FEMS Yeast Research, 19, 2019, foz034

doi: 10.1093/femsyr/foz034
Advance Access Publication Date: 26 April 2019
Research Article

RESEARCH ARTICLE

Physiological and comparative genomic analysis of new isolated yeasts *Spathaspora* sp. JA1 and *Meyerozyma caribbica* JA9 reveal insights into xylitol production

Débora Trichez[1], Andrei S. Steindorff[1,†], Carlos E. V. F. Soares[1,2], Eduardo F. Formighieri[1] and João R. M. Almeida[1,2,*]

[1]Embrapa Agroenergia. Parque Estação Biológica, PqEB – W3 Norte Final, Postal code 70.770–901, Brasília-DF, Brazil and [2]Graduate Program in Chemical and Biological Technologies, Institute of Chemistry, University of Brasília, Campus Darcy Ribeiro, Postal code 70.910-900, Brasília-DF, Brazil

[*]**Corresponding author:** Embrapa Agroenergia, Parque Estação Biológica, PqEB – W3 Norte Final - s/n°, 70.770-901 - Brasília, DF – Brasil. Tel: +55 61 3448-2337; E-mail: joao.almeida@embrapa.br

[†]Present address: U.S. Department of Energy (DOE) Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, US.

One sentence summary: Physiological and comparative genomic analysis of two novel xylose-assimilating yeast strains isolated from Brazilian Cerrado ecosystem.

Editor: Pascale Daran-Lapujade

ABSTRACT

Xylitol is a five-carbon polyol of economic interest that can be produced by microbial xylose reduction from renewable resources. The current study sought to investigate the potential of two yeast strains, isolated from Brazilian Cerrado biome, in the production of xylitol as well as the genomic characteristics that may impact this process. Xylose conversion capacity by the new isolates *Spathaspora sp*. JA1 and *Meyerozyma caribbica* JA9 was evaluated and compared with control strains on xylose and sugarcane biomass hydrolysate. Among the evaluated strains, *Spathaspora* sp. JA1 was the strongest xylitol producer, reaching product yield and productivity as high as 0.74 g/g and 0.20 g/(L.h) on xylose, and 0.58 g/g and 0.44 g/(L.h) on non-detoxified hydrolysate. Genome sequences of *Spathaspora* sp. JA1 and *M. caribbica* JA9 were obtained and annotated. Comparative genomic analysis revealed that the predicted xylose metabolic pathway is conserved among the xylitol-producing yeasts *Spathaspora* sp. JA1, *M. caribbica* JA9 and *Meyerozyma guilliermondii*, but not in *Spathaspora passalidarum*, an efficient ethanol-producing yeast. Xylitol-producing yeasts showed strictly NADPH-dependent xylose reductase and NAD^+-dependent xylitol-dehydrogenase activities. This imbalance of cofactors favors the high xylitol yield shown by *Spathaspora* sp. JA1, which is similar to the most efficient xylitol producers described so far.

Keywords: *Spathaspora*; xylitol; xylose fermentation; bagasse hydrolysate; comparative genomics; *Meyerozyma*

Received: 17 August 2018; **Accepted:** 25 April 2019

INTRODUCTION

Xylose is an abundant five-carbon sugar present in lignocellulosic biomass, that can be employed as raw material for the production of a range of biotechnological products, including ethanol, propanediol, glycolic acid and xylitol (Karhumaa *et al.* 2007; Albuquerque *et al.* 2014; Alkim *et al.* 2016; Xin *et al.* 2016; Li *et al.* 2017). Xylitol is a polyol of significant interest in the dental, pharmaceutical and food industries, due to its high anticariogenic and sweetening proprieties, as well as other characteristics such as high solubility, stability and low glycemic index (Hyvonen, Koivistoinen and Voirol 1982; Ravella *et al.* 2012; Canilha *et al.* 2013; Albuquerque *et al.* 2014; Pal, Mondal and Sahoo 2016). At the industrial scale, xylitol is currently produced by chemical hydrogenation of purified xylose obtained from hemicellulosic hydrolysates at high temperature and pressure, using nickel metal as catalyst. The production process, and need for several purification steps make the chemical synthesis route very laborious and expensive (Hyvonen, Koivistoinen and Voirol 1982; Ravella *et al.* 2012). The xylitol market is expanding and the demand for this product is estimated at 125 000 tons per year worldwide (Albuquerque *et al.* 2014). In this context, microbiological fermentation processes represent an interesting alternative for lower cost xylitol production.

Several microorganisms, including bacteria, yeasts and filamentous fungi, are able to convert xylose to xylitol. Among those, yeasts have been reported to be the most efficient xylitol producers (Albuquerque *et al.* 2014; Pal, Mondal and Sahoo 2016). Previous studies involving approximately 40 species, belong to 18 different yeast genera, have identified *Candida guilliermondii* (currently *Meyerozyma guilliermondii* (Kurtzman and Suzuki 2010)) and *Candida tropicalis* as prominent xylitol producers, with production yields up to 0.59 g/g and 0.67 g/g, respectively (Barbosa *et al.* 1988; Guamán-Burneo *et al.* 2015). Recently, new species described as *Cyberlindnera galapagoensis*, *Scheffersomyces amazonensis* and *Spathaspora* spp. strains have shown a significant ability to produce xylitol from xylose, achieving xylitol yields of 0.50 g/g, 0.75 g/g and 0.13–0.50 g/g of xylose, respectively (Cadete *et al.* 2012, 2016a ,2016b; Guamán-Burneo *et al.* 2015; Lopes *et al.* 2016). However, optimization of fermentative conditions such as aeration, biomass concentration, media pH and supplementation with nutrients, is necessary in order to obtain high xylitol yields and productivities when biomass hydrolysates are used as a substrate (Rao *et al.* 2006; de Arruda *et al.* 2011; Hernández-Pérez *et al.* 2016; Martini *et al.* 2016; Dalli *et al.* 2017).

Genetic improvement of yeast strains can be facilitated by the availability of genetic and physiologic information about native producing strains. Recently, genome sequencing of several yeasts has been reported, including for *Spathaspora passalidarum* (Wohlbach *et al.* 2011), *Scheffersomyces stipitis* (Jeffries *et al.* 2007), *Candida tenuis* (Wohlbach *et al.* 2011) and *M. guilliermondii* (Butler *et al.* 2009). The genus *Spathaspora* has been described recently (Nguyen *et al.* 2006), and only a few species have been identified of which only a few genomes are available (Lobo *et al.* 2014; Lopes *et al.* 2017). Moreover, among these, only *S. passalidarum* presents a high quality genome annotation (Wohlbach *et al.* 2011). Thus, physiological and genomic data from different yeast species may allow faster development of biological processes, such as ethanol or xylitol production.

In the current work, two natural yeast strains isolated from the Brazilian Cerrado biome were physiologically and genetically characterized for xylitol production. Evaluation of the strains under oxygen-limited conditions, using xylose and a glucose-xylose mixture as carbon sources, demonstrated that they are capable of producing xylitol with significant yields and productivity. Besides, the ability of these yeasts to convert sugar in non-detoxified hydrolysate was also compared with control strains of *S. passalidarum* NRRL Y-27 907, *M. guilliermondii* NRRL Y-324 and *C. tropicalis* CMAA1716. Through genome sequencing and phylogenetic analysis, the strains were placed in the CUG-Ser clade, more specifically to *Spathaspora* and *Meyerozyma* genera. *De novo* genome sequencing allowed identification of genes implicated in xylose metabolism, such as the genes encoding the xylose catabolic pathway enzymes xylose reductase (XR), xylitol dehydrogenase (XDH), xylulokinase and others. Results of enzymatic activity measurements in crude cell extracts for the different yeasts were used to support fermentation and genomic data.

MATERIALS AND METHODS

Yeast strains and cultivation conditions

The yeast strains *Spathaspora* sp. JA1 and *Meyerozyma caribbica* JA9 were previously isolated from decaying wood samples collected in Brasilia, Distrito Federal, Brazil (S15°35'36.1" WO47°44'28.2") by their capacity to grow on xylose (unpublished data). Strains *M. guilliermondii* NRRL Y-324 and *S. passalidarum* NRRL Y-27 907 were kindly provided by the ARS (NRRL) culture collection (Peoria, USA). *Candida tropicalis* CMAA1716 strain was previously isolated by our group (unpublished data).

YPD agar plates (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose and 20 g/L agar) were used to recover yeast cells from stocks stored at −80°C before starting the experiments. Growth and fermentation assays were carried out in defined mineral medium as described previously (Verduyn *et al.* 1985). Briefly, the defined mineral medium contained 12.5 g/L $(NH_4)SO_4$, 1.25 g/L $MgSO_4$. $7H_2O$, 7.5 g/L KH_2PO_4, vitamins, trace elements and D-xylose or a mixture of D-glucose and D-xylose as the carbon source. Filter-sterilized vitamins, trace elements and sugar solution were added to the media after heat sterilization.

Fermentation under oxygen-limited condition

Strains inoculated on YPD plates were transferred to 100 mL of defined mineral medium supplemented with 40 g/L D-xylose or a mixture of 20 g/L D-xylose and 20 g/L D-glucose, in a rotary shaker at 28°C and 200 rpm. Pre-cultures in the exponential phase were harvested by centrifugation at 6000 x *g* for 5 min (Heraeus Megafuge 16R Centrifuge, Thermo Fisher Scientific), suspended in 10 mL of the fermentation medium, and used to inoculate at an initial OD_{600} ~0.5 (< 0.3 g/L) in an INFORS bioreactor (INFORS HT Multifors) with a working volume of 1 L. The composition of the fermentation medium was similar to the preculture medium, containing in addition 10 mg/L ergosterol, 420 mg/L Tween-80, a few drops of polypropylene glycol as antifoaming agent and 40 g/L D-xylose or a mixture of 20 g/L D-xylose and 20 g/L D-glucose. Air flow was kept at 0.1 L min^{-1} in the bioreactor and pH was maintained at 5.5 by addition of 3 M KOH. Temperature was set to 28°C, and the stirrer speed was 400 rpm. Cell densities were determined by absorbance measurements at 600 nm and correlated to cell dry weight. For dry weight measurements, 5 mL samples of cell suspension were centrifuged, washed with distilled water and dried for 48 h at 65°C. The measurements were carried out in duplicate and cell weight was calculated as the weight difference between samples with and without cells. In addition, samples were taken for further analysis of metabolites. All fermentation assays were performed in two biological replicates.

Fermentation of sugarcane biomass hydrolysate

Sugarcane biomass hydrolysate was kindly provided by the Sugarcane Technology Center, São Paulo, Brazil. It was prepared by steam explosion followed by acid hydrolysis of the hemicellulosic fraction, and the composition was determined to be 1.8 g/L glucose, 27.8 g/L xylose, 0.7 g/L cellobiose, 6.0 g/L acetic acid, 0.66 g/L furfural and 0.11 g/L hydroxymethylfurfural (HMF). Inoculum cultures were grown in YP medium containing 40 g/L D-xylose in Erlenmeyer flasks, in a rotary shaker at 28°C and 200 rpm, for approximately 24 h. Cells in the exponential phase were recovered by centrifugation at 4000 x *g* for 10 min (Heraeus Megafuge 16R Centrifuge, Thermo Fisher Scientific) and suspended in fermentation medium to reach ∼10 g/L of initial biomass. Fermentation medium was composed of sugarcane biomass hydrolysate supplemented with 2 g/L YNB, 4 g/L urea, and glucose and xylose to obtain a final concentration of 8 g/L and 40 g/L, respectively. The pH value was adjusted to 5.5 with KOH. Flasks containing 50 mL culture were incubated at 28°C and 200 rpm. Samples were taken during fermentation for biomass and metabolite quantification. The experiments were performed in biological duplicates.

Analytical methods

Samples collected from the fermentation cultures were centrifuged at 10 000 x *g* for 10 min in a benchtop centrifuge (Eppendorf Mini Spin Plus), and the supernatants were stored at −20°C until further analysis. Extracellular metabolite quantification was performed by high performance liquid chromatography on a Waters System (Waters Acquity UPLC H Class, USA) equipped with an Aminex HPX-87H column (300 × 7.8 mm, 9 μm, Bio-rad) coupled to a RI detector. Five mM H_2SO_4 was used as the mobile phase at a flow rate of 0.6 mL/min. Column temperature was maintained at 45°C. Calibration curves were used to quantify the following metabolites: glucose, xylose, xylitol, glycerol, acetate and ethanol.

Enzymatic activity assays

For the enzymatic activity measurements, yeasts were grown in YP medium containing 40 g/L xylose as a carbon source. After 20 h of cultivation, cells were collected, washed with sterile water and used to prepare crude extracts with Y-PER® – Yeast Protein Extraction Reagent (Pierce, Rockford, USA) and glass beads. Protein concentration in the cell-free preparations was determined using Quick Start Bradford Protein Assay Kit (Bio-Rad Laboratories Ltd., USA), following the manufacturer's instructions. XR and XDH activities were measured in the cell-free extracts as described previously (Smiley and Bolen 1982; Almeida *et al.* 2008; Veras, Parachin and Almeida 2017). The XR reaction mixture contained 100 mM triethanolamine buffer (pH 7.0), 0.2 mM NADH or NADPH and 350 mM xylose. Xylitol oxidation (XDH activity) was determined in a reaction mixture containing 100 mM triethanolamine buffer (pH 7.0), 0.3 mM NAD^+ and 300 mM xylitol. Reduction rates of furfural and HMF were determined in 100 mM phosphate buffer (pH 7.0), 0.1 mM NADH or NADPH and 10 mM of furfural or HMF (Almeida *et al.* 2008). All reactions were performed at 30°C and started with the addition of the respective substrates. Oxidation or reduction of the coenzymes was monitored as the change in absorbance at 340 nm in 1.0 cm pathlength cuvettes. A value of 6.22 mL $(\mu mol\ cm)^{-1}$ was used as molar absorption coefficient of NADH and NADPH. One enzyme unit was defined as the amount of enzyme required to oxidize or reduce 1 μmol of cofactor per minute. Specific activities were given in units per mg of protein (U/mg). The experiment was performed in biological duplicates.

Yeast identification

Yeast identification was performed by analysis of the D1/D2 domain and the ITS-5.8S region from the large-subunit of the rRNA gene, using polymerase chain reaction as previously described (White *et al.* 1990; Kurtzman and Robnett 1998). The D1/D2 domain was amplified using primers NL1 (5′-GCATATCAATAAGCGGAGGAAAAG-3′) and NL4 (5′-GGTCCGTGTTTCAAGACGG-3′) and the ITS region was amplified using primers ITS-1 (5′-TCCGTAGGTGAACCTGCGG-3′) and ITS-4 (5′-TCCTCCGCTTATTGATATGC-3′), with genomic DNA as template. The amplified DNA was purified and sequenced (Eurofins Genomics, in Louisville, KY, USA). The partial sequences of the D1/D2 domain and ITS region have been deposited at DDBJ/EMBL/GenBank under accession numbers MK193860 and MK193858 (JA1), and MK193861 and MK193859 (JA9), respectively.

Phylogenetic analysis

The sequences for the phylogenetic tree were selected based on similarity of the D1/D2 and ITS regions downloaded from the YeastIP database (Weiss *et al.* 2013). To build the tree, all sequences were aligned using MAFFT v7.245 (Katoh and Standley 2013) with the option '–auto' for each region separately, then concatenated for phylogeny using the FastTree program (Price, Dehal and Arkin 2009).

Genome sequencing, assembly and annotation

Genomic DNA was isolated from yeast cells grown on 5 mL YPD, at 28°C for 16 h, using FastDNA® SPIN Kit for Soil, following the manufacturer´s instructions (MP Biomedicals, LLC, Santa Ana, CA, USA). Paired-end Illumina sequencing was performed by Eurofins (www.eurofins.com.br) using a fragmented library for Miseq (2 × 250 bp) and Hiseq2000 3 kb jumping library (2 × 125 bp). Sequencing data of the two strains are available from the NCBI Sequence Read Archive: Bioproject accession PRJNA385175. These Whole Genome Shotgun projects have been deposited at DDBJ/EMBL/GenBank under accession numbers NSGQ00000000 (*Spathaspora* sp. JA1) and NSDU00000000 (*M. caribbica* JA9).

Fragment library and Jumping library paired-end Illumina reads were used as input for AllPaths-LG assembler (Gnerre *et al.* 2011), using maximum coverage of 80X for each library. Genome structural and functional annotations were performed with the MAKER pipeline (Holt and Yandell 2011) using the GeneMark-ES (Ter-Hovhannisyan *et al.* 2008) Augustus (Stanke *et al.* 2008), and SNAP (Korf 2004) predictors. Prediction of functional domains (PFAM, InterPro and Gene Ontology) from gene models was performed using InterProScan v.5.21.60 (Finn *et al.* 2017). Completeness of each genome was assessed using BUSCO v2.0.1 (Simão *et al.* 2015) based on the Saccharomycetales dataset.

Specific classes of enzymes, such as carbohydrate-active enzymes (CAZymes) (Lombard *et al.* 2014), were predicted based on dbCAN HMMs (Yin *et al.* 2012); transporters were predicted based on Transporter Classification DataBase (http://www.tcdb.org/browse.php), and transcription factors based on DBD—Transcription factor prediction database (http://www.transcriptionfactor.org/index.cgi?Home). Transporters and transcription factors were predicted using blastp with e-value $< 10^{-10}$.

Genome similarity and synteny analysis

Orthologous clustering and Venn diagram construction were performed using OrthoVenn (Wang *et al.* 2015). Predicted protein datasets from each genome were used as input for OrthoMCL (Li 2003) with minimum similarity of 10^{-5}, and 2 as inflation value. Synteny analysis of genomes, *XYL1* and *XYL3* genes, were performed using the SyMAP v4.2 software package (Soderlund, Bomhoff and Nelson 2011) in order to detect synteny blocks for each pair of species: *Spathasphora* sp. JA1/*S. passalidarum*, *M. caribbica* JA9/*M. guilliermondii* and *C. tropicalis*. In order to better display the *XYL1* and *XYL3* synteny between yeasts and generate figures, the same data were used as input for mGSV (Revanna *et al.* 2011).

RESULTS AND DISCUSSION

Yeast identification

Yeast strains were previously isolated according to their ability to grow on xylose from decaying wood samples obtained in the Brazilian Cerrado. Growth profiles of the new isolated strains were evaluated on minimal media supplemented with xylose as the sole carbon source (unpublished data), and based on these results, strains JA1 and JA9 were selected for xylitol production. Analyses of the D1/D2 domains and ITS-5.8S region of the large-subunit rRNA gene revealed that the new isolates JA1 and JA9 belong to the *Spathaspora* and *Meyerozyma* clades respectively (Fig. 1). JA1 showed an identical D1/D2 domain and ITS region differing by only 14 nucleotides compared to the new species represented by *Spathaspora* sp. nov. UFMG-CM-Y426 (accession number KX101231.1 for ITS and KX097022.1 for D1/D2). Moreover, the D1/D2 sequence from the JA1 strain differed by just four nucleotide substitutions from its closest described relative *Candida materiae* CBS 10975^T (Barbosa *et al.* 2009) (Fig. 1). The JA9 strain showed an identical D1/D2 domain to that of *M. caribbica* CBS 2022 and *M. guilliermondii* ATCC 6260^T (Vaughan-Martini *et al.* 2005; Kurtzman and Suzuki 2010), whereas analysis of the ITS region showed a higher similarity with *M. caribbica* CBS 2022 (4 nucleotide substitutions). Based on these data, the JA9 strain was identified as *M. caribbica* JA9, and JA1 was treated as undescribed species and designated as *Spathaspora* sp. JA1, until further taxonomic experiments can be performed.

A phylogenetic tree based on concatenated alignment of ITS and D1/D2 sequences showed the placement of the new isolates into the independent clades *Spathaspora* and *Meyerozyma* (Fig. 1). The *Spathaspora* genus was previously described by Nguyen *et al.* (2006) and contains D-xylose-fermenting yeasts generally associated with rotting wood substrates and insects (Nguyen *et al.* 2006; Cadete et al., 2009, 2012, 2013, 2016a; Cadete and Rosa 2017). The *Meyerozyma* genus was proposed by Kurtzman and Suzuki (2010) to accommodate yeasts such as *M. guilliermondii* and *M. caribbica*, species that are widely dispersed in natural environments (Vaughan-Martini *et al.* 2005; Kurtzman and Suzuki 2010; Romi *et al.* 2014; Corte *et al.* 2015). Both genera belong to the CUG-Ser clade, as detected by the Bagheera web server (Muhlhausen and Kollmar 2014), which uses an alternative yeast nuclear code to translate CUG as serine instead of leucine (Riley *et al.* 2016). This is in agreement with previous work (Wohlbach *et al.* 2011) that described the phylogenetic placement of xylose-assimilating yeasts within the CUG clade.

Genome features

To better understand the physiological profile of *Spathaspora* sp. JA1 and *M. caribbica* JA9, their genetic backgrounds related to xylose metabolism were investigated. Genomes of both strains were sequenced and assembled (Table 1). The genera *Spathaspora* and *Meyerozyma* have been recently described (Nguyen *et al.* 2006; Kurtzman and Suzuki 2010) and the only genomes that are completely annotated with high quality are from *S. passalidarum* NRRL Y-27 907 and *M. guilliermondii* ATCC 6260 (Butler *et al.* 2009; Wohlbach *et al.* 2011). Considering this aspect, a comparative analysis was performed between *Spathaspora* sp. JA1 and *S. passalidarum* and *M. caribbica* JA9 and *M. guilliermondii*. The sequenced genome size of *Spathaspora* sp. JA1 was estimated at 14.5 Mb, consisting of 181 scaffolds and a total of 5268 predicted genes, while a 10.4 Mb genome was obtained from the *M. caribbica* JA9 strain, containing 5033 predicted genes over 27 scaffolds. The JA1 and JA9 strains showed similar features (genome size, gene content and percentage GC.) when compared with the respective reference strains *S. passalidarum* and *M. guilliermondii*. Assessment of genome assembly completeness was performed using BUSCO with the Saccharomycetales dataset. The *Spathaspora* sp. JA1 and *M. caribbica* JA9 genomes showed high percentages of genome completeness (both with 97.3% complete BUSCOs). General features of the yeast genomes are presented in Table 1.

A comparative analysis was performed on predicted protein dataset from each genome as described in Table 1. The orthologous cluster distribution (Fig. 2) shows that 2785 protein clusters are common for all five genomes and 1001 are shared for the xylose-consuming yeasts *Spathaspora* sp. JA1, *S. passalidarum*, *M. caribbica* JA9 and *M. guilliermondii*, but are not found in the *S. cerevisiae* genome. Some enriched gene ontology terms shared between the four xylose-consuming strains are related to NADH dehydrogenase (ubiquinone) activity, lipase activity, filamentous growth, cell surface proteins, cellular response to nutrients, cellulose catabolic process and beta-glucosidase activity. Examination of PFAM domains revealed an increased occurrence of putative proteins related to membrane transport (PF07690, PF00083) and epimerase/oxidoreductase activities (PF01370) for the xylose-assimilating yeasts evaluated. A total of 482 gene clusters were found specifically in both *Spathaspora* strains, including predicted proteins involved in acetate transport and oxidoreductase activity. 554 genes were uniquely present in the *M. caribbica* JA9 and *M. guilliermondii* species, among which were several putative transmembrane transport proteins. Amplification of number of genes encoding putative sugar transporters and oxidoreductase activities suggested the relevance of these genes for yeast adaptation to their specific environments, rich in sugars, as previously noted for other xylose-assimilating yeasts (Wohlbach *et al.* 2011). Similarly, amplification of number of genes related to cell surface proteins and lipases was previously associated with pathogenicity of some *Candida* strains, as well commensalism in the case of beetle symbionts (Butler *et al.* 2009; Wohlbach *et al.* 2011).

The genome of the *Spathaspora* sp. JA1 and *M. caribbica* JA9 strains contained 186 and 169 genes encoding putative CAZymes (Lombard *et al.* 2014), respectively. Although predictive CAZymes are similar between *M. caribbica* JA9 and *M. guilliermondii* ATCC 6260, the comparison reveals an enrichment of predicted esterases (CE1), oxidases (AA4) and the presence of a putative polysaccharide lyase (PL24) in the novel isolated strains. In addition, *Spathaspora* sp. JA1 possesses an expansion

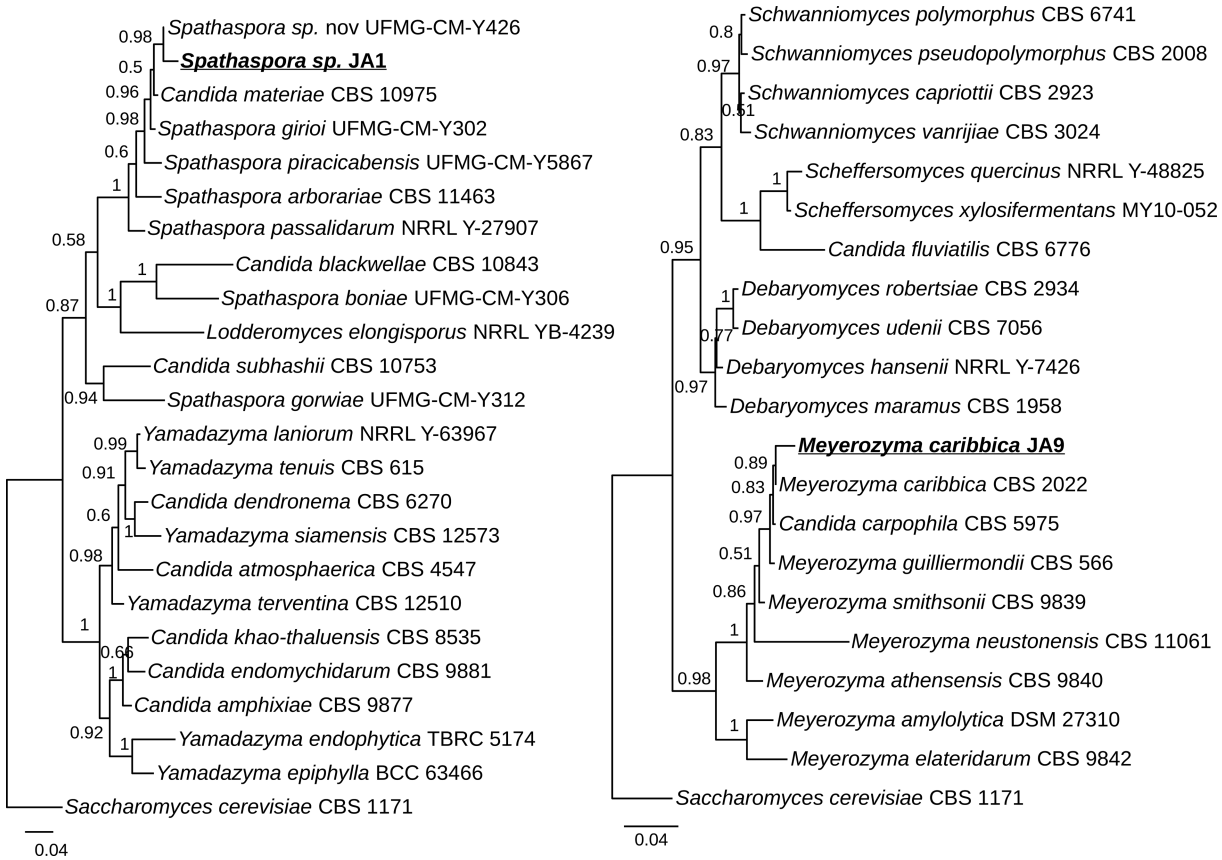

Figure 1. Phylogenetic placement of *Spathaspora* sp. JA1 and *M. caribbica* JA9 based on ITS regions and the D1/D2 domains of the LSU rRNA gene. Sequences were aligned using MAFFT and phylogenetic tree inferred using FastTree. Support values are shown next to the respective node. Bar shows number of substitutions per nucleotide position. All represented strains are type strains, with exception of *Spathaspora* sp. JA1, *Spathaspora* sp. nov UFMG-CM-Y426 and *M. caribbica* JA9.

Table 1. Yeast genomes features.

	Spathaspora sp. JA1	*S. passalidarum* NRRL Y-27 907	*M. caribbica* JA9	*M. guilliermondii* ATCC 6260	S. cerevisiae S288C
Total length (Mb)	14.5	13.2	10.4	10.6	12.7
Scaffolds	181	9	27	8	16
N50 (Mb)	0.21	2.07	1.26	1.7	0.92
GC%	34.6	37.1	47.1	43.8	38.3
Number of gaps	77	62	27	19	0
Percentage of genome in gaps	0.34	0.78	0.04	0.33	0
Gene content	5268	5983	5033	5920	6575
BUSCO completeness	97.3%	91.70%	97.3%	91.30%	100%

of predicted chitinases (CBM18, CBM50) and glycoside hydrolases (GH109, GH18, GH3) when compared to the closely related species. Additionally, all xylose-assimilating strains showed a diversified repertoire of predicted glycoside hydrolases, including GH3, GH2, GH20 and GH109, which could contribute to the degradation of xylan, giving them an advantage in their natural environments (Riley *et al.* 2016).

Orthologous genes related to xylose assimilation in yeasts are present in *Spathaspora* sp. JA1 and *M. caribbica* JA9. A large number of genes encoding putative proteins involved in the uptake of different kinds of substrates, such as carbohydrates, amino acids and peptides were detected; interestingly, among them, the Major Facilitator Superfamily was notable. Many genes were predicted to encode transporters similar to the characterized xylose transporters Xut1–7p, Sut1p and Hxt2.6p from *S. stipitis* and *S. passalidarum* (Jeffries *et al.* 2007; Wohlbach *et al.* 2011; Young *et al.* 2011; de Sales *et al.* 2015). Another important factor in xylose metabolism are the enzymes responsible for xylose assimilation. In yeasts, D-xylose is reduced to xylitol by a XR (Xyl1p encoded by *XYL1*), which is, in general, a NADPH-dependent enzyme. Subsequently, xylitol can be oxidized to D-xylulose by a XDH NAD^+-dependent enzyme (Xyl2p encoded by *XYL2*), which enters the Pentose Phosphate Pathway and Glycolysis after phosphorylation to xylulose-5P by the xylulokinase enzyme (Xyl3p encoded by *XYL3* or *XKS1*) (Bruinenberg *et al.* 1983; Verduyn *et al.* 1985; Karhumaa *et al.* 2007; Moysés *et al.*

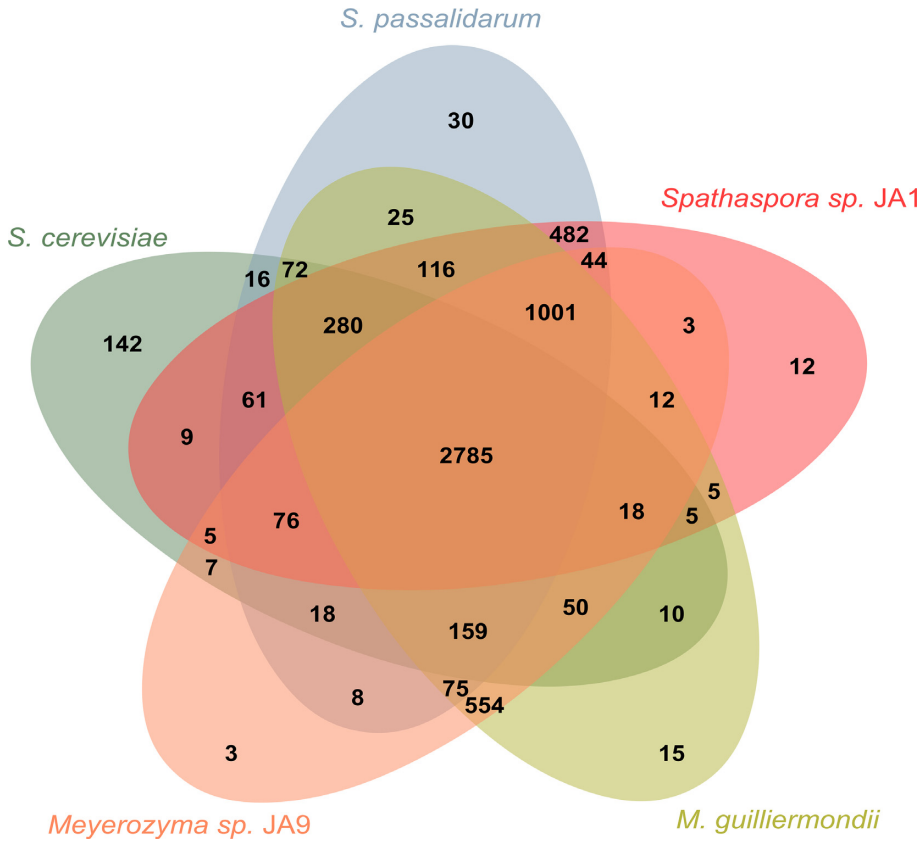

Figure 2. Distribution of orthologous clusters from *Spathaspora* sp. JA1, *M. caribbica* JA9, *M. guilliermondii*, *S. passalidarum* and *S. cerevisiae*.

2016). The preference for different co-factors in the first two reactions causes a redox imbalance that has a significant impact on xylose fermentation and xylitol production (Bruinenberg *et al.* 1983; Albuquerque *et al.* 2014; Veras, Parachin and Almeida 2017).

Xylose assimilation

Initial screening showed that the *Spathaspora* sp. JA1 and *M. caribbica* JA9 strains can consume D-xylose and produce xylitol. To better evaluate the yeast physiology, fermentation assays were performed using the new isolated yeasts and two other known yeast strains, *S. passalidarum* NRRL Y-27 907 and *M. guilliermondii* NRRL Y-324. These two strains were selected as controls based on their high production yields of ethanol (poor xylitol producer), and xylitol, respectively (Barbosa *et al.* 1988; de Arruda *et al.* 2011; Hou 2012; Cadete *et al.* 2016a). Fermentations were carried out in bioreactors, under oxygen-limited conditions, using 4% xylose or a mixture of 2% xylose and 2% glucose as substrates. Both JA1 and JA9 were able to use glucose and xylose as a carbon source for growth (Fig. 3). However, ethanol production was observed only when glucose was present in the medium (Fig. 3B,D,F,H). Among the microorganisms evaluated here, only *S. passalidarum* was able to ferment both substrates efficiently and produce high quantities of ethanol (ethanol yield on xylose, $Y_{et} = 0.48 \pm 0.01$, and on sugar mixture, $Y_{et} = 0.43 \pm 0.01$).

In the xylose assimilation assays, *S. passalidarum* showed the highest xylose consumption rate. This strain consumed all sugar in approximately 50 h, with a specific xylose consumption of 0.138 g/g.h, whereas *Spathaspora* sp. JA1, *M. caribbica* JA9 and *M. guilliermondii* showed xylose consumption rates of 0.082, 0.036 and 0.029 g/g.h respectively, and residual sugar was present at the end of cultivation. Xylitol was the major product obtained from the isolated yeasts, whereas acetate and glycerol were produced in minor concentrations (below 1.0 g/L) (Fig. 3A and C). At the end of cultivation, *Spathaspora* sp. JA1 produced 22.62 ± 6.33 g/L of xylitol, which is almost two times higher than the amount produced by *M. caribbica* JA9, the second most efficient producing strain (Table 2). In addition, xylitol yields were significantly higher for *Spathaspora* sp. JA1 (0.75 ± 0.01 g/g) and *M. caribbica* JA9 (0.54 ± 0.11 g/g) than for the control strain *M. guilliermondii* (0.44 ± 0.02 g/g) (Table 2), which is recognized as a strong xylitol producer (Barbosa *et al.* 1988; de Arruda *et al.* 2011; Cadete *et al.* 2016a). Previous studies have already reported that different species of *Spathaspora* are able to assimilate xylose with different efficiencies (Nguyen *et al.* 2006; Cadete et al., 2009, 2012, 2016a; Lopes *et al.* 2016). Although some strains, such as *S. passalidarum, S. arborariae* and *S. xylofermentans* mainly ferment xylose and produce ethanol under oxygen-limited conditions, others convert xylose to xylitol (Table 2). Among these, the best xylitol yields reported so far were 0.56 g/g xylose for the *S. roraimanensis* and 0.47 g/g xylose for *S. brasiliensis* (Table 2) (Cadete *et al.* 2016a). As such, our findings indicate that the *Spathaspora* sp. JA1 strain is a promising candidate for xylitol production. Its high xylitol yield is close to the values obtained from the best xylitol producers identified so far, such as *C. tropicalis*, *M. guilliermondii* FTI20037 and *S. amazonensis* (Table 2), even without employing optimized process conditions.

In the co-fermentation assays, xylitol production decreased slightly for the three xylitol producer strains *Spathaspora* sp. JA1, *M. caribbica* JA9 and *M. guilliermondii*, whereas ethanol production increased, especially during the glucose consumption

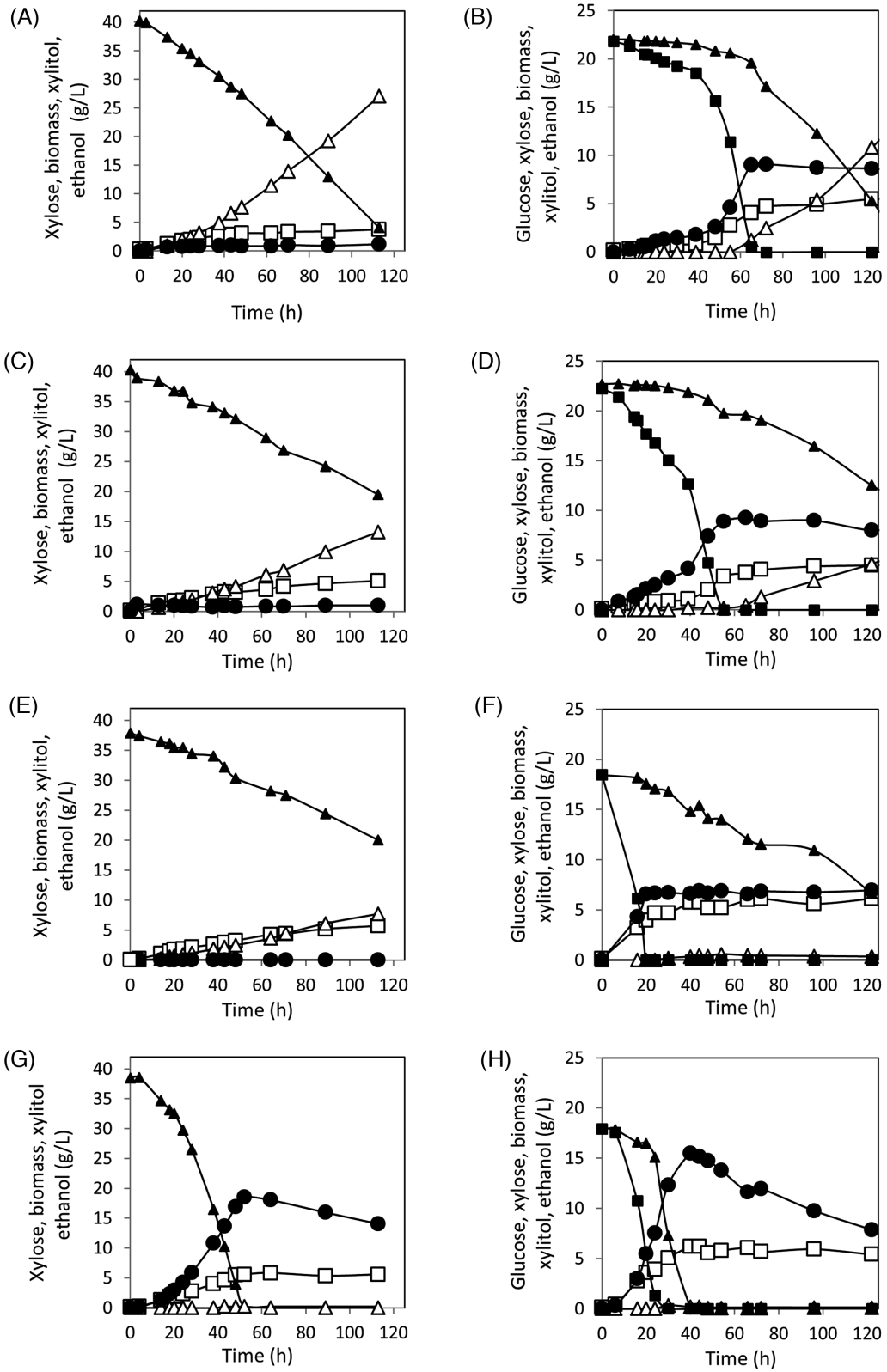

Figure 3. Sugar consumption and product formation by the yeasts *Spathaspora* sp. JA1 **(A)** and **(B)**, *M. caribbica* JA9 **(C)** and **(D)**, *M. guilliermondii* **(E)** and **(F)** and *S. passalidarum* **(G)** and **(H)** during fermentation under oxygen-limited conditions on xylose (left column) and on glucose and xylose mixture (right column). Xylose (closed triangle); glucose (closed square); xylitol (open triangle); ethanol (closed circle); biomass (open square). Experiments were performed in biological duplicates, and one profile is shown.

Table 2. Production of xylitol by different yeasts.

Microorganism	Experimental conditions	Final xylitol conc.[g/L]	Xylitol yield [g/g]	Productivity [g/(l h)]	References
Spathaspora sp. JA1	Bioreactor mineral medium + 40 g/L xylose low initial cell density	22.62 ± 6.33	0.75 ± 0.01	0.20 ± 0.06	This work
M. caribbica JA9	Bioreactor mineral medium + 40 g/L xylose low initial cell density	11.37 ± 2.60	0.54 ± 0.11	0.10 ± 0.03	This work
S. passalidarum	Bioreactor mineral medium + 40 g/L xylose low initial cell density	0.19 ± 0.06	0.01 ± 0.01	0.003 ± 0.00	This work
M. guilliermondii	Bioreactor mineral medium + 40 g/L xylose low initial cell density	8.25 ± 1.92	0.44 ± 0.02	0.063 ± 0.01	This work
Spathaspora sp UFMG-XMD-16.2	Shake flasks YP + 50 g/L xylose low initial cell density	7.8	0.21	0.16	(Cadete *et al.* 2012)
	Sugarcane bagasse hydrolysate detoxified low initial cell density	18.2	0.57	0.19	
S. brasiliensis UFMG-CM-Y353	Shake flasks YP + 40–50 g/L xylose low initial cell density	11.9	0.47	–	(Cadete *et al.* 2016a)
S. roraimanensis UFMG-CM-Y477	Shake flasks YP + 40–50 g/L xylose low initial cell density	27.4	0.56	–	(Cadete *et al.* 2016a)
S. amazonensis	Shake flasks YP + 50 g/L xylose low initial cell density	34.24	0.75	0.476	(Cadete *et al.* 2016b)
	Rice hull hydrolysate detoxified high initial cell density (DO 600 nm ~10)	6.30	1.04	0.053	
C. guilliermondii FTI20037	Shake flasks sugarcane bagasse hydrolysate detoxified low initial cell density	50.5	0.81	0.60	(de Arruda *et al.* 2011)
C. guilliermondii FTI20037	Shake flasks low initial cell density YNB + 5 g/L urea + xylose 40 g/L	23	0.59	–	(Barbosa *et al.* 1988)
	High initial cell density YNB + 5 g/L urea + 104 g/L xylose	77.2	0.74		
C. tropicalis 1004	Shake flasks low initial cell density YNB + 10 g/L yeast extract + xylose 30 g/L	17	0.57	–	(Barbosa *et al.* 1988)
C. tropicalis	Shake flasks sugarcane bagasse hydrolysate detoxified	–	0.45	–	(Rao *et al.* 2006)
	Corn fiber detoxificated	–	0.43	–	
C. tropicalis MTCC 25 057	Shake flasks YP + 100 g/L of xylose aerobic condition	–	0.67	0.70	(Mattam *et al.* 2016)
Debaryomyces hansenii	Shake flasks high initial cell density MYP + 40 g/L xylose	22.4	0.63	0.23	(Prakash *et al.* 2011)
	MYP + 100 g/L xylose	68.6	0.76	0.44	
	MYP + sugarcane bagasse hydrolysate	–	0.69	–	

phase (Fig. 3). *Spathaspora passalidarum* produced ethanol from xylose and glucose, reaching a production level of 14.94 ± 0.79 g/L. Indeed, *S. passalidarum* is capable of co-fermenting xylose and glucose, producing high concentrations of ethanol (Long *et al.* 2012). *Spathaspora* sp. JA1 still showed the best xylitol-producing performance among the evaluated strains. However, its xylitol yield was decreased by approximately 25% when compared with the results obtained when xylose was used as the sole carbon source. The strains *Spathaspora* sp. JA1, *M. caribbica* JA9 and *M. guilliermondii* showed lower xylose consumption rates (Fig. 3B,D,F) in the glucose-xylose co-fermentation experiments. These results can be explained by the adaptation phase necessary for yeasts to switch from glucose to xylose metabolism (de Arruda *et al.* 2011; Moysés *et al.* 2016; Reider Apel *et al.* 2016).

Fermentation of sugarcane biomass hydrolysate

To evaluate the capacity of the new isolates to ferment biomass hydrolysates, fermentation assays were performed using non-detoxified sugarcane biomass hydrolysate containing glucose (8 g/L) and xylose (40 g/L), with high acetic acid content (6.0 g/L). Acetic acid is a major inhibitory compound in sugarcane biomass hydrolysate (Almeida *et al.* 2007; Deparis *et al.* 2017). The concentration of this compound depends on the feedstock source and the pretreatment method employed (Deparis *et al.* 2017). In addition to acetic acid, the hydrolysate may contain other inhibitors such as furfural and HMF, which combined can affect yeast growth and fermentation (Almeida *et al.* 2008; Deparis *et al.* 2017). Since hydrolysate tolerance is species and strain-dependent (Modig *et al.* 2008), and since many studies have already shown that *C. tropicalis* can produce xylitol with high yields and productivities on fermentative medium containing hydrolysates (Cadete *et al.* 2012; Guamán-Burneo *et al.* 2015; Mattam *et al.* 2016; Dalli *et al.* 2017), a *C. tropicalis* strain isolated by our group (*C. tropicalis* CMAA1716 strain, unpublished data) was used as a control in this experiment.

Under the tested conditions, all strains, except *S. passalidarum*, were able to grow and consume all of the available glucose in the sugarcane biomass hydrolysate, whereas only *C. tropicalis* and *Spathaspora* sp. JA1 were able to consume the available xylose completely (Table 3). *Candida tropicalis* showed the highest xylose consumption rate, followed by *Spathaspora* sp. JA1, *M. caribbica* JA9 and *M. guilliermondii*. *Spathaspora* sp. JA1 and *C. tropicalis* showed the highest xylitol yields (0.58 ± 0.08 and 0.54 ± 0.01, respectively) and final concentration (22.62 ± 0.59 and 23.66 ± 1.52 g/L, respectively) among the five strains evaluated. In turn, xylitol yields for *M. caribbica* JA9 and *M. guilliermondii* were of 0.44 ± 0.06 and 0.42 ± 0.07 g/g xylose, respectively. Ethanol was also one of the main products obtained during the hydrolysate fermentation. The highest ethanol value was obtained for *Spathaspora* sp. JA1 (10.67 g/L), followed by *C. tropicalis* (8.60 g/L), *M. guilliermondii* (8.57 g/L) and *M. caribbica* JA9 (5.69 g/L). *Spathaspora passalidarum* appeared to be the most sensitive to the hydrolysate inhibitors since virtually no growth or sugar consumption were observed during 72 h of cultivation. This result is in agreement with a previous report, which showed delayed sugar consumption by *S. passalidarum* during fermentation of AFEX hydrolysate containing 1.5 g/L of acetic acid (Long *et al.* 2012). It may be an effect of acetic acid since *S. passalidarum* was able to maintain its fermentation performance in lignocellulosic hydrolysates with a lower acetic acid concentration (Long *et al.* 2012; Nakanishi *et al.* 2017).

The *Spathaspora* sp. JA1 performance in defined mineral medium (Table 2; Fig. 3) and non-detoxified biomass hydrolysate (Table 3), suggests that this strain has a great potential as a xylitol producer using renewable resources. Indeed *Spathaspora* sp. JA1 showed similar xylitol production yields to that obtained for *C. tropicalis* evaluated under the same conditions (Table 3). Similar or even higher xylitol yields were observed for other *Spathaspora* spp. (Cadete *et al.* 2012), *S. amazonensis* (Cadete *et al.* 2016b), *C. tropicalis* (Dalli *et al.* 2017) and *C. guilliermondii* (de Arruda *et al.* 2011) yeasts (Table 2). However, this occurred only when employing optimized fermentation parameters and/or a detoxified hydrolysate that leads to reduced concentrations of inhibitors such as acetic acid and others. In this way, the xylitol production by *Spathaspora* sp. JA1 strain may be further improved using optimized process conditions, such as temperature, pH and aeration.

Activity of enzymes related to xylose metabolism and hydrolysate tolerance

The first steps of xylose metabolism in yeast involve reduction of xylose to xylitol by a NAD(P)H-dependent XR, followed by the oxidation of xylitol by a NAD^+-dependent XDH (Verduyn *et al.* 1985; Karhumaa *et al.* 2007). Therefore, the activity levels and cofactor preference of XR and XDH in the new isolated yeasts were determined in crude-cell extracts obtained from *Spathaspora* sp. JA1, *M. caribbica* JA9, *S. passalidarum*, *M. guilliermondii* and *C. tropicalis* cells (Fig. 4A and B). The XDH activity was strictly NAD^+-dependent, whereas the XR activities were NADPH-dependent for most of the yeasts (Fig. 4A and B). Among the analyzed strains, only *S. passalidarum* presented XR activity with both NADH and NADPH as cofactors, showing a ratio of 1.11 NADH/NADPH XR activity. The apparent XR preference for NADH over NADPH was previously observed for *S. passalidarum* (Cadete *et al.* 2016a; Cadete and Rosa 2017; Veras, Parachin and Almeida 2017), and it was correlated with fermentative performance and increased ethanol productivity for this yeast. This is because oxidation of NADH would regenerate NAD^+ used by XDH for xylitol oxidation (Cadete *et al.* 2016a; Cadete and Rosa 2017; Veras, Parachin and Almeida 2017). In contrast, xylitol producers generally have almost exclusively NADPH-dependent XR activities (Lopes *et al.* 2016; Cadete *et al.* 2016a), concordant with the results obtained in this study (Fig. 3, Table 2). Among the xylitol-producing yeasts, *C. tropicalis* showed the highest NADPH-XR-specific activity, which was approximately two times higher than for the other strains. These observations may explain the higher xylose consumption rate and xylitol productivity shown by this yeast in the hydrolysate fermentation. High XR and XDH activities were shown to improve xylose the consumption rate (Karhumaa *et al.* 2007), and the cofactor imbalance between NADPH-dependent XR and NAD^+-dependent XDH favors xylitol secretion (Cadete et al. 2016a, Cadete et al. 2016a, 2016b; Veras, Parachin and Almeida 2017).

In addition to acetic acid, the furaldehydes HMF (5-hydroxymethyl furfural) and furfural (2-furaldehyde) are known to hinder sugar metabolism in yeast (Almeida *et al.* 2007, 2008). While acetic acid tolerance mechanisms are not directly linked with detoxification activities, yeast tolerance to HMF and furfural is achieved through the capacity to reduce these compounds to their corresponding alcohols: HMF-alcohol and furfuryl alcohol. The detoxification activities have been mainly studied in *Saccharomyces*, and are linked to the reduction of the aldehyde group of these compounds by NADH and NADPH-dependent enzymes (Petersson *et al.* 2006; Almeida *et al.* 2008; Modig *et al.* 2008; Deparis *et al.* 2017; Mukherjee *et al.* 2017). Thus, the furaldehyde-reducing activities present in the crude extracts of the selected yeasts using NADH and NADPH as cofactors were measured. All yeast strains were able to reduce both HMF and furfural (Fig. 4C,D). HMF reduction was coupled with both NADH and NADPH in *S. passalidarum*, *Spathaspora* sp. JA1 and *M. caribbica* JA9 cells, while the reduction activity was coupled only with NADPH in *M. guilliermondii* and *C. tropicalis*. However, *S. passalidarum* showed higher NADH-dependent HMF reduction activity than NADPH-dependent, which correlated with its highest level of NADH-dependent XR activity among the evaluated yeasts (Fig. 4). Previously, it was shown that *S. stipitis* XR, which has 75% identity to *S. passalidarum* XR at the amino acid level, is able to reduce HMF (Almeida *et al.* 2008), and these results suggest that like *S. stipitis* XR, the *S. passalidarum* enzyme might also be able to reduce HMF.

Contrary to HMF reduction activity, furfural reduction was exclusively NADH-dependent for all evaluated yeasts (Fig. 4D). *Spathaspora* sp. JA1, *S. passalidarum*, and *M. caribbica* JA9 specific activities were approximately five to ten times higher than the activity detected in *M. guilliermondii* and *C. tropicalis* crude extracts. The innate furaldehyde reductase activities in the evaluated yeasts and their fermentative performances, together with the low concentrations of HMF (0.11 g/L) and furfural (0.66 g/L) present in the hydrolysate employed in this study, indicate that furaldehydes were probably not the main inhibitors during hydrolysate fermentation.

Synteny analysis of genomes and genes involved in xylose metabolism

In order to assess the level of synteny between the strains *S. passalidarum*/*Spathaspora* sp. JA1, *M. guilliermondii*/*M. caribbica* JA9 with the high-xylitol producer *C. tropicalis*, the whole genome sequences were aligned, and a dot plot with the four combinations was generated (Additional file 2). When strains of closely

Table 3. Fermentation parameters of the xylose-fermenting yeasts in sugarcane biomass hydrolysate.

Yeast strain	Sugar consumption (%)[a]		Products (g/L)			Yield (g/g)		Productivity (g/L.h)	
	Glucose	Xylose	Biomass	Xylitol	Ethanol	Xylitol [b]	Ethanol [c]	xol	et
Spathaspora sp JA1	100	99	13.00 ± 1.15	22.62 ± 0.59	10.67 ± 0.73	0.58 ± 0.08	0.23 ± 0.02	0.44 ± 0.01	0.21 ± 0.01
M. caribbica JA9	100	89	20.69 ± 2.59	17.33 ± 0.11	5.69 ± 0.08	0.44 ± 0.06	0.12 ± 0.02	0.33 ± 0.00	0.11 ± 0.00
S. passalidarum	24	3	2.64 ± 0.13	0.22 ± 0.32	0.46 ± 0.46	–	–	–	–
M. guilliermondii	100	75	9.53 ± 2.22	12.95 ± 1.52	8.57 ± 0.46	0.42 ± 0.07	0.22 ± 0.02	0.25 ± 0.03	0.16 ± 0.01
C. tropicalis	100	100	7.58 ± 2.16	23.66 ± 1.52	8.60 ± 0.79	0.54 ± 0.01	0.16 ± 0.01	0.84 ± 0.05	0.31 ± 0.03

The values correspond to averages and standard deviations of two biological replicates, calculated over 64 h of fermentation, except for *C. tropicalis* (28 h).
[a]Percentage of initial sugar consumed (%).Initial glucose and xylose concentrations were approximately 8 and 40 g/L, respectively.
[b]Xylitol yield: correlation between xylitol produced and xylose consumed.
[c]Ethanol yield: correlation between ethanol produced and sugar consumed.

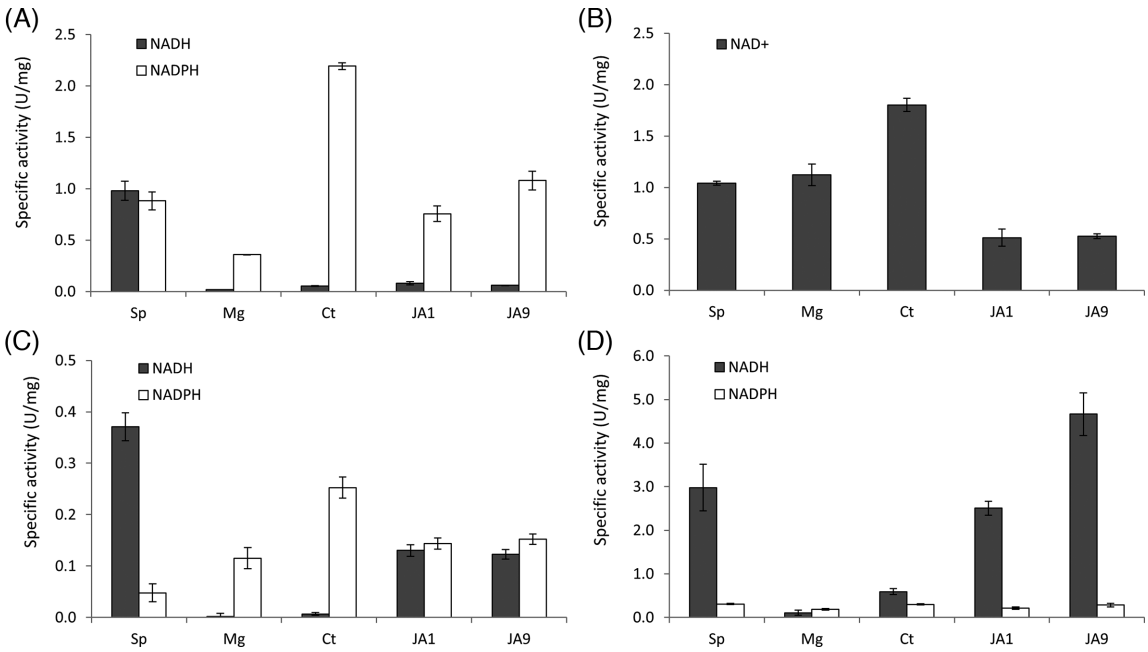

Figure 4. Xylose reduction **(A)**, xylitol oxidation **(B)**, HMF **(C)** and furfural reduction **(D)** activities expressed in units per mg of protein (U/mg), measured in crude cell extracts from *S. passalidarum* (Sp), *M. guilliermondii* (Mg), *C. tropicalis* (Ct), *Spathaspora* sp. JA1 and *M. caribbica* JA9. The values correspond to averages and standard deviations of two biological replicates.

related species are compared (Additional file 2a and c), the macrosynteny was observed as expected. On the other hand, when compared with *C. tropicalis* (Additional file 2b and d), this pattern changed to mesosynteny (Additional file 2b), and degraded mesosynteny (Additional file 2d) (Hane *et al.* 2011). There are some phenotypic similarities between these yeasts, at least in terms of xylitol production (Mattam *et al.* 2016; Dalli *et al.* 2017), but this is not reflected in the genome organization, as shown in this work. Even synteny comparisons of loci, such as for the genes *XYL1*, *XYL2* and *XYL3* (encoding Xyl1p, Xyl2p and Xyl3p respectively), among *C. tropicalis*, *Spathaspora* sp. JA1 and *M. caribbica* JA9, was not possible due to low similarity.

Synteny analysis of XR-encoding genes (*XYL1*) among the related species *S. passalidarum* and *Spathaspora* sp. JA1 or *M. guilliermondii* and *M. caribbica* JA9 revealed a significant conservation in gene order and orientation in both comparisons (Fig. 5). Our finding shows that *Spathaspora* sp. JA1 has only a single copy of the *XYL1* gene, similarly to other *Spathaspora* species already characterized, including *S. arborariae*, *S. brasiliensis*, *S. girioi*, *S. gorwiae* and *S. xylofermentans* (Lopes *et al.* 2016; Cadete *et al.* 2016a). Thus far, duplication of *XYL1* (*XYL1.1*, *XYL1.2*) was only detected in *S. passalidarum* strains (Wohlbach *et al.* 2011; Cadete *et al.* 2016a). Previous works demonstrated that *XYL1.2* encodes an NADH-preferring XR, while *XYL1.1* encodes a strictly NADPH-dependent XR. In addition, the NADH-dependent XR activity has been correlated with efficient xylose conversion to ethanol by *S. passalidarum* (Cadete *et al.* 2016a; Veras, Parachin and Almeida 2017), whereas NADPH-dependent XR activities with higher xylitol yields in different yeast species (Pal, Mondal and Sahoo 2016; Veras, Parachin and Almeida 2017). The fact that *Spathaspora* sp. JA1 has only a single copy, 97% identical to *XYL1* or *XYL1.1* from *S. brasiliensis* or *S. passalidarum*, which encodes a NADPH-dependent XR (Fig. 4), could help explain the production of xylitol over ethanol by this yeast.

Protein sequence alignment shows that Xyl1p (*Spathaspora* sp. JA1) and Xyl1.1p (*S. passalidarum*) cluster together whereas Xyl1.2p (*S. passalidarum*) is not closely related to these enzymes. *Candida tropicalis* presents a XR protein that is closely related to the Xyl1.1p proteins from *Spathaspora* species, sharing approximately 85% identity. *Meyerozyma caribbica* JA9 presents two

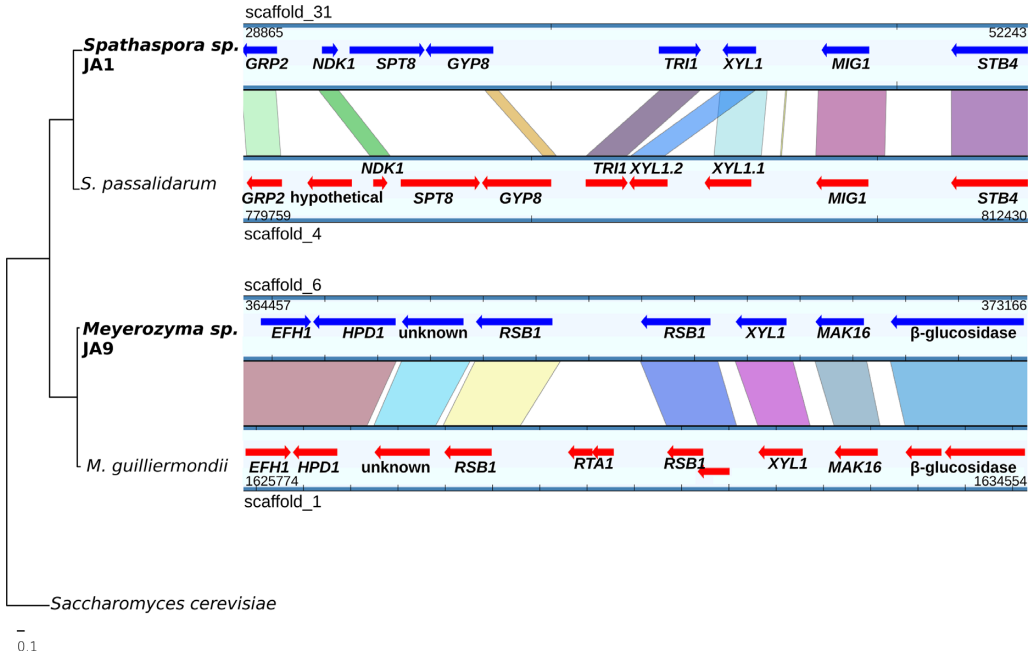

Figure 5. Synteny analysis of XR (*XYL1*) encoding genes among *Spathaspora* sp. JA1–*S. passalidarum* and *M. caribbica* JA9–*M. guilliermondii*, using blastn comparison and gff3 annotation as input for the mGSV software. *Spathaspora* synteny block shows the *XYL1* gene duplication in *S. passalidarum*. *Meyerozyma* synteny block shows the presence of RTA1 domain containing gene in *M. guilliermondii*. Phylogenetic tree obtained from a concatenated alignment of 200 conserved, single-copy orthologous genes inferred by Maximum Likelihood (ML) using RAxML software and rooted with *Saccharomyces cerevisiae*. Scale is in amino acids substitutions per site.

genes encoding putative XR. Although these proteins are closely related to the Xyl1.1p or Xyl1.2p from *M. guilliermondii*, Xyl1.1p grouped in a different branch of Xyl1.2p, displaying approximately 65% identity at the protein level (Fig. 6 and Additional file 3). Despite the similarity between the aligned sequences of the *Meyerozyma* strains (~95% identity), remarkable differences were observed in XR activity. JA9 presented approximately three times higher NADPH-XR specific activity compared with *M. guilliermondii*, which may also explain its higher HMF reduction activity, as XR was shown to reduce HMF (Almeida *et al.* 2008). In addition, both *Meyerozyma* strains showed conservation at positions 59/60 (R59) and 271/272 (N or D 271), which affects XR cofactor preference (Cadete *et al.* 2016a). Thus, the higher XR activity in *M. caribbica* JA9 is probably not due to Xyl1p differences between the *Meyerozyma* strains. All Xyl1p protein sequences analyzed contain the conserved IPKS motif (Ile-Pro-Lys-Ser) that is involved with cofactor binding, and the GXXXGXG motif related to coenzyme binding (Lee 1998).

Regarding the xylose dehydrogenase (*XYL2*) gene, the strains retain a conserved region in the paired comparison (data not shown). Both JA1 and JA9 contain two copies of the *XYL2* gene, which is consistent with other *Spathaspora* species such as *S. passalidarum*, *S. arborariae* and *S. girioi* (Lopes *et al.* 2016; Cadete *et al.* 2016a), and *Meyerozyma* strains. The *XYL3* gene, encoding a xylulose kinase, was detected in a single copy in the genomes of all strains. The examination of the syntenic region containing the *XYL3* gene (Fig. 7) showed the insertion of the *VPS10* gene (that encodes a sorting receptor that functions in targeting multiple vacuolar proteases), *FAO2* (Long-chain-alcohol oxidase) and an unknown gene (F-box domain—IPR001810) in *Spathaspora* sp. JA1. Among the *Meyerozyma* strains, the region exhibited a conserved organization between the genomes. The protein sequence alignment revealed that the Xyl3p sequences were highly conserved in the closely related species, showing an identity of 90 or 95% between the *Spathaspora* and *Meyerozyma* strains, respectively.

CONCLUSION

Two strains capable of assimilating xylose and produce xylitol in sugarcane biomass hydrolysate were isolated from the Brazilian Cerrado biome and identified as *Spathaspora* sp. JA1 and *M. caribbica* JA9. Due to the relevance of xylose utilization, their complete genomes were sequenced and assembled. To the best of our knowledge, this is the first report of genomes of xylose-assimilating yeasts from the Cerrado biome. It was shown that the xylose metabolic pathway, i.e. XR, XDH and XK, is conserved among the xylitol-producing yeasts *Spathaspora* sp. JA1, *M. caribbica* JA9 and *M. guilliermondii*, but not in *S. passalidarum*, which possess two XRs. Additionally, measurements of XR and XDH activities demonstrated a correlation between cofactor usage and ethanol and xylitol production. These results indicate that screening strategies of xylose-utilizing yeasts should take the cofactor usage during xylose reduction in consideration. The high capacity of *Spathaspora* sp.JA1 to produce xylitol opens the possibility of employment of this strain in biotechnological production processes for this compound. In this context, physiological and genetic information is essential to develop metabolic engineering strategies or to optimize the process conditions, which may lead to improved yields and productivities compatible with the industry.

Authors' contributions

DT participated in the design of the study, performed enzymatic assays and fermentation experiments, analyzed the data and wrote the manuscript. CES isolated the yeast strains, performed taxonomy analysis and helped in fermentation experiments. AS worked on genome annotation and analysis and wrote the

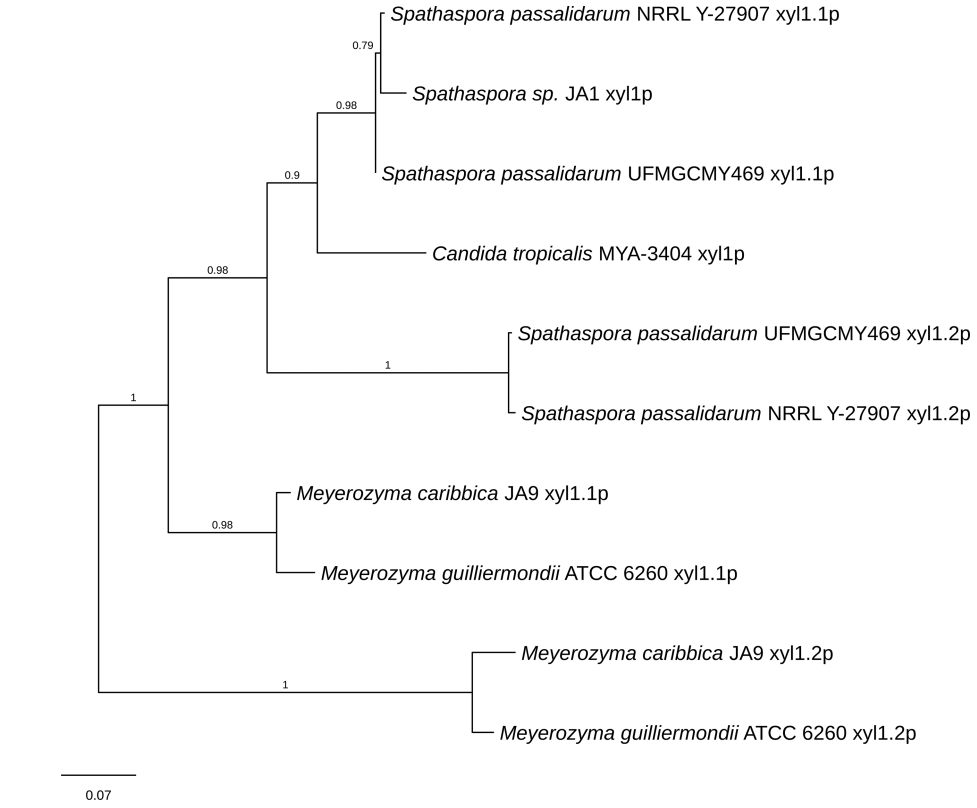

Figure 6. Phylogenetic tree obtained from alignment of the XR proteins from *Spathaspora* sp. JA1, *M. caribbica* JA9, *M. guilliermondii*, *S. passalidarum* and *C. tropicalis* using the program Muscle.

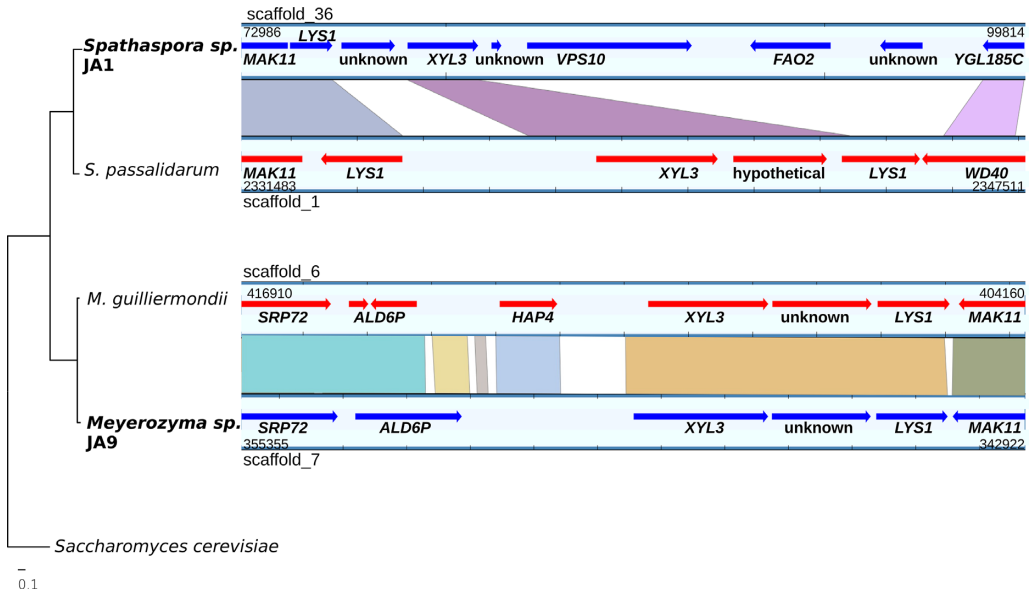

Figure 7. Synteny analysis of xylulokinase (*XYL3*) genes among *Spathaspora* sp. JA1–*S. passalidarum* and *M. caribbica* JA9–*M. guilliermondii* using blastn comparison and gff3 annotation as input for mGSV software. *Spathaspora* synteny block shows the insertion of the *VPS10* gene (which encodes a sorting receptor that functions in targeting multiple vacuolar proteases), *FAO2* (Long-chain-alcohol oxidase) and an unknown gene (F-box domain—IPR001810) in *Spathaspora* sp. JA1. Phylogenetic tree obtained from a concatenated alignment of 200 conserved, single-copy orthologous genes inferred by Maximum Likelihood (ML) using RAxML software and rooted with *S. cerevisiae*. Scale is in amino acid substitutions per site.

manuscript. JRMA and EF participated in design of the study and reviewed the manuscript. All authors read and approved the manuscript.

Availability of data and materials

Supporting data are available in additional files and the Whole Genome Shotgun projects generated during the current study are available in the DDBJ/EMBL/GenBank repository under accession numbers NSGQ00000000 (*Spathaspora* sp. JA1) and NSDU00000000 (*M. caribbica* JA9).

Acknowledgements

The authors would like to thank Lunalva P. Sallet for DNA extraction for genome sequencing.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

FUNDING

This work was supported by funding from The Brazilian Agricultural Research Corporation (EMBRAPA); Coordination for the Improvement of Higher Education Personnel (CAPES); the National Council for Scientific and Technological Development (CNPQ) and the Brazilian Development Bank (BNDES).

Competing interests

The authors declare that they have no competing interests.

REFERENCES

Alkim C, Trichez D, Cam Y *et al.* The synthetic xylulose-1 phosphate pathway increases production of glycolic acid from xylose-rich sugar mixtures. *Biotechnol Biofuels* 2016;**9**:201.

Almeida JR, Modig T, Petersson A *et al.* Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *J Chem Technol Biotechnol* 2007;**82**:340–9.

Almeida JR, Modig T, Röder A *et al. Pichia stipitis* xylose reductase helps detoxifying lignocellulosic hydrolysate by reducing 5-hydroxymethyl-furfural (HMF). *Biotechnol Biofuels* 2008;**1**:12.

Barbosa AC, Cadete RM, Gomes FCO *et al. Candida materiae* sp. nov., a yeast species isolated from rotting wood in the Atlantic Rain Forest. *Int J Syst Evol Microbiol* 2009;**59**:2104–6.

Barbosa MFS, Medeiros MB, Mancilha IM *et al.* Screening of yeasts for production of xylitol fromd-xylose and some factors which affect xylitol yield in *Candida guilliermondii*. *J Ind Microbiol* 1988;**3**:241–51.

Bruinenberg PM, Bot PHM, Dijken JP *et al.* The role of redox balances in the anaerobic fermentation of xylose by yeasts. *Eur J Appl Microbiol Biotechnol* 1983;**18**:287–92.

Butler G, Rasmussen MD, Lin MF *et al.* Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 2009;**459**:657–62.

Cadete RM, de las Heras AM, Sandstrom AG *et al.* Exploring xylose metabolism in *Spathaspora* species: XYL1.2 from *Spathaspora passalidarum* as the key for efficient anaerobic xylose fermentation in metabolic engineered *Saccharomyces cerevisiae*. *Biotechnol Biofuels* 2016a;**9**:167.

Cadete RM, Melo-Cheab MA, Viana AL *et al.* The yeast *Scheffersomyces amazonensis* is an efficient xylitol producer. *World J Microbiol Biotechnol* 2016b;**32**:207.

Cadete RM, Melo MA, Dussan KJ *et al.* Diversity and physiological characterization of d-xylose-fermenting yeasts isolated from the Brazilian Amazonian forest. *PLoS One* 2012;**7**:e43135.

Cadete RM, Melo MA, Zilli JE *et al. 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* 2013;**103**:421–31.

Cadete RM, Rosa CA. The yeasts of the genus *Spathaspora* : potential candidates for second-generation biofuel production: Spathaspora yeasts as candidates for 2 G bioethanol production. *Yeast* 2017, DOI :10.1002/yea.3279.

Cadete RM, Santos RO, Melo MA *et al. Spathaspora arborariae* sp. nov., a d-xylose-fermenting yeast species isolated from rotting wood in Brazil. *FEMS Yeast Res* 2009;**9**:1338–42.

Canilha L, de Lacerda Brambilla Rodrigues RC, Fernandes FA *et al.* Bioconversion of hemicellulose from sugarcane biomass into sustainable products. In: Chandel A (ed). *Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization*. London: InTech, 2013.

Corte L, di Cagno R, Groenewald M *et al.* Phenotypic and molecular diversity of *Meyerozyma guilliermondii* strains isolated from food and other environmental niches, hints for an incipient speciation. *Food Microbiol* 2015;**48**:206–15.

Dalli SS, da Silva SS, Uprety BK *et al.* Enhanced production of xylitol from poplar wood hydrolysates through a sustainable process using immobilized new strain *Candida tropicalis* UFMG BX 12-a. *Appl Biochem Biotechnol* 2017;**182**:1053–64.

de Albuquerque TL, da Silva IJ, de Macedo GR *et al.* Biotechnological production of xylitol from lignocellulosic wastes: a review. *Process Biochem* 2014;**49**:1779–89.

de Arruda PV, de Cássia Lacerda Brambilla Rodrigu R, da Silva DDV *et al.* Evaluation of hexose and pentose in precultivation of *Candida guilliermondii* on the key enzymes for xylitol production in sugarcane hemicellulosic hydrolysate. *Biodegradation* 2011;**22**:815–22.

Deparis Q, Claes A, Foulquié-Moreno MR *et al.* Engineering tolerance to industrially relevant stress factors in yeast cell factories. *FEMS Yeast Res* 2017;**17**, DOI:10.1093/femsyr/fox036.

de Sales BB, Scheid B, Gonçalves DL *et al.* Cloning novel sugar transporters from *Scheffersomyces* (*Pichi a*) *stipitis* allowing d-xylose fermentation by recombinant *Saccharomyces cerevisiae*. *Biotechnol Lett* 2015;**37**:1973–82.

Finn RD, Attwood TK, Babbitt PC *et al.* InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res* 2017;**45**:D190–9.

Gnerre S, MacCallum I, Przybylski D *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci* 2011;**108**:1513–8.

Guamán-Burneo MC, Dussán KJ, Cadete RM *et al.* Xylitol production by yeasts isolated from rotting wood in the Galápagos Islands, Ecuador, and description of *Cyberlindnera galapagoensis* f.a., sp. nov. *Antonie Van Leeuwenhoek* 2015;**108**:919–31.

Hane JK, Rouxel T, Howlett BJ *et al.* A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol* 2011;**12**:R45.

Hernández-Pérez AF, Costa IAL, Silva DDV *et al*. Biochemical conversion of sugarcane straw hemicellulosic hydrolyzate supplemented with co-substrates for xylitol production. *Bioresour Technol* 2016;**200**:1085–8.

Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 2011;**12**:491.

Hou X. Anaerobic xylose fermentation by *Spathaspora passalidarum*. *Appl Microbiol Biotechnol* 2012;**94**:205–14.

Hyvonen L, Koivistoinen P, Voirol F. Food technological evaluation of xylitol. *Advances in Food Research*. Vol 28. New York: Elsevier, 1982, 373–403.

Jeffries TW, Grigoriev IV, Grimwood J *et al*. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat Biotechnol* 2007;**25**:319–26.

Karhumaa K, Hahn-Hägerdal B, Garcia Sanchez R *et al*. Comparison of the xylose reductase-xylitol dehydrogenase and the xylose isomerase pathways for xylose fermentation by recombinant *Saccharomyces cerevisiae*. *Microb Cell Fact* 2007;**6**:5.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80.

Korf I. Gene finding in novel genomes. *BMC Bioinformatics* 2004;**5**:59.

Kurtzman CP, Robnett CJ. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie Van Leeuwenhoek* 1998;**73**:331–71.

Kurtzman CP, Suzuki M. Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera *Babjeviella*, *Mey erozyma*, *Millerozyma*, *Priceomyces*, and *Scheffersomyces*. *Mycoscience* 2010;**51**:2–14.

Lee H. Review: the structure and function of yeast xylose (aldose) reductases. *Yeast* 1998;**14**:977–84.

Li L. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;**13**:2178–89.

Li Y-J, Wang M-M, Chen Y-W *et al*. Engineered yeast with a CO2-fixation pathway to improve the bio-ethanol production from xylose-mixed sugars. *Sci Rep* 2017;**7**:43875.

Lobo FP, Goncalves DL, Alves SL *et al*. Draft genome sequence of the D-xylose-fermenting yeast *Spathaspora arborariae* UFMG-HM19.1AT. *Genome Announc* 2014;**2**:e01163–13.

Lombard V, Golaconda Ramulu H, Drula E *et al*. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014;**42**:D490–5.

Long TM, Su YK, Headman J *et al*. Cofermentation of glucose, xylose, and cellobiose by the beetle-associated yeast *Spathaspora passalidarum*. *Appl Env Microbiol* 2012;**78**, DOI:10.1128/AEM.00374-12.

Lopes DD, Cibulski SP, Mayer FQ *et al*. Draft genome sequence of the d-xylose-fermenting yeast *Spathaspora xylofermentans* UFMG-HMD23.3. *Genome Announc* 2017;**5**:e00815–17.

Lopes MR, Morais CG, Kominek J *et al*. Genomic analysis and D-xylose fermentation of three novel *Spathaspora* species: *Spathaspora girioi* sp. nov., *Spathaspora hagerdaliae* f. a., sp. nov. and *Spathaspora gorwiae* f. a., sp. nov. *FEMS Yeast Res* 2016;**16**:fow044.

Martini C, Tauk-Tornisielo SM, Codato CB *et al*. A strain of *Meyerozyma guilliermondii* isolated from sugarcane juice is able to grow and ferment pentoses in synthetic and bagasse hydrolysate media. *World J Microbiol Biotechnol* 2016;**32**:80.

Mattam AJ, Kuila A, Suralikerimath N *et al*. Cellulolytic enzyme expression and simultaneous conversion of lignocellulosic sugars into ethanol and xylitol by a new *Candida tropicalis* strain. *Biotechnol Biofuels* 2016;**9**:157.

Modig T, Almeida JRM, Gorwa-Grauslund MF *et al*. Variability of the response of *Saccharomyces cerevisiae* strains to lignocellulose hydrolysate. *Biotechnol Bioeng* 2008;**100**:423–9.

Moysés D, Reis V, Almeida JRM *et al*. Xylose fermentation by *Saccharomyces cerevisiae*: challenges and prospects. *Int J Mol Sci* 2016;**17**:207.

Muhlhausen S, Kollmar M. Predicting the fungal CUG codon translation with Bagheera. *BMC Genomics* 2014;**15**:411.

Mukherjee V, Radecka D, Aerts G *et al*. Phenotypic landscape of non-conventional yeast species for different stress tolerance traits desirable in bioethanol fermentation. *Biotechnol Biofuels* 2017;**10**:216.

Nakanishi SC, Soares LB, Biazi LE *et al*. Fermentation strategy for second generation ethanol production from sugarcane bagasse hydrolyzate by *Spathaspora passalidarum* and *Scheffersomyces stipitis*: fermentation strategy for second generation ethanol production. *Biotechnol Bioeng* 2017;**114**:2211–21.

Nguyen NH, Suh S-O, Marshall CJ *et al*. 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* 2006;**110**:1232–41.

Pal S, Mondal AK, Sahoo DK. Molecular strategies for enhancing microbial production of xylitol. *Process Biochem* 2016;**51**:809–19.

Petersson A, Almeida JRM, Modig T *et al*. A 5-hydroxymethyl furfural reducing enzyme encoded by the *Saccharomyces cerevisiae* ADH6 gene conveys HMF tolerance. *Yeast* 2006;**23**:455–64.

Prakash G, Varma AJ, Prabhune Aet al. Microbial production of xylitol from D-xylose and sugarcane bagasse hemicellulose using newly isolated thermotolerant yeast Debaryomyces hansenii. *Bioresour. Technol.* 2011;**102**:3304–8. 10.1016/j.biortech.2010.10.07421067918

Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;**26**:1641–50.

Rao RS, Jyothi CP, Prakasham RS *et al*. Xylitol production from corn fiber and sugarcane bagasse hydrolysates by *Candida tropicalis*. *Bioresour Technol* 2006;**97**:1974–8.

Ravella SR, Gallagher J, Fish S *et al*. Overview on commercial production of xylitol, economic analysis and market trends. In: da Silva SS, Chandel AK (eds). *D-Xylitol*. Berlin, Heidelberg: Springer, 2012, 291–306.

Reider Apel A, Ouellet M, Szmidt-Middleton H *et al*. Evolved hexose transporter enhances xylose uptake and glucose/xylose co-utilization in *Saccharomyces cerevisiae*. *Sci Rep* 2016;**6**, DOI:10.1038/srep19512.

Revanna KV, Chiu C-C, Bierschank E *et al*. GSV: a web-based genome synteny viewer for customized data. *BMC Bioinformatics* 2011;**12**:316.

Riley R, Haridas S, Wolfe KH *et al*. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci* 2016;**113**:9882–7.

Romi W, Keisam S, Ahmed G *et al*. Reliable differentiation of *Meyerozyma guilliermondii* from *Meyerozyma caribbica* by internal transcribed spacer restriction fingerprinting. *BMC Microbiol* 2014;**14**:52.

Simão FA, Waterhouse RM, Ioannidis P *et al*. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**:3210–2.

Smiley KL, Bolen PL. Demonstration of D-xylose reductase and D-xylitol dehydrogenase in *Pachysolen tannophilus*. *Biotechnol Lett* 1982;**4**:607–10.

Soderlund C, Bomhoff M, Nelson WM. SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res* 2011;**39**:e68–8.

Stanke M, Diekhans M, Baertsch R *et al*. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 2008;**24**:637–44.

Ter-Hovhannisyan V, Lomsadze A, Chernoff YO *et al*. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res* 2008;**18**:1979–90.

Vaughan-Martini A, Kurtzman C, Meyer S *et al*. Two new species in the clade: sp. nov., the ascosporic state of, and comb. nov. *FEMS Yeast Res* 2005;**5**:463–9.

Veras HCT, Parachin NS, Almeida JRM. Comparative assessment of fermentative capacity of different xylose-consuming yeasts. *Microb Cell Factories* 2017;**16**:153.

Verduyn C, van Kleef R, Frank Jzn J *et al*. NAD(P)H-dependent aldose reductase from the xylose-fermenting yeast *Pichia stipitis*. *Antonie Van Leeuwenhoek* 1985;**51**:562–62.

Wang Y, Coleman-Derr D, Chen G *et al*. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 2015;**43**:W78–84.

Weiss S, Samson F, Navarro D *et al*. YeastIP: a database for identification and phylogeny of *Saccharomycotina* yeasts. *FEMS Yeast Res* 2013;**13**:117–25.

White TJ, Bruns T, Lee S *et al*. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications*. New York: Academic Press, 1990, 315–22.

Wohlbach DJ, Kuo A, Sato TK *et al*. Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. *Proc Natl Acad Sci USA* 2011;**108**:13212–7.

Xin B, Wang Y, Tao F *et al*. Co-utilization of glycerol and lignocellulosic hydrolysates enhances anaerobic 1,3-propanediol production by *Clostridium diolis*. *Sci Rep* 2016;**6**:19044.

Yin Y, Mao X, Yang J *et al*. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 2012;**40**:W445–51.

Young E, Poucher A, Comer A *et al*. Functional survey for heterologous sugar transport proteins, using *Saccharomyces cerevisiae* as a host. *Appl Environ Microbiol* 2011;**77**:3311–9.