



OPEN

## A novel D-xylose isomerase from the gut of the wood feeding beetle *Odontotaenius disjunctus* efficiently expressed in *Saccharomyces cerevisiae*

Paulo César Silva<sup>1</sup>✉, Javier A. Ceja-Navarro<sup>2,3</sup>, Flávio Azevedo<sup>1</sup>, Ulas Karaoz<sup>4</sup>, Eoin L. Brodie<sup>4,5</sup>✉ & Björn Johansson<sup>1</sup>✉

Carbohydrate rich substrates such as lignocellulosic hydrolysates remain one of the primary sources of potentially renewable fuel and bulk chemicals. The pentose sugar D-xylose is often present in significant amounts along with hexoses. *Saccharomyces cerevisiae* can acquire the ability to metabolize D-xylose through expression of heterologous D-xylose isomerase (XI). This enzyme is notoriously difficult to express in *S. cerevisiae* and only fourteen XIs have been reported to be active so far. We cloned a new D-xylose isomerase derived from microorganisms in the gut of the wood-feeding beetle *Odontotaenius disjunctus*. Although somewhat homologous to the XI from *Piromyces* sp. E2, the new gene was identified as bacterial in origin and the host as a *Parabacteroides* sp. Expression of the new XI in *S. cerevisiae* resulted in faster aerobic growth than the XI from *Piromyces* on D-xylose media. The D-xylose isomerization rate conferred by the new XI was also 72% higher, while absolute xylitol production was identical in both strains. Interestingly, increasing concentrations of xylitol (up to 8 g L<sup>-1</sup>) appeared not to inhibit D-xylose consumption. The newly described XI displayed 2.6 times higher specific activity, 37% lower K<sub>M</sub> for D-xylose, and exhibited higher activity over a broader temperature range, retaining 51% of maximal activity at 30 °C compared with only 29% activity for the *Piromyces* XI.

Lignocellulosic material continues to be the most promising renewable raw material for the production of sustainable fuels and fine chemicals<sup>1</sup>. Xylan is the second most abundant biopolymer on earth which contains mostly the pentose sugar D-xylose<sup>2</sup>. Baker's yeast or *Saccharomyces cerevisiae* is the preferred organism for industrial transformation of sugars derived from lignocellulose due to innate resistance to fermentation inhibitors<sup>3</sup>. Expression of heterologous pathways are necessary for D-xylose utilization as it is not metabolized naturally by *S. cerevisiae*. D-xylose metabolism remains a metabolic bottleneck in *S. cerevisiae* despite the development of several types of pathways for the consumption of this sugar<sup>4</sup>.

D-xylose metabolic pathways can be classified into two main categories, D-xylose reductase/xylitol dehydrogenase (XR/XDH) and D-xylose isomerase (XI). The XR/XDH pathway converts D-xylose to xylitol by reduction with NADPH or NADH followed by an oxidation with NAD<sup>+</sup> to xylulose in an overall redox neutral process. Alternatively, the same reaction is carried out by a single XI enzyme without cofactors. The XR/XDH pathway is mainly found in fungi while the XI pathway is common in prokaryotes. The currently most promising D-xylose metabolic pathways are based on the prokaryotic XI route. The reason for this is that although the overall reaction is redox neutral, the XR/XDH pathway suffers from a NAD(P)H cofactor imbalance that has proven hard to remedy<sup>5-7</sup>. However, the XI pathway suffers from low capacity and inhibition by xylitol<sup>8</sup>. Another issue is that the XI is rather difficult to express heterologously in yeast. Several unsuccessful attempts have been made

<sup>1</sup>CBMA - Center of Molecular and Environmental Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal. <sup>2</sup>Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>3</sup>Institute for Biodiversity Science and Sustainability, California Academy of Sciences, San Francisco, CA, USA. <sup>4</sup>Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>5</sup>Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA. ✉email: pcfernandesdasilva@gmail.com; ELBrodie@lbl.gov; bjorn\_johansson@bio.uminho.pt

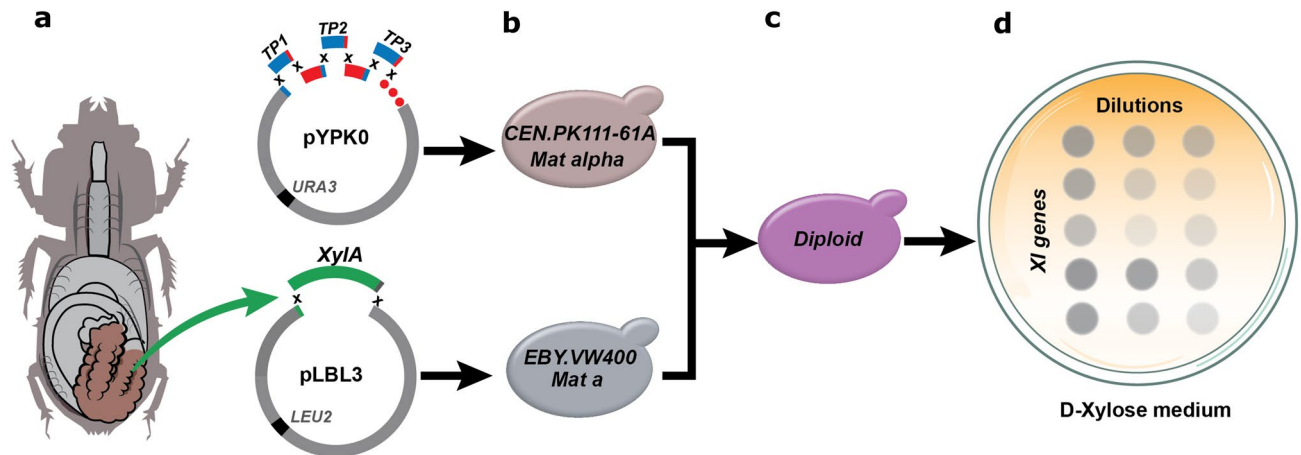
#	Source	Type	Codon optimization	$K_M$	$V_{max}$ (U mg <sup>-1</sup> )	References
1	<i>Thermus thermophilus</i>	Prokaryote, Gram–	No	NA	1.0 <sup>(a)</sup>	14
2	<i>Piromyces</i> sp. E2	Eukaryote	No	20	1.1	15
			Yes	49.85	0.0538	8
			Yes <sup>(b)</sup>	50.18	0.083	22
			No	51	0.25	23
			Yes	3.9 <sup>(c)</sup>	2.68 <sup>(c)</sup>	24
			Yes	14.90	0.0103	19
			Yes	6.2 <sup>(c)</sup>	8.3 <sup>(c)</sup>	25
			Yes	30.76	0.0112	This work
3	<i>Orpinomyces</i> sp. ukk1	Eukaryote	No	NA	1.91	26
4	<i>Clostridium phytofermentans</i>	Prokaryote, Gram+	Yes	66.01	0.0344	8
			Yes <sup>(b)</sup>	37.1	0.107	27
5	Soil–xym1 (unspecified)	Prokaryote	No	NA	0.33	17
6	Soil–xym2 (unspecified)	Prokaryote	No	NA	0.20	17
7	<i>Bacteroides stercoris</i>	Prokaryote, Gram–	No	54.03	NA	28
8	<i>Ruminococcus flavefaciens</i>	Prokaryote, Gram+	Yes <sup>(d)</sup>	66.7	1.41	29
9	<i>Prevotella ruminicola</i>	Prokaryote, Gram–	Yes <sup>(e)</sup>	34	0.81	23
10	<i>Burkholderia cenocepacia</i>	Prokaryote, Gram–	No	NA	~ 0.037	30
			No	17.08 <sup>(c)</sup>	44.97 <sup>(c)</sup>	31
11	<i>Bacteroides vulgatus</i>	Prokaryote, Gram–	No	NA	~ 1.25	32
12	Bovine rumen (unspecified)	Prokaryote	No <sup>(e)</sup>	16.8	1.31	18
13	<i>Sorangium cellulosum</i>	Prokaryote, Gram–	No <sup>(e)</sup>	17.2	0.35	18
14	Termite gut (unspecified)	Eukaryote	Yes	10.52	0.0074	19
15	Passalid beetle gut—8054_2 (unspecified)	Prokaryote	Yes	19.27	0.0296	This work

**Table 1.** Literature data on D-xylose isomerase enzymes with characterized kinetic properties that were actively expressed in *Saccharomyces cerevisiae*. Kinetic parameters were determined directly from cell lysates or otherwise stated. NA not available. <sup>a</sup>Determined by measuring the conversion of fructose to glucose. <sup>b</sup>And also random mutagenesis. <sup>c</sup>Determined from purified enzymes. <sup>d</sup>And also site-directed mutagenesis and modifications 5'-end of the gene. <sup>e</sup>And also evolutionary adaptation.

to express XIs, such as those from *Escherichia coli*<sup>9,10</sup> *Bacillus subtilis*, *Actinoplanes missouriensis*<sup>11</sup>, *Lactobacillus pentosus*<sup>12</sup> and *Clostridium thermosulfurogenes* (renamed as *Thermoanaerobacterium thermosulfurigenes*)<sup>13</sup>. The first successfully expressed XI was a thermostable enzyme from *Thermus thermophilus*<sup>14</sup> followed by an XI originating from the fungus *Piromyces* sp.<sup>15</sup>

Only fourteen different XIs with characterized kinetic parameters have been reported to actively express in *S. cerevisiae* (Table 1). Interestingly, the fungal XIs (Table 1, entry #2 and #3) both come from the same division (*Neocallimastigomycota*). These fungi are known for possessing genes originating via lateral gene transfer. Their XIs are of prokaryotic origin and have been recently incorporated in evolutionary terms<sup>16</sup>. These fungal isomerases were isolated from fungi residing in the intestines of herbivorous mammals, an ecosystem typically poised at 37 °C and with dense populations of co-occurring bacteria. Interestingly, only four XIs have been isolated directly from metagenomes and subsequently expressed in *S. cerevisiae* (Table 1, #5 and #6, #12, #14). Two XIs were derived from soil samples, one using degenerate primers and another by functional metagenomic screening using *E. coli* with a nonfunctional *xylA*<sup>17</sup>. An additional XI was amplified from bovine rumen contents using degenerate primers targeting relatively conserved XI specific sequences<sup>18</sup>, and yet another derived from protists in the hindgut of the termite *Reticulitermes speratus* using a similar approach<sup>19</sup>.

An alternative and less biased approach to retrieving novel XIs is through the assembly of shotgun metagenomic sequences from environments thought to be enriched in D-xylose utilization capacity. Following gene prediction, *in-silico* translation and annotation against appropriate databases, a subset of XI genes may then be optimized computationally for a specific host, and then synthesized in-vitro. This strategy avoids the bias associated with PCR amplification and is expected to broaden the discovery of enzymes with novel properties. Using this approach, we have identified 182 putative XIs (coded by *xylA* genes) from the microbial metagenome in the gut of a passalid beetle (*Odontotaenius disjunctus*)<sup>20</sup>, known to subsist on D-xylose rich hardwoods, and to harbor D-xylose utilizing yeasts<sup>21</sup>. We codon optimized and synthesized three XIs, one of which expressed successfully in *S. cerevisiae*, highlighting that direct metagenome reconstruction combined with in-vitro gene synthesis is a viable approach for the identification of novel XI genes for expression in *S. cerevisiae*. In the present work, this new XI was characterized and compared to the homolog from *Piromyces* sp. The D-xylose isomerization capacity of the expressed XIs were evaluated in-vivo by yeast growth and by D-xylose conversion rates. Enzyme kinetics and optimal temperature were determined in-vitro using cell extracts. Additionally, a phylogenetic analysis of the XIs that have been expressed efficiently or that could not be expressed in yeast was carried out to interpret the role of evolutionary relationships in the expression of these enzymes.



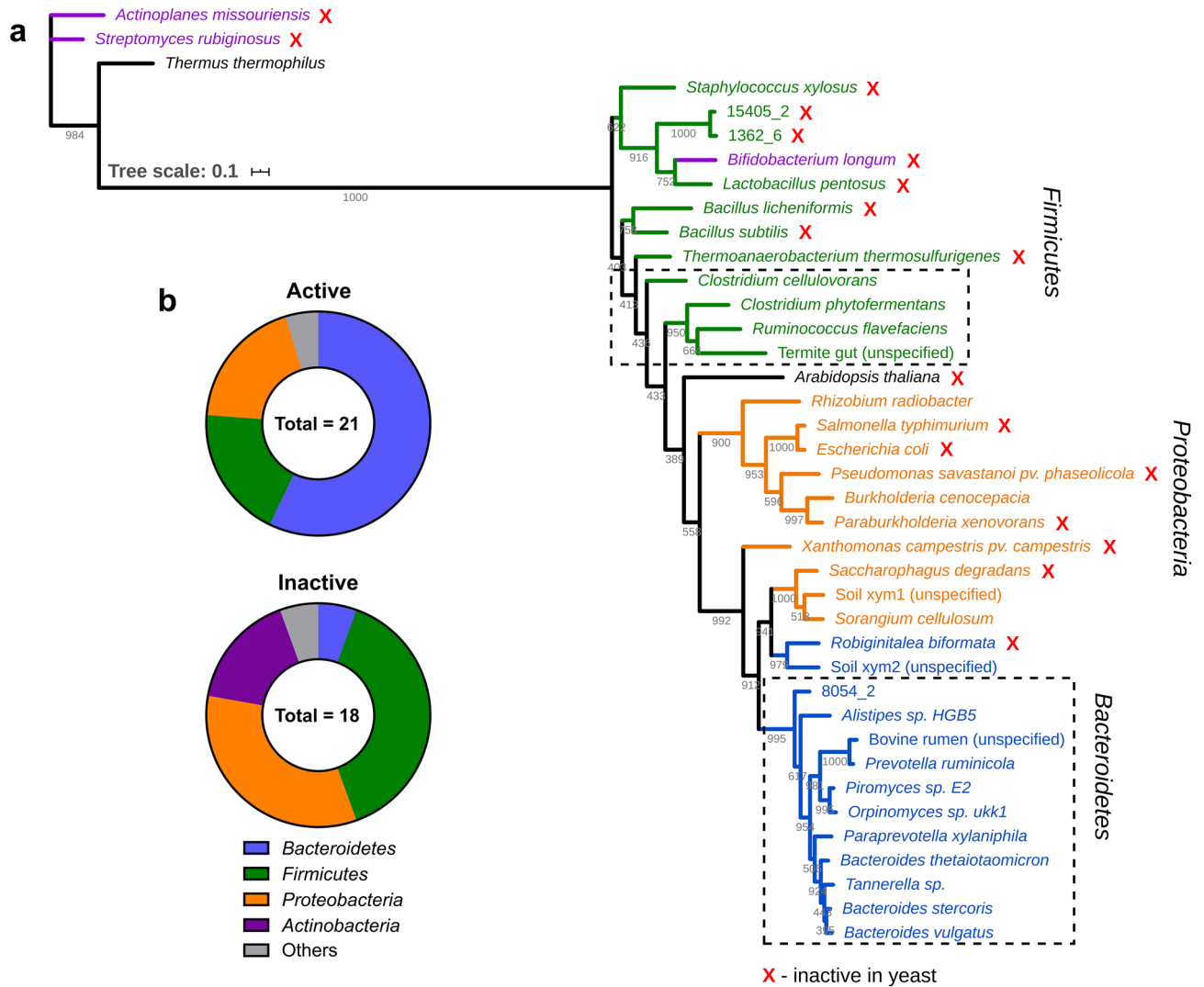
**Figure 1.** Strategy for the screening of D-xylose isomerases (XIs) in *Saccharomyces cerevisiae*. (a) Putative *xylA* genes from the microbial metagenome in the gut of a passalid beetle (*Odontotaenius disjunctus*) were codon-optimized, synthesized and cloned in the plasmid pLBL3; the partial D-xylose pathway (*TKL1*, *TAL1*, *RPE1*, *RKI1*, *XKS1*, and *Gxf1*) was constructed in the plasmid pYPK0 under the control of different terminators/promoters (TP). (b) The plasmids containing *xylA* genes (pLBL3\_XI) were carried by the strain EBY.VW400 and the plasmid containing the partial D-xylose pathway (pYPK0\_XTTRRG) was carried by the strain CEN.PK111-61A. (c) Each EBY.VW400 strain clone was mated with the CEN.PK111-61A strain. (d) Candidate XI enzymes were functionally screened by scoring yeast growth on solid media with D-xylose as the sole or main carbon source.

## Results

**Identification of XI genes from metagenomic information.** The metagenome of the wood-feeding beetle *O. disjunctus*, previously reconstructed by our team members<sup>20</sup>, was screened for the detection of D-xylose isomerases against the pfamA database. A total of 182 putative XI sequences were detected and phylogenetically placed together with the XI of *Piromyces* (see Fig. S1, Supplementary Information). Based on their phylogenetic relatedness, 3 metagenome-derived XI sequences were selected (labeled 8054\_2, 15405\_2, and 1362\_6 in Fig. S1) and further characterized using the SWISS-MODEL workspace<sup>33</sup>. Top SWISS-MODEL templates for each of the target amino acid sequences corresponded to D-xylose isomerases for which crystal structures are available (see Table S1, Supplementary Information). Sequences 1362\_6 and 14450\_2 had 55.9% and 56.7% sequence homology with the XI of *Thermotoga neapolitana*, respectively, while 8054\_2 had 82.5% homology with the XI of *Bacteroides thetaiotaomicron*. Reconstructed structure-homology models for each sequence also showed high quality scores between target and template (GMQE) and agreement between model and structure (QMEAN) (see Table S1, Supplementary Information). Based on these results, we proceeded to synthesize codon-optimized versions of these *xylA* genes for expression in *S. cerevisiae*.

The codon-optimized synthetic genes corresponding to the sequences of 8054\_2, 15405\_2, 1362\_6 and *Piromyces* sp. *xylA* were cloned into the plasmid pLBL3 under the control of a constitutive TEF promoter. Plasmid pLBL3\_XR/XDH expressing *Scheffersomyces stipitis* D-xylose reductase (XR) and D-xylitol dehydrogenase (XDH) genes (GenBank Gene ID: 4839234 and 4852013) corresponding to the initial fungal D-xylose metabolic pathway was also constructed. The resulting plasmids were transformed into *S. cerevisiae* EBY.VW4000. In parallel, a plasmid containing a partial D-xylose utilization pathway (*TKL1*, *TAL1*, *RPE1* and *RKI1*), xylulokinase (*XKS1*), and a D-xylose/glucose facilitator from *Candida intermedia* (*Gxf1*), was constructed using the yeast pathway kit<sup>34</sup> in the strain CEN.PK111-61A (Fig. 1). Yeast strains containing D-xylose isomerase genes or XR/XDH pathway were mated to the strain with the partial D-xylose utilization pathway. The resulting diploid strains were plated on solid media with D-xylose as the sole carbon source. The strains expressing the XR/XDH gene pair, the *Piromyces* sp. *xylA*, and the 8054\_2 gene produced growth clearly distinguishable from strains with the partial D-xylose utilization pathway and the empty pLBL3 vector. Strains expressing the 15405\_2 and 1362\_6 genes showed no discernible growth.

**Phylogenetic relationship among yeast (un)expressed XIs.** Phylogenetic analysis was used to further distinguish the successfully expressed from the unexpressed metagenome derived XIs and to complement the phylogenetic analysis from the SWISS-MODEL (Fig. 2a). For this purpose, we also included sequences of XIs for which activity has been reported in *S. cerevisiae* to evaluate any evolutionary relationships that might exist between successfully and unsuccessfully expressed enzymes. The sources of XIs included in the phylogenetic analysis were divided in three main prokaryotic phyla: *Firmicutes* (Fig. 2a—colored green), *Proteobacteria* (orange), and *Bacteroidetes* (blue); two other phyla that contain few representatives: *Actinobacteria* (purple; *Streptomyces rubiginosus*, *Actinoplanes missouriensis*, *Bifidobacterium longum*) and *Deinococcus-Thermus* (*Thermus thermophilus*); and one in *Plantae* (*Arabidopsis thaliana*). The XIs from the fungi *Piromyces* sp. and *Orpinomyces* sp. share 95% identity and were clustered within the *Bacteroidetes* phylum. *T. thermophilus* showed 55% and 59% identity with *A. missouriensis* and *S. rubiginosus*, respectively, but these three enzymes show very low

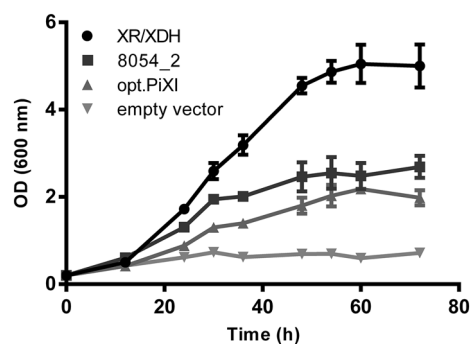


**Figure 2.** Phylogenetic analysis of the amino acid sequences of XIs that have been reported to actively express or not to express in *Saccharomyces cerevisiae*. Data from patents was not considered in the analysis. (a) In the phylogenetic tree, *Firmicutes* are colored green, *Proteobacteria* are colored orange, *Bacteroidetes* are colored blue, *Actinobacteria* are colored purple, and XIs from other sources are colored black. Dashed squares highlight two clusters of XIs belonging to *Clostridia* class (*Firmicutes*) and *Bacteroidia* class (*Bacteroidetes*) that expressed actively in yeast. (b) Ratio of active and inactive XIs expressed in yeast from the taxonomic groups presented in the panel A. GenBank accession numbers of the amino acid sequences are available in Table S2, Supplementary Information.

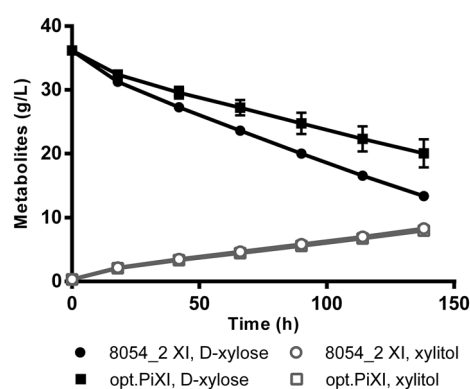
identity with the other XIs (<29%). XIs from metagenomic libraries were grouped within the prokaryotic phyla similar to previous reports<sup>17–19</sup>.

In the phylogenetic tree, 8054\_2 XI was clustered with *Bacteroidetes* sharing 82% identity with *B. thetaiotaomicron*, and 78% with *Piromyces* sp. Sequence 8054\_2 was previously reported as belonging to a metagenome assembled-genome (MAG) identified as *Parabacteroides* sp. which was highly abundant in the mostly anaerobic anterior hindgut of *O. disjunctus*<sup>20</sup>. A BLAST analysis also showed that 8054\_2 XI shares the highest identity with XIs from *Bacteroidetes* phylum, particularly with the *Porphyromonadaceae* family of the *Bacteroidia* class. *Porphyromonadaceae* are commonly found in the gastrointestinal tract and oral cavity of animals<sup>35</sup>. The 8054\_2 XI also shares high identity with the *Bacteroides* genus, such as *Bacteroides timonensis* (83%, accession number WP\_044271094.1) that is an obligate anaerobe isolated from human feces<sup>36</sup>. The 15405\_2 XI and 1362\_6 XI share 94% identity and exhibit low similarity with the other XIs displayed in the tree. A BLAST analysis revealed high identity of both 15405\_2 XI and 1362\_6 XI with XIs from *Leuconostoc* genus of the *Firmicutes* phylum (85% and 86%; accession numbers WP\_042252435.1, WP\_004912802.1, respectively).

The expression success of XIs in yeast reveals a positive tendency towards the *Bacteroidetes* phylum (Fig. 2b). The majority of the reported active XIs in yeast are from *Bacteroidetes* (12 of 21), and most of the XIs from this phylum have been expressed successfully (12 of 13). The only XI belonging to *Bacteroidetes* that was not expressed in yeast is from *Robiginitalea biformata* that, coincidentally, is also the most dissimilar enzyme of this phylum,



**Figure 3.** Aerobic growth of diploid *Saccharomyces cerevisiae* strains expressing 8054\_2 XI, the codon-optimized XI from *Piromyces* sp. E2 (opt.PiXI), XR/XDH, or the empty vector. Cultures were grown in shaking-flasks on synthetic medium with 2% (w/v) D-xylose as the sole carbon source. Data points represent an average of 3 biological replicates with standard deviation indicated.

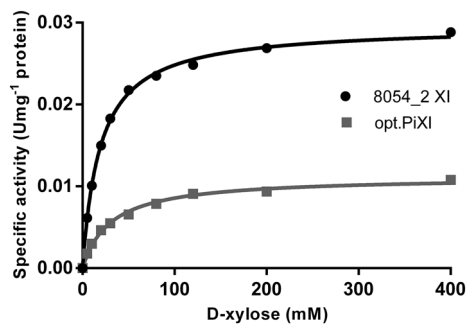


**Figure 4.** The D-xylose consumption by diploid *Saccharomyces cerevisiae* strains expressing 8054\_2 XI or the codon-optimized XI from *Piromyces* sp. E2 (opt.PiXI). Cultures were grown under high cell density ( $10 \text{ g L}^{-1}$  DCW) in shake flasks on synthetic medium with 4% (w/v) D-xylose as the sole carbon source. Data points represent an average of 3 biological replicates with standard deviation indicated.

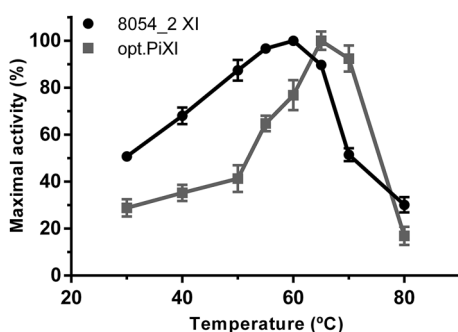
being positioned quite distant from the *Bacteroidetes* cluster. Inactive XIs (indicated by a red “X”) are comprised mostly of the other phyla: *Proteobacteria* (6 of 10), *Firmicutes* (7 of 11), and *Actinobacteria* (3 of 3). The XIs from *Proteobacteria* that have been successfully expressed in yeast (*Rhizobium radiobacter*, *Burkholderia cenocepacia*, *Sorangium cellulosum*) belong to different classes and do not share relevant ecological traits in the context of this work. The active XIs from *Firmicutes* (*Clostridium phytofermentans*, *Clostridium cellulovorans*, *Ruminococcus flavefaciens*) are from anaerobic cellulolytic organisms that belong to *Clostridia* class (Fig. 2a—*Firmicutes*’ dashed square). Similarly, actively expressed XIs clustered with *Bacteroidetes* are from anaerobic species that are commonly associated with the gastrointestinal tract of animals and belong to the same *Bacteroidia* class (*Bacteroidetes*’ dashed square), while the inactive XI belongs to a different class (*Flavobacteriia*)<sup>37</sup>.

**Growth on D-xylose under aerobic conditions.** Growth performance of yeast diploids carrying the genes 8054\_2, *xylA* from *Piromyces* sp. (opt.PiXI), and XR/XDH, or an empty vector was assessed in synthetic medium with 2% (w/v) D-xylose as the sole carbon source under aerobic conditions. The results of the growth are shown in Fig. 3. The negative control strain (Fig. 3—Empty vector), lacking XI or XR/XDH enzymes but expressing the partial D-xylose utilization pathway, did not proliferate. The cultures expressing enzymes that convert D-xylose to xylulose grew over time. The strain expressing XR/XDH proliferated at a maximal specific growth rate of  $0.10 \text{ h}^{-1}$ , the strain expressing 8054\_2 XI grew at  $0.06 \text{ h}^{-1}$ , and the strain expressing opt.PiXI grew at  $0.04 \text{ h}^{-1}$ . The XR/XDH pathway provided a considerably higher growth rate than the XI pathways as expected, and remarkably, 8054\_2 XI conferred 50% faster growth in yeast than opt.PiXI.

**D-Xylose consumption under high cell density cultures.** Cultures were inoculated at a high cell density ( $\sim 10 \text{ g L}^{-1}$  DCW) in medium with 4% (w/v) D-xylose as the sole carbon source in shake-flasks at 1:20 liquid-to-air ratio. Both cultures started consuming D-xylose immediately after incubation and showed a steady D-xylose conversion until the end of the experiment (Fig. 4). The strain carrying 8054\_2 XI consumed D-xylose at a rate of  $0.015 \text{ g D-xylose h}^{-1} \text{ g(DCW)}^{-1}$ , and the strain carrying opt.PiXI showed a consumption rate of  $0.010 \text{ g D-xylose h}^{-1} \text{ g(DCW)}^{-1}$ . The strain carrying XR/XDH depleted 96% of D-xylose after 18 h, correspond-



**Figure 5.** Enzymatic specific activity of the XI enzymes 8054\_2 XI or the codon-optimized XI from *Piromyces* sp. E2 (opt.PiXI). Kinetic parameters were determined from a Michaelis–Menten plot adjusted to the experimental data ( $R^2 > 0.99$ ). Strains were grown in synthetic medium containing 2% (w/v) D-xylose until the exponential growth phase. Crude cell extracts were prepared and used immediately in the activity assays. Data points represent an average of 3 biological replicates.



**Figure 6.** Relative enzymatic activity of XI enzymes 8054\_2 XI and the codon-optimized XI from *Piromyces* sp. E2 (opt.PiXI) at different temperatures. Strains were grown in synthetic medium containing 2% (w/v) D-xylose until the exponential growth phase. Crude cell extracts were prepared and used immediately in the activity assays. Data points represent an average of 3 biological replicates with standard deviation indicated.

ing to at least  $0.187 \text{ g D-xylose h}^{-1} \text{ g(DCW)}^{-1}$  (see Fig. S2, Supplementary Information). The higher growth rate and D-xylose consumption conferred by the XR/XDH pathway compared to the XI pathway is consistent with previous work<sup>38,39</sup>. Xylitol was produced throughout the cultivation with a yield of  $0.33 \text{ g xylitol g}^{-1} \text{ D-xylose}$  consumed for 8054\_2 XI, and  $0.47 \text{ g xylitol g}^{-1} \text{ D-xylose}$  consumed for opt.PiXI. Therefore, excluding D-xylose converted to xylitol rather than metabolized by the XI system, the rate of D-xylose isomerization was  $0.0098 \text{ g D-xylose h}^{-1} \text{ g(DCW)}^{-1}$  for 8054\_2 XI, and  $0.0057 \text{ g D-xylose h}^{-1} \text{ g(DCW)}^{-1}$  for opt.PiXI. This represents a 72% higher D-xylose isomerization by 8054\_2 XI than opt.PiXI. Absolute xylitol production was identical for both strains ( $\sim 8 \text{ g L}^{-1}$  after 138 h) due to the background activity of aldose reductases (such as *GRE3*) that reduce D-xylose to xylitol<sup>40</sup>. Xylitol has been shown to inhibit XI activity in-vitro<sup>8,41</sup>, although in our experiments, xylitol appears not to affect D-xylose consumption over time as it appears linear for the duration of the experiment for both strains.

**Kinetic parameters and optimal temperature.** Kinetic properties of the enzymes 8054\_2 XI and opt.PiXI were determined using crude cell extracts prepared from the recombinant diploid strains by a bead beating method. There seem to be no significant differences in the apparent protein expression level between both XIs in the cell extracts (see Fig. S3, Supplementary Information). A Michaelis–Menten curve was adjusted to the experimental values of both enzymes with close fit (Fig. 5). The extracts containing 8054\_2 XI converted D-xylose at a maximal rate ( $V_{\max}$ ) of  $0.0296 \text{ U mg}^{-1}$  total protein, and the extracts containing opt.PiXI showed a  $V_{\max}$  of  $0.0112 \text{ U mg}^{-1}$  total protein. The 8054\_2 XI exhibited a  $K_M$  for D-xylose of  $19.27 \pm 0.74 \text{ mM}$ , and opt.PiXI exhibited a  $K_M$  of  $30.76 \pm 2.43 \text{ mM}$ . Thus, the comparative analysis presented in this study indicates that 8054\_2 XI operates at a 2.6 times higher  $V_{\max}$  and with 37% lower  $K_M$  for D-xylose than opt.PiXI. We also assessed optimal enzyme temperature using a discontinuous colorimetric assay to measure relative activity of XIs in the cell extracts at different temperatures. The optimal temperatures of 8054\_2 XI and opt.PiXI were approximately  $60^\circ\text{C}$  and  $65^\circ\text{C}$ , respectively (Fig. 6). Relative activity of both enzymes shows a steady increase from  $30^\circ\text{C}$ , i.e., yeast optimum growth temperature, to the respective peak (100% activity) followed by a strong decrease to the lowest activity at  $80^\circ\text{C}$ . Despite the similar thermophilic behavior, 8054\_2 XI retains 51% of the maximal activity at  $30^\circ\text{C}$ , while activity of opt.PiXI was reduced to 29%.

## Discussion

Novel metagenomic sequence data as well as sequence data in public databases represent a largely untapped resource for novel catalytic capability for future biotechnological processes. Making use of these data requires accurate assembly of sequence reads, gene identification followed by gene synthesis. In this work, we searched metagenomic data we derived from the hindgut of a wood-feeding beetle<sup>20</sup> for similarity to known products of D-xylose isomerases genes, which depends on correctly translating metagenome sequence data *in-silico*. Gene synthesis remains a comparatively expensive approach relative to PCR amplification of DNA from complex extracts, with synthesis cost, time, and error rates increasing with sequence length. However, as well as eliminating PCR bias, an added benefit of gene synthesis is that its codon usage can be optimized for a specific host. Despite the expense, the use of such optimized genes is growing in popularity as evidenced by the twenty percent of sequences deposited at Addgene being composed of codon optimized genes<sup>42</sup>. Although some XI enzymes have been successfully expressed in *S. cerevisiae* allowing more or less efficient D-xylose metabolism, the screening for XIs with significant expression in yeast has proved difficult and often require a suite of optimizations<sup>8,14,15,43</sup>. Our strategy used here allowed direct selection of one very efficient XI from three synthesized candidate genes chosen from the three closest homologs to a known gene without any further evolutionary adaptation of the *S. cerevisiae* host. The separation of XI and the other genes necessary for efficient growth on D-xylose facilitate high-throughput functional screening which would enable testing of many more genes when gene synthesis costs are more permissive<sup>44</sup>.

D-xylose isomerases have been found in lignocellulosic material such as decomposing biomass, the gut of animals or their feces<sup>15,17–19</sup> presumably due to the opportunity to consume D-xylose. Little is known about the relevance of phylogenetic relationship of different XIs for their efficient expression in *S. cerevisiae*, or whether certain microbial groups contain a pool of XIs with higher probability of being expressed in yeast. One of three of the XI enzymes screened in this work was successfully expressed (8054\_2 XI) and clustered within the *Bacteroidetes* phylum (Fig. 2a). Curiously, all the previously assayed XIs from *Bacteroidia* and *Clostridia* (with exception to *T. thermosulfurigenes*) exhibited considerable activity in yeast (Fig. 2a—dashed squares). These classes of organisms have crucial ecological traits in common, such as an anaerobic metabolism and a remarkable capacity for degrading polysaccharides<sup>37,45–47</sup>. This indicates that *Bacteroidia* and *Clostridia*, particularly species residing in niches where active degradation of lignocellulosic biomass could be considered a selective pressure, may represent a valuable source of new XIs that express efficiently in yeast.

The growth rates on D-xylose presented in this work are comparable to others using non-evolved *S. cerevisiae* CEN.PK-based strains with identical genetic modifications. Comparatively, 8054\_2 XI produced 50% higher growth rates and 72% higher D-xylose isomerization than the XI from *Piromyces* sp. under identical conditions, which indicates that the 8054\_2 XI is an improved option for metabolic engineering of *S. cerevisiae* to transform D-xylose. Nevertheless, identical strains carrying the XI from *Piromyces* sp. have exhibited very distinct growth rates, which vary from 0.02 to 0.07 h<sup>-18,17,38</sup>, or 0.21 to 0.22 h<sup>-124,43</sup>. The growth rates conferred by the XI enzymes expressed in this work are closer to the first group of values. The second group is among the highest specific growth rates on D-xylose reported in literature for any XI or evolved strain<sup>22,23,32,48,49</sup>. This discrepancy may be attributed to a wide range of factors including unknown beneficial genetic modifications.

Xylitol has been identified as a competitive inhibitor of XIs<sup>41</sup>. Interestingly, contradictory results on xylitol inhibition have been reported. According to in-vitro assays, XI from *Piromyces* sp. is strongly inhibited in-vitro at a concentration of 50 mM xylitol, and an inhibition constant of 4.6 mM was determined<sup>8</sup>. Similarly, XI from *Bacteroides stercoris*, with an inhibition constant of 5.1 mM xylitol, showed a 50% decrease in activity at a concentration of 30 mM xylitol, although the fermentation rate only declined by less than 10%<sup>28</sup>. In our study, xylitol had no noticeable effect on D-xylose isomerization by either enzyme even at a concentration of approximately 50 mM (8 g L<sup>-1</sup>).

The kinetic parameters of the XI from *Piromyces* sp. reported in literature are very dissimilar, reflecting the vast experimentation with this enzyme (Table 1, entry #2). To overcome this limitation, in our study, 8054\_2 XI and opt.PiXI were compared under identical conditions to eliminate this source of uncertainty. The 8054\_2 XI showed significantly higher specific activity and affinity relative to opt.PiXI, both correlated with higher relative activity at 30 °C. As most of the XIs expressed in yeast are thermophilic, enzyme activity is dramatically reduced at industrial temperatures of fermentation that vary from 30 to 40 °C<sup>50</sup>. Only one yeast-expressed XI in the mesophilic range has been described to date<sup>26</sup>, with an optimal temperature of 37 °C and retaining 71% of activity at 30 °C, more than twice that we observed for opt.PiXI at 30 °C. We hypothesize that a mesophilic profile of XIs may be a major feature for efficient D-xylose utilization by *S. cerevisiae*. Although the 8054\_2 XI we describe here has a thermophilic temperature optimum, we note its activity over a broad thermal range that extends into mesophilic temperatures. This broad range of temperature activity is perhaps not surprising given its origin from the digestive tract of an invertebrate that cannot regulate its body temperature (i.e. is an ectotherm) unlike the endothermic (elephant) habitat of *Piromyces* sp. strain E2 from which opt.PiXI was derived<sup>51</sup>.

To our knowledge, six studies have experimentally compared the XI from *Piromyces* sp. to other XIs. One XI displayed higher affinity but lower specific activity<sup>19</sup>, and another exhibited lower affinity and activity but also lower inhibition by xylitol<sup>8</sup>. A further report identified an enzyme with lower affinity but higher specific activity and another with lower affinity and specific activity<sup>18</sup>, while two additional studies described enzymes with lower specific activity<sup>17,32</sup>. Only one new XI showed higher affinity and specific activity relative to the *Piromyces* sp. XI, which could be further enhanced by evolutionary adaptation<sup>23</sup>. However, as the XI from *Piromyces* sp. was not codon optimized, the comparison is hard to interpret directly. Our study reveals, for the first time, a new XI that exhibits clearly superior kinetics than the XI from *Piromyces* sp. E2 which translates into faster yeast growth and D-xylose consumption in-vivo.

## Conclusions

We have synthesized three putative *xylA* genes from the gut microbiota of a wood-feeding beetle and screened for D-xylulose isomerase activity in *S. cerevisiae*. One of these enzymes, 8054\_2 XI, expressed actively and showed high identity with XIs from the *Bacteroidia* class of the *Bacteroidetes* phylum. Curiously, the phylogenetic analysis revealed that all XIs from *Bacteroidia* screened in yeast have expressed successfully. The new enzyme showed higher specific activity and affinity for D-xylulose than the current gold-standard from *Piromyces* sp., as well as substantially higher relative activity at 30 °C. Superior kinetics of 8054\_2 XI correlated with higher specific growth rate and D-xylulose consumption. The novel XI represents a highly valuable addition to the *S. cerevisiae* molecular toolbox and shows promise for improved industrial conversion of carbohydrate substrates.

## Methods

**Selection of XI from the metagenome of *O. disjunctus*.** Previously published predicted genes and proteins from the metagenome of *O. disjunctus*<sup>20</sup> were screened against the pfamA-30 database<sup>52</sup> using HMMER's hmmscan<sup>53</sup>. Putative metagenome-predicted XI amino acid sequences were aligned together with the XI of *Piromyces* sp. using ClustalW and a phylogenetic tree reconstructed with IQ-TREE using the evolutionary model LG + G4. After the phylogenetic analysis, three sequences were selected and further characterized in the SWISS-MODEL workspace<sup>33</sup> against the SWISS MODEL template library<sup>54</sup> and the obtained models evaluated based on amino acid sequence identity, Global Model Quality Estimation (GMQE, evaluated from 0 to 1), and the QMEAN Z-score<sup>55</sup> (evaluated from -4 to 0).

**Plasmid constructions.** A detailed description of all vectors constructed as a part of this work is available in the form of a collection of Jupyter notebooks<sup>56</sup>. These notebooks contain Python code describing the details of the cloning with the help of the pydna package<sup>57</sup>. These notebooks allow reexamination of the cloning strategies in detail by executing the code. The executable documentation is available in a Git repository ([https://github.com/MetabolicEngineeringGroupCBMA/Silva\\_et\\_al\\_2020](https://github.com/MetabolicEngineeringGroupCBMA/Silva_et_al_2020)). The plasmid pLBL3 was used to express candidate D-xylulose isomerase genes. The pLBL3 is an expression vector with a *LEU2* auxotrophic marker with both 2 $\mu$  and pUC origins of replication. The expressed gene is controlled upstream by a TEF1 promoter (the intergenic sequence between MRL1/YPR079W and TEF1/YPR080W) and downstream by the intergenic sequence between YNL095C and RPS7B/YNL096C. Codon-optimized XI genes were synthesized by Integrated DNA Technologies, Inc (Coralville, IA, USA) and were cloned into pLBL3 by in-vivo gap repair between tailed PCR products of individual genes and the plasmid linearized with *AjiI* (Thermo Fisher Scientific Inc, Waltham, MA, USA) (Fig. 1a) in *S. cerevisiae* EBY.VW4000.

The *S. cerevisiae* CEN.PK111-61A was transformed with the pYPK0\_XTTRRG vector (Fig. 1b) expressing a partial D-xylulose utilization pathway. The vector expresses six different genes, a xylulokinase (*XKS1*), that converts D-xylulose to xylulose 5-phosphate, a D-xylulose/glucose facilitator from *Candida intermedia* (*Gxf1*), and the four genes of the non-oxidative pentose phosphate pathway *TKL1*, *TAL1*, *RPE1* and *RKII*. *TKL1* codes for a transketolase that convert xylulose-5-phosphate and ribose-5-phosphate to sedoheptulose-7-phosphate and glyceraldehyde-3-phosphate; *TAL1* codes for a transaldolase that converts sedoheptulose 7-phosphate and glyceraldehyde 3-phosphate to erythrose 4-phosphate and fructose 6-phosphate; *RPE1* codes for a D-ribulose-5-phosphate 3-epimerase that converts D-ribulose 5-phosphate to D-xylulose 5-phosphate. *RKII* codes for a ribose-5-phosphate ketol-isomerase that interconverts ribose 5-phosphate and ribulose 5-phosphate. Yeast transformations were performed as described by the high-efficiency protocol using lithium acetate, ssDNA and polyethylene glycol 3350<sup>58</sup>.

**Strains and cultivation.** *Escherichia coli* strain XL1-Blue (Stratagene, La Jolla, CA, USA) was used for routine plasmid preparation. *E. coli* strains were cultivated on lysogeny broth (LB-Lennox) containing 1% (w/v) tryptone (BD biosciences, San Jose, CA, USA), 0.5% (w/v) yeast extract (Panreact AppliChem, Darmstadt, Germany), 0.5% (w/v) sodium chloride and 100 mg L<sup>-1</sup> ampicillin (Formedium, King's Lynn, UK). *S. cerevisiae* strains were cultivated on complex media containing 2% (w/v) bacto-peptone (BD biosciences, San Jose, CA, USA), 1% (w/v) yeast extract, and 2% (w/v) glucose (YPD), maltose (YPM), or D-xylulose (YPX). Yeast strains were also cultivated in defined synthetic complete media (SC) containing 0.67% (w/v) yeast nitrogen base without amino acids (BD, Franklin Lakes, NJ, USA), 0.07% amino acid dropout Kaiser mixture<sup>59</sup>, 50 mM potassium hydrogen phthalate, and 2% (w/v) glucose, maltose or D-xylulose. SC media were adjusted to pH 5.5 using sodium hydroxide. Amino acids histidine, uracil and tryptophan were omitted as required for selection of auxotrophic markers. Agar was added to a concentration of 2% (w/v) for solid media. Liquid cultures were incubated with shaking at 200 rpm, yeast at 30 °C and *E. coli* at 37 °C. *S. cerevisiae* strains and plasmids used in this work are listed in Table 2.

**Yeast mating.** Each EBY.VW4000 (MATa) strain carrying a pLBL3\_XI plasmid was mated with the CEN.PK111-61A/pYPK0\_XTTRRG (MATa) strain (Fig. 1c), resulting in diploid yeast strains. Briefly, EBY.VW4000 (pLBL3\_XI) and CEN.PK111-61A (pYPK0\_XTTRRG) were cultivated in 5 mL YPD medium overnight. Cultures were then diluted in 5 mL YPD medium to an OD600 of 0.2 and grown for at least two generations, to a final OD600 of 0.8–1.2. Cells were harvested by centrifugation and transferred to 1-mL Eppendorf tube. Cell suspensions were washed twice with deionized water and resuspended in 1 mL rich medium. A hundred microliters of each culture were pooled, mixed and incubated overnight at room temperature. The pooled cultures were washed twice in 1 mL deionized water and resuspended in 1 mL water. A hundred microliters of the washed cultures were spread on SC medium. The resulting diploid clones were collected after two days.



Strain	Relevant genotype	References
EBY.VW4000	<i>MATa ura3-52 his3-Δ1 leu2-3,112 trp1-289</i>	60
CEN.PK111-61A	<i>MATa ura3-52 his3-Δ1 leu2-3,112 TRP1</i>	61
Plasmid	Relevant features	References
pYPK0_XTTRRG	<i>URA3; XKS1, TAL1, TKL1, RPE1, RKL1, Gxf1</i>	This work
pLBL3	<i>LEU2</i>	This work
pLBL3_8054_2	<i>LEU2; 8054_2 xylA</i>	This work
pLBL3_15405_2	<i>LEU2; 15405_2 xylA</i>	This work
pLBL3_1362_6	<i>LEU2; 1362_6 xylA</i>	This work
pLBL3_opt.PiXI	<i>LEU2; Piromyces sp. E2 xylA</i>	This work
pLBL3_XR/XDH	<i>LEU2; XR (XYL1); XDH (XYL2)</i>	This work

**Table 2.** *Saccharomyces cerevisiae* strains and plasmids used in this study.

**Screening for growth on solid medium.** Screening was performed by scoring yeast growth on solid media with D-xylose as the sole, or main, carbon source (Fig. 1d). Recombinant diploids were streaked into 5 mL SC medium containing glucose, lacking leucine, uracil and tryptophan and incubated overnight. Subsequently, 1 mL of culture was washed twice and diluted 1:10 in deionized water. Cell suspensions were incubated at 30 °C overnight and then placed on ice for 2 h. Cultures were serially diluted 1:10 in deionized water. Drops with a volume of 10 μL of dilutions 1, 10<sup>-1</sup> and 10<sup>-2</sup> were spotted on YPX and SC media with D-xylose. Plates were incubated for 5 days before scoring growth.

**Liquid culture growth assays.** Individual colonies were used to inoculate SC media containing glucose and grown overnight. Yeast cells were washed and transferred to pre-warmed (30 °C) SC medium containing D-xylose at an OD600 of 0.2 and incubated in 50-mL glass tubes for 24 h. Cells were then transferred to 5 mL identical fresh medium in 50-mL shaking-flasks at an OD600 of 0.2 to start measurements. Absorbance was measured using a NanoDrop 1000 spectrophotometer. Cell growth experiments were performed in triplicate.

**D-Xylose consumption.** For measuring specific D-xylose consumption, cells from frozen stocks (– 80 °C) were transferred into 50 mL SC media containing glucose and incubated in 500-mL culture flasks for 24 h. Cells were harvested, washed and suspended in 25 mL identical medium containing 4% (w/v) D-xylose, instead of glucose. The inoculation level was 10 g L<sup>-1</sup> dry cell weight (DCW). Cultivation was performed in 500-mL flasks and samples were taken every 24 h. D-xylose and xylitol concentrations were quantified by high performance liquid chromatography (Hitachi LaChrom Elite) and detection was done by refractive index with an Elite LaChrom L-2490 RI detector (VWR Hitachi). Samples were prepared by adding 125 μL trichloroacetic acid to a final concentration of 10% (v/v) followed by a 2 h incubation on ice. Samples were centrifuged at 4 °C for 15 min and filtered using a 0.22 μm pore hydrophilic PTFE filter. Samples were separated on a Rezex 8 μm ROA-organic acid H<sup>+</sup>(8%) column (Phenomenex) with a mobile phase of 2.5 mM H<sub>2</sub>SO<sub>4</sub>, and a flow rate of 0.5 mL min<sup>-1</sup> at 40 °C. Data was analyzed with the EZChrom Elite 3.3.2 SP2 software. D-xylose consumption experiments were performed in triplicate.

**D-xylose isomerase activity assay.** Yeast cells expressing XI enzymes (*Piromyces* sp., 8054\_2) were grown overnight in SC media containing glucose. Cultures were diluted in 50 mL same media at an OD600 of 0.3 and incubated for three generations. Cells were harvested, washed twice with water and suspended in 100 mM Tris–HCl buffer pH of 7.5, followed by disruption with glass beads (0.45 mm) using FastPrep FP120 cell disrupter (6.0 oscillations min<sup>-1</sup> for 20 s). Cell debris was removed by centrifugation at 16,000×g for 10 min and the supernatant was conserved on ice. Crude cell extracts were used in the enzyme assays immediately after preparation. Protein concentration was measured by the Bradford assay having bovine serum albumin as standard. All of the enzyme assays were performed in triplicate.

**Resorcinol method.** Resorcinol-based activity assay was performed in this work for determination of optimal temperature of activity of the XI enzymes<sup>62,63</sup>. Cell extracts were diluted in 100 mM Tris–HCl buffer at pH of 7.5. Experiments were carried out by adding 30 μL cell extracts containing 3 mg mL<sup>-1</sup> total protein to 20 μL test solution (100 mM Tris–HCl at pH of 7.5, 30 mM D-xylose, 0.3 mM MnSO<sub>4</sub>), and then incubated at the testing temperature for an appropriate period of time previously optimized. Sample incubation was carried out into BioRad T100 thermal cycler set at specific temperatures. For color development, 150 μL of a 1:1 mixture (v/v) of solution A (0.05% resorcinol in ethanol) and solution B (0.216 g FeNH<sub>4</sub>(SO<sub>4</sub>)<sub>2</sub> · 12 H<sub>2</sub>O in 1 L concentrated HCl) were added to the samples followed by incubation at 80 °C for 40 min. The absorbance measurements were determined on the microplate reader SpectraMax Plus 384 (Molecular Devices) at 630 nm.

**d-Sorbitol dehydrogenase (SDH) method.** The SDH method was employed to determine specific activity of the XIs. Assays were performed in mixtures containing 100 mM Tris–HCl (pH value of 7.5), 10 mM MgCl<sub>2</sub>, 0.15 mM NADH, 2 U SDH (Roche Diagnostics, Mannheim, Germany), and 1 mg mL<sup>-1</sup> cell extract<sup>15</sup>. The reac-

tions started with the addition of the mixture to D-xylose solutions that ranged from 5 to 400 mM. The reactions occurred at 30 °C and were monitored spectrophotometrically (SpectraMax Plus 384, Molecular Devices) at 340 nm through the depletion of NADH resulting from the reduction of xylulose to xylitol by SDH. Kinetic parameters were determined from the interpolation of the experimental data to the Michaelis–Menten curve by the least squares fit method using Graphpad Prism 6.

**Phylogenetic analysis.** The amino acid sequences from the metagenome-derived sequences and those from reported studies were aligned using Clustal Omega with default parameters<sup>64</sup>. This software generated a percent identity matrix that was used to compare similarity among the sequences of the enzymes, and an alignment file in “ClustalW” format for phylogenetic tree construction. This file was converted to a usable “Phylip” format<sup>65</sup>. The phylogenetic trees were constructed using PhyML 3.0 on the web server ATGC-Montpellier, using WAG substitution model, mix of NNI/SPR improvements, 8 substitution rate categories, and 1000 bootstrap replicates<sup>66</sup>. The resulting trees were visualized and edited with the on-line software “Interactive Tree of Life”—iTOL v4<sup>67</sup>. GenBank accession numbers of the amino acid sequences used to generate the phylogenetic tree are available in Table S2, Supplementary Information.

## Data availability

All data generated or analyzed during this study are included in this published article, public repositories, or in the Supplementary Information, as indicated in the manuscript.

Received: 16 October 2020; Accepted: 9 February 2021

Published online: 26 February 2021

## References

- Arevalo-Gallegos, A., Ahmad, Z., Asgher, M., Parra-Saldivar, R. & Iqbal, H. M. N. Lignocellulose: A sustainable material to produce value-added products with a zero waste approach—A review. *Int. J. Biol. Macromol.* **99**, 308–318 (2017).
- Dodd, D. & Cann, I. K. O. Enzymatic deconstruction of xylan for biofuel production. *Glob. Change Biol. Bioenergy* **1**, 2–17 (2009).
- Cunha, J. T., Romani, A., Costa, C. E., Sá-Correia, I. & Domingues, L. Molecular and physiological basis of *Saccharomyces cerevisiae* tolerance to adverse lignocellulose-based process conditions. *Appl. Microbiol. Biotechnol.* **103**, 159–175 (2019).
- Moysés, D. N., Reis, V. C. B., de Almeida, J. R. M., de Moraes, L. M. P. & Torres, F. A. G. Xylose fermentation by *Saccharomyces cerevisiae*: challenges and prospects. *Int. J. Mol. Sci.* **17**, 207 (2016).
- Kötter, P., Amore, R., Hollenberg, C. P. & Ciriacy, M. Isolation and characterization of the *Pichia stipitis* xylitol dehydrogenase gene, XYL2, and construction of a xylose-utilizing *Saccharomyces cerevisiae* transformant. *Curr. Genet.* **18**, 493–500 (1990).
- Kötter, P. & Ciriacy, M. Xylose fermentation by *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **38**, 776–783 (1993).
- Wahlbom, C. F., Eliasson, A. & Hahn-Hägerdal, B. Intracellular fluxes in a recombinant xylose-utilizing *Saccharomyces cerevisiae* cultivated anaerobically at different dilution rates and feed concentrations. *Biotechnol. Bioeng.* **72**, 289–296 (2001).
- Brat, D., Boles, E. & Wiedemann, B. Functional expression of a bacterial xylose isomerase in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **75**, 2304–2311 (2009).
- Briggs, K. A., Lancashire, W. E. & Hartley, B. S. Molecular cloning, DNA structure and expression of the *Escherichia coli* D-xylose isomerase. *EMBO J.* **3**, 611–616 (1984).
- Sarthy, A. V. *et al.* Expression of the *Escherichia coli* xylose isomerase gene in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **53**, 1996–2000 (1987).
- Amore, R., Wilhelm, M. & Hollenberg, C. P. The fermentation of xylose: An analysis of the expression of *Bacillus* and *Actinoplanes* xylose isomerase genes in yeast. *Appl. Microbiol. Biotechnol.* **30**, 351–357 (1989).
- Hallborn, J. *Metabolic Engineering of Saccharomyces cerevisiae: Expression of Genes Involved in Pentose Metabolism* (Lund University, Lund, 1995).
- Moes, C. J., Pretorius, I. S. & van Zyl, W. H. Cloning and expression of the *Clostridium thermosulfurogenes* D-xylose isomerase gene (xylA) in *Saccharomyces cerevisiae*. *Biotechnol. Lett.* **18**, 269–274 (1996).
- Walfridsson, M. *et al.* Ethanolic fermentation of xylose with *Saccharomyces cerevisiae* harboring the *Thermus thermophilus* xylA gene, which expresses an active xylose (glucose) isomerase. *Appl. Environ. Microbiol.* **62**, 4648–4651 (1996).
- Kuyper, M. *et al.* High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*? *FEMS Yeast Res.* **4**, 69–78 (2003).
- Murphy, C. L. *et al.* Horizontal gene transfer as an indispensable driver for evolution of neocallimastigomycota into a distinct gut-dwelling fungal lineage. *Appl. Environ. Microbiol.* **85**, 15. <https://doi.org/10.1128/AEM.00988-19> (2019).
- Parachin, N. S. & Gorwa-Grauslund, M. F. Isolation of xylose isomerases by sequence- and function-based screening from a soil metagenomic library. *Biotechnol. Biofuels* **4**, 9 (2011).
- Hou, J. *et al.* Characterization and evolution of xylose isomerase screened from the bovine rumen metagenome in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* <https://doi.org/10.1016/j.jbiosc.2015.05.014> (2015).
- Katahira, S. *et al.* Screening and evolution of a novel protist xylose isomerase from the termite *Reticulitermes speratus* for efficient xylose fermentation in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* **10**, 203 (2017).
- Ceja-Navarro, J. A. *et al.* Gut anatomical properties and microbial functional assembly promote lignocellulose deconstruction and colony subsistence of a wood-feeding beetle. *Nat. Microbiol.* **4**, 864–875 (2019).
- Suh, S.-O., McHugh, J. V., Pollock, D. D. & Blackwell, M. The beetle gut: A hyperdiverse source of novel yeasts. *Mycol. Res.* **109**, 261–265 (2005).
- Lee, S.-M., Jellison, T. & Alper, H. S. Directed evolution of xylose isomerase for improved xylose catabolism and fermentation in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **78**, 5708–5716 (2012).
- Hector, R. E., Dien, B. S., Cotta, M. A. & Mertens, J. A. Growth and fermentation of D-xylose by *Saccharomyces cerevisiae* expressing a novel D-xylose isomerase originating from the bacterium *Prevotella ruminicola* TC2-24. *Biotechnol. Biofuels* **6**, 84 (2013).
- Verhoeven, M. D. *et al.* Mutations in PMR1 stimulate xylose isomerase activity and anaerobic growth on xylose of engineered *Saccharomyces cerevisiae* by influencing manganese homeostasis. *Sci. Rep.* **7**, 46155 (2017).
- Lee, M., Rozeboom, H. J., Keuning, E., de Waal, P. & Janssen, D. B. Structure-based directed evolution improves *S. cerevisiae* growth on xylose by influencing in vivo enzyme performance. *Biotechnol. Biofuels* **13**, 5 (2020).
- Madhavan, A. *et al.* Xylose isomerase from polycentric fungus *Orpinomyces*: gene sequencing, cloning, and expression in *Saccharomyces cerevisiae* for bioconversion of xylose to ethanol. *Appl. Microbiol. Biotechnol.* **82**, 1067–1078 (2009).
- Seike, T. *et al.* Molecular evolutionary engineering of xylose isomerase to improve its catalytic activity and performance of micro-aerobic glucose/xylose co-fermentation in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* **12**, 139 (2019).

28. Ha, S.-J., Kim, S. R., Choi, J.-H., Park, M. S. & Jin, Y.-S. Xylitol does not inhibit xylose fermentation by engineered *Saccharomyces cerevisiae* expressing xylA as severely as it inhibits xylose isomerase reaction in vitro. *Appl. Microbiol. Biotechnol.* **92**, 77–84 (2011).
29. Aeling, K. A. *et al.* Co-fermentation of xylose and cellobiose by an engineered *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.* **39**, 1597–1604 (2012).
30. de Figueiredo Vilela, L. *et al.* Functional expression of Burkholderia cenocepacia xylose isomerase in yeast increases ethanol production from a glucose-xylose blend. *Bioresour. Technol.* **128**, 792–796 (2013).
31. Vieira, I. P. V. *et al.* Understanding xylose isomerase from Burkholderia cenocepacia: Insights into structure and functionality for ethanol production. *AMB Express* **9**, 73 (2019).
32. Peng, B., Huang, S., Liu, T. & Geng, A. Bacterial xylose isomerases from the mammal gut Bacteroidetes cluster function in *Saccharomyces cerevisiae* for effective xylose fermentation. *Microb. Cell Fact.* **14**, 70 (2015).
33. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
34. Pereira, F. *et al.* Yeast pathway kit: A method for metabolic pathway assembly with automatically simulated executable documentation. *ACS Synth. Biol.* **5**, 386–394 (2016).
35. Sakamoto, M. The family Porphyromonadaceae. In *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (eds Rosenberg, E. *et al.*) 811–824 (Springer, Berlin Heidelberg, 2014).
36. Ramasamy, D. *et al.* Genome sequence and description of *Bacteroides timonensis* sp. nov. *Stand. Genom. Sci.* **9**, 1181–1197 (2014).
37. Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut bacteroidetes: The food connection. *Front. Microbiol.* **2**, 93 (2011).
38. Karhumaa, K., Sanchez, R. G., Hahn-Hägerdal, B. & Gorwa-Grauslund, M.-F. Comparison of the xylose reductase-xylitol dehydrogenase and the xylose isomerase pathways for xylose fermentation by recombinant *Saccharomyces cerevisiae*. *Microb. Cell Fact.* **6**, 5 (2007).
39. Li, X. *et al.* Comparison of xylose fermentation by two high-performance engineered strains of *Saccharomyces cerevisiae*. *Biotechnol. Rep.* **9**, 53–56 (2016).
40. Träff, K. L., Otero Cordero, R. R., van Zyl, W. H. & Hahn-Hägerdal, B. Deletion of the GRE3 aldose reductase gene and its influence on xylose metabolism in recombinant strains of *Saccharomyces cerevisiae* expressing the xylA and XKS1 genes. *Appl. Environ. Microbiol.* **67**, 5668–5674 (2001).
41. Yamanaka, K. Inhibition of D-xylose isomerase by pentitols and D-lyxose. *Arch. Biochem. Biophys.* **131**, 502–550 (1969).
42. Kunjapur, A. M., Pfingst, P. & Thompson, N. C. Gene synthesis allows biologists to source genes from farther away in the tree of life. *Nat. Commun.* **9**, 4425 (2018).
43. Kuyper, M. *et al.* Metabolic engineering of a xylose-isomerase-expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation. *FEMS Yeast Res.* **5**, 399–409 (2005).
44. Carlson, R. *On DNA and Transistors. Synthesis.* [http://www.synthesis.cc/synthesis/2016/03/on\\_dna\\_and\\_transistors](http://www.synthesis.cc/synthesis/2016/03/on_dna_and_transistors) (2016).
45. Warnick, T. A. *Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil. *Int. J. Syst. Evol. Microbiol.* **52**, 1155–1160 (2002).
46. Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* **6**, 121–131 (2008).
47. Doi, R. H. & Tamaru, Y. The *Clostridium cellulovorans* cellulosome: an enzyme complex with plant cell wall degrading activity. *Chem. Rec.* **1**, 24–32 (2001).
48. Shen, Y. *et al.* An efficient xylose-fermenting recombinant *Saccharomyces cerevisiae* strain obtained through adaptive evolution and its global transcription profile. *Appl. Microbiol. Biotechnol.* **96**, 1079–1091 (2012).
49. Lee, S.-M., Jellison, T. & Alper, H. S. Systematic and evolutionary engineering of a xylose isomerase-based pathway in *Saccharomyces cerevisiae* for efficient conversion yields. *Biotechnol. Biofuels* **7**, 122 (2014).
50. Carlos, L., Olitta, T. & Nitsche, S. Ethanol production in Brazil: the industrial process and its impact on yeast fermentation. In *Biofuel Production—Recent Developments and Prospects* (ed. Dos Santos Bernardes, M. A.) (InTech, Rijeka, 2011).
51. Harhangi, H. R. *et al.* Xylose metabolism in the anaerobic fungus *Piromyces* sp. strain E2 follows the bacterial pathway. *Arch. Microbiol.* **180**, 134–141 (2003).
52. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
53. Eddy, S. R. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Comput. Biol.* **4**, e1000069 (2008).
54. Bienert, S. *et al.* The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* **45**, D313–D319 (2017).
55. Studer, G. *et al.* QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics* **36**, 2647 (2020).
56. Kluyver, T. *et al.* Jupyter Notebooks: A publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (ed. Loizides, F.) 87–90 (IOS Press, Amsterdam, 2016).
57. Pereira, F. *et al.* Pydna: a simulation and documentation tool for DNA assembly strategies using python. *BMC Bioinform.* **16**, 142 (2015).
58. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87–96 (2002).
59. Kaiser, C., Michaelis, S. & Mitchell, A. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual* (CSHL Press, New York, 1994).
60. Wiczorke, R. *et al.* Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.* **464**, 123–128 (1999).
61. Entian, K.-D. & Kötter, P. 25 Yeast genetic strain and plasmid collections. *Methods Microbiol.* **36**, 629–666 (2007).
62. Schenk, M. & Bisswanger, H. A microplate assay for D-xylose/ D-glucose isomerase. *Enzyme Microb. Technol.* **22**, 721–723 (1998).
63. Kulka, R. G. Colorimetric estimation of ketopentoses and ketohexoses. *Biochem. J.* **63**, 542–548 (1956).
64. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
65. Bikandi, J., San Millán, R., Rementería, A. & Garaizar, J. In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics* **20**, 798–799 (2004).
66. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
67. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

## Acknowledgements

This work was supported by the project FatVal PTDC/EAM-AMB/32506/2017 (POCI-01-0145-FEDER-032506), co-funded by the European Regional Development Fund (ERDF), through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Portugal 2020, and by the Fundação para a Ciência e a Tecnologia—FCT I.P through national funds. CBMA was supported by the “Contrato-Programa” UIDB/04050/2020 funded by national funds through the FCT I.P. PCS is recipient of a FCT PhD fellowship

(SFRH/BD/140039/2018), and was supported by a Fulbright Scholarship Portugal grant from January to May 2020 at Lawrence Berkeley National Laboratory, Berkeley, CA, USA. BJ was awarded a Fulbright grant from The Swedish Fulbright Commission for Visiting Lecturers and Research Scholars between September 2014 and January 2015 visiting Lawrence Berkeley National Laboratory, Berkeley, CA, USA. This work was supported in part by the United States Department of Energy's Genomic Science Program (grant SCW1039). Part of this work was performed at Lawrence Berkeley National Laboratory under US Department of Energy contract number DE-AC02-05CH11231. DNA sequencing was performed at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley, supported by NIH S10 Instrumentation grants S10RR029668 and S10RR027303.

### Author contributions

All authors contributed to the design of the project. P.C.S., J.C.N., F.A. and U.K. performed the experiments and analyzed the data. P.C.S., J.C.N., E.L.B. and B.J. drafted the manuscript. E.L.B. and B.J. revised and edited the final version. All authors read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-83937-z>.

**Correspondence** and requests for materials should be addressed to P.C.S., E.L.B. or B.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021