,^a Davi L. Gonçalves,^b Sergio L. Alves, Jr,^{b*} Alexandra L. Gerber,^c Ana Tereza R. de Vasconcelos,^c Luiz C. Basso,^d Glória R. Franco,^e Marco A. Soares,^f Raquel M. Cadete,^f Carlos A. Rosa,^f Boris U. Stambuk^b

Laboratório Multiusuário de Bioinformática, Embrapa Informática Agropecuária, Campinas, São Paulo, Brazil^a; Departamento de Bioquímica, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil^b; Laboratório de Bioinformática, Unidade de Genômica Computacional Darcy Fontoura de Almeida, Laboratório Nacional de Computação Científica, Petrópolis, Rio de Janeiro, Brazil^c; Departamento de Ciências Biológicas, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil^d; Departamento de Bioquímica, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil^e; Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil^f

* Present address: Sergio L. Alves, Jr, Unidade Bom Pastor, Campus Chapecó, Universidade Federal da Fronteira Sul, Chapecó, Brazil.

The draft genome sequence of the yeast *Spathaspora arborariae* UFMG-HM19.1A^T (CBS 11463 = NRRL Y-48658) is presented here. The sequenced genome size is 12.7 Mb, consisting of 41 scaffolds containing a total of 5,625 predicted open reading frames, including many genes encoding enzymes and transporters involved in D-xylose fermentation.

Received 4 December 2013 Accepted 7 December 2013 Published 16 January 2014

Citation Lobo FP, Gonçalves DL, Alves SL, Jr, Gerber AL, de Vasconcelos ATR, Basso LC, Franco GR, Soares MA, Cadete RM, Rosa CA, Stambuk BU. 2014. Draft genome sequence of the D-xylose-fermenting yeast *Spathaspora arborariae* UFMG-HM19.1A^T. *Genome Announc.* 2(1):e01163-13. doi:10.1128/genomeA.01163-13.

Copyright © 2014 Lobo et al. 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 Boris U. Stambuk, boris.stambuk@ufsc.br.

Yeasts of the genus *Spathaspora* have recently been discovered to be associated with rotting wood substrates and insects that occupy this ecological niche, and they produce typical elongated ascospores with curved ends (1–4). These yeasts are biotechnologically important because they can ferment D-xylose and other sugars present in plant biomass hydrolysates (2–6). The yeast *Spathaspora arborariae* was first isolated from rotting wood samples collected in the Atlantic rainforest and the Cerrado ecosystem in Brazil, and it not only ferments D-xylose efficiently, but it can also be used for bioethanol production from lignocellulosic hydrolysates (2, 7). Since D-xylose-fermenting yeasts are an important source of genomic information for biotechnological traits, including genes, enzymes, and/or sugar transporters to engineer industrial strains for the efficient production of bioethanol from renewable biomass (8), the draft genome of *S. arborariae* strain UFMG-HM19.1A^T was determined and automatically annotated.

Genome sequencing was performed by a whole-genome shotgun strategy using Roche 454 paired-end (insert size of 3 kb) technology. The raw sequence data comprise 915,700 reads with 657,682 mate-pairs totaling 291,670,584 nucleotides. The reads were assembled using Newbler (9) with default parameters into 439 contigs and 41 scaffolds, with a total length of 12,708,019 bp. The assembled genome has an N₅₀ of ~679 kb (6 scaffolds) and an N₉₀ of ~202 kb (18 scaffolds), with an average G+C content of 31.7%, which is comparable to those of other yeast genomes (10). We performed genome annotation with MAKER2 (11) and found 5,625 putative open reading frames (ORFs) >100 nucleotides (nt), in agreement with the genome content of other yeast species of the CTG clade (this taxon translates the CTG codon as serine instead of leucine [12]). Automatic gene annotation using the BLAST software (13) revealed 5,402 ORFs similar to sequences in the nonredundant protein database from the National Center for

Biotechnology Information (e-value cutoff of 10⁻¹⁰). We identified 185 tRNA genes scattered across the scaffolds and located the 28S and 18S rRNA genes at scaffold 9 using tRNAscan-SE 1.3 (14) and RNAmmer (15) softwares, respectively. The *S. arborariae* mitochondrial genome was assembled into a 22,709-bp fragment (scaffold 29). Pulsed-field gel electrophoresis (16) revealed seven chromosomal bands, with one probably migrating as a doublet. Thus, the 40 scaffolds with approximately 12.7 Mb are likely to represent most of the genomic content found on the eight chromosomes present in this yeast.

The draft genome of *S. arborariae* revealed several interesting biotechnological genes for bioethanol production, including genes coding for sugar transporters and D-xylose reductase, xylitol dehydrogenase, and xylulokinase, responsible for the conversion of D-xylose to D-xylulose-5P, which is subsequently channeled into the pentose-phosphate pathway for ethanol production. The genes for D-xylose reductase from *Spathaspora* yeasts (*Spathaspora passalidarum* and *Spathaspora arborariae*) are particularly interesting, as these enzymes accept both NADH and NADPH as cofactors, allowing anaerobic D-xylose fermentation (5, 17).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [AYLH010000000](https://www.ncbi.nlm.nih.gov/nuccore/AYLH010000000). The version described in this paper is version AYLH01000000.

ACKNOWLEDGMENTS

This work was supported in part by grants and fellowships from the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (process no. 552877/2007-7, 490029/2009-4, 560715/2010-2, and 551392/2010-0), Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), and Fundação de Apoio à Pesquisa Científica

fica e Tecnológica do Estado de Santa Catarina (FAPESC) (process no. 17293/2009-6).

REFERENCES

- Nguyen NH, Suh SO, Marshall CJ, Blackwell M. 2006. Morphological and ecological similarities: wood-boring beetles associated with novel xylose-fermenting yeasts, *Spathaspora passalidarum* gen. sp. nov. and *Candida jeffriesii* sp. nov. *Mycol. Res.* 110:1232–1241. <http://dx.doi.org/10.1016/j.mycres.2006.07.002>.
- Cadete RM, Santos RO, Melo MA, Mouro A, Gonçalves DL, Stambuk BU, Gomes FC, Lachance MA, Rosa CA. 2009. *Spathaspora arborariae* sp. nov., a D-xylose-fermenting yeast species isolated from rotting wood in Brazil. *FEMS Yeast Res.* 9:1338–1342. <http://dx.doi.org/10.1111/j.1567-1364.2009.00582.x>.
- Cadete RM, Melo MA, Dussán KJ, Rodrigues RC, Silva SS, Zilli JE, Vital MJ, Gomes FC, Lachance MA, Rosa CA. 2012. Diversity and physiological characterization of D-xylose-fermenting yeasts isolated from the Brazilian Amazonian forest. *PLoS One* 7:e43135. <http://dx.doi.org/10.1371/journal.pone.0043135>.
- Cadete RM, Melo MA, Zilli JE, Vital MJ, Mouro A, Prompt AH, Gomes FC, Stambuk BU, Lachance MA, Rosa CA. 2013. *Spathaspora brasiliensis* sp. nov., *Spathaspora suhii* sp. nov., *Spathaspora roraimanensis* sp. nov. and *Spathaspora xylofermentans* sp. nov., four novel D-xylose-fermenting yeast species from Brazilian Amazonian forest. *Antonie Van Leeuwenhoek* 103:421–431. <http://dx.doi.org/10.1007/s10482-012-9822-z>.
- Hou X. 2012. Anaerobic xylose fermentation by *Spathaspora passalidarum*. *Appl. Microbiol. Biotechnol.* 94:205–214. <http://dx.doi.org/10.1007/s00253-011-3694-4>.
- Long TM, Su YK, Headman J, Higbee A, Willis LB, Jeffries TW. 2012. Cofermentation of glucose, xylose, and cellobiose by the beetle-associated yeast *Spathaspora passalidarum*. *Appl. Environ. Microbiol.* 78:5492–5500. <http://dx.doi.org/10.1128/AEM.00374-12>.
- da Cunha-Pereira F, Hickert LR, Sehnem NT, de Souza-Cruz PB, Rosa CA, Ayub MA. 2011. Conversion of sugars present in rice hull hydrolysates into ethanol by *Spathaspora arborariae*, *Saccharomyces cerevisiae*, and their co-fermentations. *Bioresour. Technol.* 102:4218–4225. <http://dx.doi.org/10.1016/j.biortech.2010.12.060>.
- Wohlbach DJ, Kuo A, Sato TK, Potts KM, Salamov AA, Labutti KM, Sun H, Clum A, Pangilinan JL, Lindquist EA, Lucas S, Lapidus A, Jin M, Gunawan C, Balan V, Dale BE, Jeffries TW, Zinkel R, Barry KW, Grigoriev IV, Gasch AP. 2011. Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. *Proc. Natl. Acad. Sci. U. S. A.* 108:13212–13217. <http://dx.doi.org/10.1073/pnas.1103039108>.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380. <http://dx.doi.org/10.1038/nature03959>.
- Liu X, Kaas RS, Jensen PR, Workman M. 2012. Draft genome sequence of the yeast *Pachysolen tannophilus* CBS 4044/NRRL Y-2460. *Eukaryot. Cell* 11:827. <http://dx.doi.org/10.1128/EC.00114-12>.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491–505. <http://dx.doi.org/10.1186/1471-2105-12-491>.
- Maguire SL, O'Heigeartaigh SS, Byrne KP, Schröder MS, O'Gaora P, Wolfe KH, Butler G. 2013. Comparative genome analysis and gene finding in *Candida* species using CGOB. *Mol. Biol. Evol.* 30:1281–1291. <http://dx.doi.org/10.1093/molbev/mst042>.
- 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>.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
- Gerring SL, Connelly C, Hieter P. 1991. Positional mapping of genes by chromosome blotting and chromosome fragmentation. *Methods Enzymol.* 194:57–77. [http://dx.doi.org/10.1016/0076-6879\(91\)94007-Y](http://dx.doi.org/10.1016/0076-6879(91)94007-Y).
- Mouro A. 2012. Xylose and cellobiose fermentation by yeasts isolated from the Brazilian biodiversity. M.Sc. thesis. Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil.