Comparative genomics of xylose-fermenting fungi for enhanced biofuel production

Dana J. Wohlbach\textsuperscript{a,b}, Alan Kuo\textsuperscript{c}, Trey K. Sato\textsuperscript{d}, Katlyn M. Potts\textsuperscript{a}, Asaf A. Salamov\textsuperscript{e}, Kurt M. LaButti\textsuperscript{f}, Hui Sun\textsuperscript{g}, Alicja Clum\textsuperscript{h}, Jasmin L. Pangilinan\textsuperscript{i}, Erika A. Lindquist\textsuperscript{j}, Susan Lucas\textsuperscript{k}, Alla Lapidus\textsuperscript{l}, Mingjie Jin\textsuperscript{m,n}, Christa Gunawan\textsuperscript{o,p}, Venkatesh Balan\textsuperscript{q,r}, Bruce E. Dale\textsuperscript{s,t}, Thomas W. Jeffries\textsuperscript{b}, Robert Zinkel\textsuperscript{b}, Kerrie W. Barry\textsuperscript{b}, Igor V. Grigoriev\textsuperscript{v}, and Audrey P. Gasch\textsuperscript{b,w,x,y,z}

\textsuperscript{a}Department of Genetics, University of Wisconsin, Madison, WI 53706; \textsuperscript{b}Great Lakes Bioenergy Research Center, Madison, WI 53706; \textsuperscript{c}US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598; \textsuperscript{d}Biomasq Conversion Research Laboratory, Department of Chemical Engineering and Materials Science, Michigan State University, Lansing, MI 48910; and \textsuperscript{e}Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48824

Edited by Chris R. Somerville, University of California, Berkeley, CA, and approved July 5, 2011 (received for review February 24, 2011)

Cellulosic biomass is an abundant and underused substrate for biofuel production. The inability of many microbes to metabolize the pentose sugars abundant within hemi-cellulose creates specific challenges for microbial biofuel production from cellulosic material. Although engineered strains of \textit{Saccharomyces cerevisiae} can use the pentose xylose, the fermentative capacity pales in comparison with glucose, limiting the economic feasibility of industrial fermentations. To better understand xylose utilization for subsequent microbial engineering, we sequenced the genomes of two xylose-fermenting, beetle-associated fungi, \textit{Spathaspora passalidarum} and \textit{Candida tenuis}. To identify genes involved in xylose metabolism, we applied a comparative genomic approach across 14 Ascomycete genomes, mapping phenotypes and genotypes onto the fungal phylogeny, and measured genomic expression across five Hemiascomycete species with different xylose-consumption phenotypes. This approach implicated many genes and processes involved in xylose assimilation. Several of these genes significantly improved xylose utilization when engineered into \textit{S. cerevisiae}, demonstrating the power of comparative methods in rapidly identifying genes for biomass conversion while reflecting on fungal ecology.

bioenergy | genome sequencing | transcriptomics

Biofuel production from cellulosic material uses available substrates without competing with food supplies and therefore presents an economic and environmental opportunity (1). In lignocellulosic plant stocks, which include agricultural residues and wood waste, the second-most abundant sugar after glucose is the pentose xylose. Native \textit{Saccharomyces cerevisiae} (Scer) does not consume xylose but can be engineered for xylose consumption with a minimal set of assimilation enzymes, including xylose reductase (Xyl1) and xylitol dehydrogenase (Xyl2) from the xylose-fermenting \textit{Pichia stipitis} (Psti) (Fig. 1A) (2, 3). However, xylose fermentation remains slow and inefficient in Scer, especially under anaerobic conditions when NADH cannot be recycled for NAD\textsuperscript{+}-dependent Xyl2 (2, 4, 5). Therefore, improving xylose utilization in industrially relevant yeasts is essential for producing economically viable biofuels from cellulosic material.

A handful of Hemiascomycete yeasts naturally ferment pentose sugars (2, 3, 6). Best known is the xylose-fermenting yeast Psti, associated with wood-boring beetles that may rely on fungi to release nutrients from wood (7, 8). Other related yeasts cannot ferment pentoses, suggesting that xylose fermentation has evolved in this unique fungal environment (9). Although some details are known (Fig. 1A) (2, 3, 10–12), much of the mechanism of xylose fermentation remains unresolved.

To elucidate genetic features that underlie xylose utilization, we sequenced the genomes of two additional xylose-fermenting species and applied a cross-species comparative genomic approach. Comparing phenotypic and genomic differences in diverse Hemiascomycete species indicated many genes important for xylose utilization and also reflected the unique niche experienced by these beetle-associated species. Here we present the comparative analysis of the genomes and transcriptomes of these yeasts, highlighting aspects of pentose assimilation as well as the ecological significance of these interesting fungi. In the process, we identified several genes that, when expressed in Scer, significantly improve xylose-dependent growth and xylose assimilation. By harnessing the power of nature and comparative genomics, this work provides a key improvement to xylose utilization, a significant roadblock to cellulosic biofuel production.

Results and Discussion

We sequenced the genomes of two xylose-fermenting yeasts, \textit{Spathaspora passalidarum} (Spas, NRRL Y-27907) and \textit{Candida tenuis} (Cten, NRRL Y-1498), for comparison with the existing \textit{Psti} genome (Materials and Methods and SI Appendix, Materials and Methods and Table S1) (13). The \textit{Spas} genome was sequenced to 43.77X coverage over 13.1 Mb arranged in eight scaffolds. The \textit{Cten} genome was sequenced to 26.9X coverage, generating 10.7 Mb in 61 scaffolds representing eight chromosomes. Compared with other sequenced Hemiascomycetes, genome size and genetic composition in the xylose-fermenting yeasts span the range from compact (5,533 genes in the 10.7-Mb \textit{Cten} genome) to among the largest (5,841 genes in the 15.4-Mb genome of \textit{Psti} and 5,983 genes in the 13.2-Mb genome of \textit{Spas}) (Table 1 and SI Appendix, Table S2). Sixty-seven percent of \textit{Spas} and 74% of \textit{Cten} genes are orthologs located in syntenic regions (SI Appendix, Fig. S1), and about half of all genes in \textit{Spas}, \textit{Cten}, and \textit{Psti} show three-way synteny.

Xylose Consumers Are Members of the CUG Clade of Commensal Fungi. We selected 11 other Ascomycetes with available genome sequences (Table 1) for comparison with \textit{Spas}, \textit{Cten}, and \textit{Psti} (Fig. 1B and C). Whole-genome phylogenetic analysis placed both \textit{Spas} and \textit{Cten} within the CUG clade of yeasts (Fig. 1B and SI Appendix, Materials and Methods) named for the alternative decoding of the CUG codon as serine instead of leucine (14–16). We compared tRNA sequences across the 14 species in our


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The assemblies and annotations reported in this paper have been deposited in the GenBank database [accession nos. AE000000000 (\textit{Spathaspora passalidarum}) and AE000000000 (\textit{Candida tenuis})]. Microarray data have been deposited in the Gene Expression database (accession no. GSE24858).

Freely available online through the PNAS open access option.

The authors declare no conflict of interest.

This article contains supporting information online at www.pnas.org/cgi/doi/10.1073/pnas.1103039108.
Interestingly, most other species in this CUG group are commensal with humans but can emerge as opportunistic pathogens (17, 18). Thus, commensalism, albeit in association with different hosts, appears to be a feature common to this clade.

Clade-Specific Patterns of Gene Presence. To identify genes associated with xylose utilization, we compared gene content in the 14 Ascomycetes in our phylogeny by assigning orthology and paralogy relationships among the metaset of 81,907 predicted fungal protein-coding genes (SI Appendix, Materials and Methods). More than 12,000 orthologous gene groups (OGGs) were resolved, with 5,749 OGGs (91% of all genes) found in at least two species (SI Appendix, Table S3 and Fig. S3A). In contrast, the other OGGs (52% of all OGGs representing 9% of all genes) are species-specific paralogs that are distributed nonrandomly throughout the phylogeny (SI Appendix, Fig. S3B). Within the CUG clade, Debaryomyces Hansenii and Pichia guilliermondii have the most single-species expansions, and the xylose-fermenting fungi (Spas, Cten, and Pse) have some of the fewest. Interestingly, amplifications in the xylose fermenters include sugar transporters and cell-surface proteins, possibly related to their unique sugar environment (SI Appendix, Tables S4 and S5).

We analyzed conservation patterns of the 5,749 multispecies OGGs through a clustering approach, which identified clade-specific OGGs enriched for different functional properties (Fig. 2A and SI Appendix, Table S6). Approximately half of the multispecies OGGs are common to all 14 Ascomycetes. These ubiquitous OGGs are significantly enriched for essential metabolic processes including nucleic acid (\(p = 1.32e-42\), hypergeometric distribution), small molecule (\(p = 6.28e-35\)), and protein (\(p = 2.51e-14\)) metabolism as well as for transcription (\(p = 2.76e-23\)) and response to stress (\(p = 1.30e-31\)).

The remaining OGGs can be clustered into five major clade-specific groups. Remarkably, the majority of clade-specific OGGs (including those unique to well-studied fungi such as Scer) are significantly enriched for unclassified and uncharacterized proteins (\(p = 4.27e-21\)). This finding reveals a general bias in our understanding of gene function and highlights the dearth of information on species-specific processes, even for the best-characterized organisms like Scer.

OGGs unique to the CUG clade are enriched for genes encoding lipases and cell-surface proteins (\(p = 1.306e-6\) and 6.665e-6, respectively), as previously noted in Candida species (19). Although enrichment of these genes in Candida species was interpreted previously as being important for pathogenicity (19), their presence in beetle symbionts suggests they may be relevant.

Table 1. Strain sources and genome statistics

<table>
<thead>
<tr>
<th>Organism</th>
<th>Strain</th>
<th>Genome size (Mb)</th>
<th>% GC</th>
<th>Total ORFs</th>
<th>Sequencing coverage</th>
<th>Data source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sp. passalidarum</em> (Spas)</td>
<td>NRRL Y-27907</td>
<td>13.2</td>
<td>42.0</td>
<td>5983</td>
<td>44X</td>
<td>DOE JGI</td>
<td>This work</td>
</tr>
<tr>
<td><em>C. tenuis</em> (Cten)</td>
<td>NRRL Y-1498</td>
<td>10.7</td>
<td>42.9</td>
<td>5533</td>
<td>27X</td>
<td>DOE JGI</td>
<td>This work</td>
</tr>
<tr>
<td><em>P. stipitis</em> (Pst)</td>
<td>CBS 6054</td>
<td>15.4</td>
<td>42.3</td>
<td>5841</td>
<td>Complete</td>
<td>DOE JGI</td>
<td></td>
</tr>
<tr>
<td><em>C. albicans</em> (Calb)</td>
<td>WO-1</td>
<td>14.4</td>
<td>33.5</td>
<td>6157</td>
<td>10X</td>
<td>Broad Institute</td>
<td>49</td>
</tr>
<tr>
<td><em>C. tropicalis</em> (Cro)</td>
<td>MYA-3404</td>
<td>14.6</td>
<td>33.1</td>
<td>6258</td>
<td>10X</td>
<td>Broad Institute</td>
<td>19</td>
</tr>
<tr>
<td><em>C. lusitaniae</em> (Clus)</td>
<td>ATCC 42720</td>
<td>12.1</td>
<td>46.8</td>
<td>5936</td>
<td>9X</td>
<td>Broad Institute</td>
<td>19</td>
</tr>
<tr>
<td><em>Debaryomyces hansenii</em> (Dhan)</td>
<td>CBS767</td>
<td>12.2</td>
<td>37.5</td>
<td>6887</td>
<td>10X</td>
<td>Genolevures</td>
<td>50</td>
</tr>
<tr>
<td><em>L. elongisporus</em> (Leol)</td>
<td>NRRL Y-4239</td>
<td>15.5</td>
<td>40.4</td>
<td>5796</td>
<td>9X</td>
<td>Broad Institute</td>
<td>19</td>
</tr>
<tr>
<td><em>P. guilliermondii</em> (Pgu)</td>
<td>ATCC 6260</td>
<td>10.6</td>
<td>44.5</td>
<td>5920</td>
<td>12X</td>
<td>Genolevures</td>
<td>50</td>
</tr>
<tr>
<td><em>C. glabrata</em> (Cgl)</td>
<td>CBS 138</td>
<td>12.3</td>
<td>40.5</td>
<td>5215</td>
<td>8X</td>
<td>Genolevures</td>
<td>50</td>
</tr>
<tr>
<td><em>Kluyveromyces lactis</em> (Klac)</td>
<td>NRRL Y-1140</td>
<td>10.7</td>
<td>40.1</td>
<td>5327</td>
<td>11X</td>
<td>Genolevures</td>
<td>50</td>
</tr>
<tr>
<td><em>S. cerevisiae</em> (Scer)</td>
<td>5288c</td>
<td>12.1</td>
<td>34.4</td>
<td>5695</td>
<td>Complete</td>
<td>SGD</td>
<td>51</td>
</tr>
<tr>
<td><em>Yarrowia lipolytica</em> (Ylip)</td>
<td>CLIB122</td>
<td>20.5</td>
<td>53.7</td>
<td>6436</td>
<td>10X</td>
<td>Genolevures</td>
<td>50</td>
</tr>
<tr>
<td><em>Schizosaccharomyces pombe</em> (Spom)</td>
<td>972h-</td>
<td>12.5</td>
<td>39.6</td>
<td>5004</td>
<td>8X</td>
<td>Weillcome Trust</td>
<td>52</td>
</tr>
</tbody>
</table>

DOE JGI, Department of Energy Joint Genome Institute; SGD, Saccharomyces Genome Database.

Fig. 1. Overview of xylose assimilation and phylogeny of xylose-fermenting fungi. (A) The simplified pathway includes genes that have been engineered in Scer via overexpression (red text) for improved xylose fermentation. GND1, 6-phosphogluconate dehydrogenase; RKT1, ribose-5-phosphate ketol-isomerase; RPE1, ribulose-5-phosphate 3-epimerase; XKS1/XYL3, xylulokinase; xylA, xylose isomerase. (B) Maximum likelihood phylogeny from concatenated alignment of 136 universal orthologs, with bootstrap values. (C) Electron microscopy images of Cten (Top), Pst (Middle), and Spas (Bottom). (Scale bars: 2 μm.)

analysis and confirmed that Spas and Cten harbor the serine tRNA evolved to recognize the CUG codon (14), whereas there were no identifiable sequences similar to standard Scer serine tRNAs (SI Appendix, Fig. S2A and B). Likewise, a genome-wide scan revealed that the majority of CUG codons from Candida and related species (including Spas and Cten) are decoded as serine in Scer orthologs; CUG codons from species outside the CUG clade are decoded as leucine in orthologous Scer genes (SI Appendix, Fig. S2C). Together, these results support the phylogenetic placement of xylose-fermenting species within the CUG clade.
to commensals rather than to pathogenicity per se. Additionally, many genes unique to CUG yeasts are involved in de novo NAD\(^+\) biosynthetic processes (\(p = 0.00891\)), suggesting metabolism that may reflect a more complex environment of these commensal organisms.

Surprisingly, orthologs of known xylose-utilization genes are present in all 14 Ascomycetes, even though most Hemiascomycetes cannot use xylose (6). This group includes orthologs of \(Pst\) \(XYL1\) (11), \(XYL2\) (12), and xylulokinase (\(XYL3\)) (10), the minimal set of genes required to engineer \(Scer\) for xylose assimilation (Fig. 2\(A\)) (2, 3, 5). However, these genes show no evolutionary signatures of selection or constraint to suggest functional modification in the xylose-using species (SI Appendix, Fig. S4). Thus, factors other than the mere presence of this minimal gene set must contribute to phenotypic differences in xylose consumption.

**Conservation of Orthologous Gene Groups Points to Xylose Utilization Genes.** To identify genes relevant to xylose fermentation, we devised a phylogenetic approach to correlate genotype to phenotype across the Ascomycetes. First, we examined xylose growth and fermentation (Fig. 2\(B\) and SI Appendix, Figs. S5 and S6). \(Pst\), \(Spas\), and \(Cten\) were the only species able to ferment xylose measurably in our assay (SI Appendix, Fig. S6). Intriguingly, these species also are the yeasts associated with beetles, many of which are attracted to fermentation byproducts (20). Only three genes are found uniquely in these xylose-fermenting species, one of which contains an \(\alpha\)-glucuronidase domain and a signal peptide sequence indicative of secretion (SI Appendix, Fig. S7). Although its connection to xylose utilization is not clear, this protein may be secreted for degradation of complex carbohydrates in woody biomass.

We expanded our analysis to consider xylose assimilation. Notably, \(L.\) lenispora (\(Lelo\)) is the lone member of the CUG clade unable to grow on xylose (Fig. 2\(B\)), suggesting that the phenotype was present in the group’s common ancestor but lost in this lineage. Because genes involved in sugar metabolism are not maintained in the absence of selection (21, 22), we reasoned that species unable to grow on xylose may have lost key assimilation genes. We therefore looked for genes whose presence and absence across the fungi correlated with the ability to grow on xylose.

Forty-three genes were absent in xylose nongrowers but common to all xylose fermenters, with varying conservation across species that could assimilate xylose (Fig. 2\(C\) and Dataset S1). Fifteen showed presence and absence patterns that correlated strictly with xylose assimilation. These genes include orthologs of a putative \(Pst\) xylose transporter and several endoglucanases that break down higher-order sugars in hemicellulose. Most other genes are unannotated and fungal specific; 10 also are found in other fungi capable of plant cell wall degradation. However, two of the proteins have signal peptide sequences: an oxidoreductase and a putative glycoside hydrolase, both of which potentially could be useful for biomass degradation (see SI Appendix, Fig. S7 for protein domain and signal peptide analysis). Although the conservation of these genes suggests functional importance, we did not detect any signatures of constraint within the xylose fermenters.

**Cross-Species Genomic Expression Identifies Additional Xylose-Responsive Genes.** As a second approach to identify xylose metabolism genes, we characterized genomic expression during glucose versus xylose growth in five species, including the three xylose fermenters, xylose-growing \(C\). albicans (Calb), and \(Lelo\), which is unable to grow on xylose (Materials and Methods). We performed a comparative analysis of orthologous gene expression via hierarchical clustering (Fig. 3 and SI Appendix, Fig. S8) and significance testing (SI Appendix, Tables S7 and S8). The
xylose response was strikingly dissimilar across species (Fig. 3A). In particular, Lelo altered the expression of thousands of genes, including orthologs of the yeast environmental stress response that are induced when Scer is stressed (23) or experiences xylose (SI Appendix, Fig. S8A and Table S9) (24). This massive expression pattern in Lelo likely represents a starvation response to carbon limitation and demonstrates that the environmental stress response is conserved in this species. In addition, Lelo, along with Cten and Calb, induced ∼90 OGGs enriched for fatty acid and lipid catabolism, suggesting reliance on fatty acids as a carbon source (SI Appendix, Fig. S8B and Table S10). We also identified two clusters of genes induced by xylose in most or all species, regardless of their xylose growth phenotypes (Fig. 3B and C). These clusters include genes whose expression is required for optimal xylose utilization in engineered Scer [e.g., XYL1, XYL2, XYL3, transketolase (TKL1), and transaldolase (TAL1)] (Fig. L4). Strikingly, several of these genes were strongly induced in Lelo, even though it cannot use xylose. Thus, remnants of the xylose-signaling cascade persist in Lelo, despite recent loss of xylose assimilation.

In addition to known xylose-metabolism genes, others relating to carbohydrate transport and metabolism were highly induced in xylose growers specifically. Genes encoding β-glucosidases and cellulases were strongly induced, suggesting that xylose participates in a positive-feedback loop to catalyze its own release from hemicellulose. Orthologs of genes metabolizing other carbohydrates (including galactose, maltose, and glucose) were up-regulated also. Thus, in their native environment these species may not encounter free xylose in the absence of complex sugars and are unlikely to rely on it as a sole carbon source. Additionally, the xylose-fermenting species induced several genes linked to redox regeneration, a well-known bottleneck in Scer engineered for xylose fermentation (2, 3). Genes encoding NADPH-generating steps of the pentose phosphate pathway [glucose-6-phosphate dehydrogenase (ZWF1) and phosphoglucone isomerase (PGI1)] were up-regulated, perhaps to feed NADPH-consuming xylose reductase. Other genes implicated in NAD(P)⁺/H⁺ recycling or oxido-reduction were also induced and may function to maintain redox balance during xylose assimilation.

Candidate Genes Improve Xylose Utilization. We tested 10 of the implicated genes for their ability to enhance xylose utilization in two different engineered Scer strains (Materials and Methods and SI Appendix, Note S1). Genetic background influenced the effect of overexpression, and several genes improved growth on both xylose and glucose (SI Appendix, Fig. S9), including a putative hexose transporter, SpHXT, and a glucose-6-phosphate dehydrogenase, SpGPD. Importantly, two genes had a specific positive effect on xylose utilization in one or both strain backgrounds: a Cten aldo/keto reductase, CtAKR, and a Spas unannotated protein, SpINA, with homology to uncharacterized fungal-specific proteins (Fig. 4 and SI Appendix, Fig. S10). Expression of plasmid-born CtAKR significantly improved xylose consumption during both aerobic and anaerobic growth (Fig. 4B). Notably, xylose consumption increased by 32% after 72 h of anaerobic fermentation (p = 0.0369, t test). At the same time, xylitol production relative to xylose consumption was 73% lower (Fig. 4C), indicating improved flux through the xylose-assimilation pathway. Glycerol production, which represents a significant drain on ethanol production under anaerobic con-
ditions (25, 26), was not increased significantly (SI Appendix, Fig. S11). However, acetate production was reduced 42% (Fig. 4C). Because acetate is a weak acid stress for yeast, lower acetate levels could facilitate increased cell growth. Indeed, some of the increased xylose utilization went into biomass production (SI Appendix, Fig. S11); however, the improved xylose utilization did not increase ethanol titers, revealing that ethanol production was limited by factors other than carbon availability. Nonetheless, the significant effect of pCtAKR on anaerobic xylose assimilation and concomitant reduction in xylitol represents a major advance in cellulose biomass conversion by Scer.

CtAKR is a member of the large protein family that includes xylose reductases (SI Appendix, Fig. S124). However, CtAKR is most similar to the NADP+-dependent glycerol dehydrogenase Gcy1 from Scer, which functions in an alternative pathway for glycerol catabolism (Fig. L1) (27). Notably, CtAKR contains residues known to establish NADP+ binding (SI Appendix, Fig. S128) (reviewed in ref. 28), suggesting CtAKR also may function in a NADP+-specific manner. We examined the effect of pCtAKR expression on glycerol metabolism in a Scer mutant lacking three functionally redundant aldo/keto reductases (Gcy1, Ypr1, and GRE3) (Materials and Methods). Glycerol levels increased in the mutant strain but were restored to wild-type levels by pCtAKR (Fig. 4D). Together, these data suggest that CtAKR functions as a NADP+-dependent glycerol dehydrogenase in Scer. Indeed, like CtAKR, overexpression of Scer Gcy1 or Ypr1 had a positive effect on xylose utilization (SI Appendix, Fig. S13), further supporting our hypothesis concerning CtAKR function.

Conclusions

Previous work aimed at improving Scer xylose fermentation focused on metabolic modeling (29), single-species genome and expression analysis (29, 30), or directed evolution (31). In this study, we used a comparative genomics approach to understand xylose utilization in several different beetle-associated fungi. Our approach reveals that these species share some features with other commensal fungi but display specific traits (e.g., the ability to ferment xylose and expression of genes involved in cellulose degradation) that may be specific to their relationship with wood-boring insects. The ability to assimilate xylose is associated with altered expression of several genes central to glycolysis, xylose catabolism, and the pentose phosphate shuttle, revealing that decades of directed evolution largely have recapitulated the natural expression response in these species. That some aspects of this response were observed in species that cannot assimilate xylose (namely Lelo) indicates that remnants of the genomic expression program can remain long after the ability to consume the sugar has been lost.

Additionally, several induced genes are related to reducing potential. Indeed, one of the biggest challenges for xylose fermentation in Scer engineered with Pts YXY1, 2, 3 is the cofactor imbalance that emerges under anaerobic conditions. During anaerobic growth, NADH cannot be recycled through respiration, leading to a shortage of NAD+ to supply Xy2 and thus an accumulation of xylitol (2). To reduce this redox imbalance, Scer increases NADH-dependent glycerol production. We found that overexpression of a Cten glycerol dehydrogenase significantly increased flux through the xylose assimilation pathway without the typical xylitol accumulation. We hypothesize that CtAKR increases cycling through the glycerol metabolic pathway, producing NADPH through alternative glycerol catabolism, which in turn promotes glycerol production and NADH recycling. That glycerol levels do not change significantly in strains engineered with pCtAKR is consistent with this cycling hypothesis. The combined effects may promote the first two steps of xylose assimilation, which require NADPH and NAD+, by helping alleviate cofactor imbalance. Decreased acetate levels also may result from increased glycerol cycling, because otherwise acetate is generated as a fermentation byproduct to alleviate cofactor imbalance (4). Although the precise mechanism will be the subject of future study, our ability to identify genes that improve xylose assimilation shows the promise of harnessing ecology and evolution through comparative genomics for biofuel research.

Materials and Methods

Genome and EST Sequencing, Assembly, and Annotation. We sequenced Spas and Cten using Sanger (40-kb fosmid library) and 454 (standard and paired-end libraries) sequencing platforms. Newbler (v. 2.3; Roche) was used to produce hybrid 454/Sanger assemblies. Gaps were closed by gapResolution (http://www.jgi.doe.gov), PCR and fosmid clone primer walks, or editing in Consed (32). Illumina reads improved the final consensus quality with Polisher (33). mRNA was purified using the Absolutely MrNA purification kit (Stratagene) and was reverse transcribed with SuperScriptIII using dT17VN primer. cDNA was synthesized with Escherichia coli DNA ligase, polymerase I, and RNaseH (Invitrogen), nebulized, and gel purified for fragment sizes between 500–800 bp. Fragments were end repaired, adapter ligated, and made into single-stranded DNA libraries using the GS FLX Titanium library kit. Single-stranded DNA libraries were amplified in bulk and sequenced using a 454 Genome Sequencer FLX. Reads from each EST library were filtered, screened, and assembled using Newbler. Both genomes were annotated using the JGI annotation pipeline (SI Appendix, Materials and Methods) and can be accessed through the JGI Genome Portal (http://www.jgi.doe.gov/spathaspora/ and http://www.jgi.doe.gov/tenuis). SI Appendix, Tables S1, S12, S13, and S14 list genome-sequencing statistics.

Species Phylogeny and Orthology. We estimated the phylogeny using protein sequences of 136 single-copy orthologs present in all species (SI Appendix, Materials and Methods). Phylogenies were constructed with MrBayes v. 3.1.2 (34, 35). We created OGGs using a modified reciprocal smallest distance (RSD) (36) and OrthoMCL (37) method with the RSD parameters significance threshold, 10−3; alignment threshold, 0.3; and the OrthoMCL parameters significance threshold, 10−5; inflation parameter, 1.5. Pairwise one-to-one orthologs were assigned with RSD between each species and one of four reference species: Scer, Pts YXY1, 2, 3, and Lelo. The two methods were compared and combined using a custom Prac- tical Extraction and Reporting Language (perl) script to maximize high-confidence assignments (true positives) and minimize low-confidence assignments (false positives) (SI Appendix, Materials and Methods).

Fungal Strains. All fungal species used in this study are sequenced strains and are listed in Table 1. Heterologous overexpression of selected Spas or Cten genes was conducted in two different Scer strain backgrounds: BY4741 or a wild diploid strain (GLBRCYOA). A codon-optimized DNA cassette (Gen- script) containing the Scer PFK1 promoter (Pr), TDS3 terminator (tr), PrTDS3, tTEF2, PrTEF2, and t, followed by the kanamycin (KanMX) selection marker (38) was synthesized with or without codon-optimized Pts YXY1, 2, 3, or XYL1, XYL2, and XYL3 genes between each promoter–terminator pair (in order). The cassette was integrated at its single-copy locus. Ten individual transformants in single-copy EcoRI clones from each Cten lacking CUG codons that were induced in a majority of the xylose fermenters were cloned between PrTDS3 and tUTB1 in a 2-μm pRS426 vector (39) modified with an Hyg selection marker (see SI Appendix, Note S1 for more details on engineered strains). Gene deletions were created by homologous recombination to replace the coding sequence with KanMX or hygromycin (HgMX) drug-resistance cassettes. All constructs were confirmed by diagnostic PCR and/or DNA sequencing.

Phenotypic Assays. For all assays, cultures were grown in 1% yeast extract, 2% peptone, 2% glucose or in synthetic complete (SC) medium (1.7 g/L yeast nitrogen base, essential amino acids, and 1 g/L ammonium sulfate or monosodium glutamate when mixed with Geneticin), with 2% glucose (SCD) at 30 °C for at least 16 h to early-mid log phase. Cells were washed once in SC (no sugar), diluted, and transferred to either liquid or solid medium containing 10% glucose or xylose. For solid growth assays, plates were scored after 2 d at 30 °C. For liquid growth assays, OD600 was monitored with Spectronic 20D+ (Thermo Scientific), or TECAN F500 or M1000 plate readers. For fermentation, 50 mL of washed cells were transferred to an airlocked 125-mL Erlemeyer flask and were incubated at 30 °C in an orbital shaker at 100 rpm. Supernatant was filtered through a 0.22-μm filter before analysis by HPLC with a Bio-Rad Aminex HPX-87H column (40). Concentrations of ethanol also were determined using an Agilent Technologies 7890A gas chromatograph with a 7693 autosampler and flame ionization detector (SI Appendix, Materials and Methods).
Microarrays. Cells were collected at OD600 0.5–0.6 after growth for three generations in 2% glucose or xylose. Cell lysis and total RNA isolation were performed as described (43). RNA was purified further with LiCl and the Qiagen RNeasy kit. Sample labeling was performed as previously described (41) using cyanine dyes (Amersham), SuperScript III (Invitrogen), and amino-alloy-dUTP (Ambion). Whole-genome, species-specific 37K microarrays (Roche-NimbleGen) were designed with chipD (SI Appendix, Table 11) (42). Arrays from three biological replicates were hybridized in a NimbleGen hybridization system 12 (BioMicro), washed, and scanned using a scanning laser (GenePix 4000B; Molecular Devices) according to NimbleGen protocols (http://www.nimblegen.com/). Data normalization and statistical analyses were performed using Bioconductor (43) and custom perl scripts. The affy() package (44) was used to apply probe-level quantile normalization to the log signal of RNA versus a species-specific genomic DNA control. Genes with significant expression differences in response to xylose were identified separately for each species by performing paired t tests using the Bioconductor package Limma v. 2.9.8 (45) with a false-discovery rate (FDR) correction of 0.05 (46). For cross-species comparisons, genes within OGGs were evaluated for differences in expression.

When an OGG contained more than one gene from a particular species, genes with the smallest phylogenetic distance [determined with PAML v. 4.3 (47)] were compared directly. Hierarchical clustering of gene expression across species was performed with Cluster 3.0 using the uncentered Pearson correlation as the distance metric (48).

ACKNOWLEDGMENTS. We thank Meredith Blackwell, Avig Regev, Dawn-Anne Thompson, and Cletus Kurtzman for strains; Yann Dufour for assistance in microarray design; Cécile Ané for bioinformatic and phylogenetic analysis; and Todd Higbee and Gwendolyn Bone for GC and HPLC analysis; Thomas Kuster for images of yeasts; and Rebecca Breuer and Ben Bice for technical assistance. This work was performed under the auspices of the US Department of Energy (DOE) and was funded in part by the DOE Great Lakes Bioenergy Research Center (DOE Biological and Environmental Research Office of Science Grant DE-F02-07ER64494), National Library of Medicine Traineeship TSTLM055421 and National Science Foundation Career Award 0447887 (to A.P.G.). The work conducted by the US DOE Joint Genome Institute is supported by the Office of Science of the US DOE under Contract No. DE-AC02-05CH11231.