Draft genome sequence of *Sugiyamaella xylanicola* UFMG-CM-Y1884^T^, a xylan-degrading yeast species isolated from rotting wood samples in Brazil

Thiago M. Batista\(^a\), Rennan G. Moreira\(^c\), Heron O. Hilário\(^a\), Camila G. Morais\(^b\), Glória R. Franco\(^a\), Luiz H. Rosa\(^b\), Carlos A. Rosa\(^b,⁎\)

\(^a\) Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG CEP 31270-901, Brazil
\(^b\) Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, CEP 31270-901, Brazil
\(^c\) Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, CEP 31270-901, Brazil

**A B S T R A C T**

We present the draft genome sequence of the type strain of the yeast *Sugiyamaella xylanicola* UFMG-CM-Y1884^T^ (＝UFMG-CA-32.1^T^＝CBS 12683^T^), a xylan-degrading species capable of fermenting D-xylose to ethanol. The assembled genome has a size of ~13.7 Mb and a GC content of 33.8% and contains 5971 protein-coding genes. We identified 15 genes with significant similarity to the D-xylose reductase gene from several other fungal species. The draft genome assembled from whole-genome shotgun sequencing of the yeast *Sugiyamaella xylanicola* UFMG-CM-Y1884^T^ (＝UFMG-CA-32.1^T^＝CBS 12683^T^) has been deposited at DDBJ/ENA/GenBank under the accession number MQSX00000000 under version MQSX01000000.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

**Keywords:** *Sugiyamaella xylanicola*, genome sequence, D-xylose-fermenting yeast, Xylan-degrading species

1. Direct link to deposited data

   http://www.ncbi.nlm.nih.gov/bioproject/PRJNA354640

2. Introduction

   The genus *Sugiyamaella* comprises yeast species that inhabit the soil or insect guts or live in association with rotting plant materials [1,2]. In the present study, the yeast *Sugiyamaella xylanicola* UFMG-CM-Y1884^T^ was isolated from rotting wood samples in the private Natural Heritage Reserve of the Caraça, Minas Gerais state in Brazil [3].

   The sequenced *S. xylanicola* strain exhibits xylanolytic activity and is capable of producing ethanol from xylose, two important characteristics that are important for the production of lignocellulosic ethanol [2,3]. *S. xylanicola* harbors genes and enzymes involved in important pathways, such as xylose metabolism, and is thus a relevant source of genomic information for mining biotechnological traits engineering of industrial strains to achieve efficient production of second generation ethanol from renewable biomass.

3. DNA extraction, library construction, and sequencing

   Genomic DNA of the type strain of *S. xylanicola* UFMG-CM-Y1884^T^ (＝UFMG-CA-32.1^T^＝CBS 12683^T^) was isolated via phenol:chloroform (1:1) extraction. DNA quality was assessed via gel electrophoresis, and purity and quantity were determined using the NanoDrop 1000 UV–vis spectrophotometer and Qubit 2.0 fluorometer using the Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific). Paired-end libraries were constructed with Nextera XT DNA Library Preparation Kit (Illumina). Generated fragments with a mean length of 983 bp were sequenced using a MiSeq instrument, whereas fragments with a mean size of 482 bp were sequenced on a HiSeq 2500 instrument.

4. Data analysis and results

   A total of 2,582,982 reads (2 × 301) were generated by MiSeq at an estimated coverage of 52× and 63,873,820 reads (2 × 101) generated
by HiSeq 2500 with coverage estimated of 921 ×. De novo assembly was performed using MaSuRCA [4] version 3.2.1, using the reads produced by MiSeq. Resulting contigs of MaSuRca were used with the parameter “−trust-contigs” in SPAdes assembler [5] version 3.9.0. The assembled draft genome consisted of 13,714,239 bp distributed across 1251 contigs longer than 272 bp and a GC content of 33.8%. The longest contig had a length of 638,759 bp, and the N50 contig length was 180,392 bp. CEGMA [6] analysis showed that the assembly is 96.77% complete, whereas BUSCO [7] analysis using the fungi lineage dataset indicated that the assembly is 90% complete (Table 1). Quality assessment of the assembly was performed using Quast software [8]. Gene prediction using Maker2 [9] identified 5971 predicted protein-coding genes. Sequence similarity searching using Blastx [10] version 2.2.31 + (−e-value cutoff: 1e−20) returned matches with 5638 proteins (94.42%) against NCBI’s non-redundant database. A total of 321 tRNAs were identified using tRNAscan [11]. Alcohol fermentation from lignocellulosic substrates is dependent on efficiency of D-xylose conversion. The main enzyme involved in this pathway is NAD(P)H-dependent α-xylose reductase (XR), which is encoded by the XYL1 gene. The XYL1 gene of Scheffersomyces stipitis has been successfully used to produce Schacharomyces cerevisiae strains capable of xylose fermentation [12]. Thus, we used the XYL1 gene from S. stipitis (Uniprot: P31897) as query for searching orthologous clusters from fungi in the OrthoDB database [13]. The cluster EOG092C324N was found to consist of 1950 orthologs searching orthologous clusters from fungi in the OrthoDB database. A total of 1240 orthologs were identified that the core dataset comprises 1438 orthologous proteins. CEGMA assessment returned 248 core orthologous proteins. BUSCO assessment indicated that the core dataset comprises 1438 orthologous proteins. The authors declare no competing interests.

**Conflict of interest**

**Acknowledgments**

This study was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - Grant 457499/2014-1) and Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG - Grant APQ-01525-14).

**References**

[1] H. Urbina, J. Schuster, M. Blackwell. The gut of Guatemalan passalid beetles: a habitat colonized by cellulose- and xylose-fermenting yeasts. Fungal Ecol. 6 (2013) 339–355, http://dx.doi.org/10.1016/j.fuseneo.2013.06.005.

[2] L.M.F. Sena, C.G. Moreira, M.R. Lopes, R.O. Santos, A.P.T. Uetanabaro, P.B. Morais, M.J.S. Vital, M.A. de Morais, M.-A. Lachance, C.A. Rosa, α-xylose fermentation, xylitol production and xylanase activities by seven new species of Sugiyamaella. Antonie van Leeuwenhoek 110 (2016) 53–67, http://dx.doi.org/10.1007/s10482-016-0775-5.

[3] C.G. Morais, C.A. Lara, S. Marques, C. Fonseca, M.A. Lachance, C.A. Rosa. Sugiyamaella xylanicola sp. nov., a xylan-degrading yeast species isolated from rotting wood. Int. J. Syst. Evol. Microbiol. 63 (2013) 2356–2360, http://dx.doi.org/10.1099/ijsem.0.050856-0.

[4] A.V. Zimin, G. Marçais, D. Puiu, M. Roberts, S.L. Salzberg, J.A. Yorke, The MaSuRCA genome assembler. Bioinformatics 29 (2013) 2669–2677, http://dx.doi.org/10.1093/bioinformatics/btt478.

[5] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, AS. Kulikov, VM. Lesin, S.I. Nikolenko, S. Pham, AD. Prjibelski, AV. Pyshkin, AV. Sirotkin, V.Ya. Vyahhi, G. Tesler, MA. Alekseyev, P.A. Pavlnin, SPAdeS: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (2012) 455–477, http://dx.doi.org/10.1089/cmb.2012.0021.

[6] G. Parra, K. Bradnam, I. Korf, CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23 (2007) 1061–1067, http://dx.doi.org/10.1093/bioinformatics/btm071.

[7] F.A. Simão, RM. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31 (2015) 3210–3212, http://dx.doi.org/10.1093/bioinformatics/btv351.

[8] A. Gurevich, V. Saveliev, V. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies. Bioinformatics 29 (2013) 1072–1075, http://dx.doi.org/10.1093/bioinformatics/bts386.

[9] C. Holt, M. Yandell, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinf. 12 (2011) 491, http://dx.doi.org/10.1186/1471-2105-12-491.

[10] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Keeler, TL. Madden, BLAST + architecture and applications. BMC Bioinf. 10 (2009) 421, http://dx.doi.org/10.1186/1471-2105-10-421.

[11] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25 (1997) 955–964.

[12] B.C.H. Chu, H. Lee, Genetic improvement of Saccharomyces cerevisiae for xylose fermentation. Biotechnol. Adv. 25 (2007) 425–441, http://dx.doi.org/10.1016/j.biotechadv.2007.04.001.

[13] E.V. Kriventseva, F. Tegenfeldt, T.J. Pettry, RM. Waterhouse, F.A. Simão, L.A. Pozdnayakov, P. Ioannidis, E.M. Zdobnov, OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res. 43 (2015) D250–D256, http://dx.doi.org/10.1093/nar/gku1220.

[14] P. Jones, B. Binn, H.Y. Chang, M. Fraser, W.I. C. McNaula, H. McWilliam, M. Maslen, A. Mitchell, G. Nuka, S. Peseat, A.F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (2014) 1236–1240, http://dx.doi.org/10.1093/bioinformatics/btu231.

---

Table 1

| No. of contigs | Total length | Length of longest contig | Mean length | N50 | GC content | Completeness by CEGMA* | (No. of core genes/% completeness) | Completeness by BUSCO* | (No. of core genes/% completeness) | No. of predicted genes | Blastx × nr hits |
|----------------|-------------|--------------------------|-------------|-----|------------|------------------------|-------------------------------|------------------------|-------------------------------|------------------------|-------------------|
| 1251           | 13,714,239 bp | 638,759 bp               | 10,962 bp   | 180,392 | 33.8%      | 240/96.77%              | 1308/90%                      | 5971                   | 5638                      |                        |                   |