APJ1 and *GRE3* Homologs Work in Concert to Allow Growth in Xylose in a Natural *Saccharomyces sensu stricto* Hybrid Yeast

Katja Schwartz,[1] Jared W. Wenger,[1,2] Barbara Dunn, and Gavin Sherlock[3]
Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305-5120

ABSTRACT Creating *Saccharomyces* yeasts capable of efficient fermentation of pentoses such as xylose remains a key challenge in the production of ethanol from lignocellulosic biomass. Metabolic engineering of industrial *Saccharomyces cerevisiae* strains has yielded xylose-fermenting strains, but these strains have not yet achieved industrial viability due largely to xylose fermentation being prohibitively slower than that of glucose. Recently, it has been shown that naturally occurring xylose-utilizing *Saccharomyces* species exist. Uncovering the genetic architecture of such strains will shed further light on xylose metabolism, suggesting additional engineering approaches or possibly even enabling the development of xylose-fermenting yeasts that are not genetically modified. We previously identified a hybrid yeast strain, the genome of which is largely *Saccharomyces uvarum*, which has the ability to grow on xylose as the sole carbon source. To circumvent the sterility of this hybrid strain, we developed a novel method to genetically characterize its xylose-utilization phenotype, using a tetraploid intermediate, followed by bulk segregant analysis in conjunction with high-throughput sequencing. We found that this strain's growth in xylose is governed by at least two genetic loci, within which we identified the responsible genes: one locus contains a known xylose-pathway gene, a novel homolog of the aldo-keto reductase gene *GRE3*, while a second locus contains a homolog of *APJ1*, which encodes a putative chaperone not previously connected to xylose metabolism. Our work demonstrates that the power of sequencing combined with bulk segregant analysis can also be applied to a nongenetically tractable hybrid strain that contains a complex, polygenic trait, and identifies new avenues for metabolic engineering as well as for construction of nongenetically modified xylose-fermenting strains.

LIGNOCELLULOSIC biomass, an untapped feedstock for biofuel production, is rich in five-carbon sugars such as xylose and arabinose; the metabolism of these sugars to ethanol or other economically important molecules is thus crucial for the cost-effective use of such biomasses (Buckeridge *et al.* 2011; Chandel *et al.* 2011). However, a fundamental problem in moving toward industrial-level production of cellulosic ethanol is that currently used strains of the predominant microorganism utilized in industrial fermentations—the budding yeast *Saccharomyces cerevisiae*—do not use xylose as a fermentable substrate (Chiang and Knight 1960). Significant progress has been made over the past 30 years to address this issue, and through the use of metabolic engineering and directed evolution (Ho *et al.* 1998; Sonderegger and Sauer 2003; Kuyper *et al.* 2004; Matsushika *et al.* 2009; Kim *et al.* 2010; Ha *et al.* 2011) strains of *S. cerevisiae* that have the capability to ferment xylose to ethanol now exist. Despite this progress, problems remain to be solved before these strains come into widespread industrial use, including the fact that most current xylose-fermenting strains are genetically modified—a notion that continues to remain unpopular in many countries (Byrne 2006).

Traditionally it has been thought that *S. cerevisiae* does not metabolize xylose, despite the fact that its genome contains genes putatively encoding the requisite enzymes for the two-step redox conversion of xylose to the fermentable-intermediate xylulose (Chiang and Knight 1960; Toivari *et al.* 2004). To date, only two studies have demonstrated the existence of natural *Saccharomyces* isolates capable of

doi: 10.1534/genetics.112.140053
Manuscript received October 1, 2011; accepted for publication February 26, 2012
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/content/suppl/2012/03/16/genetics.112.140053.DC1.
Illumina data from this article have been deposited with the Short Read Archive (NCBI) under accession no. SRA045682.1.
[1]These authors contributed equally to this work.
[2]Present address: Amyris, Inc., Emeryville, CA 94608.
[3]Corresponding author: Room S201, Grant Bldg., Stanford University Medical School, 259 Campus Dr., Stanford, CA 94305-5120. E-mail: gsherloc@stanford.edu

xylose metabolism (Attfield and Bell 2006; Wenger *et al.* 2010), of which only the latter included genetic characterization of the trait. In Wenger *et al.* (2010) we described a gene, *XDH1*, that is present in many wine yeast strains but not found in the reference yeast S288c genome, which encodes a putative xylitol dehydrogenase sufficient to allow otherwise wild-type *S. cerevisiae* laboratory strains to grow slowly in xylose. However, there is no evidence that these strains grow anaerobically in xylose or that they produce any ethanol, and the observed growth is modest at best. Due to this poor xylose utilization, work on creating industrially viable, xylose-metabolizing *Saccharomyces* yeasts has largely focused on metabolic engineering, often combined with directed evolution.

Metabolic engineering of xylose fermentation in *Saccharomyces* yeasts takes advantage of the fact that other fungi and bacteria, while often not industrially suitable for large-scale ethanol fermentations, are nevertheless capable of xylose metabolism via one of two pathways. Fungi such as *Scheffersomyces stipitis* (formerly *Pichia stipitis*), *Pachysolen tannophilus*, and *Candida shehatae* metabolize xylose to its keto-isomer xylulose via a two-step reduction oxidation pathway involving xylose reductase (XR) and xylitol dehydrogenase (XDH) (Jeffries 2006). In most bacteria and some fungi, however, xylose is directly isomerized to xylulose by xylose isomerase (XI) (Jeffries 1983). In both cases, xylulose is subsequently phosphorylated to xylulose-5-phosphate and metabolized via the nonoxidative pentose phosphate pathway (PPP) (Wang *et al.* 1980). Introduction into *S. cerevisiae* of the genes from other organisms encoding the two oxidoreductases or the isomerase has produced strains that can utilize xylose, but these approaches have been plagued by various problems (Chandel *et al.* 2011). These include issues such as poor expression of genes encoding XR, XDH, and xylulokinase (XK) activities, redox imbalances due to different cofactor specificities of XR/XDH enzymes, glucose catabolite repression, low affinity of the hexose transporters for xylose, and low flux through the PPP. Others have attempted to address these problems with metabolic engineering and directed evolution of engineered strains; see (Buckeridge *et al.* 2011) for a recent review. On the basis of these strategies, a strain has been recently developed that shows rapid cofermentation of cellobiose and xylose (Ha *et al.* 2011) and other xylose-fermenting strains continue to show improvement. Despite these advances, to our knowledge no *Saccharomyces* strains are currently utilized for xylose fermentation in large-scale industrial settings.

In light of the remaining challenges in the xylose metabolic-engineering field, we believe much can still be learned from studying natural *Saccharomyces* yeasts that are capable of xylose utilization. Characterization of the genetics and physiology of these natural xylose-utilizing yeasts will provide testable hypotheses for use in further modification and improvement of existing engineered strains. Toward this goal, we have characterized the genetic basis of a polygenic, xylose-metabolism phenotype in a *Saccharomyces sensu stricto* hybrid yeast that we previously identified as capable of xylose utilization (Wenger *et al.* 2010). In pursuit of the loci that contribute to this strain's growth in xylose, we have developed a novel method for generating progeny from this otherwise genetically intractable hybrid strain and utilized high-throughput sequencing in conjunction with bulk segregant analysis for the identification of quantitative trait loci. Among these loci we have identified a new homolog of a known xylose pathway gene, *GRE3*, as well as a homolog of *S. cerevisiae APJ1*, which encodes a putative chaperone, which was previously unconnected to xylose metabolism.

Materials and Methods

Yeast strains and techniques

All *S. uvarum* and hybrid yeast strains used in this study are shown in Table 1. GSY1063 was derived from CBS7001 by introducing *ho*::*kanMX* (see primers in Supporting information, Table S1). GSY2712 is a Leu$^+$ derivative of JRY8145, while GSY2719 was derived from a cross between JRY8153 and GSY1063. *apj1Δ*::*URA3* (GSY4341) and *gre3Δ*::*URA3* (GSY4324) strains were generated in GSY2719 by transformation with a fusion PCR product (see Table S1 for details). Yeast transformation was performed by the lithium acetate method (Schiestl and Gietz 1989). Preparation of yeast genomic DNA was performed as described previously (Treco 1987). All strains were grown at 25°.

For long-term growth curves, strains were first grown to saturation in yeast extract/peptone/2% glucose (YPD) medium, after which, to start the growth curve, they were diluted 100-fold into a total of 5 ml of yeast extract/peptone medium containing either 2% xylose or no carbon source. All strains were grown at 25° with aeration in a roller drum. Optical density (OD_{600}) was measured at 600 nm in a Biomate 3 spectrophotometer. Cell concentration was measured in Z2 Beckman Coulter Counter. For bulk segregant analysis, progeny were pooled as described in the *Results* section.

Molecular cloning techniques

Standard molecular biology techniques were used for all plasmid construction and cloning. High-fidelity Phusion DNA polymerase (Finnzymes) was used for DNA amplifications according to the manufacturer's recommendations. Plasmids are listed in Table 2 and details of their construction are available upon request. Briefly, pGS35 and pGS36 were constructed from YCplac22 (Gietz and Sugino 1988) by replacing the *TRP1* gene with the kanMX or hphMX cassette, respectively. pGS37 and pGS38 contain the *GAL1/GAL10* promoter and *ACT1* transcriptional terminator from pTS210 (Marschall *et al.* 1996). The *HO* endonuclease open reading frame was amplified from the *S. cerevisiae* Simi White wine yeast strain (GSY788) using primers GSP1 and GSP545 (Table S1) and cloned into the *Xba*I sites of pGS37 and pGS38 to make pGS39 and pGS40, respectively. To make pGS131 and pGS132 we amplified the *APJ1* gene from

Table 1 *Saccharomyces uvarum* strains used in this study

Strain	Genotype	Origin
CBS7001	*MAT***a**/α	CBS Fungal Biodiversity Centre
CBS1502	*MAT***a**/α (hybrid)	CBS Fungal Biodiversity Centre
JRY8145	*MAT***a** *ho::natMX leu1-1*	Gallagher *et al.* (2009)
JRY8153	*MAT***a** *ho::natMX his3-1 lys2-5 trp?-1 ura3-1*	Gallagher *et al.* (2009)
GSY2607	*MAT***a**/**a**/α/α (hybrid)	This study
GSY2612	*MAT***a**/α	This study
GSY1063	*MAT*α *ho::kanMX*	This study
GSY2694	*MAT***a**/α (used for BSA)	This study
GSY4318	*MAT*α *ho::kanMX GRE3*^{CBS1502} *APJ1*^{CBS1502}	This study
GSY2712	*MAT***a** *ho::natMX*	This study
GSY4319	*MAT***a**/α *ho::kanMX/ho::natMX B*^{CBS7001}/*B*^{CBS7001} *GRE3*^{CBS1502}/*GRE3*^{CBS7001} *APJ1*^{CBS1502}/*APJ1*^{CBS7001}	This study
GSY2719	*MAT***a**/α *ho::kanMX/ho::natMX ura3-1/ura3-1*	This study
GSY4340	*MAT*α *ho::natMX GRE3*^{CBS1502} *APJ1*^{CBS1502}	This study
GSY4342	*MAT***a** *ho::natMX GRE3*^{CBS1502} *APJ1*^{CBS1502}	This study
GSY4341	*MAT*α *ho::kanMX ura3-1 apj1*Δ*::URA3*	This study
GSY4322	*MAT*α *ho::natMX apj1*Δ*::URA3 GRE3*^{CBS1502}	This study
GSY4324	*MAT***a** *ho::kanMX ura3-1 gre3*Δ*::URA3*	This study
GSY4326	*MAT***a** *ho::natMX gre3*Δ*::URA3 APJ1*^{CBS1502}	This study
GSY4327	*MAT*α *ho::natMX gre3*Δ*::URA3 APJ1*^{CBS1502}	This study

CBS7001 and from GSY4318 using primers GSP546 and GSP547 and then cloned it into the *Bam*HI site of pGS35. Similarly, to make pGS149 and pGS159, the *GRE3* gene was amplified from the same strains using primers GSP535, 536, 538, and 539 and cloned into the *Bam*HI site of pGS35. To create a promoter swap construct in pGS156, the promoter region of *GRE3* was amplified from CBS7001, fused by PCR to the *GRE3* coding and the 3′ region from GSY4318 using a primer that contained overlapping sequence from the 3′ end of the promoters and the 5′ end of the open reading frame (GSP533), and cloned into pGS35. pGS170 and pGS171 were constructed by swapping the *Nae*I fragment containing the last 435 amino acids from *APJ1* coding sequence between plasmids pGS131 and pGS132.

Array comparative genomic hybridization

Genomic DNA from CBS1502 was prepared using Zymo Research YeaStar columns according to the manufacturer's recommendations, and then digested with *Hae*III. We then labeled 350 ng of this DNA with Cy5 (red); we similarly labeled reference DNA (an equimolar mix of *S. uvarum* (CBS7001 strain) and *S. cerevisiae* (S288c strain) sheared genomic DNA) with Cy3 (green). The two labeled DNAs were then mixed together and hybridized to microarrays containing probes densely covering both the *S. uvarum* (CBS7001 strain) and *S. cerevisiae* (S288c strain) genomes; the microarrays and the hybridization methods used were exactly as described in Dunn and Sherlock (2008).

Preparation of genomic DNA for High Throughput Sequencing

Segregants for bulk segregant analysis were frozen in a sorbitol solution (0.9 M sorbitol, 0.1 M EDTA, and 0.1 M Tris-HC1, pH 8.0), and then combined for DNA isolation, as described (Treco 1987). DNA was prepared for sequencing on the Illumina platform as follows. Paired-end Illumina adapters were preannealed in a 50-μl reaction containing 1× T4 DNA ligase buffer (NEB no. B0202S) and each

Table 2 Plasmids used in this study

Plasmid	Plasmid details	Origin
pGS35	*CEN/ARS* kanMX	Wenger *et al.* (2010)
pGS36	*CEN/ARS* hphMX	Wenger *et al.* (2010)
pGS37	*CEN/ARS* kanMX, *GAL1/10* promoter, *ACT1* terminator	This study
pGS38	*CEN/ARS* hphMX, *GAL1/10* promoter, *ACT1* terminator	This study
pGS39	*HO* under *GAL1/10* promoter in pGS35	This study
pGS40	*HO* under *GAL1/10* promoter in pGS36	This study
pGS131	*APJ1*^{CBS7001} in pGS35	This study
pGS132	*APJ1*^{CBS1502} in pGS35	This study
pGS149	*GRE3*^{CBS7001} in pGS35	This study
pGS150	*GRE3*^{CBS1502} in pGS35	This study
pGS156	*GRE3*^{CBS1502} under CBS7001 promoter in pGS35	This study
pGS170	*APJ1*^{G234D} in pGS35	This study
pGS171	*APJ1*$^{short\ (18)\ polyQ}$ in pGS35	This study

adapter at a concentration of 40 μM by incubating at 94° for 5 min, and then 70°, 60°, 50°, 40°, 30°, and 25°, each for 1 min. Five micrograms of genomic DNA was sheared by sonication to approximately 500 bp in a COVARIS sonicator. Thirty microliters of sheared DNA was subjected to end repair in a 50-μl reaction (1× T4 DNA ligase buffer, 0.8 μM dNTPs (NEB no. N0447S), 2.5 μl of T4 DNA polymerase (NEB no. M0203L), 0.5 μl Klenow (large fragment) (NEB #M0210L), and 2.5 μl of T4 PNK (NEB no. M0201L) by incubation at 20° for 30 min. End-repaired DNA was purified using a QIAquick PCR purification column, eluting in 33 μl of buffer EB. Addition of a dATP to end-repaired DNA was performed by incubation at 37° for 30 min (32 μl of end-repaired DNA, 5 μl of buffer 2 (NEB no. B7002S), 1 μl 10 mM dATP (Invitrogen no. 18252-015), 3 μl Klenow Exo-Fragment (NEB no. M0212L)). After addition of dATP, reactions were purified using a QIAgen MinElute column, eluting in 11 μl of buffer EB. Illumina adapter ligation was performed in a 20-μl reaction by incubation at 20° for 15 min followed by 65° for 10 min (10 μl of DNA from previous step, 1× T4 DNA ligase buffer, 1 μl T4 DNA ligase (NEB no. M0202S), 1 μl 40 μM adapter mix from preannealing). Following adapter ligation, size selection was performed on the Invitrogen E-gel system, targeting 600 bp fragments. Following size selection, the library was amplified using PCR in a 20-μl reaction (1.25 μM primers PE1 and PE2, 2-μl size-selected DNA, 0.25 μM dNTPs, 1× HF buffer, and 0.5 μl Phusion DNA polymerase (NEB no. F-530L). DNA was amplified using the following program: 98° for 30 sec; 12 cycles of 98° for 10 sec, 65° for 30 sec, and 72° for 30 sec and a final 72° extension time of 5 min. The amplified library was purified using a QIAquick PCR purification column, eluting in a final volume of 30 μl buffer EB. The final library concentration and size estimates were determined using Qubit (Invitrogen) and Bioanalyzer (Agilent). Flow cells for the Illumina GAII platform were prepared according to manufacturer's instructions and sequencing was performed for 36 cycles.

Analysis of high-throughput sequencing data

Sequence reads with their qualities (FASTQ) were mapped to the *S. uvarum* genome (available at http://saccharomycessensustricto.org) (Scannell *et al.* 2011) using Stampy v. 1.0.13 (Lunter and Goodson 2011) in conjunction with BWA v. 0.5.9-r16 (Li and Durbin 2009), with default parameters. Whole-genome pileup files were created using the Samtools v. 0.1.16 "pileup" command (Li *et al.* 2009) with option –c. Custom perl scripts were written to calculate allele frequency differences between positive and negative pools and to determine positions with significantly different frequencies. For SNP calling, we required a position to be covered by at least 20 sequencing reads. The majority SNP call from the Samtools "pileup" was used to calculate an allele frequency in the positive and negative pools, and we calculated a T statistic, on the basis of Craig *et al.* (2009) as

$$\frac{(\text{Allele Freq. Pos.} - \text{Allele Freq. Neg.})^2}{(\text{Variance Pos.} + \text{Variance Neg.})},$$

where the binomial Variance is

$$\frac{(\text{Allele Freq.} \times (1\text{-Allele Freq}))}{(2 \times \text{Pool Size})}.$$

P-values were then calculated assuming the *T* statistic follows a χ^2 distribution (Craig *et al.* 2009). *P*-value cutoffs were determined using a Bonferroni correction of the alpha significance value, 0.01 divided by the number of SNPs tested. False discovery rates (FDR) were estimated empirically by permuting the pool labels of the SNP calls at each position, recalculating allele frequencies, and generating *P*-values, as described, from the permuted data. Each pool was permuted 500 times, and the FDR was determined by (median number of false positives after 500 permutations) divided by ("true" positives from unpermuted data). Data were plotted using R. Sequence data are available in the Short Read Archive with accession no. SRA045682. Perl scripts are available upon request.

Quantitative RT–PCR

Strains were pregrown in YPD overnight and diluted 100 fold into 20 ml YP medium containing 2% xylose. After 3 days of growth, cells were harvested by filtering and frozen in liquid nitrogen until RNA purification. Hot phenol RNA preparation was performed as described previously (Lee *et al.* 2008) and followed by treatment with Ambion TURBO-DNA*free* treatment using manufacturer's recommendations (Life Technologies). Two micrograms of total RNA were reverse transcribed using oligo(dT) primer and Superscript II according to the manufacturer's instructions (Invitrogen). Real-time PCR was performed on a Bio-Rad CFX96 cycler using SsoFast EvaGreen Supermix (Bio-Rad). *S. uvarum* YDR458C and YJL088W were used as reference genes, with primer pairs for those genes as described in Bullard *et al.* (2010). The primer pair for the *GRE3* gene was designed to recognize both $GRE3^{CBS7001}$ and $GRE3^{CBS1502}$. Primers used for qPCR (GSP556-561) are listed in Table S1. To calculate the relative quantification value we used average $\Delta\Delta C_t$ values, normalizing relative expression to the ΔC_t in the $APJ1^{CBS7001}$ $GRE3^{CBS7001}$ strain (Relative quantification value RQ $= 2^{\Delta\Delta C_t}$, with upper limit UL $= 2^{\Delta\Delta C_t - \text{SD}}$, and lower limit LL $= 2^{\Delta\Delta C_t + \text{SD}}$).

Results

A Saccharomyces sensu stricto hybrid that grows in xylose

To identify naturally occurring yeasts that have the ability to grow in xylose as a carbon source, we previously reported a screen of 647 *Saccharomyces* yeasts where strains that could reproducibly grow in minimal and/or rich media supplemented with 2% xylose were designated "xylose positive"

(Wenger *et al.* 2010). We identified 38 xylose-positive yeasts in this screen, 29 of which were *S. cerevisiae* wine strains whose modest growth in xylose was controlled by a single locus, *XDH1*. Of the 9 other xylose-positive yeasts that we identified, the strain with the most robust xylose phenotype was CBS1502, which showed a reproducible increase in both optical density and cell number relative to a xylose-negative, *S. uvarum* control strain (Figure 1). CBS1502 is also known as Yorkshire Haze 1, and its CBS-KNAW Fungal Biodiversity Centre record reports its provenance as either an *S. bayanus* or *S. pastorianus* yeast isolated from cloudy beer. Because of the uncertainty in classification of this brewing contaminant, we first determined the genomic makeup of this strain by array comparative genomic hybridization (aCGH) using custom DNA microarrays that contain specific probes that distinguish the *S. cerevisiae* and *S. uvarum* genomes (Dunn and Sherlock 2008). These data show that the vast majority of this strain's genome is derived from *S. uvarum*, but also contains regions derived from *S. cerevisiae* and the recently discovered *S. eubayanus* (Libkind *et al.* 2011) (Figure S1, blue circled regions).

To determine if this strain was genetically tractable, we characterized its sporulation efficiency and determined that it was too low for standard genetic analysis (<1%, as would be expected for a hybrid) and, therefore, developed a novel approach for characterization of the xylose phenotype.

Strategy for analyzing a genetically intractable strain

Because CBS1502 has both poor sporulation and spore viability we developed a novel strategy using a tetraploid intermediate to segregate and identify genetic factors that contribute to its growth in xylose (see Figure 2 and Figure S2). First, CBS1502 and CBS7001 (the sequenced reference *S. uvarum* strain) were transformed with plasmids that express the site-specific endonuclease encoding-gene *HO* under control of the galactose inducible *GAL1* promoter and contain one of two different selectable markers (*KanMX* and *HphMX*). These strains were then individually grown to saturation in rich medium supplemented with 2% raffinose and then shifted to galactose-containing medium to induce expression of *HO*. *HO*-induced strains were then combined and plated onto solid YEPD-based media containing both Geneticin (Invitrogen, 200 mg/liter) and Hygromycin B (Cellgro, 150 mg/liter). Transient expression of *HO*—which is normally repressed in **a**/α diploids—allows recombination at the mating-type locus, and at a low frequency will allow the formation of diploids with mating type **a**/**a** or α/α. These diploids are mating competent and are able to form CBS1502/CBS7001 **a**/**a**/α/α tetraploids. One tetraploid (GSY2607) was then put through one round of meiosis and tetrad dissection to generate heterozygous diploids that could be **a**/**a**, α/α, or **a**/α mating type. Because both CBS7001 and CBS1502 contain a wild-type *HO*, any **a**/**a** or α/α spores will switch mating type following cell division and then self-fertilize to produce tetraploids, while the **a**/α spores would remain as stable, nonmating diploids. We selected several of these **a**/α diploid spores and then sporulated them again to produce haploid spores; these spores were then able to switch mating type and self-fertilize to become homozygous **a**/α diploids, our desired end products for further genetic characterization.

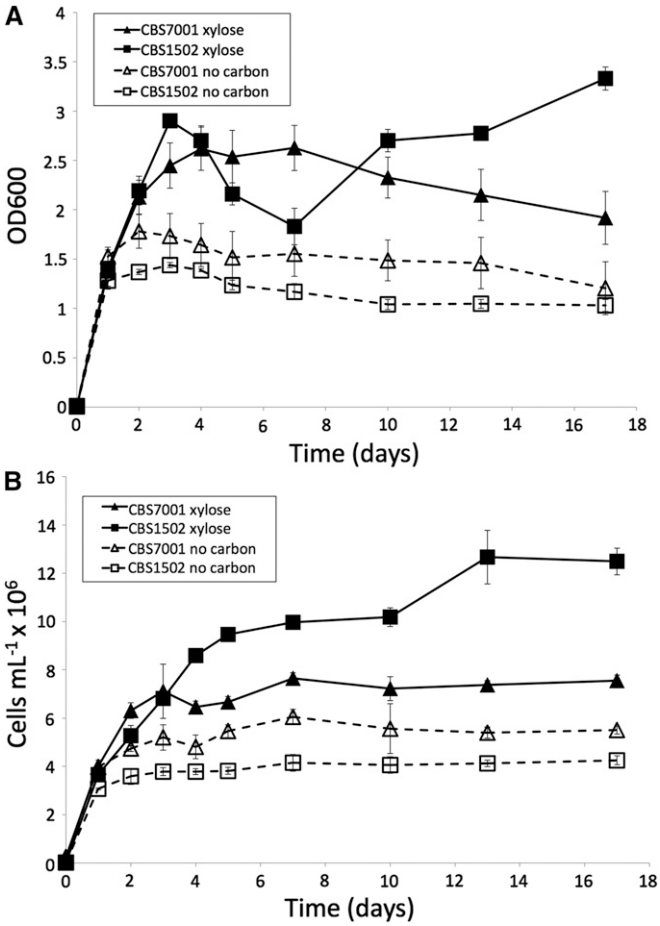

Figure 1 Xylose growth phenotype of CBS1502. Increase in OD600 (A) and cell concentration (B) of CBS1502 and a control xylose-negative *S. uvarum* strain (CBS7001) were measured over 17 days of aerobic culture in 2% xylose. Error bars represent standard deviations of three biological replicates for CBS1502 and four biological replicates of CBS7001.

These "double-reduced" strains were assessed for growth on xylose, and we selected one of the resulting strains as our xylose-positive lineage of interest. This strain (GSY2612) was then backcrossed two more times to CBS7001, using the same tetraploid-intermediate method, to further increase spore viability (Figure S2). At each backcross, the best xylose-positive spore product was chosen to proceed into the next round of crossing. After these backcrosses, some of the xylose-positive progeny were determined to be stable haploid strains—presumably containing a mutation in the *HO* gene or another gene involved in mating-type switching—and were crossed one additional time to a haploid (*ho*Δ::*KanMX*) CBS7001 derivative (GSY1063). The haploid segregants from this diploid strain (GSY2694) were screened for growth in xylose and used for bulk segregant analysis. See Table 1 and Figure S2 for strain names and crossing details.

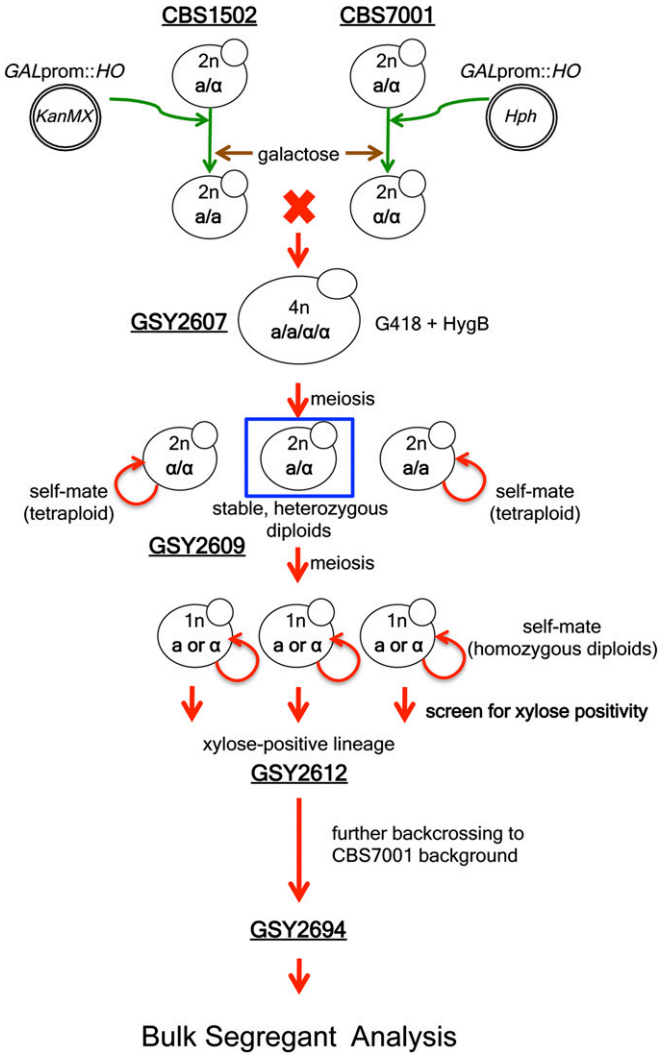

Figure 2 Diagram of genetic analysis of a usually intractable strain via a tetraploid intermediate. Diploids were mated to form a tetraploid, which was subsequently sporulated. See text for details.

From this final backcross, the segregation pattern of growth in xylose was roughly three xylose-negative spores to one xylose-positive spore. This pattern is consistent with a hypothesis that two unlinked genes contribute to growth in xylose, both of which are required for the most robust xylose-positive phenotype. This observation is also consistent with the diversity of xylose phenotypes seen in CBS1502 spores.

Bulk segregant analysis by sequencing reveals a polygenic xylose phenotype

Bulk segregant analysis (BSA), originally developed using microarrays but more recently adapted for high-throughput sequencing, has been proven to quickly and specifically identify candidate loci on the basis of a strategy that pools progeny of a cross between two polymorphic strains based on a phenotype of interest (Quarrie *et al.* 1999; Brauer *et al.* 2006). To determine the loci contributing to growth in xylose in the derivative of CBS1502 described above, we created one pool containing 21 xylose-positive segregants (from tetrad dissection of GSY2694) and one pool containing 21 xylose-negative segregants from the same cross, where xylose positivity was defined as an increase in both cell number (as measured by Coulter counter) and an increase in optical density relative to a negative control, *S. uvarum* CBS7001. DNA was isolated from each pool and genomic DNA libraries were prepared for sequencing on the Illumina GAIIx platform (see *Materials and Methods*).

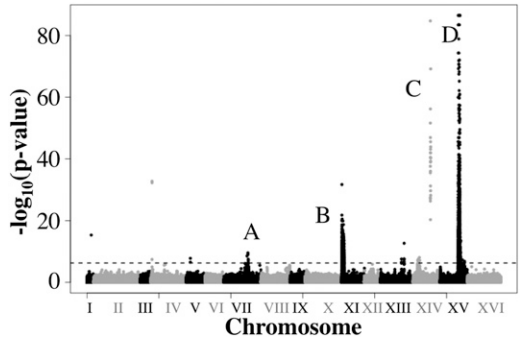

Figure 3 Bulk Segregant Analysis of CBS1502 lineage GSY2694. $-\log^{10}(P\text{-values})$ are plotted for each SNP across all 16 chromosomes (alternating shading) and represent significance of the difference between allele frequencies in the xylose-positive and xylose-negative pools. See *Materials and Methods* for derivation of *P*-values. The dotted line indicates the Bonferroni-corrected (α) significance cutoff. Genomic intervals for each lineage are lettered A–D.

We mapped the resulting sequence reads to the *S. uvarum* (CBS7001) genome (Scannell *et al.* 2011); we then called SNPs and quantified their allele frequencies at polymorphic sites across the genome and determined if each site had a significantly different frequency between the positive and negative pools (see *Materials and Methods* for further details). The results of this analysis are shown in Figure 3, with false discovery rates estimated to be $<0.2\%$. After performing BSA on GSY2694 progeny, we observed three genomic intervals in which the CBS1502 alleles significantly segregate with the xylose-positive phenotype: one on chromosome VII that is approximately 13 kb (region A), one on chromosome XIV that is approximately 10 kb (region C), and one on chromosome XV that is approximately 76 kb (region D). We also observed a 65-kb region on chromosome XI (region B), in which the CBS7001 alleles segregate with the xylose-positive phenotype, suggesting that there is a genomic region in CBS1502 that is detrimental to growth on xylose.

On the basis of the segregation pattern of growth in xylose in GSY2694 we had expected to find two genes that were both required for the phenotype. The BSA data support this hypothesis because of the pattern of allele frequencies that we observed in the four genomic intervals. In both regions C and D, we observed that the positive pool contained nearly 100% CBS1502 alleles, while the negative pool contained only 50% or less of CBS1502 alleles. This is

consistent with the causative genes in these regions both being required for the phenotype to be observed. In region A, however, we observed that the positive pool contained approximately 60–70% CBS1502 alleles, while the negative pool contained approximately 20–30% CBS1502 alleles. Region B represented a third category, where the positive pool contained <10% CBS1502 alleles, while the negative pool contained ~50% CBS1502 alleles. These data suggest that regions C and D are the two main causative alleles for the phenotype, while regions A and B enhance or modify the phenotype but are not necessary for it to be observed. The CBS1502 alleles in peak B presumably negatively affect growth in xylose and were thus selected against in our xylose positives.

To genetically confirm that each of these four genomic intervals segregates as predicted on the basis of the BSA data, we chose SNPs that created a restriction fragment length polymorphism within each region and tested each of the 21 xylose-positive and 21 xylose-negative GSY2694 segregants for which polymorphism they contained. The data (not shown) confirmed that these four regions segregate nonrandomly in the positive and negative pools as predicted by the sequence data. A χ^2 goodness-of-fit test significantly rejected a null hypothesis of random segregation between the pools for all four peaks ($P < 0.01$ for peak A, $P < 0.001$ for peaks B through D).

Regions $C^{CBS1502}$ and $D^{CBS1502}$ are both required for growth in xylose

To confirm that regions C and D—the two hits from the bulk segregant analysis that we predicted were both required for growth in xylose—were the responsible intervals, we selected a single segregant of GSY2694 that contained regions C and D from CBS1502 and region B from CBS7001 and crossed it to a haploid derivative of CBS7001. Note that both copies of region B are derived from CBS7001, while regions C and D, which are unlinked, are heterozygous and thus segregating. We then tested haploid strains containing all four pairwise combinations of regions C and D for their ability to grow in xylose. We observed that the presence of region $C^{CBS1502}$ results in increased optical density in xylose relative to CBS7001 (Figure 4A). The presence of region $D^{CBS1502}$ does not result in a significant phenotype on its own; however, the presence of regions $C^{CBS1502}$ and $D^{CBS1502}$ together results in increased optical density in xylose that is greater than the sum of the individual regions $C^{CBS1502}$ and $D^{CBS1502}$ xylose phenotypes (Figure 4A). Interestingly, this synergistic interaction, indicative of positive epistasis between the two genes, is more noticeable when we measured growth in xylose as an increase in cell number, as only strains that contain both region C and D from CBS1502 show significant increases in cells per milliliter at the end of the time course (Figure 4B). These data show that the genes within these two intervals interact via positive epistasis to contribute to an increase in cell number and cell size in xylose in CBS1502, confirming the hypothesis that both are required for the most robust xylose-positive phenotype.

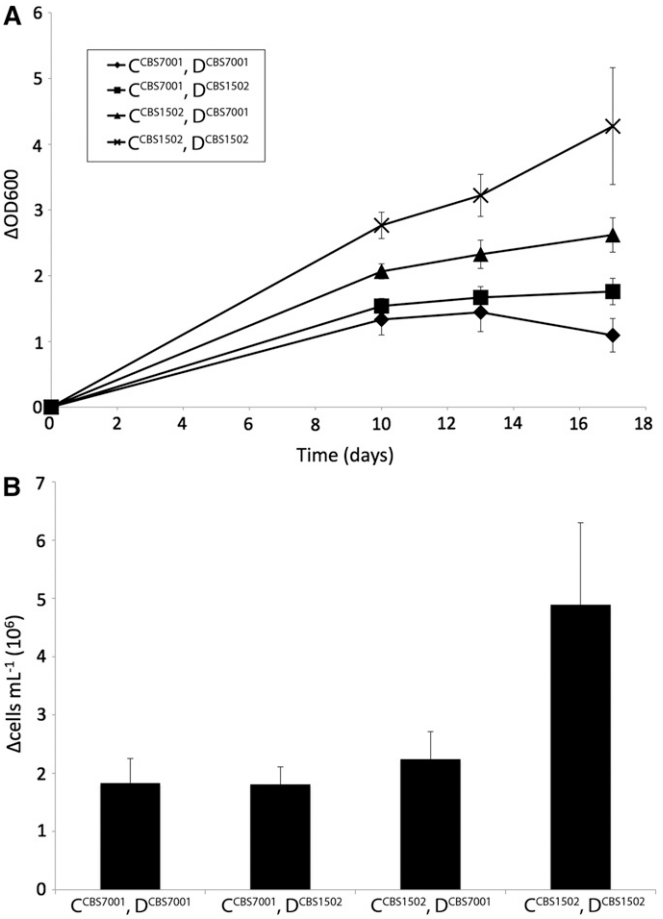

Figure 4 Regions C and D interact epistatically and are required for growth in xylose. (A) Data are OD (600 nm) values and show the difference (Δ) in OD between growth in xylose and the absence of an added carbon source. (B) Data are cells/ml × 10^6 (again, the difference between growth in xylose and no added carbon source) and were measured at day 17 of the time course of growth in xylose as shown in A. Data are averages of at least 6 biological replicates, with error bars showing standard deviation. All segregants are derived from GSY4319.

We also tested whether loci within peaks A and B affected the xylose-positive phenotype in segregants containing both the $C^{CBS1502}$ and $D^{CBS1502}$ regions. We redissected GSY2694 and by PCR identified and selected 19 spores containing CBS1502 alleles for both regions C and D and then genotyped them for regions A and B. We then measured OD_{600} and cell density at the end of a xylose growth experiment. Comparing growth between these spores on the basis of their genotypes for regions A and B revealed a subtle but statistically insignificant ($P > 0.5$) difference between the different A and B genotypes (data not shown). We did not pursue regions A and B further because of their lack of a significant phenotype.

GRE3 and APJ1 are loci responsible for growth in xylose

Having confirmed that regions C and D positively and synergistically contribute to growth in xylose, we wanted to determine the specific genes that are causal for this phenotype.

The sequence of region C was found to contain two genes with nonsynonymous changes: *APJ1* (YNL077W) and *NIS1* (YNL078W). To determine which of these two genes might be responsible, we approached the problem with the assumption that the responsible allele may be recessive (having determined that region C is homozygous is CBS1502, at least consistent with this notion; data not shown). We transformed GSY4340, which contains regions $C^{CBS1502}$ and $D^{CBS1502}$, with plasmids containing either $APJ1^{CBS7001}$ or $NIS1^{CBS7001}$ and screened the resulting transformants for growth in xylose. We observed that transformation with the $APJ1^{CBS7001}$ plasmid reduced growth in xylose, whereas the $NIS1^{CBS7001}$ plasmid had no effect on the xylose phenotype, suggesting that the CBS1502 allele in this region is a recessive allele of *APJ1* (data not shown). The protein sequence of $Apj1^{CBS1502}$ is shown in Figure S3.

To determine if this *APJ1* homolog is a loss-of-function allele in addition to being recessive, we constructed a haploid derivative of GSY2719 with *APJ1* deleted (GSY4341) and then genetically introduced region $D^{CBS1502}$ into this deletion background via mating and dissection, producing GSY4322 (*apj1*Δ::*URA3* $D^{CBS1502}$). GSY4322 was transformed with empty vector, plasmids expressing either $APJ1^{CBS1502}$ (pGS132), or $APJ1^{CBS7001}$ (pGS131). Interestingly, GSY4322 transformed with the empty vector had a phenotype similar—albeit not identical—to that of the strains containing both $APJ1^{CBS1502}$ and $D^{CBS1502}$, indicating that $APJ1^{CBS1502}$ might be a loss-of-function allele (Figure 5A). More specifically, because the phenotypes are not identical, $APJ1^{CBS1502}$ may be a hypomorphic allele (partial loss-of-function). Adding credence to this supposition, when GSY4322 was transformed with a plasmid containing $APJ1^{CBS1502}$, the xylose phenotype was comparable to both the same strain transformed with the empty vector and to the parental strain. Conversely, transformation with $APJ1^{CBS7001}$ inhibited growth in xylose (Figure 5A). Since the coding region of $APJ1^{CBS1502}$ contained only two changes from $APJ1^{CBS7001}$ (a shorter polyglutamine repeat and a G234D substitution; Figure S3), we separated these two changes and generated plasmids containing either $APJ1^{shorter(18)\ polyQ}$ (pGS171) or $APJ1^{G234D}$ (pGS170) and tested their effect in GSY4322. While the phenotype conferred by $APJ1^{shorter(18)\ polyQ}$ was indistinguishable from that conferred by $APJ1^{CBS7001}$, the $APJ1^{G234D}$ allele still resulted in xylose-positive growth (Figure 5B). Taken together, these data show that $APJ1^{CBS1502}$ acts as a recessive, perhaps partial loss-of-function allele to allow growth on xylose and that the G234D substitution is responsible for this phenotype.

Region D is 100 kb long and contains 37 genes (32 with nonsynonymous changes), including an obvious candidate in the *S. uvarum* homolog of *GRE3* (YHR104W), which in *S. cerevisiae* is a known aldo-keto (xylose) reductase (Traff *et al.* 2001). To determine whether this was the specific gene within this interval responsible for increased growth in xylose, we cloned both of the alleles from CBS1502, which is heterozygous at the *GRE3* locus. We Sanger sequenced both

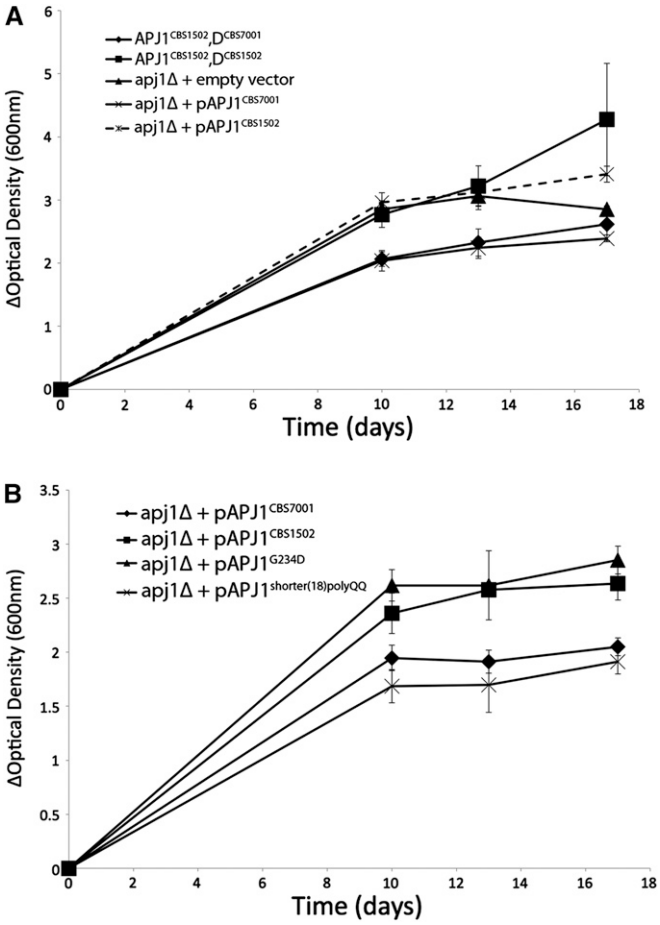

Figure 5 G234D substitution in *APJ1* gene in region C is responsible for growth in xylose. Data are optical density values (600 nm) of the difference (Δ) between growth in 2% xylose and no added carbon source. Average values for four biological replicates are plotted, with error bars showing standard deviation. (A) GSY4322 (*apj1Δ*::*URA3* $D^{CBS1502}$transformed with pGS35, pGS131 (p$APJ1^{CBS7001}$) and pGS132 (p$APJ1^{CBS1502}$). Controls ($APJ1^{CBS1502}$ $GRE3^{CBS1502}$ and $APJ1^{CBS1502}$ $GRE3^{CBS7001}$) are the data from Figure 4, shown for comparison. (B) GSY4322 transformed with pGS131, pGS132, pGS170 (p$APJ1^{G234D}$), and pGS171 (p$APJ1^{shorter(18)polyQ}$).

GRE3 alleles from the resulting plasmids. One *GRE3* allele is identical to the *GRE3* gene found in CBS7001, while the other (hereafter referred to as $GRE3^{CBS1502}$) is identical to that found in the recently discovered *S. eubayanus* (Libkind *et al.* 2011). The protein sequence of $Gre3^{CBS1502}$, as well as a phylogenetic tree of closely related aldo-keto reductases, is shown in Figure S4. We constructed a haploid derivative of GSY2719 with $GRE3^{CBS7001}$ deleted (GSY4324) and then genetically introduced region $C^{CBS1502}$ into this deletion background via mating and dissection, producing GSY4326 and GSY4327. Transformation of GSY4326 and GSY4327 (*gre3Δ*::*URA3* $APJ1^{CBS1502}$) with a plasmid containing the $GRE3^{CBS7001}$ allele partially rescued growth in xylose, whereas transformation with a plasmid containing $GRE3^{CBS1502}$ increased growth in xylose by almost twofold relative to the CBS7001 allele, indicating that the $GRE3^{CBS1502}$ allele contributes to the CBS1502 xylose growth phenotype (Figure 6).

Because there are changes in the promoter region of *GRE3*^{CBS1502} compared to *GRE3*^{CBS7001} in addition to the ~20 amino acid changes between the two putative protein sequences, we replaced the promoter sequence (~1.5 kb) from the hybrid strain with the promoter sequence from *S. uvarum* to determine whether changes in this region affect growth in xylose. The xylose phenotype in GSY4326 or GSY4327 transformed with a plasmid containing this construct is the same as *GRE3*^{CBS1502} with its own promoter (Figure 6), indicating that changes in the promoter sequence are not responsible for the increased growth in xylose. More work is needed to determine what specific change(s) in *GRE3*^{CBS1502} coding sequence or in the 3′ region result in enhanced growth in xylose in the hybrid CBS1502.

To further investigate the requirement for *GRE3* and to determine whether it encodes the sole/major xylose reductase, we analyzed the phenotype of *GRE3* deletion in the presence of either *APJ1* allele. We showed that *GRE3* deletion eliminates growth in xylose in backgrounds containing either the *APJ1*^{CBS1502} or *APJ1*^{CBS7001} alleles (data not shown). This is in contrast to *GRE3*^{CBS7001} or *GRE3*^{CBS1502}, when paired with *APJ1*^{CBS7001}, allowing moderate growth in xylose (Figure 4A). These data show that *GRE3* is the main xylose reductase in CBS1502 and that the improvement in the xylose phenotype provided by *APJ1* requires the presence of a *GRE3* allele, be it the allele from CBS1502 or CBS7001. Taken together, these data show that *GRE3*^{CBS1502} encodes the major functional xylose reductase in CBS1502 and the *APJ1*^{CBS1502} xylose-positive phenotype requires its presence. In summary, our data show that *APJ1*^{CBS1502} (loss-of-function) and *GRE3*^{CBS1502} (gain-of-function) interact epistatically to contribute to the robust xylose-positive phenotype of CBS1502 and are the causative genes in the regions C and D genomic intervals from BSA of GSY2694.

GRE3 expression is higher in APJ1^{CBS1502} *strains*

Because transcription of *S. cerevisiae GRE3* is known to be responsive to stress (Garay-Arroyo and Covarrubias 1999) and because *APJ1* (at least in *S. cerevisiae*) is a heat-shock protein, we decided to test whether transcript abundances of *GRE3*^{CBS7001} or *GRE3*^{CBS1502} are altered in strains carrying the *APJ1*^{CBS1502} allele. Four tetratype tetrads from GSY4319 representing four biological replicates for each genotypic combination were grown in xylose-containing medium for 72 hr. We prepared total RNA and performed quantitative RT–PCR as described in the *Materials and Methods*. C_t values for *S. uvarum* YDR458C and YJL088W were used as controls because we did not observe significant variation of their levels of expression between different genotypes (Figure 7A). As shown in Figure 7B, relative mRNA levels of both *GRE3* alleles were increased in the presence of *APJ1*^{CBS1502}. The increase in *GRE3* transcript abundance in the *APJ1*^{CBS1502} *GRE3*^{CBS1502} strain compared to wild type is significant (*P*-value < 0.01, as determined by a *t*-test, Bonferroni correcting for 6 tests, each genotype using both controls). Neither of the other two genotypes showed a significant difference in *GRE3* transcript abundance compared to wild type.

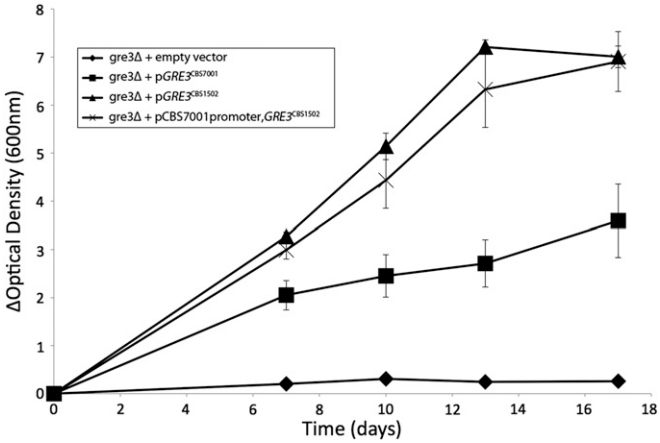

Figure 6 *GRE3* is the gene in region D responsible for growth in xylose. Data are optical density values (600 nm) of the difference (Δ) between growth in 2% xylose and no added carbon source. Average values for at least five biological replicates are plotted, with error bars showing standard deviation.

Discussion

In this study we characterized the genomic architecture of a polygenic xylose phenotype in a *Saccharomyces* hybrid yeast strain. Applying high-throughput sequencing to BSA of this phenotype revealed at least four loci that contribute to the phenotype; two are homologs of *S. cerevisiae GRE3* and *APJ1*, while the remaining two loci have yet to be identified.

Array CGH and sequencing revealed that this strain is a complex interspecific hybrid between *S. uvarum*, *S. cerevisiae*, and the recently described and sequenced species *S. eubayanus* (Libkind *et al.* 2011); its hybrid nature was further supported by its low levels of sporulation and spore viability, as is typical of hybrids (Greig *et al.* 2002). It is also possible that there are recessive lethal alleles that also contribute to the observed poor spore viability. Because of the complex nature of this strain, straightforward genetic techniques were not feasible, and we therefore developed a novel approach to performing genetic analyses in this hybrid, utilizing a tetraploid intermediate. Our method of generating tetraploids by transient expression of *HO* can be applied to any strain that cannot normally be sporulated for various reasons, and simply requires that the strain be amenable to DNA transformation, and that it is capable of mating to a closely related but polymorphic strain. Notably, this method may have applications for commercial yeasts, or yeasts isolated from industrial environments, which themselves are often hybrids or have poor or no sporulation (Tsuboi and Takahashi 1988).

We identified four loci in the CBS1502 hybrid that contribute to xylose utilization (including one that negatively affects growth in CBS1502) and identified two of the

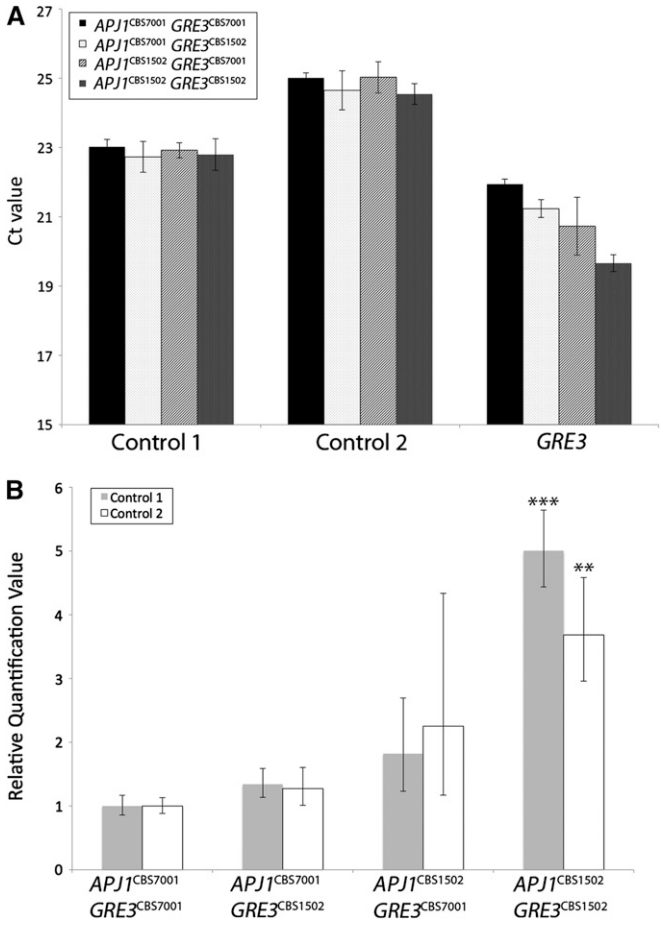

Figure 7 *GRE3* expression is higher in *APJ1*^{CBS1502} strains. (A) Average C_t values for reference genes (control 1 for YDR458C and control 2 for YJL088W) and for *GRE3* for four biological replicates, with error bars showing standard deviation. (B) Relative quantification values for *GRE3* between pairwise genotypical combinations for four biological replicates. Upper and lower limits (calculated as described in *Materials and Methods*) are shown as error bars. (***) *P*-value compared to wild type < 0.001; (**) *P*-value < 0.01.

genes that contribute to the xylose-positive phenotype: homologs of the *S. cerevisiae* genes *GRE3* and *APJ1*. In *S. cerevisiae*, *GRE3* encodes a nonspecific aldo-keto reductase that has NADPH-dependent activity on xylose as a substrate (Toivari *et al.* 2004). Our previous work has shown that endogenous *S. cerevisiae GRE3* contributes to xylose utilization in *S. cerevisiae* carrying the *XDH1* gene (Wenger *et al.* 2010). However, Gre3p in *S. uvarum* appears to be the major functional xylose reductase, unlike in *S. cerevisiae*.

S. cerevisiae Apj1p is a putative member of the Hsp40/DnaJ family of chaperone proteins. These proteins are involved in regulation of the heat-shock protein Hsp70 (Cyr *et al.* 1994) via direct interaction with Hsp70 through their conserved J domains. While we do not know the specific role of Apj1p during growth in xylose, we speculate that it might act as a negative regulator of *GRE3* expression. We have demonstrated that a hypomorphic *APJ1*^{CBS1502} allele results in higher *GRE3* transcript abundance compared to the presumably fully functional *APJ1*^{CBS7001} allele. Interestingly, the effect on *GRE3* transcript abundance is more pronounced for the CBS1502 allele than the CBS7001 allele, likely responsible for the epistatic interaction we observed. Because we have ruled out the promoter as being responsible for the allelic difference between *GRE3*^{CBS1502} and *GRE3*^{CBS7001} with respect to the xylose-positive phenotype, the allelic specificity may be due to *APJ1*^{CBS1502}-dependent increased stability of *GRE3*^{CBS1502} mRNA rather than direct transcriptional regulation. Indeed, it has been demonstrated not only that *GRE3* is induced under stress, but that *GRE3*'s transcript stability is also increased under stress (Castells-Roca *et al.* 2011). Perhaps our APJ1 hypomorphic allele somehow mimics a stress condition, either directly or indirectly affecting *GRE3*. Further work is required to determine the exact mechanism of increased transcript abundance of *GRE3* in the presence of the *APJ1*^{CBS1502} allele. We have also determined that the G234D substitution in Apj1^{CBS1502} is responsible for the xylose-positive phenotype; this glycine is conserved throughout the *Saccharomyces sensu stricto* and lies within Apj1's zinc finger domain (Walsh *et al.* 2004).

We previously identified the *XDH1* gene—which exists in some *S. cerevisiae* wine strains but not in laboratory strains—and found that it encodes a putative xylitol dehydrogenase and is sufficient to confer xylose utilization on a laboratory strain (Wenger *et al.* 2010). We tested for its presence in the other 38 xylose-positive strains identified in our original screen and showed by PCR that it is present in CBS1502 (Wenger *et al.* 2010). We have mapped *XDH1* in CBS1502 to the right end of chromosome IX (data not shown); this is striking in the context of the array comparative genomic hybridization data, which show a loss of *S. uvarum* sequence in this same location (Figure S1, black circled region), possibly suggesting that the *XDH1*-containing region of the CBS1502 genome introgressed from another species and replaced that portion of the *S. uvarum* genome. Sanger sequencing of the *XDH1* locus from CBS1502 revealed that this gene's DNA sequence is identical to the *XDH1* gene identified in various wine strains of *S. cerevisiae* (Wenger *et al.* 2010), suggesting that this sequence is identical by descent in CBS1502 and other *S. cerevisiae* strains that contain this region (Novo *et al.* 2009). Surprisingly, the presence or absence of *XDH1* has no effect on growth in xylose in CBS1502 progeny that contain *GRE3*^{CBS1502} and *APJ1*^{CBS1502} (data not shown). This suggests that there are other functional xylitol dehydrogenases encoded by the *S. uvarum* genome.

One drawback of our method to genetically analyze an otherwise intractable strain is that the BSA resulted in a large range of interval sizes for the identified loci, from as narrow as 10 kb to as large as 76 kb. This disparity in size of genomic intervals reinforces the notion that achieving specificity in BSA requires high meiotic recombination rates. The small pool size derived from GSY2694 (21 each of xylose positives and negatives), combined with potential recombination problems such as the possible low sequence

similarity, or the presence of inversions or translocations between the strains used, is likely responsible for the large interval sizes. These results suggest that adapting a strategy similar to X-QTL (Ehrenreich *et al.* 2010)—in which very large numbers of segregants are selected for opposite, extreme phenotypes—might be useful in cases such as this. Alternatively, or perhaps in combination, multiple rounds of segregation could also be useful in decreasing interval size (Parts *et al.* 2011).

This drawback notwithstanding, our BSA screen for loci associated with xylose growth identified the *APJ1* gene, a gene with no previously known connection to xylose metabolism. This demonstrates that the study of natural *Saccharomyces* xylose-utilizing yeasts still offers new discoveries for the improvement of currently existing, genetically modified *S. cerevisiae* xylose-fermenting strains. Identifying and understanding the genetic basis of novel xylose-metabolism phenotypes can uncover new enzymes or enzyme variants in the canonical xylose pathway or in other aspects of metabolism or cell biology that are important in xylose utilization, and modifications in these genes or pathways may help move these strains into industrial use.

Acknowledgments

The authors thank Yuya Kobayashi, Dan Kvitek, and Sasha Levy for critical reading of this manuscript and Travis Maures for assistance with RT-qPCR. We also thank Phil Lacroute and Ghia Euskirchen at the Stanford Center for Genomics and Personalized Medicine for high-throughput sequencing. This project was supported by the Stanford Global Climate and Energy Project (GCEP) (grant 33450) as well as the National Institutes of Health– National Institute of General Medical Sciences (NIH-NIGMS) Genetics & Developmental Biology Training Program (NIH GM007790).

Literature Cited

Attfield, P. V., and P. J. L. Bell, 2006 Use of population genetics to derive nonrecombinant *Saccharomyces cerevisiae* strains that grow using xylose as a sole carbon source. FEMS Yeast Res. 6: 862–868.

Brauer, M. J., C. M. Christianson, D. A. Pai, and M. J. Dunham, 2006 Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. Genetics 173: 1813–1816.

Buckeridge, M. S., and G. H. Goldman (Editors), 2011 *Routes to Cellulosic Ethanol*. Springer, New York.

Bullard, J. H., Y. Mostovoy, S. Dudoit, and R. B. Brem, 2010 Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. Proc. Natl. Acad. Sci. USA 107: 5058–5063.

Byrne, P. F., 2006 Safety and public acceptance of transgenic products. Crop Sci. 46: 113.

Castells-Roca, L., J. García-Martínez, J. Moreno, E. Herrero, G. Bellí *et al.*, 2011 Heat shock response in yeast involves changes in both transcription rates and mRNA stabilities. PLoS ONE 6: e17272.

Chandel, A. K., G. Chandrasekhar, K. Radhika, R. Ravinder, and P. Ravindra, 2011 Bioconversion of pentose sugars into ethanol: a review and future directions. Biotechnol. Mol. Biol. Rev. 6: 008–020.

Chiang, C., and S. G. Knight, 1960 Metabolism of D-xylose by moulds. Nature 188: 79–81.

Craig, J. E., A. W. Hewitt, A. E. McMellon, A. K. Henders, L. Ma *et al.*, 2009 Rapid inexpensive genome-wide association using pooled whole blood. Genome Res. 19: 2075–2080.

Cyr, D. M., T. Langer, and M. G. Douglas, 1994 DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. Trends Biochem. Sci. 19: 176–181.

Dunn, B., and G. Sherlock, 2008 Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. Genome Res. 18: 1610–1623.

Ehrenreich, I. M., N. Torabi, Y. Jia, J. Kent, S. Martis *et al.*, 2010 Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464: 1039–1042.

Gallagher, J. E. G., J. E. Babiarz, L. Teytelman, K. H. Wolfe, and J. Rine, 2009 Elaboration, diversification and regulation of the Sir1 family of silencing proteins in Saccharomyces. Genetics 181: 1477–1491.

Garay-Arroyo, A., and A. A. Covarrubias, 1999 Three genes whose expression is induced by stress in *Saccharomyces cerevisiae*. Yeast 15: 879–892.

Gietz, R. D., and A. Sugino, 1988 New yeast-*Escherichia coli* shuttle vectors constructed with *in vitro* mutagenized yeast genes lacking six-base pair restriction sites. Gene 74: 527–534.

Greig, D., E. J. Louis, R. H. Borts, and M. Travisano, 2002 Hybrid speciation in experimental populations of yeast. Science 298: 1773–1775.

Ha, S.-J., J. M. Galazka, S. R. Kim, J.-H. Choi, X. Yang *et al.*, 2011 Engineered *Saccharomyces cerevisiae* capable of simultaneous cellobiose and xylose fermentation. Proc. Natl. Acad. Sci. USA 108: 504–509.

Ho, N., Z. Chen, and A. Brainard, 1998 Genetically engineered *Saccharomyces* yeast capable of effective cofermentation of glucose and xylose. Appl. Environ. Microbiol. 64: 1852–1859.

Jeffries, T., 1983 Utilization of xylose by bacteria, yeasts, and fungi. Adv. Biochem. Eng. Biotechnol. 27: 1–32.

Jeffries, T. W., 2006 Engineering yeasts for xylose metabolism. Curr. Opin. Biotechnol. 17: 320–326.

Kim, S. R., K.-S. Lee, J.-H. Choi, S.-J. Ha, D.-H. Kweon *et al.*, 2010 Repeated-batch fermentations of xylose and glucose-xylose mixtures using a respiration-deficient *Saccharomyces cerevisiae* engineered for xylose metabolism. J. Biotechnol. 150: 404–407.

Kuyper, M., A. Winkler, J. Dijken, and J. Pronk, 2004 Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: a proof of principle. FEMS Yeast Res. 4: 655–664.

Lee, A., K. D. Hansen, J. Bullard, S. Dudoit, and G. Sherlock, 2008 Novel low abundance and transient RNAs in yeast revealed by tiling microarrays and ultra high-throughput sequencing are not conserved across closely related yeast species. PLoS Genet. 4: e1000299.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25: 1754–1760.

Li,, H., B. Handsaker, A. Wysoker, T. Fennel, J. Ruan *et al.*, 2009 The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25: 2078–2079.

Libkind, D., C. T. Hittinger, E. Valerio, C. Goncalves, J. Dover *et al.*, 2011 Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc. Natl. Acad. Sci. USA 108: 14539–14544.

Lunter, G., and M. Goodson, 2011 Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21: 936–939.

Marschall, L. G., R. L. Jeng, J. Mulholland, and T. Stearns, 1996 Analysis of Tub4p, a yeast gamma-tubulin-like protein:

implications for microtubule-organizing center function. J. Cell Biol. 134: 443–454.
Matsushika, A., H. Inoue, T. Kodaki, and S. Sawayama, 2009 Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. Appl. Microbiol. Biotechnol. 84: 37–53.
Novo, M., F. Bigey, E. Beyne, V. Galeote, F. Gavory *et al.*, 2009 Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc. Natl. Acad. Sci. USA 106: 16333–16338.
Parts,, L., F. Cubillos, J. Warringer, K. Jain, and F. Salinas *et al.*, 2011 Revealing the genetic structure of a trait by sequencing a population under selection. Genome Res. 21: 1131–1138.
Quarrie, S., V. Lazic-Jancic, D. Kovacevic, A. Steed, and S. Pekic, 1999 Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. J. Exp. Bot. 50: 1299–1306.
Scannell, D. R., O. A. Zill, A. Rokas, C. Payen, M. J. Dunham *et al.*, 2011 The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. G3: Genes, Genomes, Genetics 1: 11.
Schiestl, R. H., and R. D. Gietz, 1989 High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr. Genet. 16: 339–346.
Sonderegger, M., and U. Sauer, 2003 Evolutionary engineering of *Saccharomyces cerevisiae* for anaerobic growth on xylose. Appl. Environ. Microbiol. 69: 1990–1998.
Toivari, M. H., L. Salusjärvi, L. Ruohonen, and M. Penttilä, 2004 Endogenous xylose pathway in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 70: 3681–3686.
Träff, K. L., R. R. Otero Cordero, W. H. van Zyl, and B. Hahn-Hägerdal, 2001 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.
Treco, D. A., 1987 Preparation of yeast DNA, pp. 13.11.1–13.11.2 in *Current Protocols in Molecular Biology*, edited by F. M. Ausubel, R. Brent, R. Kingston, D. D. Moore, J. G. Seidman *et al*. John Wiley, New York.
Tsuboi, M., and T. Takahashi, 1988 Genetic analysis of the non-sporulating phenotype of brewer's yeasts. J. Ferment. Technol. 66: 605–613.
Walsh, P., D. Bursać, Y. C. Law, D. Cyr, and T. Lithgow, 2004 The J-protein family: modulating protein assembly, disassembly and translocation. EMBO Rep. 5: 567–571.
Wang, P. Y., C. Shopsis, and H. Schneider, 1980 Fermentation of a pentose by yeasts. Biochem. Biophys. Res. Commun. 94: 248–254.
Wenger, J. W., K. Schwartz, and G. Sherlock, 2010 Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. PLoS Genet. 6: e1000942.

Communicating editor: J. Boeke

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2012/03/16/genetics.112.140053.DC1

APJ1 and *GRE3* Homologs Work in Concert to Allow Growth in Xylose in a Natural *Saccharomyces sensu stricto* Hybrid Yeast

Katja Schwartz, Jared W. Wenger, Barbara Dunn, and Gavin Sherlock

DOI: 10.1534/genetics.112.140053

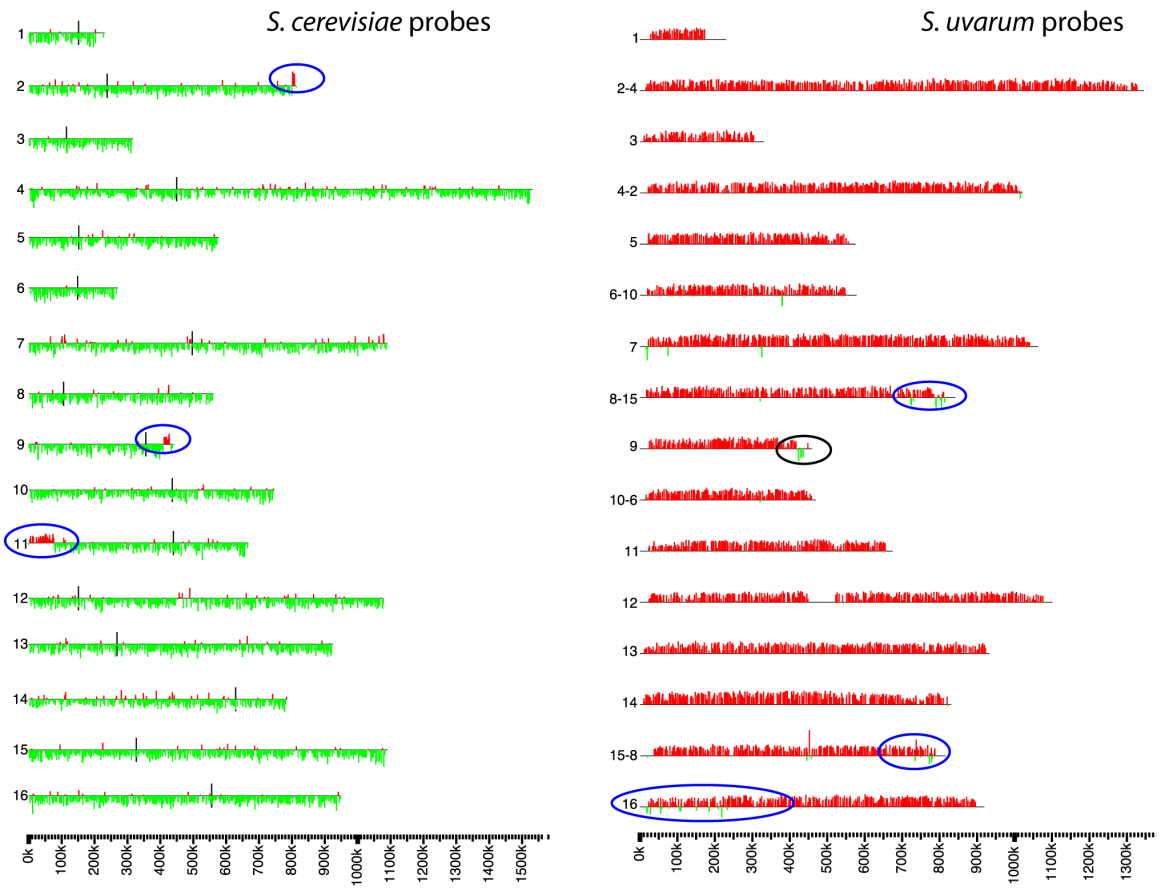

Figure S1 Array Comparative Genomic Hybridization of CBS1505. Genomic content of GSY2680 was assayed using custom microarrays that distinguish *S. cerevisiae* genomic content from *S. uvarum*. Red and green bars represent the log2 ratio of probe intensity of GSY2680 relative to a reference DNA pool of equimolar *S. cerevisiae* and *S. uvarum*. Blue circles denote regions of introgression or hybridization regions from *S. cerevisiae* (left panel) or possibly *S. eubayanus* (right panel; weaker hybridization to the *S. uvarum* probes). The black circle indicates the location of the genomic region containing *XDH1*.

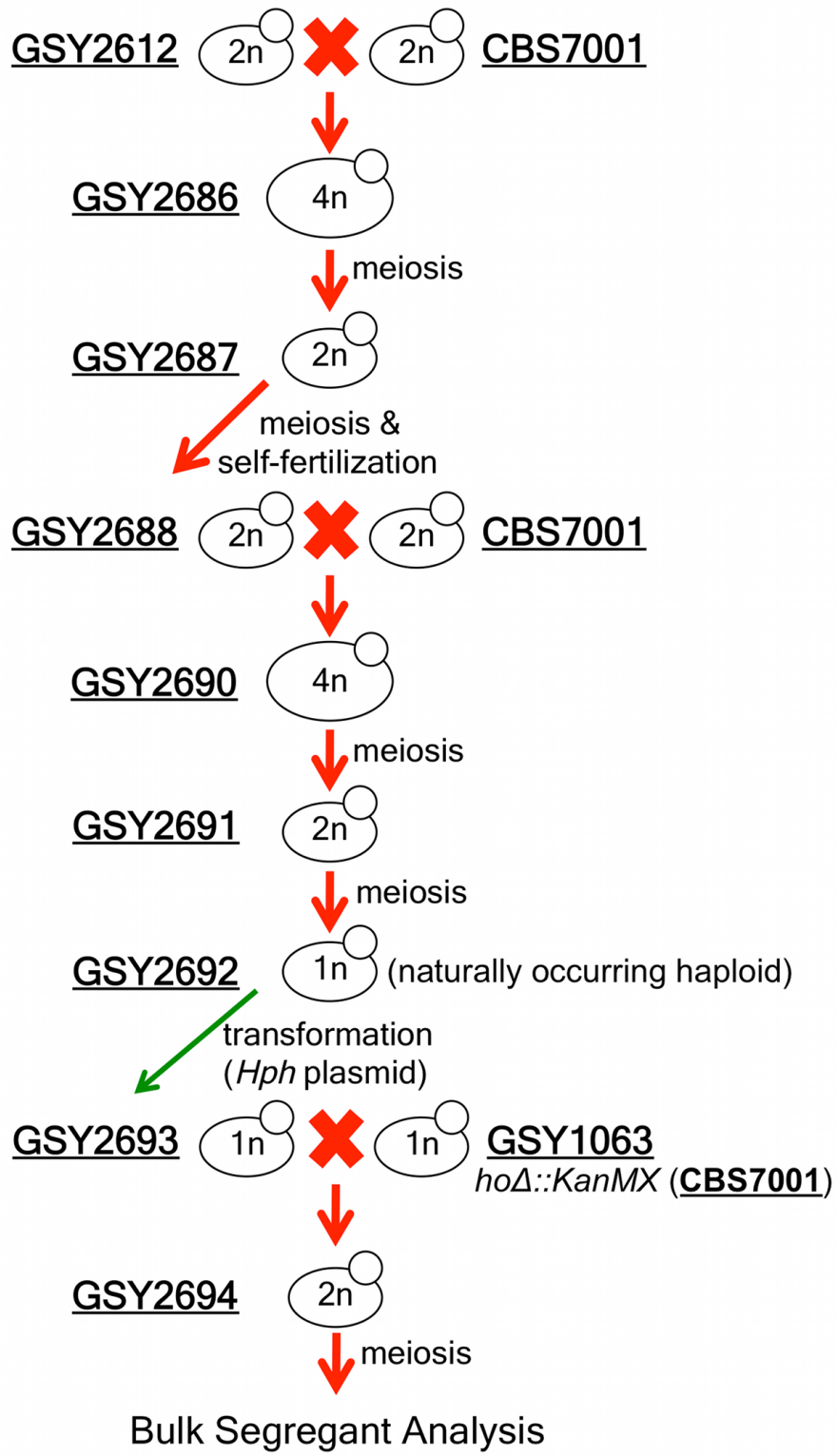

Figure S2 Diagram of Backcrosses for Derivation of GSY2694 Lineage of CBS1502. Schematic of the generation of the lineage used for Bulk Segregant Analysis.

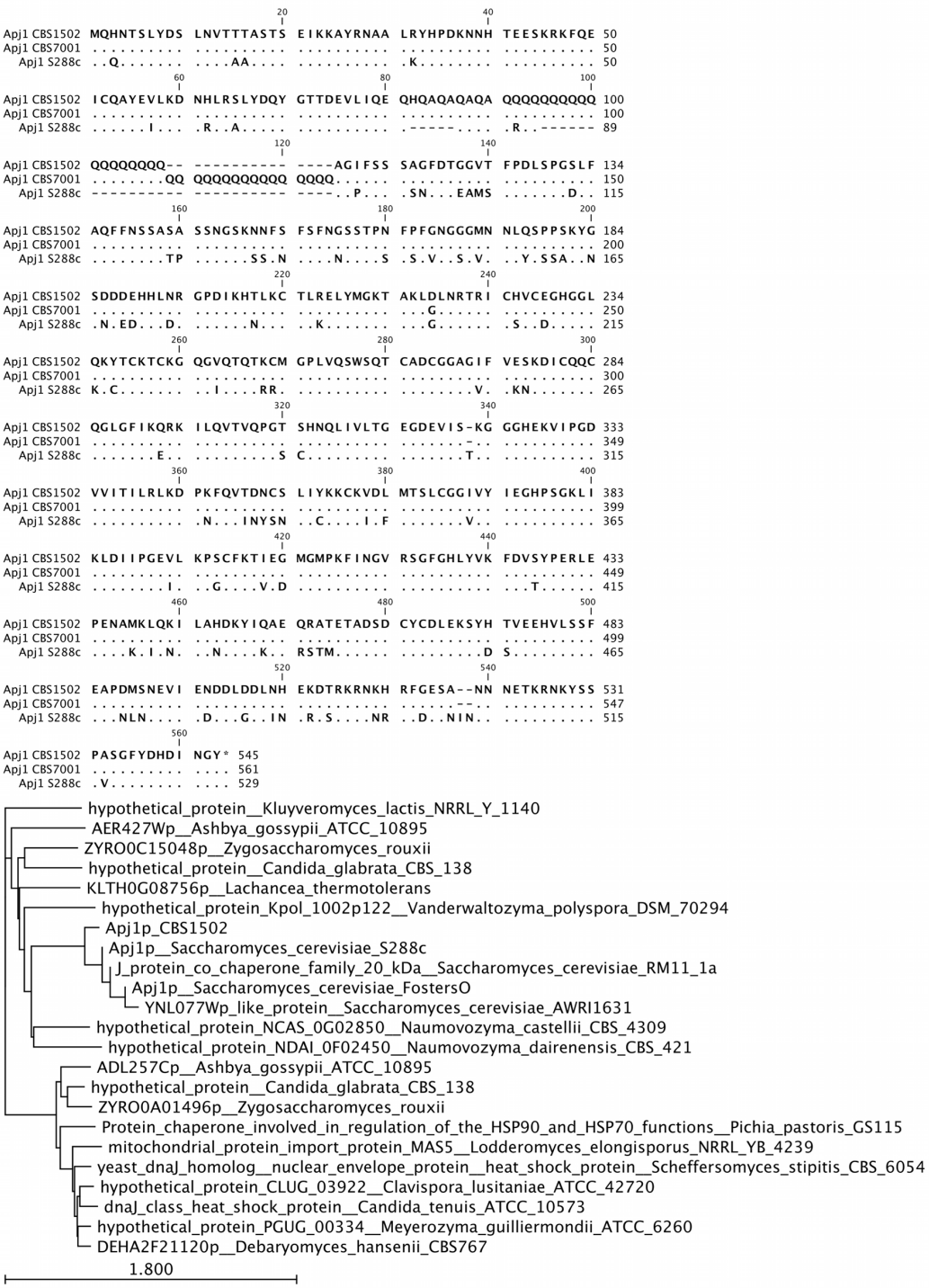

Figure S3 Multiple Sequence Alignment and Phylogenetic Tree of *APJ1. APJ1* was sequenced from CBS1502 via Sanger sequencing, and the predicted protein sequence was aligned to the *S. cerevisiae* and *S. uvarum* predicted protein sequences using CLC Sequence Viewer. The phylogenetic tree was constructed using the COBALT tool with default parameters available from NCBI. BLAST hits were included that had an e value less than or equal to 10^{-50}.

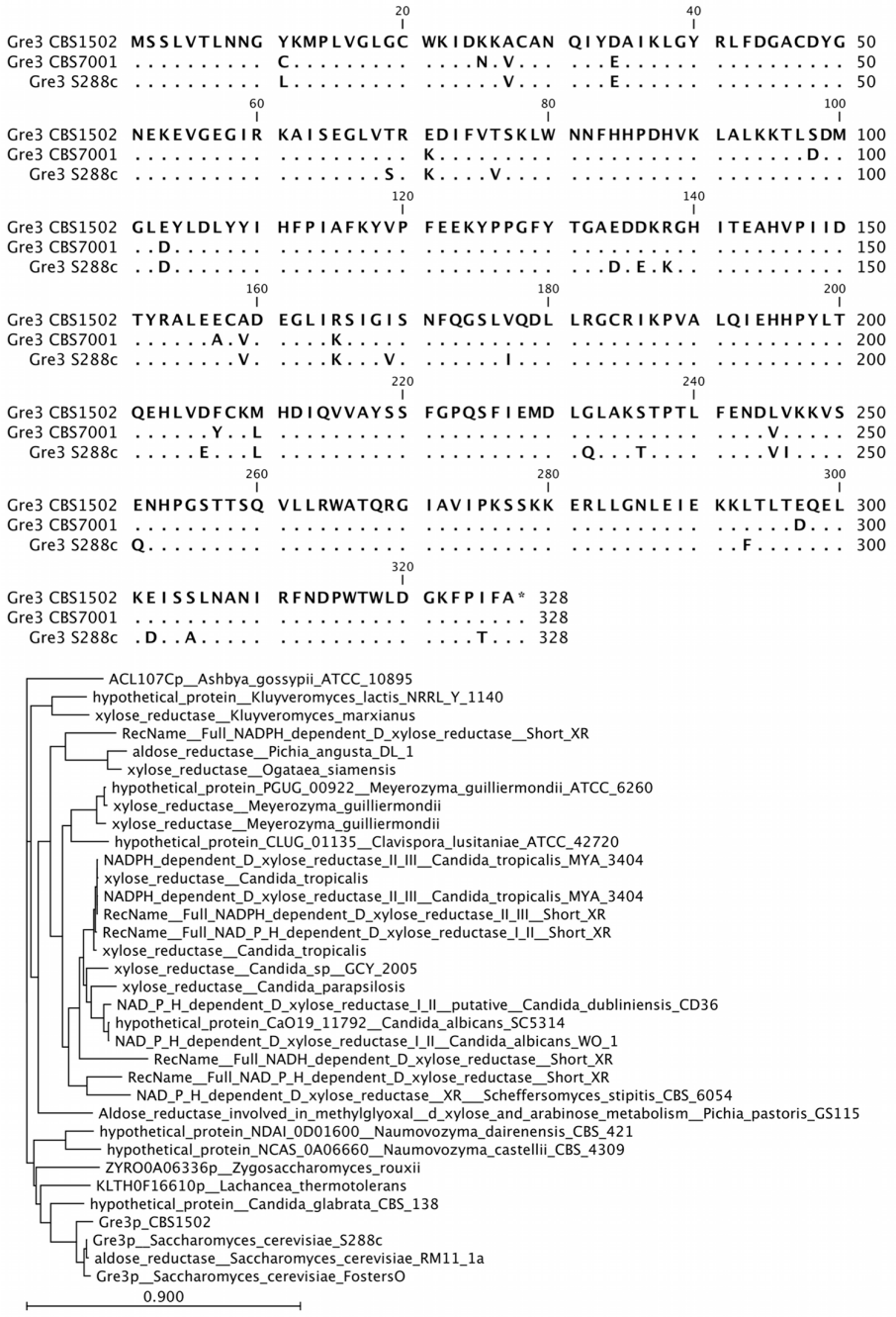

Figure S4 Multiple Sequence Alignment and Phylogenetic Tree of *GRE3. GRE3* was sequenced from CBS1502 via Sanger sequencing, and the predicted Gre3^{CBS1502} protein sequence was aligned to the *S. cerevisiae* and *S. uvarum* predicted protein sequences using CLC Sequence Viewer. The phylogenetic tree was constructed using the COBALT tool with default parameters available from NCBI. The top 33 BLAST hits were included to show Xyl1p from *Scheffersomyces stipitis.*

Table S1 Primers used in this study

Primer	Sequence (5'-3')	Description
GSP12	GTCTGCAGGAGTACCACG	*HO* disruption in *S.uvarum*, amplifies 5'
GSP13	*GTCGTGACTGGGAAAACCCTGGCG*GGCTAGATGGAAGTGATTGC	*HO* disruption in *S.uvarum*, amplifies 5' to fuse to m13-*kanMX*
GSP14	*TCCTGTGTGAAATTGTTATCCGC*TGCGTGTCCTAGAAAGGGC	*HO* disruption in *S.uvarum*, amplifies 3' to fuse to m13-*kanMX*
GSP15	CAATGGAGTGACCGTATTGG	*HO* disruption in *S.uvarum*, amplifies 3'
GSP1	*GCGTCTAGA*ATGCTTTCTGAAAACACGACTATTCTGATGGCT	*HO* locus amplification with XbaI site
GSP545	*CGCTCTAGA*CACCAAGGCCATGTCTTCTCG	*HO* locus amplification with XbaI site
GSP546	*GCGGGATCC*CCGTTCCAAAGCTGAGAAACCAG	*APJ1* locus amplification with BamHI site
GSP547	*CGCGGATCC*CTTTGAAGGGTTATGCTAGGTTCG	*APJ1* locus amplification with BamHI site
GSP535	*CGCGGATCC*AACTAGATGCAGTAAGTCACTCAAGGC	$GRE3^{CBS7001}$ locus amplification with BamHI site
GSP538	*CGCGGATCC*TGTGAGTCAATTTCGAATTCAGG	$GRE3^{CBS7001}$ locus amplification with BamHI site
GSP536	*CGCGGATCC*AACTGGATGCAGTAAGTCACTCAAG	$GRE3^{CBS1502}$ locus amplification with BamHI site
GSP539	*CGCGGATCC*CTCCTGCTGAGAGTCAACTTCGAAT	$GRE3^{CBS1502}$ locus amplification with BamHI site
GSP533	CCAAAAAAAAAAAAAAGAGAAATTCAAAATGTCATCACTAGTCACTTTGAACA	primer to fuse $GRE3^{CBS1502}$ ORF with $GRE3^{CBS7001}$ promoter
GSP548	*GTCGTGACTGGGAAAACCCTGGCG*CTCCGTACTTCTGTATTACCTAATAG	*APJ1* disruption, amplifies 5' to fuse to m13-*URA3*
GSP549	*TCCTGTGTGAAATTGTTATCCGC*TGAAGGCTATATGTTGGCATTAATTGC	*APJ1* disruption, amplifies 3' to fuse to m13-*URA3*
GSP550	CGCTCAAAGGACACAACTCCGGC	*APJ1* disruption, amplifies 5'
GSP551	CTTTGAAGGGTTATGCTAGGTTCG	*APJ1* disruption, amplifies 3'
GSP552	GCATCCCCGTATTCATCGACGATGAAC	*GRE3* disruption, amplifies 5'
GSP553	*GTCGTGACTGGGAAAACCCTGGCG*GTTCAAAGTGACTAGTGATGACAT	*GRE3* disruption, amplifies 5' to fuse to m13
GSP554	*TCCTGTGTGAAATTGTTATCCGC*GACCCATGGACTTGGTTGGACGGT	*GRE3* disruption, amplifies 3' to fuse to m13
GSP555	GTAATTGGGGTGCTTCGGCGGCTGCC	*GRE3* disruption, amplifies 3'
GSP556	GAAGCCCATGTCCCTATCAT	*GRE3* qPCR primer, forward
GSP557	TACACCCCCGCAGTAAATCT	*GRE3* qPCR primer, reverse
GSP558	TCGTTTGGTTGCCTTTGGAC	*YDR458c,* qPCR primer, forward (Bullard *et al*. 2010)
GSP559	ACTTCTTTAGCCCCTCCTGA	*YDR458c,* qPCR primer, reverse (Bullard *et al*. 2010)
GSP560	ATTGGGTGTCAACGAATCCT	*YJL088w,* qPCR primer, forward (Bullard *et al*. 2010)
GSP561	CAAACCGCTTGTAAGGGATG	*YJL088w,* qPCR primer, reverse (Bullard *et al*. 2010)