Genetic Engineering of Yeast with Novel Properties Using Localized Hypermutation

Bledar Xhialli

Supervisor: Dr. Kin Chan

This thesis is submitted to the University of Ottawa in partial fulfillment of the requirement of the M.Sc. in Biochemistry program.

May 17, 2020

Department of Biochemistry, Microbiology, & Immunology

Faculty of Medicine

University of Ottawa

ABSTRACT

Bioethanol is a promising alternative to fossil fuels, but is currently too cost prohibitive. Bioethanol is produced from the fermentation of sugars derived from plant biomass by bacteria and yeasts such as *S. cerevisiae*. Improving the fermentation step by engineering more efficient microorganisms can help reduce costs. The gold standards for strain development in biofuel research are UV or chemical mutagenesis and laboratory evolution. These approaches are limited by their low mutation rates and nonspecificity, which make engineering of polygenic traits difficult. This study investigated using an inducible double-strand break and human APOBEC3A to generate localized hypermutation as a method for targeting genes of interest at high mutation rates in *S. cerevisiae*. I hypothesized that localized hypermutation can be a useful tool for strain development that complements traditional techniques. Expression of APOBEC3A in *S. cerevisiae* increased the single gene inactivation frequency in a reporter system by 2 orders of magnitude and the double gene inactivation frequency by 1 order of magnitude. *S. cerevisiae* expressing *XYL1* and *XYL2* from *S. stipitis* and subjected to localized hypermutation yielded mutants with 1.8-fold higher growth rate on xylose over adapted-only controls. Using localized hypermutation also facilitated the generation of glucose and galactose co-utilizing mutants. By producing mutants with improved growth on xylose and mixed-sugar media, these results demonstrate the applicability of localized hypermutation to not only the biofuel industry, but to biotechnology in general.

ACKNOWLEDGEMENTS

I owe many thanks to my parents for their never-ending patience, support and the numerous sacrifices they made for my education. To my supervisor for giving me this amazing opportunity to contribute to his research and for arranging this project to suit my research interests. I haven't expressed enough gratitude for not only being a great research mentor, but also taking an active interest in preparing me for my future career. To my colleagues, Reena, Nathan, Suzana, Salma, Kassidy, Ghadir, and Emma for their constant companionship and for making the lab such an enjoyable place to be. I've never had such stimulating discussions about science before and I'm grateful for that. To my thesis advisory committee and my thesis examiners, Dr. Baetz, Dr. Downey, and Dr. Kaern for taking time out of their busy schedule to provide insightful advice on my research. Thanks to all of you for making graduate school such a pleasant and unforgettable experience.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	Adenine
2-DG	2-Deoxyglucose
5-FOA	5-Fluoroorotic acid
A3A	APOBEC3A
Ade^-	Adenine auxotrophy phenotype
bp	Base pair
C	Cytosine
Can^R	Canavanine resistance phenotype
Chr.	Chromosome
CI	Confidence interval
CEN	Centromere element
CORE	Counter-selectable reporter cassette
DNA	Deoxyribonucleic acid
DNS	3,5-dinitrosalicylic acid
Doxy	Doxycycline
DSB	Double-strand break
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
ER	Ethanol Red
G	Guanine

gDNA	genomic DNA
I-SceIcs	I-SceI cut site
Indel	insertion and deletion
kb	kilobase
Kl	*Kluyveromyces lactis*
LH	Localized hypermutation
NAD+	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced)
NADP+	Nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
nt	Nucleotide
OD	Optical density
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
P_{GAL1}	Promoter of *GAL1*
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PES	Polyethersulfone
pH	Potential of hydrogen ion
SBS	Single-base substitution
SC	Synthetic complete

SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
Ss	*Scheffersomyces stipitis*
ssDNA	Single-stranded DNA
T	Thymine
Ura^-	Uracil auxotrophy phenotype
XI	Xylose isomerase
XDH	Xylitol dehydrogenase
XR	Xylose reductase
YP	Yeast extract peptone
YPAc	Yeast extract peptone 20 g/L acetate
YPD	Yeast extract peptone 20 g/L glucose
YPGal	Yeast extract peptone 20 g/L galactose
YPR	Yeast extract peptone 20 g/L raffinose
wt	Wild type

LIST OF FIGURES

LIST OF SUPPLEMENTARY FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

1. INTRODUCTION

1.1. Climate change and bioethanol

Over the last decade anthropogenic climate change has attracted increasing national attention and is being considered a worldwide crisis according to the United Nations. Society's dependence on fossil fuels has been attributed to global temperature increase, increased frequency of extreme weather, drought, acidification of the oceans, and melting of the arctic ice sheets. The link between fossil fuel usage and climate change is that combustion of gasoline produces carbon dioxide, which is accumulating in the atmosphere and contributes to the greenhouse effect (climate.nasa.gov). Bioethanol is a promising alternative to fossil fuels in part because the feedstock is plant biomass, which is cheap, abundant, sustainable, and not limited to any geographical location, but also because combustion of bioethanol is considered carbon neutral (eia.gov). Furthermore, key aspects of bioethanol production are agriculture and fermentation, both of which have a long history of research and development.

In Canada, bioethanol is mainly produced from wheat and corn (nrcan.gc.ca). Using food crops such as wheat and corn as feedstocks for vehicle fuel presents an ethical dilemma thus current research is focused on using inedible plant biomass such as corn husks. This type of biomass is mainly lignocellulose and is composed of lignin, cellulose, and hemicellulose polymers. The cellulose and hemicellulose are made of mainly hexose and pentose sugars. The production of bioethanol starts with pretreatment of lignocellulose to remove the nonfermentable lignin, then the remaining

cellulose and hemicellulose are hydrolyzed into simple sugars typically by acid treatment or enzymatic treatment (Mohd Azhar et al., 2017). The composition of the remaining sugars is about 60% glucose and 30% xylose (Okeke and Obi, 1994). These sugars are fed to microorganisms such as *S. cerevisiae* and fermented into ethanol fuel. The main advantage of bioethanol over fossil fuels is that carbon dioxide produced from combustion of ethanol fuel is recaptured through photosynthesis by growing more crops. This recycling of carbon dioxide is what makes bioethanol considered carbon neutral; however, crop harvesting, transport, and treatment are energy consuming and greenhouse gas emitting processes. Despite this drawback, life cycle assessment studies are optimistic that bioethanol provides net savings in greenhouse gas emissions (Hanaki and Portugal-Pereira, 2018). With more efficient fertilizer production, bioethanol is projected to provide 25% less global warming potential than fossil fuels (Wang et al., 2013).

According to the U.S. Department of Energy, the price of 85% ethanol fuel was on average equal to the price gasoline in 2019 (afdc.energy.gov), but since ethanol contains 30% less energy content per volume than gasoline, it is currently more expensive on an energy basis. In order to reduce the price of bioethanol enough to be competitive with gasoline, improvements must be made in the pretreatment, hydrolysis, and fermentation steps. The focus of this study is on strain development of *S. cerevisiae* for improving fermentation of xylose and glucose mixtures for bioethanol production.

1.2. Xylose utilization in *S. cerevisiae*

D-xylose is a naturally occurring pentose commonly found as a constituent of the polysaccharide cell wall of plants. Since xylose makes up about 30% of the sugars in lignocellulosic hydrolysates, considerable research has been devoted to efficient fermentation of this pentose. Many species of bacteria and fungi can use xylose as a carbon source, but wild *S. cerevisiae* cannot. Despite this shortcoming, *S. cerevisiae* is still the standard microorganism used in research for two reasons: its long history in genetics research facilities strain development through genetic modification, and its long history in brewing has produced strains with high ethanol tolerance. In xylose-utilizing yeasts such as *Scheffersomyces stipitis*, xylose is isomerized to xylulose in a two-step reaction. First, xylose reductase (XR) reduces xylose to xylitol in an NADPH dependent reaction, then xylitol dehydrogenase (XDH) oxidizes xylitol to xylulose in an NAD+ dependent reaction. Xylulose is then phosphorylated by xylulokinase and enters central carbon metabolism through the pentose phosphate pathway and ultimately leads to ethanol production (Figure 1). Since *S. cerevisiae* can grow on xylulose , but not on xylose it is evident that unlike other yeasts, *S. cerevisiae* has defective genes for converting xylose into xylulose (Chiang et al., 1981).

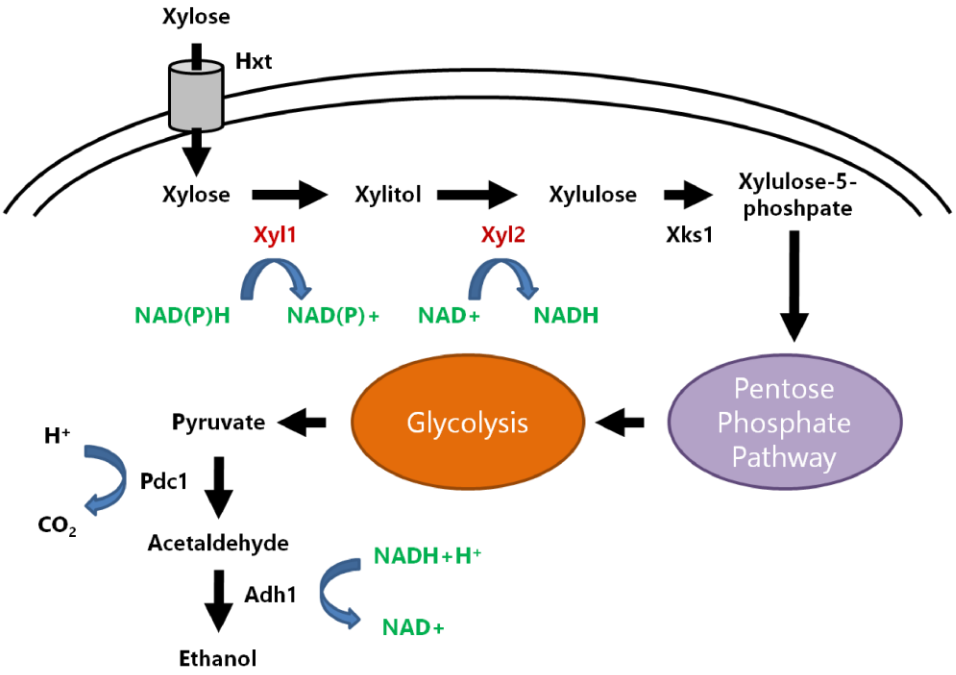


Figure 1. Xylose-utilization pathway in yeasts. Xylose is transported across the plasma membrane through hexose/pentose transporters (Hxt) and is converted into xylitol by xylose reductase (Xyl1 or XR) then into xylulose by the *S. stipitis* xylitol dehydrogenase (Xyl2 or XDH). Xylulokinase Xks1 converts xylulose into xylulose-5-phosphate, which enters central carbon metabolism through the anaerobic part of the pentose phosphate pathway. Glycolytic intermediates produced by the pentose phosphate pathway are converted into pyruvate through glycolysis. Pyruvate then gets converted to ethanol through the sequential reactions catalyzed by pyruvate decarboxylase (Pdc1) and alcohol dehydrogenase (Adh1).

S. cerevisiae contains several general reductases with up to 72% amino acid similarity to XR and dehydrogenases with up to 65% amino acid similarity to XDH. Overexpression of these genes was shown to enable xylose utilization in *S. cerevisiae* (Toivari et al., 2004). Attfield and Bell (2006) isolated a nonrecombinant xylose-utilizing *S. cerevisiae* mutant with a doubling time of 6 hours after over 1,400 days of selection; however, neither of these strains are suitable for efficient bioethanol production.

Currently, bioethanol-producing *S. cerevisiae* strains are engineered with heterologous expression of genes from natural xylose-utilizing yeast or bacteria. Both approaches have advantages and disadvantages. With the yeast XR/XDH pathway, the enzymes express better in *S. cerevisiae* since the genes come from hosts that are more closely related. The downside is that the accumulation of xylitol can significantly reduce ethanol yield if XDH activity is not closely coupled to XR activity. In the bacterial pathway, xylose isomerase (XI) converts xylose to xylulose in a single reaction without xylitol production. The downside is that bacterial XIs exhibit protein-folding problems when expressed in yeast without additional bacterial chaperone proteins (Temer et al., 2017). A comparison of the top performing strains revealed that backgrounds with XR/XDH had higher ethanol productivity and xylitol production, while the XI backgrounds had higher ethanol yield and less xylitol production (Li et al., 2016).

While both approaches are commonly used, this study will focus on the development of the XR/XDH pathway. The activity of XR and XDH from *S. stipitis* was first characterized in the 1990s and shown to be sufficient for enabling xylose utilization

in *S. cerevisiae,* albeit with low expression (Amore et al., 1991; Kötter et al., 1990). *S. stipitis* is the preferred donor as it exhibits the highest ethanol yields and xylose uptake rates among common xylose-utilizing yeasts (Veras et al., 2017). Even with heterologous expression of XR and XDH, *S. cerevisiae* has 3 major hurdles that must be overcome for efficient xylose utilization: (1) xylose transport, (2) XR/XDH cofactor imbalance, and (3) low fermentation capacity (Hou et al., 2017). Monosaccharides enter the cell by facilitated diffusion using hexose transporters, but *S. cerevisiae* lacks dedicated xylose transporters. In mixed-sugar media such as that from lignocellulose feedstocks, xylose is competitively inhibited by glucose as the transporters have over 100-fold affinity for glucose over xylose (Kötter and Ciriacy, 1993). Through rational mutagenesis, directed mutagenesis, and laboratory evolution single amino acid changes in the endogenous transporters have been discovered that change the affinity preference from glucose to primarily xylose transport (Farwick et al., 2014; Nijland et al., 2014; Oreb et al., 2012; Young et al., 2014).

Cofactor imbalance is a major problem with the XR/XDH pathway. The *S. stipitis* XR requires either NADH or NADPH as a cofactor, but strongly prefers NADPH (Verduyn et al., 1985). Conversely XDH from *S. stipitis* uses strictly NAD+ as a cofactor. Thus the conversion of xylose to xylulose will perturb the NADP+/NADPH and NAD+/NADH ratio, which is maintained at a steady level in the cell (Canelas et al., 2008). Regeneration of NADPH occurs in the pentose phosphate pathway via a decarboxylation step, thus ethanol yield is reduced as some carbon is lost to CO_2 production. Moreover, when

grown on glucose, *S. cerevisiae* will oxidize excess NADH by fermentation to ethanol or by production of glycerol. Analysis by Kuyper et al. (2004) suggests that fermentation alone is not sufficient to maintain a redox balance and aerobic respiration is necessary to oxidize the net NADH produced by XDH. Production of glycerol has a net energy cost, therefore *S. cerevisiae* resorts to respiring xylose in the presence of oxygen and fails to grow anaerobically. Different approaches for rewiring cofactor regeneration have been successful in improving xylose utilization, such as using heterologous expression of NADH dependent enzymes to oxidize excess NADH (Henningsen et al., 2015; Zhang et al., 2012, 2016). An innovative approach by Zhang et al. (2016) utilized bacterial acetyl-CoA synthetase (ACS) and NADH dependent acetylating acetaldehyde dehydrogenase (AADH) to convert acetate into ethanol. This provides three benefits: it detoxifies the acetate inhibitor found in lignocellulosic hydrolysates, it regenerates NAD+, and it converts acetate to ethanol.

A second reason why xylose is preferentially respired is because *S. cerevisiae* does not recognize xylose as a fermentable carbon source (Jin et al., 2004; Osiro et al., 2019). Transcriptional studies reveal that when *S. cerevisiae* is grown on xylose, there is an upregulation in respiratory pathways such as the tricarboxylic acid cycle (Salusjärvi et al., 2008), but pathways central to xylose utilization like the pentose phosphate pathway are not significantly upregulated (Jin et al., 2004; Zeng et al., 2016).

Because efficient xylose utilization requires the coordination of many pathways: respiratory response, fermentative growth, pentose phosphate pathway, cofactor

regeneration, and transport, strain development is often done through random mutagenesis and laboratory evolution. High aerobic growth rates in *S. cerevisiae* have been achieved (Demeke et al., 2013a; Liu and Hu, 2010; Scalcinati et al., 2012). Analysis of the mutations from different xylose-utilizing strains does not point to a key regulator, but rather many global changes (Hou et al., 2017).

Current research is focused on improving fermentative growth under anaerobic conditions, which is preferable because *S. cerevisiae* tends to respire xylose under room air (Wahlbom et al., 2003), and it eliminates aeration costs in industrial bioreactors. Enabling anaerobic growth is more challenging for *S. cerevisiae* and adaptation periods can take over 80 days in already engineered strains (Kuyper et al., 2004; Sonderegger and Sauer, 2003). Among top performing strains ethanol yields greater than 80% of the theoretical yield under oxygen-limited conditions have been attained (Cunha et al., 2019; Hector et al., 2013).

Overall, great strides have been made in engineering *S. cerevisiae* for efficient use of xylose for aerobic and anaerobic growth; however, there is a distinct lack of genetic studies to broaden the understanding of the mechanisms behind adaptation to xylose. Studies typically pursue only the best mutants and therefore sample sizes are small. Furthermore, *XYL1* and *XYL2* are implicated in 2 of the main problems with xylose utilization, but there is a lack of research in investigating mutants in *XYL1* and *XYL2*. In this study *XYL1* and *XYL2* were chosen for hypermutation to investigate if novel allele combinations can be produced that allow for more efficient growth on xylose.

1.3. Glucose repression and co-utilization of sugars in *S. cerevisiae*

S. cerevisiae prefers glucose over other types of carbon sources and has evolved sophisticated regulatory systems to ensure the efficient breakdown of glucose before other carbon sources (Busti et al., 2010; Gancedo, 2008; Santangelo, 2006). In addition to regulating carbon metabolism, glucose also signals for global changes in cell state such as regulation of ribosome biogenesis, stress response, cell cycle, budding, and more (Broach, 2012; Busti et al., 2010; Schneper et al., 2004). When glucose is sensed by the cell it generates a signal cascade that ultimately leads to transcriptional upregulation of genes for fermenting glucose and downregulation of genes for utilizing alternative carbon sources such as sucrose, galactose, ethanol, and glycerol. This phenomenon is called *glucose repression* or *catabolite repression*. The minimum concentration for glucose repression depends on the second carbon source, but can be as little as 2.5 g/L in the case of sucrose (Meijer et al., 1998). There are 3 main pathways involved in glucose repression: (1) the Snf3/Rgt2 pathway, (2) the Snf1 pathway, and (3) the Gpr1-Gpa2 pathway (Schneper et al., 2004). These pathways are cross-talking and generally converge on the kinase SNF1, which when activated signals for the upregulation of alternative carbon source metabolic genes. There is also evidence of translational regulation in response to glucose (Ashe et al., 2000).

In the Snf3/Rgt2 pathway glucose is sensed extracellularly to regulate both alternative carbon source metabolism and glucose flux into the cell. The glucose sensors Snf3 and Rgt2 respond to low and high glucose concentrations and upregulate high-

and low-affinity hexose transporters respectively. There 18 hexose transporters in *S. cerevisiae*; *HXT1* through *HXT17* and *GAL2*. In *snf3*Δ *rgt2*Δ cells that lack both types of glucose sensors, growth is impaired on glucose due to lack of glucose transport, and cells exhibit glucose derepression (Ozcan et al., 1998).

While the Snf3/Rgt2 pathway has been studied for glucose transport and glucose flux, the Snf1 pathway has been studied for the induction of alternative carbon source metabolic genes. SNF1 is a multisubunit kinase that serves as a master regulator of alternative carbon source metabolism. The complex consists of Snf1 (catalytic subunit), Snf4 (regulatory subunit), and either Gal83/Sip1/Sip2, which are thought to confer substrate specificity. In the absence of glucose, the phosphorylation of this complex occurs through a not completely elucidated mechanism and results in the upregulation of over 400 genes (Young et al., 2003). In the presence of glucose, SNF1 becomes deactivated through dephosphorylation by the Glc7-Reg1 phosphatase complex. Reg1 is a key regulatory unit that directs Glc7 towards SNF1 deactivation in the presence of the glucose (Tu and Carlson, 1995). Hxk2 is the main hexokinase isoform expressed on glucose and is known to interact with SNF1 (Ahuatzi et al., 2007). Additionally, there is evidence of interaction between Snf1, Hxk2, and Reg1 (Fernández-García et al., 2012; Sanz et al., 2000). Moreover, deletions in *HXK2* and *REG1* also reduce expression of hexose transporter genes (Ozcan and Johnston, 1995). This means that *HXK2* and *REG1* affect signalling in both the Snf1 and Snf3/Rgt2 pathway.

The third glucose-sensing pathway in *S. cerevisiae* has an overlapping and

largely redundant role with the Snf3/Rgt2 pathway (Belinchón and Gancedo, 2007). The Gpr1-Gpa2 pathway is a nitrogen and carbon starvation sensing pathway that signals for the increase of cAMP levels to activate protein kinase A (Gancedo, 2008). Gpr1 is thought to be a glucose sensor that upon activation regulates growth (Lorenz et al., 2000), ribosome production (Klein and Struhl, 1994), *HXK2* expression, and hexose transporter expression (Palomino et al., 2006).

The takeaway from these findings is that *HXK2* and *REG1* are central players in glucose repression as they are involved in all 3 or 2 of the glucose signalling pathways respectively. The importance of *HXK2* and *REG1* has been corroborated by genetic studies in which only *hxk2*Δ and *reg1*Δ cells provided the strongest glucose derepression phenotype (McCartney et al., 2014). Surprisingly, no studies have investigated *HXK2* and *REG1* in the context of glucose and xylose co-utilization for bioethanol production. In some studies, spontaneous mutants in *HXK2* arose during laboratory evolution. Lane et al. (2018) adapted *S. cerevisiae* for co-utilization on glucose and xylose and found a mutant with a single-base substitution (SBS) in all three hexokinase isoforms (*HXK2*, *HXK1*, *GLK1*). In another study xylose and glucose co-utilizing mutants also harboured SBSs in *HXK2* as well as other genes (Papapetridis et al., 2018). No studies were found that isolated mutants in *REG1* or mutants in both genes. Both of these studies produced mutants that could co-utilize glucose and xylose, but they consumed glucose at slower rates than normal. In this study *HXK2* and *REG1* were chosen for hypermutation to

investigate if novel allele combinations could be generated that allow for more efficient co-utilization of mixed-sugar media.

1.4. Localized hypermutation and APOBEC3A

In this study human APOBEC3A (A3A) was used in yeast to generate localized hypermutation as a method of generating yeast mutants with novel properties. There are 11 genes in the human genome that are part of the apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) family of genes (Salter et al., 2016). Included in this family is the activation induced deaminase (AID) that is involved in somatic hypermutation in antibody diversification. The subfamily of APOBEC3 enzymes are cytidine deaminases that convert cytidine bases into uridine in single-stranded DNA (ssDNA) only and are thought to provide defence against pathogens by mutating foreign DNA (Stavrou and Ross, 2015). APOBEC3A in particular is expressed in monocytes/macrophages and has been shown to protect against HIV-1 infection (Berger et al., 2011). More recently the APOBEC3 genes have been implicated in the mutation patterns of certain cancers (Alexandrov et al., 2013; Roberts et al., 2013). In these cancers, there exist genome-wide clusters of many mutations that span a region of about 10 kb (Nik-Zainal et al., 2012). This phenomenon called kataegis has been attributed to hypermutation by APOBEC3 enzymes and has been corroborated by yeast studies (Chan et al., 2015).

In another study by Chan et al. (2012), APOBEC3G was expressed in an *S. cerevisiae* background with the temperature-sensitive *cdc13-1* allele. In this system, the telomere-capping protein Cdc13 becomes inactive at a restrictive temperature and unbinds from the telomeres, which leads to telomere resection and ssDNA formation. Hypermutation by APOBEC3G revealed as many as 7 mutations within the span of 10 kb. This system has been successful in studying the mutation profiles of DNA damaging agents, but the nature of this technique presents some challenges in its application as a mutagenesis technique. First, this system limits the temperature at which cells can be grown without inducing arrest. Since temperature is an important parameter affecting growth rate and metabolism, the *cdc13-1* restriction would make it difficult to compare mutants to those grown at the standard 30 °C. Second, this system would make all the chromosome ends susceptible to hypermutation, which is undesirable as it increases off-target mutations. In this study an improvement on the *cdc13-1* system was investigated by using an inducible double-strand break (DSB) system. Upon induction of the DSB, DNA resection occurs analogously to telomere-end resection, except limited to only 2 DNA ends. The second advantage is that this system is not temperature dependent.

The method is summarized in Figure 2. The principle behind this approach is that A3A's specificity for ssDNA can be used to hypermutate genes of interest by causing those genes to be in ssDNA form at a greater frequency than the rest of the genome (Chan et al., 2012). This can be achieved by using an inducible double-strand break system. In this example the I-SceI endonuclease is placed under a *GAL1* promoter and

induces a DSB at its recognition site in the presence of galactose. In *S. cerevisiae* DSBs lead to 5′ end resection by exonucleases that can span up to at least 20 kb on either side of the DSB (Storici et al., 2006). An inducible A3A (in this example placed under a doxycycline regulated promoter) will deaminate cytidines in the ssDNA, which encompass the genes of interest. Uracil bases are excised by uracil-DNA glycosylase to generate abasic sites. During DNA repair of the resected strand, translesion synthesis occurs via polymerase ζ (zeta), which inserts a base across the abasic site. The error-prone activity of polymerase zeta generates mainly C to T (equivalently G to A) and C to G mutations (Chan et al., 2012).

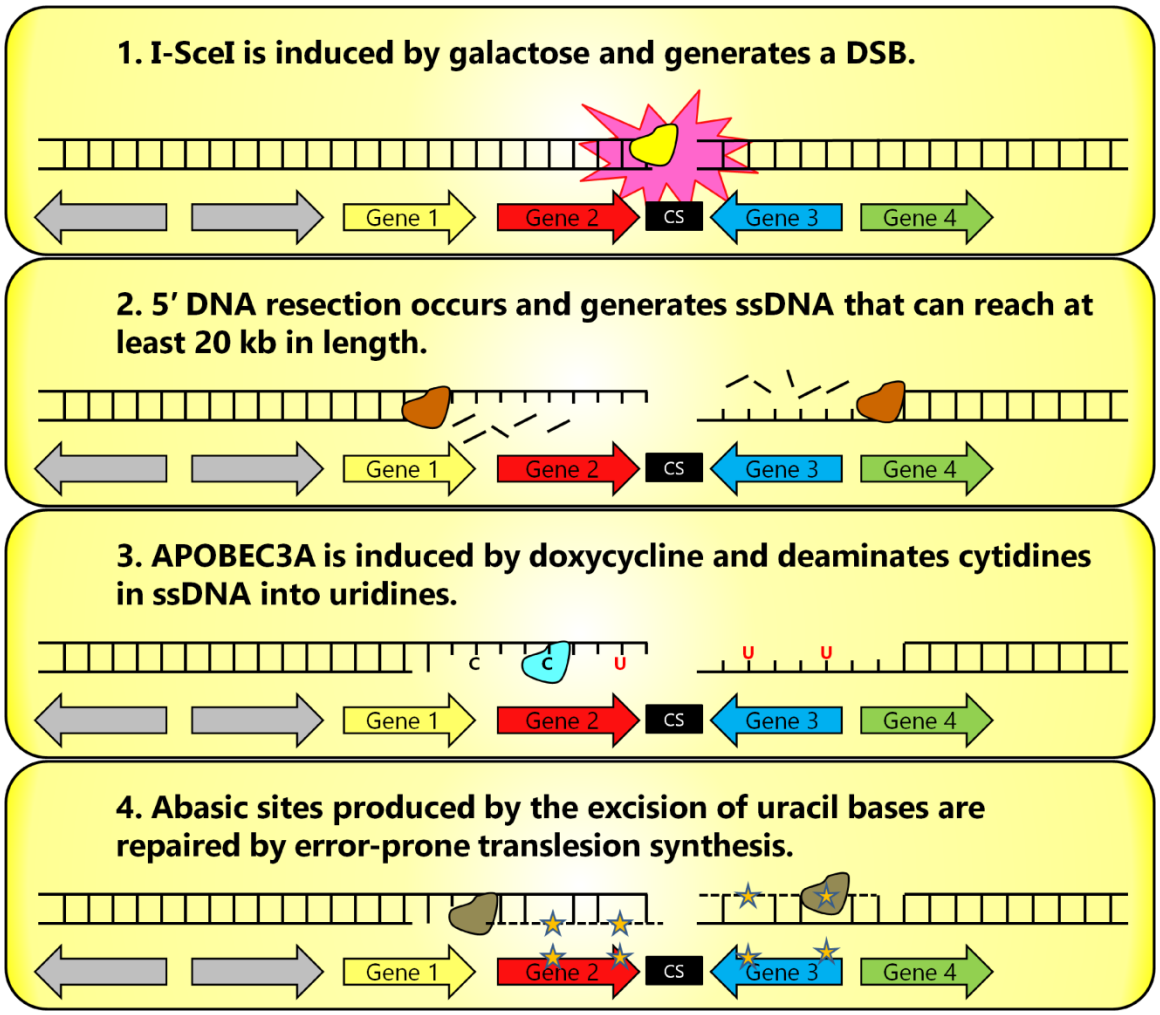


Figure 2. Mechanism for localized hypermutation with APOBEC3A. 1. Genes of interest are placed next to an inducible DSB. Induction of I-SceI by galactose generates a DSB. 2. Exonucleases cause DNA resection that can span up to 20 kb in both directions and converts the genes of interest into ssDNA form. 3. Induction of A3A with doxycycline leads to deamination of cytidines in the ssDNA. Uracil-DNA glycosylase excises uracils in ssDNA and creates abasic sites. 4. During DNA synthesis polymerase zeta is employed to extend past abasic sites in an error-prone manner. Mutations of mainly cytosine to thymine and cytosine to guanine are created at sites deaminated by A3A.

1.5. Rationale

The relationship between evolutionary fitness and changes in genotype are often described by a fitness landscape as shown in Figure 3 (Visser and Krug, 2014). For some measure of fitness, such as growth of yeast on a particular substrate, the elevations in the fitness landscape describe different fitness levels; peaks represent increased growth rate and valleys represent decreased growth rate. The path between any two points on the landscape represents the mutational path to go from one state to another. As shown in Figure 3, a change in initial fitness (point A) to improved fitness (point C) can be achieved through several paths. The mutations acquired through paths 1 and 2 either confer no change or improved fitness at all times. In some cases the shortest path to improved fitness may require passing through a fitness valley as indicted by path 3. This can occur if several particular mutations improve growth when present together, but individually are deleterious and thus the cell is at a lower fitness until it acquires all the mutations (Kvitek and Sherlock, 2011). The implication of fitness valleys is that they act as hurdles in adaptation and may limit paths towards improved fitness.

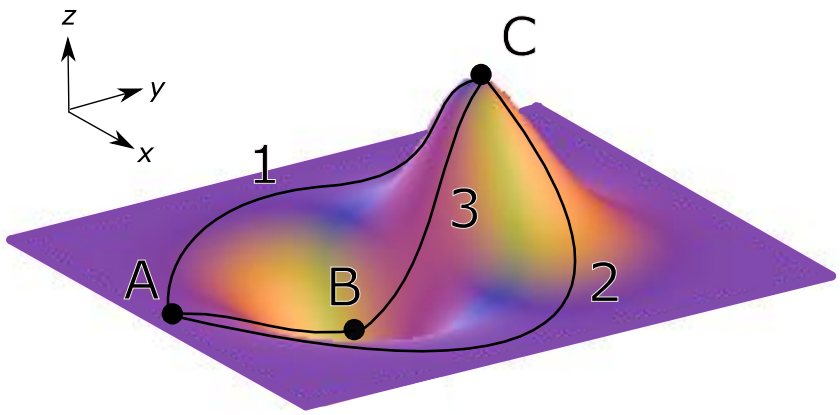


Figure 3. Representation of a fitness landscape. Intuitively, the fitness landscape can be visualized as a surface in 3 dimensions that maps the genotype space in the xy-plane to some measure of fitness in the z-axis. Point A represents the initial fitness level and the peak at point C represents a genotype conferring higher fitness. Likewise, the valley at point B represents a genotype conferring lower fitness. Paths 1, 2, and 3 represent different mutation paths that can lead from point A to point C.

In strain development two of the most widely used techniques for generating yeast mutants are radiation and chemical mutagenesis combined with laboratory evolution (Steensels et al., 2014). These techniques have low mutation rates and studies that employ them frequently isolate mutants with usually a single mutation in few genes (Lane et al., 2018; Nijland et al., 2014; Papapetridis et al., 2018; Reider Apel et al., 2016). Consequently these techniques may restrict adaptation to paths limited with the fewest nondeleterious mutations. For example if a cell gains a deleterious mutation and enters a fitness valley (point B in Figure 3), and if the mutation rate is too low then the cell may get outcompeted and drop out from selection before it can cross the valley. Alternatively, with a low mutation rate paths that go around fitness valleys, such as paths 1 and 2 in Figure 3, may require numerous mutations and thus very long adaptation periods. In addition, the non-specific nature of techniques like UV mutagenesis make it difficult to sample diverse phenotypes for polygenic traits. For example Abt et al. (2016) generated yeast mutants carrying upwards of 900 mutations, but only obtained mutations in 3 genes relevant to their objective. A hypermutation approach for generating yeast mutants may be useful for strain development in biotechnology because the high mutation rates will allow for jumping over fitness valleys and quicker adaptation periods. This may allow for sampling of phenotypes that could be extremely rare with traditional methods.

1.6. Hypothesis and aims

1.6.1. Hypothesis

Localized hypermutation with APOBEC3A can be used to simultaneously mutate several genes of interest at high frequencies and allows sampling of phenotypes that may be rare for traditional methods.

1.6.2. Aims

1. Assess the ability of localized hypermutation with APOBEC3A to increase the mutation frequency of a triple gene reporter system.
2. Test if localized hypermutation can be used to improve polygenic traits by targeting the genes *XYL1* and *XYL2* for hypermutation in an engineered *S. cerevisiae* background and screening for xylose-utilizing mutants.
3. Test if localized hypermutation can be used to improve polygenic traits by targeting the genes *HXK2* and *REG1* for hypermutation in *S. cerevisiae* and screening for mixed-sugar co-utilizing mutants.

2. MATERIALS AND METHODS

2.1. Strains, media, and culture conditions.

All yeast strains used in this study are listed in Table 1. The *S. cerevisiae* strain ySR128 was genetically modified prior to this work to contain an *ADE2 URA3 CAN1* triple reporter system in the *LYS2* locus of chromosome 2 (Hoopes et al., 2016). This strain was the progenitor of strains used in experiments for quantifying the inactivation frequency of the reporter system by APOBEC3A. The *S. cerevisiae* strain Ethanol Red was a gift from Lesaffre Advanced Fermentation (USA) and was the progenitor of all strains used in experiments for generating xylose-utilizing, and glucose and galactose co-utilizing mutants. All strains were kept in 20% v/v glycerol stocks and stored at -80 °C. Cultivation of yeasts was done in yeast extract peptone (YP), synthetic complete (SC), or mineral media. Liquid YP media was prepared with 20 g/L peptone (BD, USA), 10 g/L yeast extract (BD, USA), and 20 g/L of various carbon sources (glucose, galactose, xylose, raffinose, or sucrose) from Biobasic (Canada). Synthetic media was prepared with 20 g/L glucose (Biobasic, Canada), 5 g/L ammonium sulfate (Biobasic, Canada), 1.7 g/L yeast nitrogen base (without amino acids and ammonium sulfate) purchased from BD (USA), 0.06 g/L adenine sulfate, 0.05 g/L L-arginine HCL, 0.075 g/L L-aspartic acid, 0.1 g/L L-glutamic acid, 0.02 g/L L-histidine HCL, 0.05 g/L L-isoleucine, 0.1 g/L L-leucine, 0.12 g/L L-lysine HCL, 0.02 g/L L-methionine, 0.05 g/L L-phenylalanine, 0.375 g/L L-serine, 0.1 g/L L-threonine, 0.05 g/L L-tryptophan, 0.05 g/L L-tyrosine, 0.15 g/L L-valine, 0.06 g/L uracil (amino acids and nucleotides were from BioBasic, Canada), and pH adjusted to 5.8 with

sodium hydroxide. Mineral media consists of the same components as SC media, but without amino acids and nucleotides. For solid media, an additional 20 g/L of agar A (Biobasic, Canada) was added. Liquid media was filter sterilized with 0.2 μm PES bottle top filters (Fisher Scientific, USA) and solid media was sterilized by autoclave at 121 °C for 30 min. Selective media was supplemented with filter-sterilized solutions of concentrated reagents as necessary (after autoclaving and cooling in the case of solid media). These reagents were filter sterilized using 0.2 μm PES syringe filters (Fisher Scientific, USA).

For routine cultivation, frozen stocks were streaked on 2% w/v glucose yeast extract peptone (YPD) plates and incubated at 30 °C for 2 to 4 days or until single colony isolates appeared. For experiments involving growth on xylose or galactose, 2% w/v raffinose mineral media plates were used. Tubes with 5 mL of YPD or YP Raffinose (YPR) cultures were prepared using a flame-sterilized wire to transfer a small amount of yeast mass from an isolated colony as inoculum. Cultures were incubated in an Excella E24 Incubator Shaker (New Brunswick Scientific, USA) that was set to a temperature of 30 °C and a shaking speed of 200 rpm.

Table 1. List of yeast strains used in this study. Strains used in the gene inactivation frequency experiments were constructed in the ySR128 background. Strains used in the xylose and sugar co-utilization experiments were constructed in the Ethanol Red background. The genotype and phenotype of Ethanol Red was largely unknown besides being prototrophic on mineral media. Gene deletions and translocations were done using plasmid pCORE-UK to screen for the presence of G418 resistance and 5-FOA sensitivity.

***S. cerevisiae* strain**	**Parent Strain**	**Genotype**	**Source**
ySR128	CG379	*MATα his7-2 ura3*Δ *can1*Δ *ade2*Δ *leu2-3,112 trp1-289 lys2*::*ADE2-URA3-CAN1*	(Hoopes et al., 2016)
ySR128D	ySR128	*ura3*Δ::pGAL1-*SCEI-kanMX*-(I-SceIcs)-Kl*URA3*	This work
ySR128A	ySR128	*ura3*Δ::pGAL1-*SCEI-kanMX*-Kl*URA3* pSR435	This work
YSR128DA	ySR128D	pSR435	This work
ER	-	*MATa/MATα*	Fermentis (Lesaffre)
yBX27	ER	*MATa*	
yBX76	yBX27	*ura3*Δ::pGAL1-*SCEI-kanMX*-(I-SceIcs)-Kl*URA3* Chr. XVI 850475::Ss*XYL1*-(I-SceIcs)-Ss*XYL2*	This work
yBX76DA	yBX76	pSR435	This work
*hxk2*Δ	yBX27	*ura3*Δ::*hph hxk2*Δ::*kanMX*-Kl*URA3*	This work
*reg1*Δ	yBX27	*ura3*Δ::*hph reg1*Δ::*kanMX*-Kl*URA3*	This work
*hxk2*Δ *reg1*Δ	yBX27	*ura3*Δ::*hph hxk2*Δ *reg1*Δ::*kanMX*-Kl*URA3*	This work
yBX130	yBX27	*reg1*Δ *hxk2*Δ *ura3*Δ::pGAL1-*SCEI-kanMX*-(I-SceIcs)-Kl*URA3* Chr.XVI 850475::*HXK2*-(I-SceIcs)-*REG1*	This work
yBX130DA	yBX130	pSR435	This work

2.2. Plasmids and DNA sequences. The plasmids and DNA sequences used for genetic manipulation are listed in Table 2 and Supplementary Table 1 respectively. The plasmids were housed in *E. coli* DH5α competent cells (Invitrogen, USA) and stored at -80 °C. Plasmid extraction was done using plasmid purification kits on ampicillin resistant *E. coli* isolates. The plasmid pSR435 was purified using the QIAGEN Plasmid Midi Kit (Qiagen, USA), and the rest were purified using the EZ-10 Spin Column Plasmid DNA Miniprep kit (BioBasic, Canada). DNA oligonucleotides of desalted purity were synthesized by Thermo Fisher Scientific (USA) and were resuspended in tris buffer (pH 8) and stored at -20 °C. Plasmid pSR435 carried the codon-optimized human APOBEC3A. The sequence for the APOBEC3A ORF was verified by Sanger sequencing done by the StemCore Laboratories (Canada).

Table 2. List of plasmids used in this study. Plasmids pGSKU, pCORE-UK, and pAG32 were used as template for PCR amplification. Plasmid pSR435 is based on the pCM252 centromeric vector and contains the doxycycline inducible A3A. This plasmid was introduced in all yeasts for expression of A3A. The A3A ORF was verified by Sanger sequencing before use.

Plasmid	Description	Source
pGSKU	Amp-R Kl*URA3* kanMX pGAL1-10-*SCEI*	(Storici et al., 2003)
pCORE-UK	Amp-R Kl*URA3* kanMX	(Storici and Resnick, 2003)
pSR435	Amp-R *hph* ARS416 pCMV-tTA p6XTET-R-APOBEC3A I-SceIcs	(Chan et al., 2015)
pAG32	Amp-R P_{TEF}-*hph*-T_{TEF}	(Goldstein and McCusker, 1999)

2.3. Genetic construction.

Genetic manipulations were done using the *delitto perfetto* method (Storici and Resnick, 2006). This method utilizes a counterselectable reporter (CORE) cassette to facilitate markerless gene insertion via homologous recombination. Strain ySR128D was constructed by replacing *URA3* with a cassette containing *kanMX*, *KlURA3* (*URA3* from *Kluyveromyces lactis*) and P_{GAL1}*SCEI* (*SCEI* under the *GAL1* promoter). The cassette was amplified from pGSKU using primer pair oBX1+oBX2; one of which contains the I-SceI cut site. Strain ySR128DA was constructed by transforming ySR128D with pSR435. The strain ySR128A was constructed similarly to ySR128DA except using primer pair oBX1+oBX11 that lacks the I-SceI cut site for integration of the *kanMX-KlURA3*-P_{GAL1}*SCEI* cassette. Genomic integration was verified by PCR using primer pair oBX3+oBX4.

The CORE cassette used for generating strains yBX76 and yBX130 was amplified from pCORE-UK and consists of kanMX that confers resistance to G418 and Kl*URA3* that confers sensitivity to 5-fluoroorotic acid (5-FOA) in *ura3*Δ backgrounds. Strain construction in these backgrounds was done in a sequential manner by first integrating the CORE cassette at the landing site for a gene of interest and screening for G418 resistance and uracil prototrophy in *ura3*Δ cells. Next, the gene of interest was targeted to the CORE cassette and transformants were screened for G418 sensitivity and 5-FOA resistance.

Construction of the *XYL1 XYL2* background was done as follows. First, *URA3* was replaced with the hygromycin B resistance gene *hph* using primer pair oBX20+oBX21

and PCR verified with primer pair oBX22+oBX23. Next, an I-SceI cut site was introduced in chromosome 16 using primer pairs oBX61+oBX62, oBX65+oBX66, and oBX63+oBX64 for placing the CORE cassette, integrating the I-SceI cut site, and PCR verification respectively. The *XYL1* gene including the entire intergenic region upstream and downstream of the ORF was amplified from genomic DNA (gDNA) of *S. stipitis* strain Y-11545 (acquired from NRRL, USA) using oBX70-1+oBX71. The gene was placed upstream of the I-SceI cut site using oBX61+oBX67 and oBX63+oBX64 for CORE cassette placement and PCR verification respectively. *XYL2* was amplified along with its upstream and downstream intergenic regions using oBX76+oBX77 and integrated downstream of the I-SceI cut site using oBX72+oBX73 and oBX74+oBX64 for CORE cassette placement and PCR verification respectively. A cassette consisting of I-SceI under a GAL1-10 promoter and Kl*URA3* was amplified from pGSKU using oBX20+oBX21 and targeted to the *ura3*Δ::*hph* locus and verified with oBX22+oBX23.

Construction of the relocated *HXK2* and *REG1* background was performed similarly. *HXK2* and *REG1* were amplified from Ethanol Red gDNA using oBX68+oBX69 and oBX78+oBX79 respectively. The *HXK2* ORF including the upstream and downstream intergenic region was integrated in chromosome 16 using oBX61+oBX67 and oBX63+oBX64 for CORE placement and PCR verification respectively. For *REG1*, primer pairs oBX72+oBX73 and oBX75+oBX64 were used for CORE placement and PCR verification respectively. Deletion of *HXK2* and *REG1* at their native loci was achieved by replacement with a CORE cassette then subsequent deletion of the CORE cassette by

transformation with a pair of 100 bp complementary oligonucleotides. For *HXK2* deletion, primer pairs and oligonucleotides oBX38+oBX39, oBX42+oBX43, and oBX40+oBX41 were used for CORE placement, CORE deletion, and PCR verification respectively. For *REG1* deletion, oBX80+oBX81, oBX84+oBX85, and oBX82+oBX83 were used for CORE placement, CORE deletion, and PCR verification. Successful transformation with the plasmid pSR435 harbouring APOBEC3A was verified by PCR using primer pair oBX13+oBX14 to amplify the APOBEC3A ORF.

2.4. Polymerase chain reaction.

Template DNA was amplified using a T100 Thermal Cycler (Bio-Rad Laboratories, USA). Taq polymerase (Neb, USA) was used for routine PCR verification and amplification of CORE cassettes. Reactions were prepared in a volume of 25 µL following the manufacturer's protocol. Template DNA was supplied in the form of purified gDNA or a small amount of colony biomass (just enough to cover the opening of a typical 1 µL pipette tip). Annealing temperatures were determined using NEB's online T_M Calculator (v 1.12.0) and the extension time was set to 1 min/kb (rounding up). For amplification of *XYL1*, *XYL2*, *HXK2*, *REG1*, and *SCEI* the high-fidelity polymerases Phusion or Q5 were used (Neb, USA) with an extension time of 30 s/kb. Primers were designed using Primer3 version 0.4.0 (Untergasser et al., 2012). PCR products used for transformation were purified and concentrated using an EZ-10 Spin Column PCR Purification kit (BioBasic, Canada)

2.5. Gel electrophoresis.

Agarose gels were run for visualization of PCR amplicon size in a Mini-Sub Cell GT Cell (Bio-Rad Laboratories, USA). Gels were typically 0.7% w/v agarose (BioBasic, Canada) and prepared in 50 mL of 1X TBE (BioBasic, Canada) and microwaved for 1 minute to dissolve the agarose. Prior to casting, 0.5 µg/mL of ethidium bromide (BioBasic, Canada) was added. PCR reactions were mixed with 6X purple gel loading dye (Neb, USA) and loaded at various dilutions along with a 1 kb or 100 bp DNA ladder (Neb, USA). Gels were run with 1X TBE running buffer typically at 120 V for 30 min to 1.5 hours depending on the expected band sizes. Imaging was done using a Gel Doc XR+ System (BioRad, USA) with automatic UV exposure settings.

2.6. Yeast transformation.

Yeasts were transformed using a standard lithium acetate protocol (Storici and Resnick, 2006). Briefly, an overnight YPD culture grown from a patch of cells was used for a 3:100 dilution into 50 mL of fresh YPD media. This culture was grown for 3 h at 30 °C and 200 rpm. The culture was then spun in a Sorvall ST 16 centrifuge (Thermo Scientific, USA) at 3,000 rpm for 2 min. The pellet was washed with sterile water and resuspended in 250 µL of buffer (lithium acetate 0.1 M, tris 10 mM pH 8.0, EDTA 1 mM, pH 7.5). An aliquot was mixed at a ratio of 1:5 with 300 µL of 50% w/v PEG 4000 buffer (lithium acetate 0.1 M, tris 10 mM pH 8.0, EDTA 1 mM, pH 7.5). Concentrated PCR product (or plasmid) and freshly boiled salmon sperm DNA were added at ratios of 1:5 and 1:10 respectively relative to

the volume of cell suspension added. The mixture was incubated at 30 °C for 30 min then transferred to a water bath at 42 °C for 15 min. The cells were washed with sterile water and plated on YPD media. After 1 to 2 days of incubation at 30 °C, imprints were made on selective plates YPD (G418 200 µg/mL, hygromycin B 300 µg/mL, or 5-flouroorotic acid 1 g/L) using sterile velvet clothes. The antibiotics and uracil analogue were purchased from BioBasic (Canada). Positive candidates that grew on selective plates were restreaked onto YPD plates. Single colony isolates from YPD plates were tested for the following phenotypes: mitochondrial function, DNA repair defects, G418 resistance, hygromycin B resistance, uracil prototrophy, and 5-FOA sensitivty. These phenotypes were tested by imprinting candidate patches onto 2% v/v glycerol , 6×10^{-3}% w/v L-canavanine, 200 µg/mL G418, 300 µg/mL hygromycin B, uracil dropout, and 1 g/L 5-FOA plates respectively.

2.7. Genomic DNA extraction and quantification.

Yeasts were grown in 5 mL YPD for 3 days at 30 °C and 200 rpm. The cell pellet was collected by centrifugation at 14,200 rpm for 2 min using a MiniSpin Plus centrifuge (Eppendorf, USA) and the supernatant was discarded. The dry pellet was left at -20 °C for overnight. The pellet was resuspended in 500 µL buffer (sorbitol 1.2 M, EDTA 20 mM, tris 100 mM, pH 8) with 6 µL zymolyase 100T (20 mg/mL) purchased from Nacalai Tesque (Japan) and 10 µL RNase A (10 mg/mL) purchased from BioBasic (Canada). The reaction mixture was kept at 37 °C for 1.5 h and mixed by inversion every 30 min. The pellet was

then collected by spinning at 8,000 rpm for 5 min and resuspended in 500 µL of buffer (tris 50 mM, EDTA 10 mM, pH 8) with 50 µL SDS (10% v/v), and 10 µL RNase (10 mg/mL). The mixture was incubated in a 65 °C water bath for 25 min followed by addition of 200 µL of 5 M KAc and was mixed by inversion until the mixture turned opaque. The mixture was incubated in an ice-water bath for 40 min then spun at 14,200 rpm for 10 min. From the supernatant a volume of 700 µL was added to 700 µL of propan-2-ol and mixed until a precipitate formed. The precipitated DNA was isolated then transferred to 70% v/v ethanol and stored at 4 °C overnight. The DNA mixture was then spun at 14,200 rpm and the supernatant was discarded. The pellet was left in a flow hood for 30 min to evaporate any remaining ethanol then resuspended in 121 µL tris (pH 8.0). DNA concentration was quantified using the Qubit dsDNA High-Sensitivity Assay kit (Thermo Fisher Scientific, USA) for the Qubit 4 Fluorometer (Thermo Fisher Scientific, USA) following the manufacturer's protocol.

2.8. DNA sequencing, variant calling, and annotation.

Genomic DNA samples were diluted to 40 ng/µL and sent to Genome Québec (Canada) for whole-genome sequencing. Paired-end sequencing was done using the Illumina HiSeq X PE150 sequencer with shotgun (PCR free) libraries. Local alignment was done using Bowtie2 (Langmead and Salzberg, 2012) and variants were called using the SAMtools package (Li, 2011). An initial reference sequence for yBX76 was constructed by modifying the reference sequence of ySR128 (GenBank accession number

GCA_004328465.1) to contain the modifications made in yBX76. The reads from sequencing of yBX76 were aligned to the modified ySR128 reference and variants were called based on the following filter: quality score ≥30, depth ≥10, number of forward alternate reads >0, number of reverse alternate reads >0, and the fraction of alternate reads to total reads >0.9. An initial consensus reference was constructed by substituting the variants into the reference sequence (multi-allelic variants were ignored). The reads were realigned to the consensus reference sequence in an iterative process until no mono-allelic variants were called. The alignment converged on the 12th iteration in which only multi-allelic variants were called and were ignored. A final alignment was performed to verify that the consensus sequence was in agreement with the yBX76 reads and this alignment was designated as the reference sequence for yBX76. Xylose-utilizing mutants were aligned to the yBX76 reference using the same variant filtering scheme described before. The gene model used for annotation was that of the S288C reference genome, which was edited to reflect the modifications in strain yBX76. The gene annotation coordinates were lifted to the yBX76 reference using the software RATT (Otto et al., 2011). Variant annotation was done using the gene-based annotation command from the software ANNOVAR (Wang et al., 2010).

2.9. Localized hypermutation protocol.

Strains were streaked from frozen stock onto YPD with hygromycin B (300 µg/mL) and incubated at 30 °C for 2 to 3 days. For the triple reporter experiment, single colonies

were used to inoculate 5 mL of YPD and were grown for 3 days at 30 °C with shaking (200 rpm). For the xylose utilization and sugar co-utilization experiments the colonies were grown in YPR instead. An aliquot of this culture was diluted at a ratio of 1:37 into the reaction media consisting of 10 µg/mL doxycycline in YPGal (2% w/v galactose) and was grown for up to 4 days at the same conditions as before. Each day an aliquot was taken for plating onto selective media.

2.10. Triple reporter system screening.

Canavanine resistant (Can^R) mutants were screened on canavanine supplemented synthetic media (6×10^{-3}% w/v L-canavanine) and viability was assessed on SC plates. Samples from the reaction media were diluted at a ratio of 1:100 in sterile water and cell density was determined using a Neubauer hemocytometer. On canavanine media and SC media, 10^6 (10^7 on day 0) and 300 cells were plated respectively. Each sample was plated on 3 plates of both media types from which an average inactivation frequency and viability was calculated. The plates were typically incubated at 30 °C for 3 days. Double inactivation phenotypes Can^R Ura^- and Can^R Ade^- were screened by replica plating the canavanine plates onto uracil dropout and adenine dropout synthetic media respectively. Viability was calculated by dividing the plating efficiency on each day by the plating efficiency on day 0 (Equation 1). The plating efficiency is defined as the number of observed colonies on a plate divided by the number of expected colonies (Equation 2). Plates were imaged using an Epson V800 (Epson America, USA) flatbed scanner in

transmission mode. Colonies were counted using the Colony Counter plugin for ImageJ (Schneider et al., 2012). The minimum circularity was set to 0.75 and the minimum size was 10 pixels. Each genotype was tested with 3 independent transformants obtained from transformation with the A3A-carrying plasmid.

Equation 1

$$Viability\ (day\ n) = \frac{Plating\ Efficiency\ on\ day\ n}{Plating\ Efficiency\ on\ day\ 0}$$

Equation 2

$$Plating\ Efficiency = \frac{Numberof\ colonies\ observed}{Number\ of\ colonies\ expected}$$

2.11. Sporulation and mating type determination.

Haploids of Ethanol Red were generated by random spore isolation as described in Yeast Genetics (2014) with slight modification. Ethanol Red was grown in YPD overnight then diluted at a ratio of 1:100 into YPAc (2% w/v acetate) and incubated at 30 °C overnight. The cells were washed with sterile water, resuspended in 1% KAc, and incubated at 23 °C with shaking for 1 week. From the culture a 50 µL aliquot was spun at 5,000 rpm for 4 min in a MiniSpin Plus centrifuge and resuspended in 1 M sorbitol with 1 mg/mL zymolyase. The mixture was incubated at 37 °C for 4 hours then diluted at a ratio of 1:20 in 0.1% Tween 80 (BioBasic, Canada). The mixture was then subjected to 8 rounds of sonication for 30 s using an FB120 Sonicator (Fisher Scientific, USA) set to the minimum amplitude. The mixture was then spun at 10,000 rpm for 2 min, the pellet was

resuspended in 200 µL of sterile water, plated on YPD plates at a density of about 200 cells/plate, and incubated at 30 °C for 3 days. Mating type was verified by *MAT* locus PCR and by growth inhibition assay with α-factor and a-factor. For PCR verification, the *MAT* locus was amplified from gDNA using the primer set oBX58+oBX59+oBX60 (Huxley et al., 1990). For the growth inhibition assay, haploid candidate colonies were first deleted for *BAR1* with a kanMX marker using primer pair oKC192+oKC193 and verified with primer pair oKC200+oKC201. *BAR1* deletions were plated on YPD and grown into a lawn. The plate was replica plated onto two more YPD plates to reduce the cell density. On the replica plates 20 µL of both yeast pheromones were loaded onto 1 cm diameter Whatman Grade 1 filter papers (Fisher Scientific, USA). The concentrations used were 5 µM of α-factor and 10 µg/mL to 1 mg/mL of a-factor (Zymo Research, USA). Plates were incubated at 30 °C for up to 3 days to identify the presence of a zone of inhibition.

2.12. Optical density and biomass calibration curve.

Cells were grown in 5 mL YPD media overnight at 30 °C with shaking (200 rpm) then diluted at a ratio of 1:100 into 100 mL raffinose mineral media and grown for another 3 days. Various concentrations of the cultures were made ranging from 20% to 200% v/v using sterile media. A 1 mL aliquot was taken for absorbance measurement (OD) at 600 nm using a Synergy H1 plate reader (BioTek, USA). The plate was shaken in double orbital mode for 1 minute at 1 mm amplitude prior to reading. OD values were read in

replicates of five per well. The remaining culture volumes were spun at 3,000 rpm for 2 min in a Sorvall ST 16 centrifuge. The supernatant was discarded and the pellets were resuspended in 200 µL of water. The mixture was transferred onto Whatman grade 42 filter papers (Fisher Scientific, USA) and placed in a drying oven set to 105 °C for 20 min. The weight of the dried biomass was determined by subtracting the weight of dried filter paper from the total weight measured on an electronic balance (VWR, USA). A calibration curve relating biomass concentration and optical density was constructed by grouping the data from 2 fast-growing and 2 slow-growing mutants isolated for growth on xylose (Supplementary Figure 1). This sample had mutants that grew in liquid media as aggregates as well as mutants that grew as single cells. This was done to include the effect of cell morphology on optical density. This plot also included the biomass data for the parental strain yBX27 grown on YPR and YPD. The data was fit to a second-order polynomial using the linear model function in R.

2.13. Xylose utilization screen and growth rate determination.

Xylose-utilizing yeast were screened on 20 g/L xylose mineral media plates. The plates were incubated at 30 °C for approximately 30 days. Colonies that grew above the background were restreaked onto fresh xylose plates. One representative colony from the restreaked plate was tested for the growth rate in liquid xylose media. Samples were first grown on YPR for 3 days at 30 °C with shaking at 200 rpm. An aliquot was spun at 5,000 rpm for 4 min and the pellet was resuspended in sterile water and then diluted at a

ratio of 1:50 in xylose media. The culture was grown for 4 days at 30 °C with shaking at 200 rpm. Each day an aliquot was taken for OD_{600} readings in triplicate using a Synergy H1 plate reader (BioTek, USA).

For experiments involving growth under 1% and 0% oxygen, yeasts were grown in tubes sealed with rubber stoppers. A gas line was connected to the culture by a hypodermic needle inserted through the rubber stopper. A second hypodermic needle was inserted into the stopper as a gas outlet. The gas flow rate was maintained at 20 mL/min and monitored by a rotameter. The gases used for the hypoxic and anoxic conditions were composed of 1% O_2 (balance N_2) and 100% N_2 respectively and were purchased from Praxair (Canada). The specific growth rate was calculated from the secant of the biomass-time data for consecutive days as shown in Equation 3.

Equation 3

$$\mu_{net,n} = \frac{\ln(X)_{n+1} - \ln(X)_{n-1}}{t_{n+1} - t_{n-1}}$$

$$X = biomass\ concentration\ \left[\frac{g}{L}\right],\ t = time\ [h]$$

$$\mu_{net,n} = net\ specific\ growth\ rate\ at\ time\ point\ n\ \left[\frac{1}{h}\right]$$

2.14. Quantification of xylose utilization.

Determination of xylose concentration in media was based on the dinitrosalicylic acid (DNS) assay (Miller, 1959). Stocks of the DNS reagent (1% w/v dinitrosalicylic acid, 0.05% w/v sodium sulfite, 1% w/v sodium hydroxide) and the Rochelle salt reagent (40% w/v

sodium tartrate) were prepared and stored away from sunlight. The media was assayed by mixing 300 µL of the media with 300 µL of DNS reagent and incubating in a water bath at 90 °C for 3 min. Afterwards, 100 µL of Rochelle salt reagent was added to the mixture and incubated in a room temperature water bath for 5 min. The mixture was equilibrated in room air for 5 min after which the absorbance at 575 nm was read using a Synergy H1 plate reader (BioTek, USA). For yeast cultures, a 1 mL aliquot was spun at 14,500 rpm for 1 min in a MiniSpin Plus centrifuge and the supernatant was used for measurement. A calibration curve was made using standard solutions of xylose in mineral media (Supplementary Figure 2). The data was fit to a line using the linear model function in R. This calibration curve was used to infer the change in the xylose levels during growth of different yeast samples.

2.15. Mixed-sugar co-utilization screen and growth rate determination.

Glucose and galactose co-utilizing yeasts were screened on Gal+2-DG mineral media (20 g/L galactose, 0.3 g/L 2-deoxyglucose) plates. The plates were incubated at 30 °C for up to a week or until colonies were visible. Colonies were restreaked onto raffinose mineral media plates. Two representative colonies from each mutant candidate were patched out and grown on raffinose plates. The patches were replica plated onto galactose and Gal+2-DG plates to test the inheritance of 2-DG resistance. Patches that retained 2-DG resistance were tested for their growth rates on glucose liquid media and glucose and galactose media. For growth rate determination, each sample was grown in YPR at 30 °C

with shaking (200 rpm) for 3 days. A 1 mL aliquot was spun at 5,000 rpm for 4 min and the pellet was resuspended in sterile water and then diluted at a ratio of 1:50 in glucose mineral media (20 g/L glucose) and G+Gal mineral media (5 g/L glucose, 20 g/L galactose). The samples were loaded onto a clear flat-bottom 96-well plate and the wells were sealed with a transparent film. Two technical replicates were loaded per sample. The plate was incubated in a Synergy H1 plate reader (BioTek, USA) at 30 °C with continuous shaking (double-orbital mode with 1 mm amplitude) and absorbance readings were measured every 20 min for 14 h and 30 min.

2.16. I-SceI cutting assay.

Strains carrying the inducible DSB were grown in YPR at 30 °C with shaking (200 rpm) for 3 days. The cultures were diluted at a ratio of 1:37 into YPGal media for each time point (1, 3 and 5 h) and incubated at the same conditions. The pellets were collected at each time point (including a 0 h time point from the pregrowth) by spinning the cultures at 3,000 rpm for 2 min in a Sorvall ST 16 centrifuge. The pellets were resuspended in 750 µL of sterile water and spun again at 14,5000 rpm for 2 min in a MiniSpin Plus centrifuge. The supernatant was discarded and the pellets were stored at -20 °C overnight followed by DNA extraction. DNA concentration was measured using a Nanodrop spectrophotometer (Thermofisher, USA). The cut site and *HO* locus were amplified by PCR with the Q5 polymerase (NEB, USA) using primer pairs oBX74+oBX98 and oBX6+oBX7 respectively. Each PCR was loaded with 20 ng of template DNA. The PCRs

were then run out on a 0.8% agarose TAE gel. Relative abundance of PCR amplicons were inferred from band intensity and was measured using the Gel Analyzer plugin for ImageJ (Schneider et al., 2012).

2.17. Software and statistical tests.

Tukey's HSD test was used for comparison of more than 2 means. Welch's t-test was used for comparison of 2 means. Unless stated otherwise, values from a single biological replicate are reported as the mean of technical replicates and values from samples with more than one biological replicate are reported as the mean ± the 95% confidence interval (CI). The confidence interval was based on a one sample t-test. Statistics were done using the built-in R packages. Plots were generated with the ggplot2 package in R (Wickham, 2009). Symbols: NS, not significant; *, $p<0.05$; **, $p<0.005$;***, $p<0.001$.

3. RESULTS

3.1. Multiple gene inactivation using a triple-reporter system.

The capacity of APOBEC3A (A3A) to mutate 3 genes of interest simultaneously was tested by screening for inactivation in the genes *CAN1*, *ADE2*, and *URA3* that were located within 20 kb of a double-strand break (DSB) site as shown in Figure 4. A galactose inducible endonuclease I-SceI, and the I-SceI cut site (I-SceIcs) were integrated at the *LYS2* locus of strain ySR128. This background was denoted as the DSB genotype, which contains the three selectable markers within proximity to an inducible DSB site (I-SceIcs). A derivative of this strain denoted as the DSB A3A genotype carried A3A on a CEN plasmid, which was stably inherited for at least 2 days (Supplementary Figure 3).

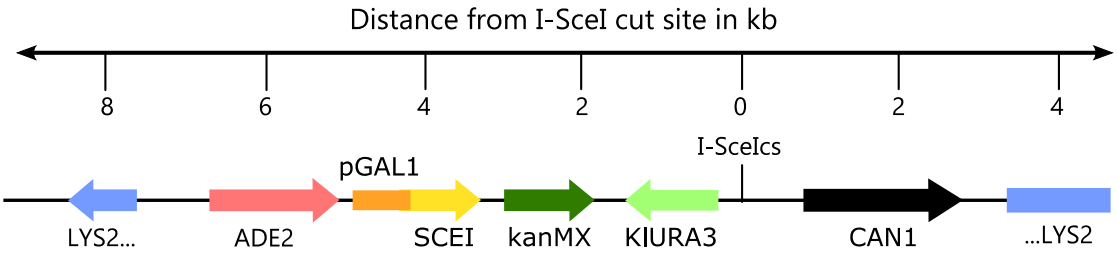


Figure 4. Position of the reporter genes *CAN1*, *ADE2*, and Kl*URA3* relative to the DSB site. Strain ySR128 was transformed with a PCR amplicon containing *SCEI* under a *GAL1* promoter, I-SceI cut site, kanMX, and Kl*URA3*. The amplicon was targeted to the *LYS2* locus in chromosome 2, which was previously modified with insertion of *ADE2* and *CAN1.* Symbols: I-SCEIcs, I-SceI recognition site (DSB site).

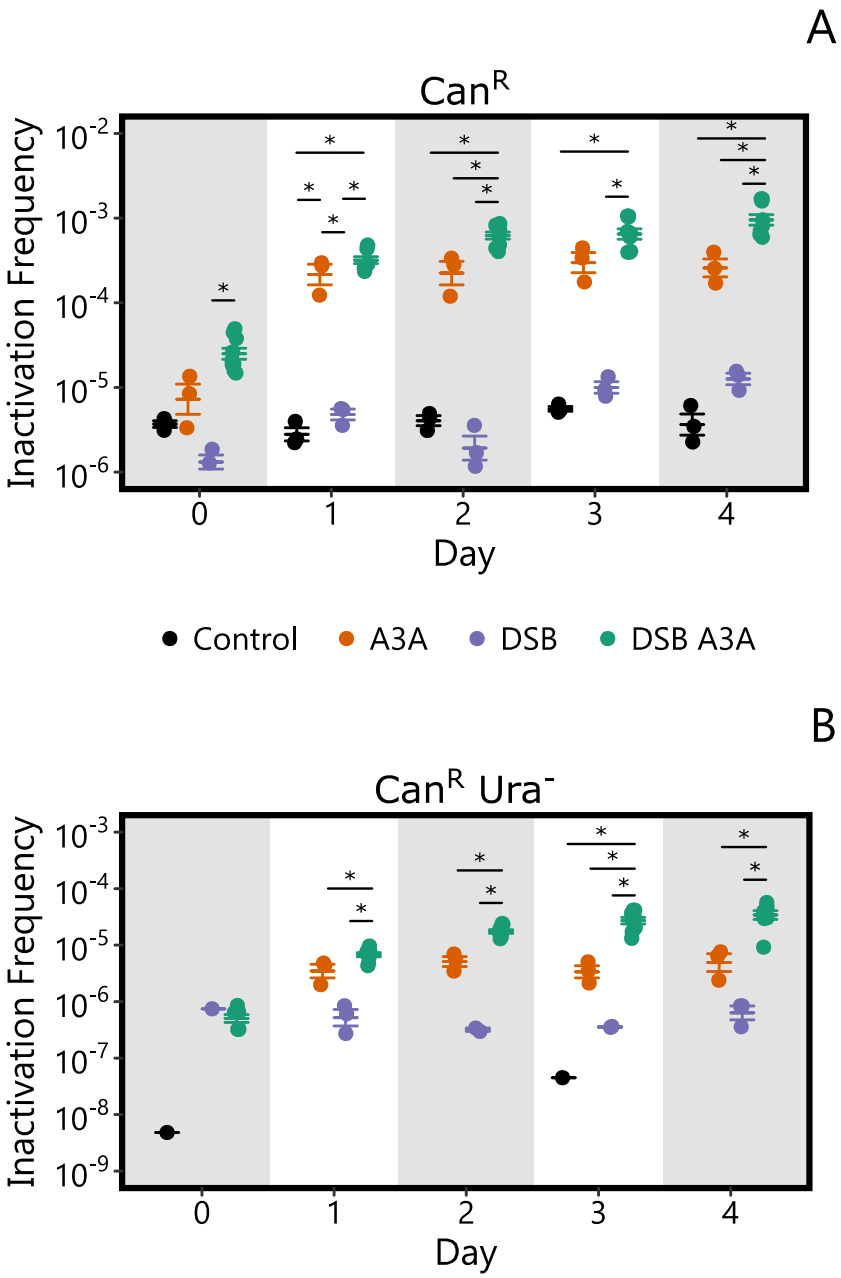
3.1.1. Single and double gene inactivation frequency by A3A.

Cells were mutagenized by incubating in the reaction media consisting of 2% w/v galactose (for induction of I-SceI) and 10 μg/mL of doxycycline (for induction of A3A). The average inactivation frequency of *CAN1* over 4 days of incubation was significantly higher in backgrounds carrying A3A than in the controls lacking A3A ($p<0.05$, Figure 5A). The observed proportion of Can^R colonies in A3A backgrounds increased the most after 1 day of incubation and began to stabilize onwards. The highest *CAN1* inactivation frequency was $1.0\pm0.36\times10^{-3}$ (±95% CI) on day 4 of incubation in the DSB A3A background, which was 2 orders of magnitude greater than the parental control ($p<0.001$). In the parental control, *CAN1* inactivation remained steady over 4 days, and in the DSB background there was a slight increase on day 3 and 4.

A similar pattern was observed for the double gene inactivation frequency of *CAN1 URA3* and *CAN1 ADE2* (Figure 5BC). Again, the average frequency experienced the highest increase after 1 day of mutagenesis and began to plateau on day 4 towards a maximum of $3.8\pm1.1\times10^{-5}$ and $2.9\pm1.6\times10^{-5}$ for *CAN1 URA3* and *CAN1 ADE2* respectively. The presence of Can^R Ura^- and Can^R Ade^- mutants were often undetectable in the control when plating 10^6 cells per plate. Thus for comparison purposes it was assumed that the double gene inactivation frequency in the parental background was $<1\times10^{-6}$. With this assumption, DSB A3A exhibited a double gene inactivation frequency of at least an order of magnitude greater than the control. Triple Can^R Ura^- mutants were

extremely rare even in A3A backgrounds with at most 1 colony per day in the pooled replicates.

There was no difference in survival between backgrounds with A3A and controls after 1 day of mutagenesis (Figure 5D). At 2 days of incubation the mean viability in the DSB A3A background dropped to 39±24% relative to the pre-induction survival rate on day 0. The presence of A3A was associated with reduced survival compared to the parental control ($p<0.05$). In A3A backgrounds without a DSB the mean viability was 58±52% relative to pre-induction, and there was no difference in viability between A3A-only backgrounds and no-A3A controls at the 5% significance level. The lack of significance may have been a product of the higher variability in A3A-only backgrounds relative to DSB A3A as evidenced by the much wider confidence intervals and the smaller sample size. By day 4, a drop in viability was observed in A3A backgrounds as well as DSB backgrounds relative to the parent control ($p<0.05$). The mean viability after 4 days of induction reached 9±4% in DSB A3A and was lower than the DSB-only background, which had a survival of 59±17% ($p<0.05$). No difference in survival could be attributed to A3A versus DSB A3A ($p>0.05$).


A
Can^R
Inactivation Frequency
10^-2
10^-3
10^-4
10^-5
10^-6
0
1
2
3
4
Day
Control
A3A
DSB
DSB A3A
B
Can^R Ura^-
Inactivation Frequency
10^-3
10^-4
10^-5
10^-6
10^-7
10^-8
10^-9
0
1
2
3
4
Day

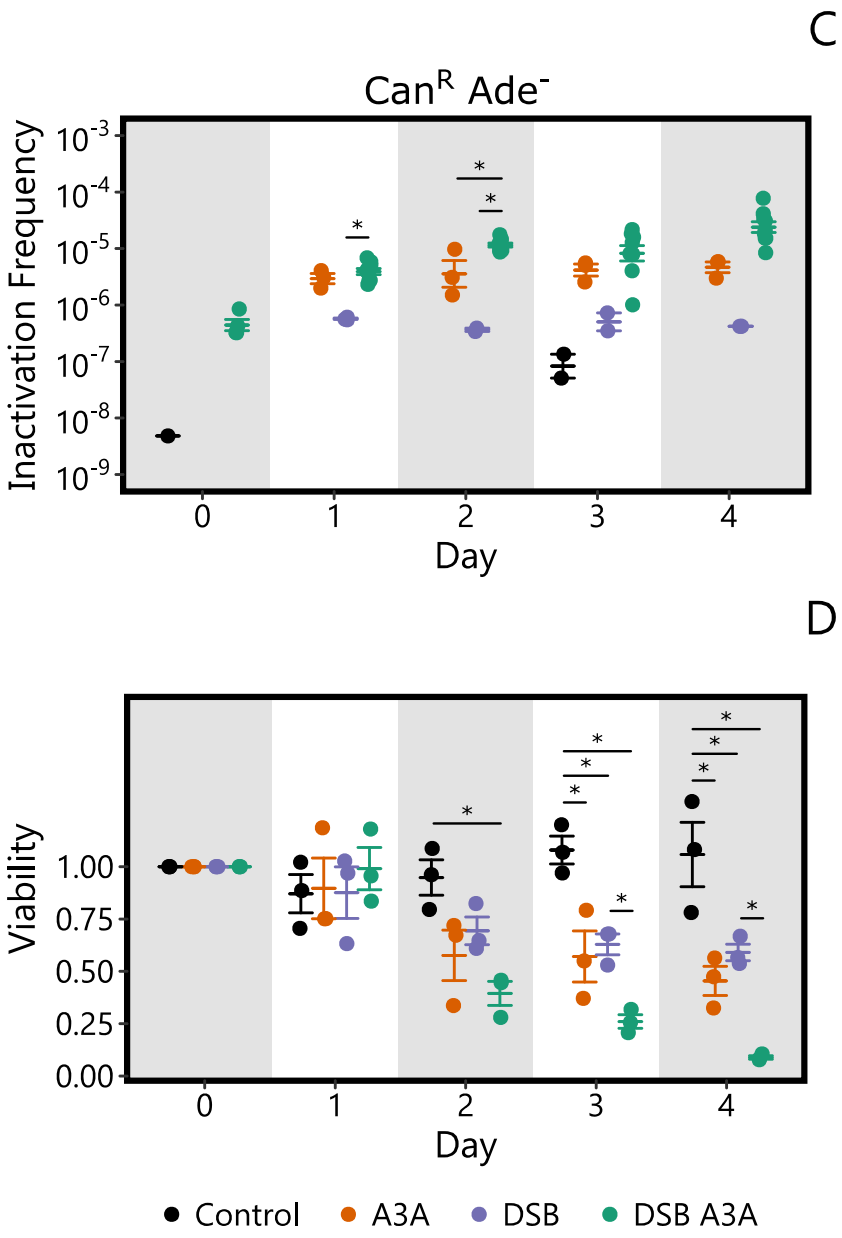

C
Can^R Ade^-
Inactivation Frequency
10^-3
10^-4
10^-5
10^-6
10^-7
10^-8
10^-9
0
1
2
3
4
Day
D
Viability
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
Day
Control
A3A
DSB
DSB A3A

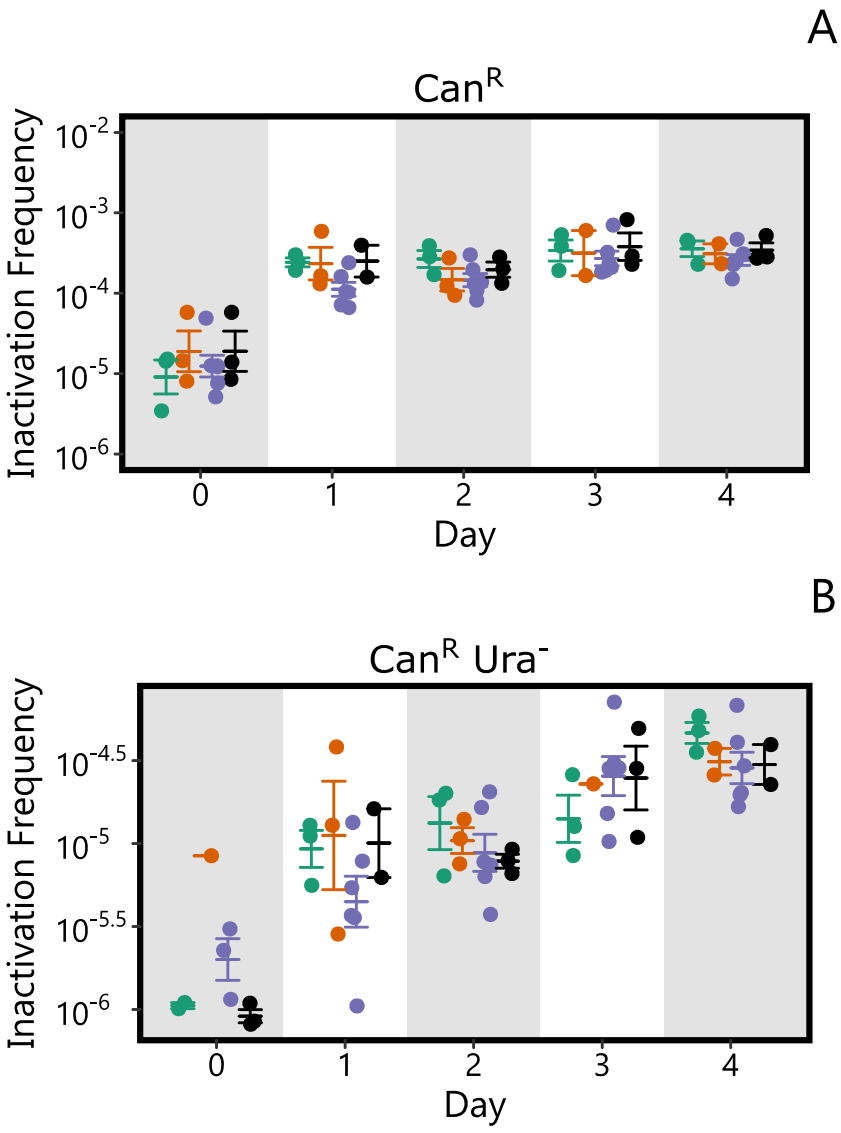
Figure 5. Inactivation frequency of *CAN1*, *CAN1 ADE2,* and *CAN1 URA3* was higher in backgrounds carrying A3A. Single colonies of ySR128, ySR128D, ySR128A, and ySR128DA were pregrown in YPR media and transferred to 20 g/L galactose YP media with 300 µg/mL hygromycin B, 10 µg/mL doxycycline, and cultured for 4 days. Each day cells were plated on diagnostic media and incubated for 2 to 3 days. **(A)** Each day 10^6 cells were plated on 6×10^{-3}% w/v canavanine synthetic media to assess Can^R frequency. **(B)** Can^R colonies on the canavanine plates were replica plated onto uracil dropout media to assess the frequency of Can^R Ura^-. **(C)** Can^R colonies on the canavanine plates were also replica plated onto adenine dropout media to assess the frequency of Can^R Ade^-. **(D)** Each day 300 cells were plated on SC media to assess viability. Survival was calculated as the plating efficiency on synthetic media normalized to day 0 (pregrowth). Each point is the average of 3 technical plating replicates. Three independent colony isolates of each genotype were used for n = 3 biological replicates, except for DSB A3A, which had n=9 biological replicates for **A-C**. Horizontal bars represent the log mean. Error bars represent the SEM. Symbols: Control, ySR128; A3A, ySR128A; DSB, ySR128D; DSB A3A, ySR128DA.

3.1.2. Comparison of mutation rate under different doxycycline treatments.

The effect of A3A induction on mutation frequency was investigated. The concentration of doxycycline used in the previous experiment was 10 µg/mL and is based on previous work using the same expression system in yeast (Chan et al., 2012). Doxycycline has been reported to have first-order degradation kinetics in liquid media with a half-life of 91 h (Mazur et al., 1999). When adding an additional 2.3 µg/mL doxycycline (Doxy Boost) each day to counteract degradation, no general difference was observed in gene inactivation or viability (Figure 6). No dose-dependent effect on inactivation frequency was observed when the initial doxycycline concentration was doubled to 20 µg/mL (2X Doxy). One difference was that cells treated with double the initial doxycycline concentration exhibited lower survival on day 3 and 4 of induction compared to the normal dosage ($p<0.05$).

The third treatment for increasing the mutation rate was to increase the proportion of cells carrying the A3A plasmid by pregrowing cells with the antibiotic hygromycin B. Although the plasmid was stably inherited (Supplementary Figure 3), there was notably reduced growth in the reaction media based on cell counts. Therefore, with no growth it was expected that the number of cells carrying the plasmid will decrease as cells die over time. By pregrowing with hygromycin B selection, it was hypothesized that a larger proportion of cells carrying the A3A plasmid will enter the reaction media and thus increase the proportion of mutant colonies. There was no difference observed in single or double gene inactivation frequency, but there was lower survival on day 3 and 4

(p<0.05). Since none of treatments improved mutation frequency and at worst reduced survival, the standard treatment was used for the rest of this study.


A
Can^R
Inactivation Frequency
10^-2
10^-3
10^-4
10^-5
10^-6
0
1
2
3
4
Day
B
Can^R Ura^-
Inactivation Frequency
10^-4.5
10^-5
10^-5.5
10^-6
0
1
2
3
4
Day

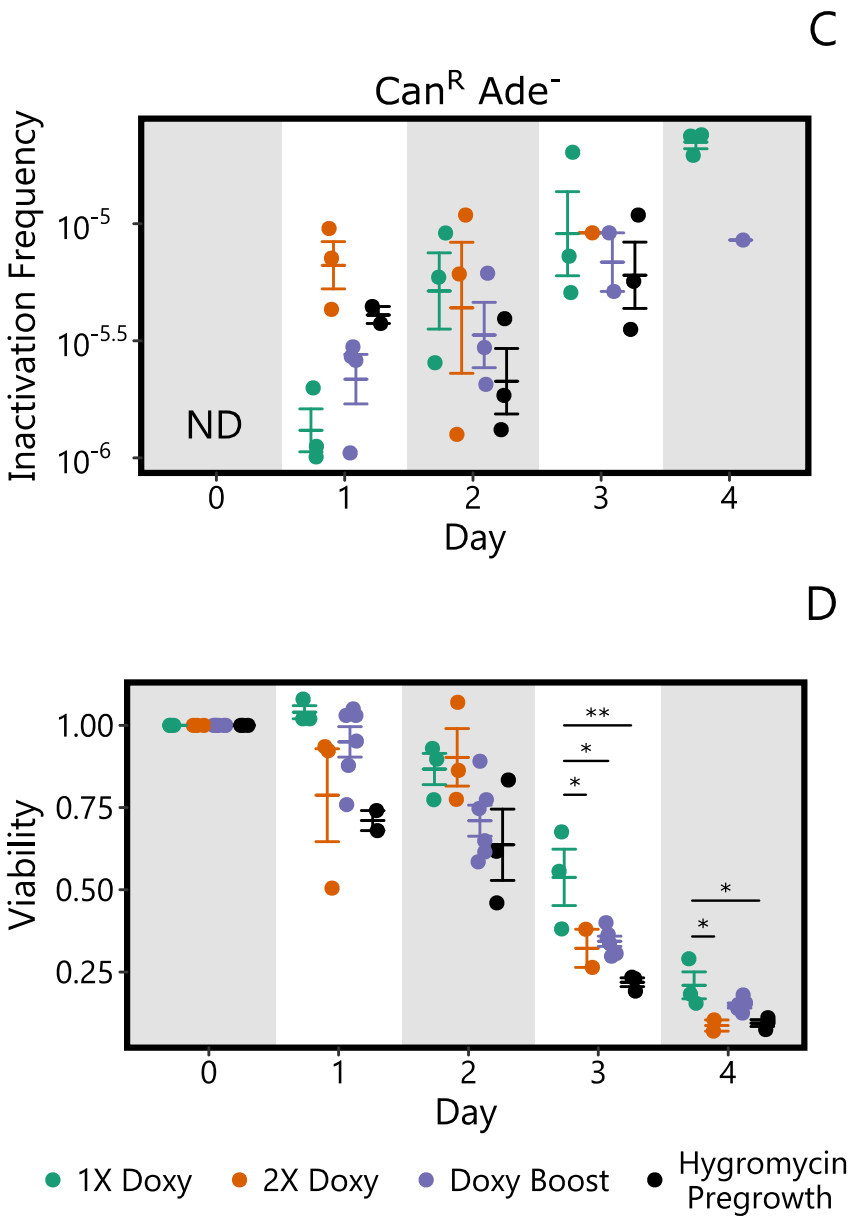

C
Can^R Ade^-
Inactivation Frequency
10^-5
10^-5.5
10^-6
ND
0
1
2
3
4
Day
D
Viability
1.00
0.75
0.50
0.25
**
*
*
*
*
0
1
2
3
4
Day
1X Doxy
2X Doxy
Doxy Boost
Hygromycin Pregrowth

Figure 6. Increasing the doxycycline concentration does not increase inactivation frequency. Single colonies of ySR128DA were pregrown in YPR media and transferred to 20 g/L galactose YP media with 300 µg/mL hygromycin B, various doxycycline amounts, and cultured for 4 days. Each day cells were plated on diagnostic media and incubated for 2 to 3 days. **(A)** Each day 10^6 cells were plated on 6×10^{-3}% w/v canavanine synthetic media to assess Can^R frequency. **(B)** Can^R colonies on canavanine plates were replica plated onto uracil dropout media to assess the frequency of Can^R Ura^-. **(C)** Can^R colonies on canavanine plates were replica plated onto adenine dropout media to assess the frequency of Can^R $Ade2^-$. **(D)** Each day 300 cells were plated on SC media to assess viability. Survival was calculated as the plating efficiency on synthetic media normalized to day 0 (pregrowth). Each point is the average of 3 technical plating replicates. Three independent colony isolates of each genotype were used for n = 3 biological replicates. Horizontal bars represent the log mean. Error bars represent the SEM. Symbols: 1X Doxy, 10 µg/mL doxycycline; 2X Doxy, 20 µg/mL doxycycline; Doxy Boost, 10 µg/mL doxycycline with additional 2.3 µg/mL each day; Hygromycin B Pregrowth, 300 µg/mL hygromycin B in pregrowth media.

3.1.3. Effect of rescue with DSB repair oligonucleotides on survival.

As shown in Figure 5D, the majority of cells do not survive 2 days of incubation in the reaction media. A major contributor to cell death is likely the persistence of the DSB that signals for temporary cell arrest, but if not repaired quickly cell growth can resume with damaged DNA (Sandell and Zakian, 1993). Cells with persistent DSBs can accumulate many kinds of mutations at the DSB site, therefore rescuing these cells before they die is advantageous for a greater mutant sampling pool. DSB repair was facilitated by transformation with oligonucleotides using a slightly modified approach described by Yang et al. (2008). The oligonucleotides bridge the DSB site with 50 bp on both sides and serve as a template to synthesize 50 bp overhangs to help re-anneal the ssDNA (Figure 7).

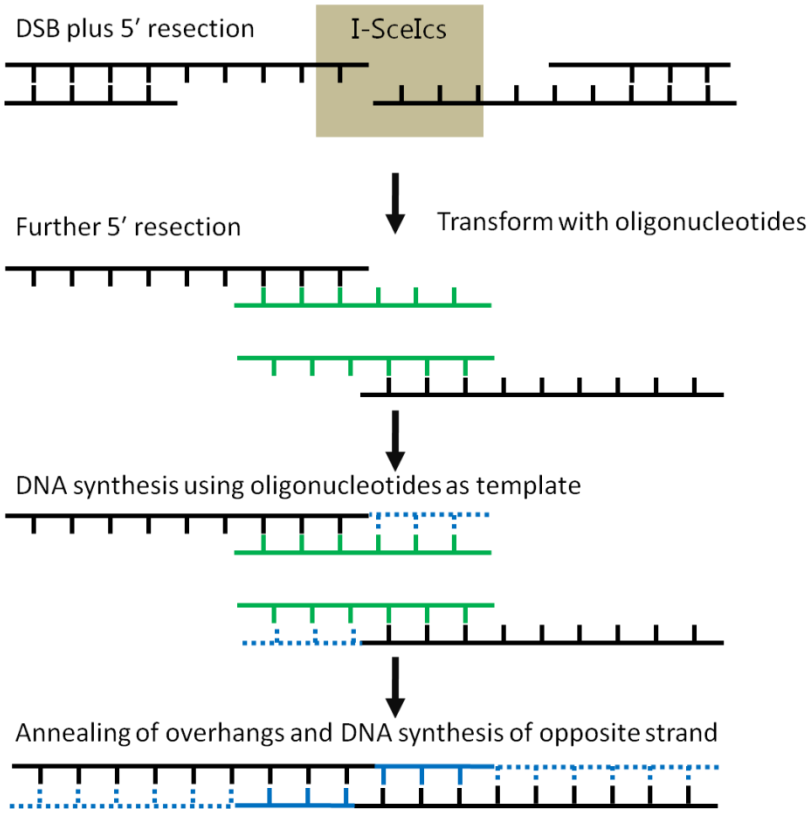


Figure 7. Mechanism for increasing LH survival through transformation with DSB repair oligonucleotides. In step 1 DNA cleavage by I-SceI generates a DSB at the I-SceI cut site (I-SceIcs). The newly generated ends become resected and signal for DNA damage induced arrest. If not repaired promptly, the damage can be lethal due to missegregation of broken chromosomes. By transforming cells with 100 bp oligonucleotides that span the DSB site, the cells are provided with templates for DNA synthesis past the DSB site (step 2). Each ssDNA end will generate 50 nt of sequence complementary to the opposite strand (step 3), which can facilitate reannealing and repair of the DSB (step 4).

In this experiment strain yBX76DA was used. This background was constructed from the biofuel strain Ethanol Red and contains the inducible DSB system and A3A as described before. After 2 days of A3A induction cultures were transformed with repair oligonucleotides and plated on SC plates to access survival. As shown in Figure 8, the mean plating efficiency for treatment with oligonucleotide was 34.9±12.8% (±95% CI) and was lower than the no transformation control plating efficiency, which was 106±43% ($p<0.001$). There was no difference in plating efficiency between transforming with repair oligonucleotides and water ($p>0.05$). Since the plating efficiency was defined as the number of observed colonies divided by the number expected colonies (Equation 2), an important consideration is accurate cell counting. Ethanol Red cells grow in clumps of highly variable morphology and also form branched-chain structures. It was observed that transformation had reduced cell aggregation and this was reflected in the reduced variability of plating efficiency for transformed cells.

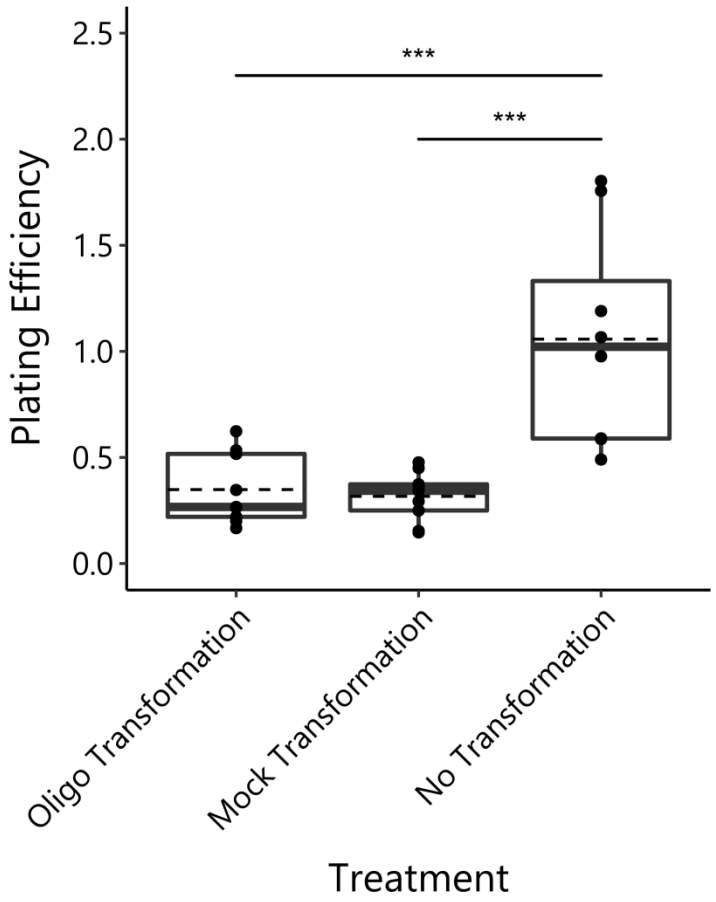


Figure 8. Transformation with DSB repair oligonucleotides reduced overall cell survival. Ethanol Red cells were incubated in 20 g/L galactose YP media with 10 µg/mL doxycycline and 300 µg/mL hygromycin B for 2 days then collected by centrifugation and transformed with a pair of complementary DSB repair oligonucleotides. Cells were counted and plated on SC media at 300 cells/plate to assess survival. Plating efficiency was calculated as the number of colonies formed divided by the number of colonies expected (300 colonies). Water was used in place of oligonucleotides for the mock transformation control. Each treatment shows the combined data for n=3 biological replicates from independent colony isolates tested at n=3 technical plate replicates. Dotted horizontal lines represent the mean and solid horizontal lines represent the median. The individual data points are shown in addition to a box plot.

3.2. Generating xylose-utilizing *S. cerevisiae* mutants with LH.

S. cerevisiae does not naturally grow on xylose without genetic manipulation or extensive adaptation (Attfield and Bell, 2006). Since the yeast can grow on xylulose, the minimal modification necessary is to provide a mechanism for converting xylose to xylulose. To achieve this, haploid cells produced from the diploid biofuel strain Ethanol Red were transformed with *XYL1* and *XYL2* from *S. stipitis* to establish a pathway for the conversion of xylose to ethanol (Figure 1). This strain was chosen over lab strains for its high ethanol productivity and inhibitor tolerance thus making it easily transferable for industrial use (Mukherjee et al., 2017; Snoek et al., 2015). An Ethanol Red haploid designated as yBX27 was used for all constructions in this study. The open reading frame for *XYL1* and *XYL2* along with their entire upstream and downstream intergenic regions were amplified from the *S. stipitis* Y-11545 strain. The genes were integrated sequentially in chromosome 16 between the retrotransposons YPRWTy1-3 and YPRCTy1-4 in the configuration as shown in Figure 9. P_{GAL1}*SCEI* was integrated at the *URA3* locus. The construction was confirmed by PCR and whole-genome sequencing. This strain containing P_{GAL1}*SCEI*, *XYL1*, *XYL2*, and the DSB site was named yBX76 and also referred to as the *XYL1 XYL2* DSB background.

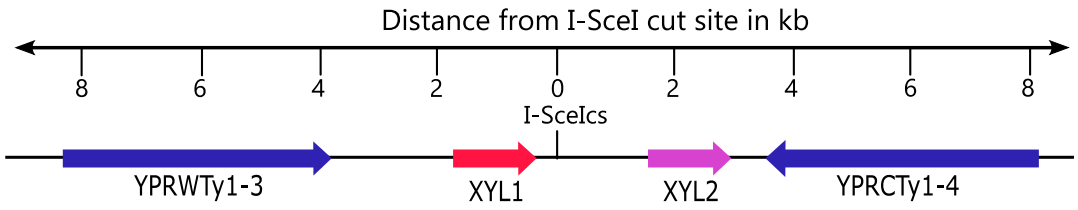


Figure 9. Relative position of *XYL1* and *XYL2* to the DSB site in chromosome 16 for localized hypermutation. The genes *XYL1* and *XYL2* encoding xylose reductase (XR) and xylitol dehydrogenase (XDH) respectively from *S. stipitis* were integrated between two LTR retrotransposons at position Chr. XVI 850475 in yBX27. The genes were amplified along with the entire upstream and downstream intergenic sequences and flank an I-SceI cut site (I-SceIcs).

3.2.1. Screening and growth rate determination of xylose-utilizing mutants.

The *XYL1 XYL2* DSB construct was transformed with the A3A-carrying plasmid and subjected to LH for up to 2 days. Cells were plated and screened for colony growth on xylose plates (20 g/L xylose, mineral media). When plating at low cell densities (300 cells per plate), all colonies from the reaction media and the pregrowth culture grew at a similar rate and positive candidates were not identifiable. To enhance selection and reduce background growth, the cell density was increased to about 100×10^6 cells per plate. The reasoning was that a high cell density would deplete the xylose faster and allow fast consumers of xylose to outgrow competitors. After about 30 days of incubation colonies were visible above the background (Figure 10).

Figure 10. Screening for xylose-utilizing mutants requires a 30 day adaption period on xylose media. Representative plates for the adaptation periods required before mutant colonies grew above the background on xylose media. A colony isolate of yBX76DA was subjected to LH for 2 days then a 2.5 mL aliquot of the reaction media (about 100×10^6 cells) was plated on 20 g/L xylose mineral media and incubated at 30°C for an average of 30 days. Colonies growing above the background were visible after about 4 weeks of incubation.

The growth rates of 153 A3A-carrying mutants were tested in liquid xylose media with the same composition as the plates (Figure 11A). There was no growth on xylose for neither the diploid Ethanol Red nor the haploid, and very little growth for *XYL1 XYL2* cells. Adaptation of the *XYL1 XYL2* background on xylose media for approximately 30 days (Figure 11A, yBX76 Adapted), yielded mutants with considerable growth on xylose—as high as 10-fold biomass increase by day 4. The best-performing mutants came from the A3A-carrying strains. In these mutants, the highest biomass increase reached 30-fold by day 4. The top performing mutant was named yBX186.

Splitting up the growth data of the A3A-carrying mutants from Figure 11A by the induction period in the reaction media showed a similar spectrum of growth profiles between A3A-carrying mutants that were subjected to LH (1 and 2 days induction) compared to mutants plated directly from pregrowth (0 days induction, Figure 11B). In all three cases mutants were isolated that either failed to grow reproducibly on xylose (like Ethanol Red), grew similar to mutants adapted without A3A, or grew exceptionally well relative to controls.

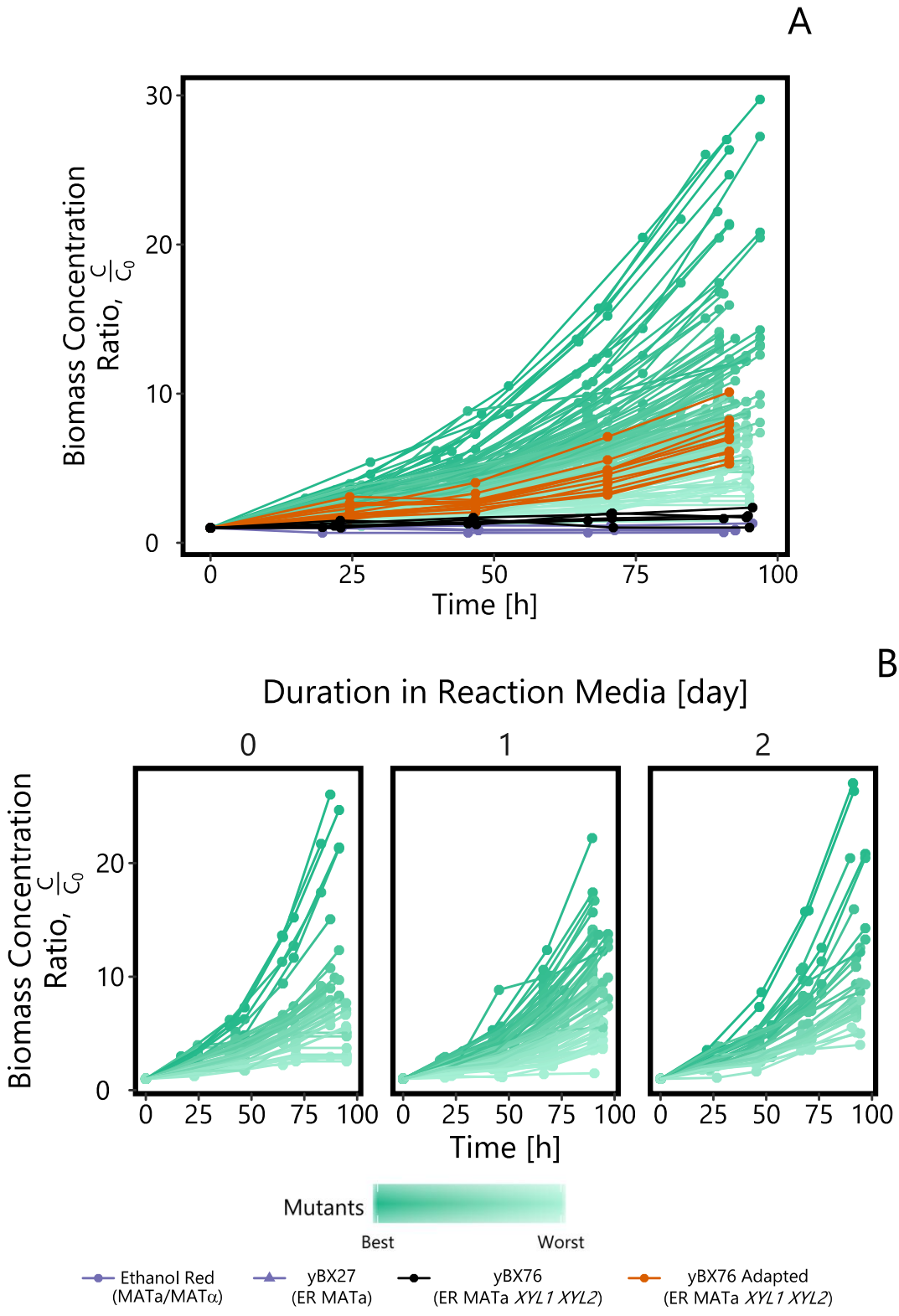

A
Biomass Concentration Ratio, $\frac{C}{C_0}$
30
20
10
0
0
25
50
75
100
Time [h]
B
Duration in Reaction Media [day]
0
1
2
Biomass Concentration Ratio, $\frac{C}{C_0}$
20
10
0
0 25 50 75 100
Time [h]
Mutants
Best
Worst
Ethanol Red (MATa/MATα)
yBX27 (ER MATa)
yBX76 (ER MATa XYL1 XYL2)
yBX76 Adapted (ER MATa XYL1 XYL2)

Figure 11. Localized hypermutation with APOBEC3A generated *S. cerevisiae* mutants with improved xylose utilization. (A) Growth curves of mutants (n = 153) on 20 g/L xylose mineral liquid. The biomass concentration was determined from OD measurement at 600 nm and converted to dry cell weight using a calibration curve. The ordinate represents the fold-change in biomass relative to the 0 h time point. The sample sizes for the controls are as follows: 4 (Ethanol Red), 8 (yBX27), 5 (yBX76), and 11 (yBX76 Adapted). All curves represent n=1 biological replicates. **(B)** Growth curves of the A3A-carrying mutants (n=153) from **A** were grouped by the duration of time spent in the reaction media. Mutants from 0 days of incubation in the reaction media were taken from the pregrowth culture and plated directly on xylose media. Each point represents the average of n=5 technical OD replicates. Points were connected by straight lines to aid in visualization.

3.2.2. Growth on xylose under different oxygen conditions.

S. cerevisiae can grow, and ferment glucose anaerobically, but can only ferment xylose anaerobically without growth. Under low oxygen conditions, *S. cerevisiae* grows on xylose slower than in room air. Moreover, the cofactor imbalance between Xyl1 and Xyl2 has been reported to contribute to the oxygen requirement for growth on xylose (Kuyper et al., 2004). Testing the growth rates under low oxygen conditions thus can be used to identify mutants with more efficient xylose utilization. Since *XYL1* and *XYL2* were targeted for mutation, it was of interest to test if any mutants had an enhanced ability to grow under different oxygen conditions. The growth rates of the top 10 mutants were evaluated at two other oxygen conditions. Mutants were grown as before, but this time under either nitrogen gas (0% oxygen) or 1% oxygen gas (Figure 12). None of the mutants or controls were able to grow under nitrogen gas. Ethanol Red (yBX27) and *XYL1 XYL2* (yBX76) did not grow under any of the oxygen conditions. Transferring the cultures that were incubated in nitrogen gas to room air allowed growth to resume in all samples except for in the controls (Supplementary Figure 4). When grown under 1% oxygen, an oxygen dependence on growth was observed for 8 out of the 10 mutants tested. These mutants attained a higher biomass concentration in room air (21% oxygen) than in 1% oxygen. For 2 mutants (yBX191 and yBX192) the growth curves for 21% and 1% oxygen were nearly identical, but they were unable to grow under a complete lack of oxygen.

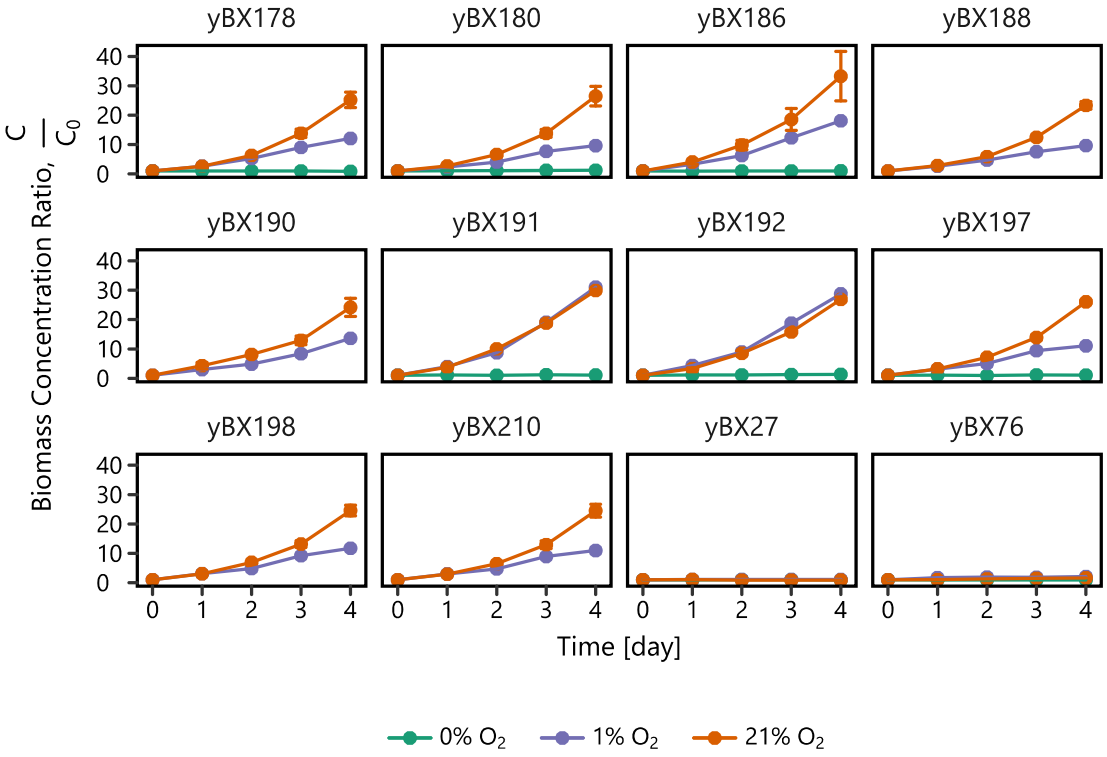


Figure 12. Growth on xylose is oxygen dependent. The top 10 performing mutants were tested for their growth on 20 g/L xylose mineral media under room air (21% oxygen), 1% oxygen (balance nitrogen), and 0% oxygen (nitrogen gas). For the 1% and 0% oxygen conditions, cultures were grown in sealed tubes connected to an inlet and outlet gas line flowing at 20 mL/min. The biomass concentration was determined from OD measurement at 600 nm and converted to dry cell weight using a calibration curve. The ordinate represents the fold-change in biomass relative to the 0 h time point. Mutants yBX188, yBX190, yBX198, and yBX210 were not tested under 0% oxygen. Points were connected by straight lines to aid in visualization. The 21% oxygen curves have n=3 independent colony biological replicates, the rest have n=1 biological replicates. Error bars represent the SEM.

Table 3 shows the highest growth rates for the top three performing mutants yBX186, yBX191, and yBX192 in room air and 1% oxygen. The growth rates were calculated between consecutive days according to Equation 3 (Materials and Methods) and were generally the highest in the initial growth phase at 1 day of incubation. The values listed in Table 3 are the maximum growth rates. The highest growth rate in room air was $4.74 \pm 1.73 \times 10^{-2}$ h^{-1} (±95% CI) by mutant yBX186, which grew 91% as fast in 1% oxygen. Mutants yBX191 and yBX192 attained similar maximum growth rates of $4.62 \pm 0.48 \times 10^{-2}$ h^{-1} and $4.45 \pm 0.90 \times 10^{-2}$ h^{-1} respectively in room air and their growth rates under 1% oxygen were 10% to 12% higher. Because the experiments under 1% oxygen and nitrogen gas were run for n=1 biological replicates, no statistical inference could be made on the growth rates in these two conditions. Under room air the growth rate of yBX186 was approximately 1.8-fold ($p<0.05$) and 5-fold ($p<0.01$) higher than that of adapted and unadapted *XYL1 XYL2* controls respectively.

Table 3. Growth rates of the top 3 mutants on xylose media. Local growth rates were calculated between successive days using Equation 3. The largest of these rates is reported as μ_{max} , which was typically on day 1. A3A status refers to whether the strain was transformed with plasmid pSR435 harbouring A3A (+) or not (-). Adapted on xylose refers to whether the strains were isolated from xylose plates incubated for about 30 days (+) or were never exposed to xylose prior to testing (-). Average growth rates are shown for n=3 replicates in room air and n=1 replicates in 1% oxygen. The errors represent the SEM.

Strain	Genotype	A3A Status	Adapted on Xylose	μ_{max} [h^{-1}] Air, $\times 10^2$ (n=3)	μ_{max} [h^{-1}] 1% O_2, $\times 10^2$ (n=1)
yBX186	*XYL1 XYL2*	+	+	4.74 ±0.401	4.30
yBX191	*XYL1 XYL2*	+	+	4.62 ±0.112	5.09
yBX192	*XYL1 XYL2*	+	+	4.45 ±0.209	5.16
yBX76	*XYL1 XYL2*	-	+	2.61 ±0.209	-
yBX76	*XYL1 XYL2*	-	-	0.95 ±0.126	0.15
yBX27	Ethanol Red MATa	-	-	0.10±0.042	0.20

3.2.3. Determination of the xylose uptake rate in the top-performing mutant.

The xylose uptake rate was evaluated for mutant yBX186 in xylose mineral media (20 g/L xylose). The sugar levels in the media were measured via reaction with 3,5-dinitrosalicylic acid (Miller, 1959). When growing cultures under the same conditions used for determining growth rate, no change in xylose level was detected after 1 week of incubation. In order to ensure a detectable change in xylose levels, the pregrowth was washed into the xylose media without dilution, which was equivalent to a 50-fold biomass concentration increase. Figure 13 shows a single representative plot of the biomass and media xylose levels over the course of 3 days. The reaction was carried out for a week, but on day 5 and onwards there was evidence of a small amount of contaminating microbe found in the pellets of centrifuged samples from yBX186 and the control. Since the contaminating microbe was absent when the pregrowth was washed, the analysis was restricted to the initial rates of reaction up to day 3. Mutant yBX186 consumed about 60±48% (±95% CI, n=2) of the xylose within 3 days, while the parental control only consumed about 22±7.9% (n=3,Figure 13A). Similarly, mutant yBX186 exhibited a greater biomass increase than the *XYL1 XYL2* control (Figure 13B). The rate of xylose disappearance was normalize by dry cell weight for better comparison (Figure 14). Because the xylose and biomass concentrations were not measured at identical time points between replicates, curve-fitting was done using cubic splines from which values could be averaged for all replicates. From this figure it was apparent that the mutant and the control displayed different trends in xylose uptake. In the mutant, the average

specific xylose utilization decreased linearly from $31.0\pm25.0\times10^{-3}$ h^{-1} at the start to $11.0\pm35.0\times10^{-3}$ h^{-1} by hour 80. For yBX76 the xylose uptake rate was more steady and only increased from $5.8\pm4.0\times10^{-3}$ h^{-1} to $6.9\pm5.0\times10^{-3}$ h^{-1} over the same time period.

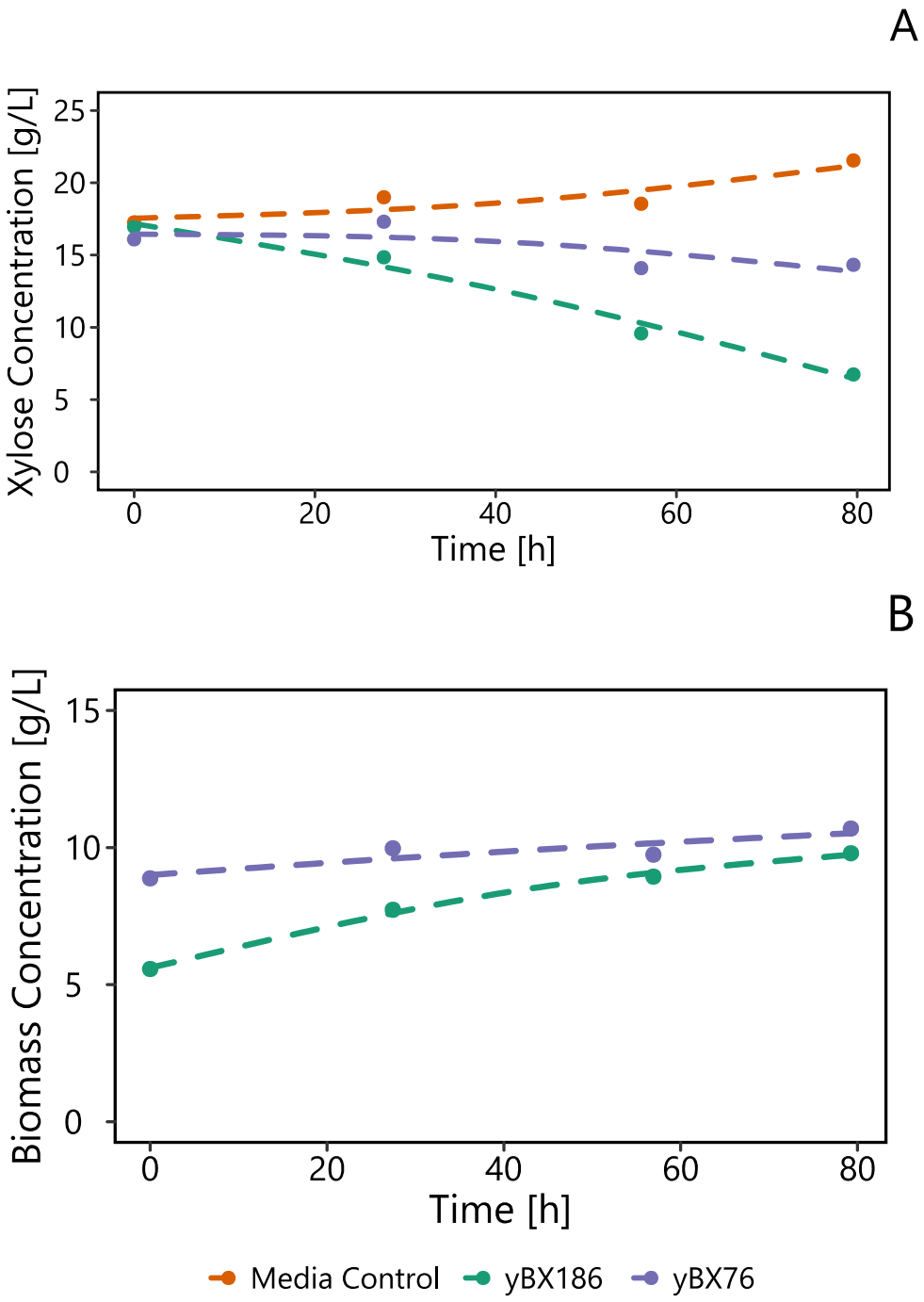

A
Xylose Concentration [g/L]
25
20
15
10
5
0
0
20
40
60
80
Time [h]
B
Biomass Concentration [g/L]
15
10
5
0
0
20
40
60
80
Time [h]
Media Control
yBX186
yBX76

Figure 13. Mutant yBX186 depleted approximately 60% of the xylose in 3 days. The xylose uptake of mutant yBX186 and control yBX76 was monitored in 20 g/L xylose mineral media. The cultures were incubated at 30 °C for 3 days and an aliquot was taken each day to measure the xylose and biomass concentration in the media. **(A)** The xylose levels were determined with the DNS assay. OD readings at 575 nm were converted to xylose concentration using a calibration curve. As a control the xylose levels of sterile 20 g/L media was measured alongside the culture (Media Control). **(B)** The biomass concentration was determined from OD measurement at 600 nm and converted to dry cell weight using a calibration curve. Each point is the average of n=3 technical OD replicates for a single representative biological replicate. The data was fit to a natural cubic spline (dashed lines).

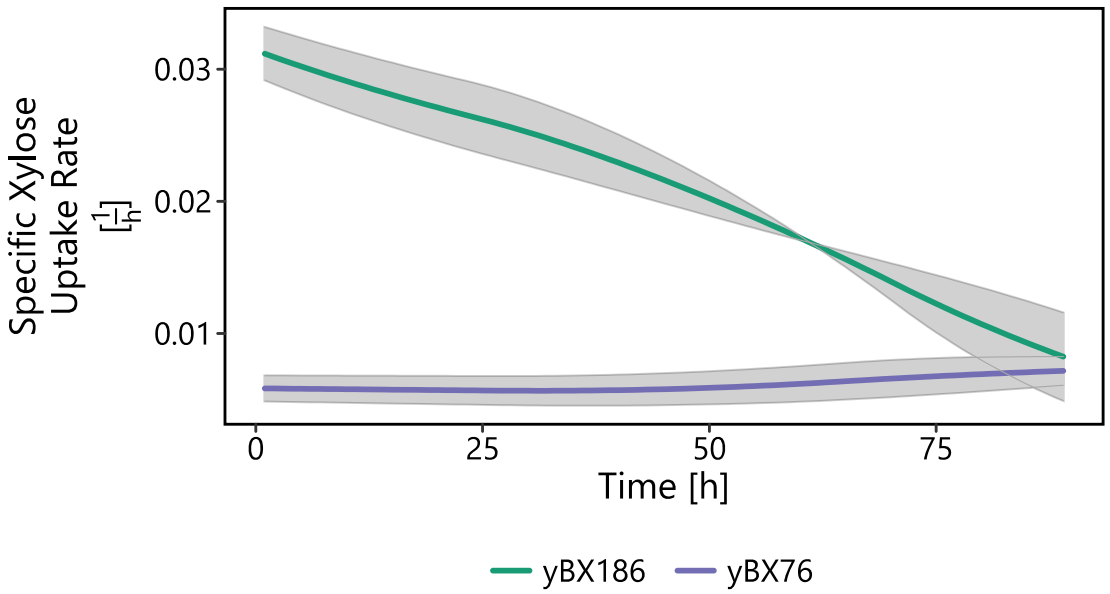


Figure 14. Mutant yBX186 has a 5-fold higher initial xylose utilization rate compared to yBX76. Using the spline models from Figure 13 to interpolate, the specific xylose utilization rate was plotted by dividing the derivative of the xylose concentration curve from Figure 13A by the biomass concentration curves in Figure 13B. The curves represent the mean of n=2 biological independent colony replicates for yBX186, and n=3 for yBX76. The shaded grey area represents the SEM.

3.2.4. Whole-genome sequencing of xylose-utilizing mutants.

Two representative mutants yBX102 and yBX128 with growth rates of 3.7×10^{-2} h^{-1} and 4.0×10^{-2} h^{-1} respectively were chosen for whole-genome sequencing to identify possible causative mutations for their enhanced growth on xylose. At the time of sequencing, these mutants were among the top-performing mutants isolated. The sequencing data from these mutants were aligned to the reference genome of the *XYL1 XYL2* control (yBX76), which was also sequenced to confirm the genetic construction. The overall local alignment rate was about 98% for both mutants. Variants were filtered for high-quality calls based on quality score, read depth, and alternate-to-reference allele ratio. yBX102 and yBX128 acquired 633 and 767 nuclear variants respectively. Of those variants there 59 and 43 indels in yBX102 and yBX128 respectively, and the rest were SBSs. Contrary to expectation, there were no mutations found in *XYL1*, *XYL2*, or in the intergenic regions near the DSB site.

To determine if there was any A3A activity in these mutants, the variants were analyzed at the trinucleotide context to see if there was an enrichment of mutations that occurred at the preferred motifs of A3A (Figure 15). As indicated in Figure 15 the most frequent mutation was the C to T transition, which is the primary mutation caused by A3A (Chan et al., 2015). Furthermore, the largest trinucleotide enrichments were in TCA and TCT motifs, which are the preferred motifs for A3A (Roberts et al., 2013). The third highest motif, TCG, is also known to be overrepresented by A3A hypermutation (Chan et al., 2015; Shi et al., 2017; Taylor et al., 2013).

Gene-based annotation revealed 102 and 109 genes with nonsynonymous mutations for yBX102 and yBX128 respectively (Supplementary Table 2 and Supplementary Table 3). In both mutants the 2 highest gene ontology term usages were for biological processes involving response to chemicals and transcription by RNA polymerase II (between 14% to 17% usage). Among the mutated genes, 39 of them were common to both mutants. There were no common genes involved in the xylose utilization pathway (Figure 1), however there were 2 genes *HAP4* and *CAT8* in which deletions are known to improve xylose utilization (Matsushika and Hoshino, 2015; Michael et al., 2016). Since all 39 common genes, including *HAP4* and *CAT8* had identical amino acid changes in both strains, these mutations were likely ancestral and were not attributed to the treatment.

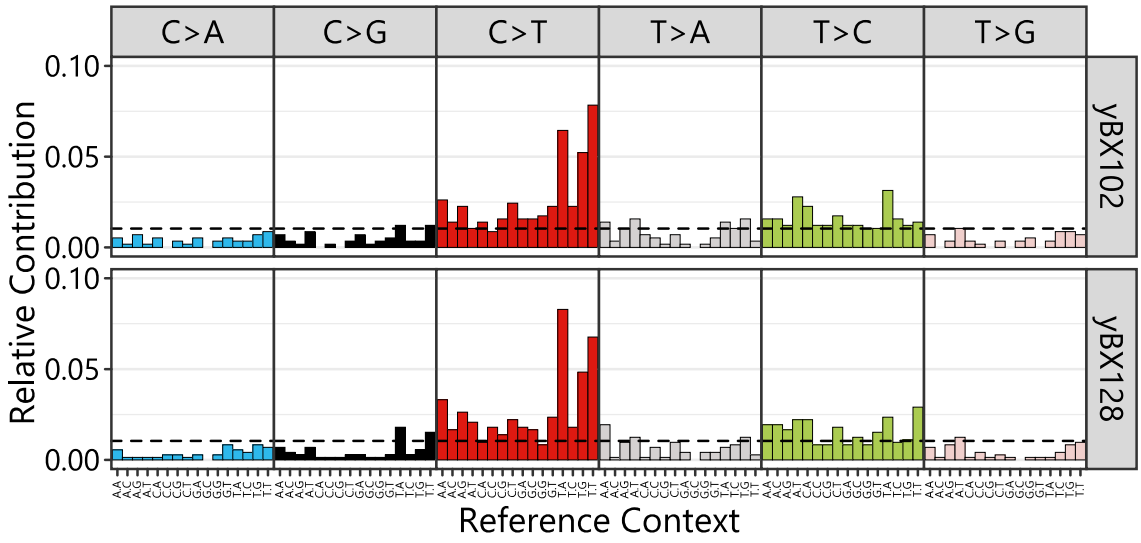


Figure 15. The mutation spectrum in mutants yBX102 and yBX128 reveals enrichment at the APOBEC3A preferred motifs. Paired-end reads generated from whole-genome sequencing of mutants yBX102 and yBX128 with 30× average coverage were locally aligned to the yBX76 reference genome. Multiallelic variants and indels were ignored. The remaining SBS variants were filtered for quality score >= 30, depth >=10, and alternate-to-reference read ratio >=0.9. The remaining variant count for yBX102 and yBX128 was 574 and 724 in the nuclear DNA. The abscissa shows the reference base with 1 base upstream and downstream context. The ordinate represents the proportion of variants that occurred at that context. The dashed line represents the expected contribution per context assuming equal likelihood of each context (1 in 98).

3.3. Improving mixed-sugar co-utilization in *S. cerevisiae* with LH.

When grown on a mixture of carbon sources, *S. cerevisiae* will preferentially consume fermentable carbon sources before nonfermentable sources because of glucose repression (Raamsdonk et al., 2001). Moreover, engineered *S. cerevisiae* strains ferment xylose, but transcriptional profiling shows that their global gene expression pattern on xylose is similar to that of nonfermentable carbon sources (Jin et al., 2004). Consequently in glucose-xylose mixtures, *S. cerevisiae* consumes xylose only when glucose is depleted (Kim et al., 2012a). Studies have shown that *HXK2* and *REG1* are key negative regulators of alternative carbon source metabolism and positive regulators of glucose repression (Herwig and von Stockar, 2002; McCartney et al., 2014). Because studies investigating simultaneous mutation in both genes are lacking, these genes were chosen for hypermutation.

The *HXK2* and *REG1* genes were deleted at their native loci and relocated with their upstream and downstream intergenic sequences to chromosome 16 (Figure 16). This construct was then transformed with an A3A-carrying plasmid and designed as yBX120DA. Before generating mutants, the screening method required optimization. In this experiment 2-deoxyglucose (2-DG) plus an alternative carbon source was used to screen for glucose derepressed mutants. The toxic glucose analogue 2-DG is known to induce glucose repression, but is not an energy source thus cells can only grow if they can consume the alternative carbon source in the presence of 2-DG (Curiel et al., 2016; Lane et al., 2018; McCartney et al., 2014).

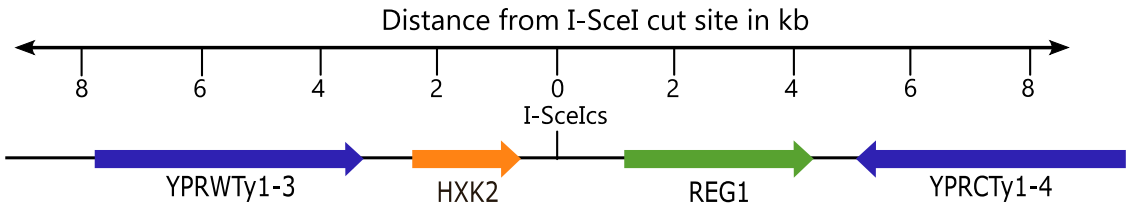


Figure 16. Relative position of *HXK2* and *REG1* to the DSB site in chromosome 16 for localized hypermutation. *HXK2* and *REG1* were relocated to chromosome 16 at position 850475 between two LTR retrotransposons. The two genes were amplified along with their entire upstream and downstream intergenic sequences and deleted at their native loci.

Since studies that use 2-DG are typically conducted in liquid media, and that the degree of glucose repression is strain dependant (New et al., 2014; Raamsdonk et al., 2001), it was necessary to test the inhibitory effect of 2-DG on Ethanol Red. In glucose repression studies, sucrose is the preferred alternative sugar as the expression of *SUC2* , which is required for sucrose breakdown, is known to be entirely regulated by glucose repression (Herwig and von Stockar, 2002; McCartney et al., 2014; Meijer et al., 1998; Ye et al., 1999). Since Ethanol Red has poor growth on xylose, sucrose was used as a substitute for xylose. Ethanol Red cells were plated on 20 g/L sucrose mineral media with various concentrations of 2-DG to assess viability (Figure 17). No colonies were observed for the wild-type background at concentrations of 0.3 g/L and 0.4 g/L 2-DG. As a positive control *hxk2*Δ cells maintained over 70% viability at all concentrations tested.

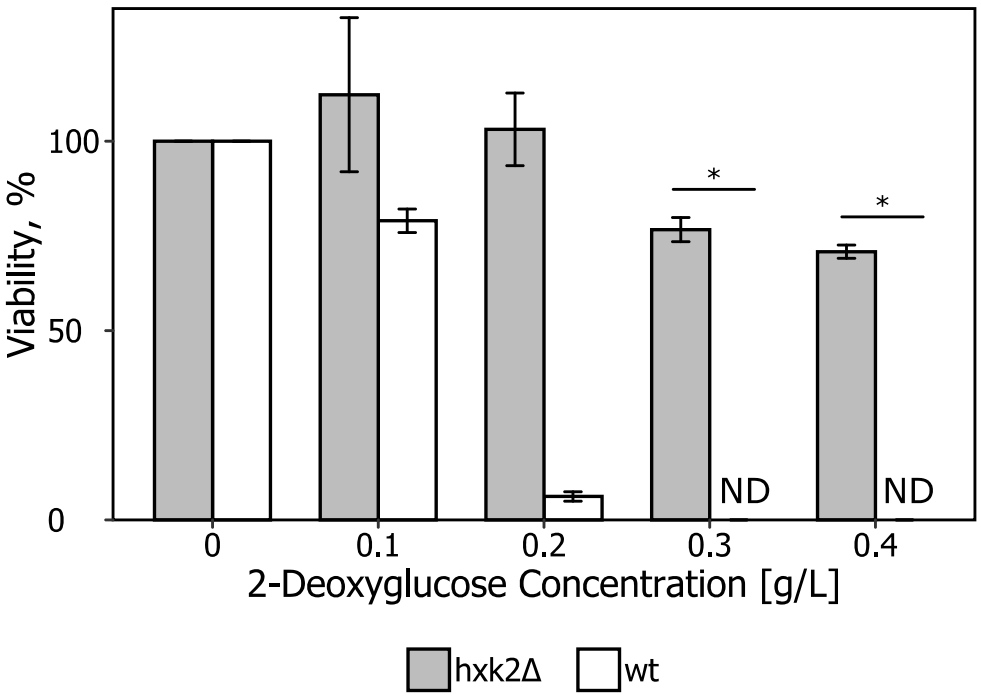


Figure 17. 2-Deoxyglucose inhibits the growth of wild-type Ethanol Red, but not *hxk2*Δ on sucrose media. Cultures of yBX27 and *hxk2*Δ were grown in YPR media and plated on 20 g/L sucrose mineral media with various concentrations of 2-DG ranging from 0 to 0.4 g/L. For each concentration 200 cells were plated and the viability was calculated as the plating efficiency relative to the 0 g/L 2-DG plates. Symbols: ND, not detected (no colonies grew up); wt, wild type (yBX27). Bars represent the average of 2 biological replicates. Error bars represent the SEM. Symbols: *, $p<0.05$.

Initially, cells were subjected to LH and were screened on 20 g/L sucrose 0.3 g/L 2-DG media; however, when plating at 10^6 cells/plate there was considerable background growth of both treated samples and parental backgrounds. The 2-DG resistance in these colonies was often not inherited when they were restreaked. To reduce the frequency of false positives, galactose was used instead of sucrose as a glucose-repressible sugar (Escalante-Chong et al., 2015; New et al., 2014). When grown on galactose and 2-DG the background growth was greatly reduced and 2-DG resistant colonies showed inheritance of 2-DG resistance when retested. The deletion genotypes *hxk2*Δ, *reg1*Δ, and *hxk2*Δ *reg1*Δ were pregrown on YPR and plated on 20 g/L galactose 0.3 g/L 2-DG (Figure 18). All three genotypes exhibited more colonies on 2-DG than wild type. The double deletion yielded substantially more colonies than the single deletions as evidenced by the lawn of cells. Additionally, *reg1*Δ cells displayed a growth defect on galactose.

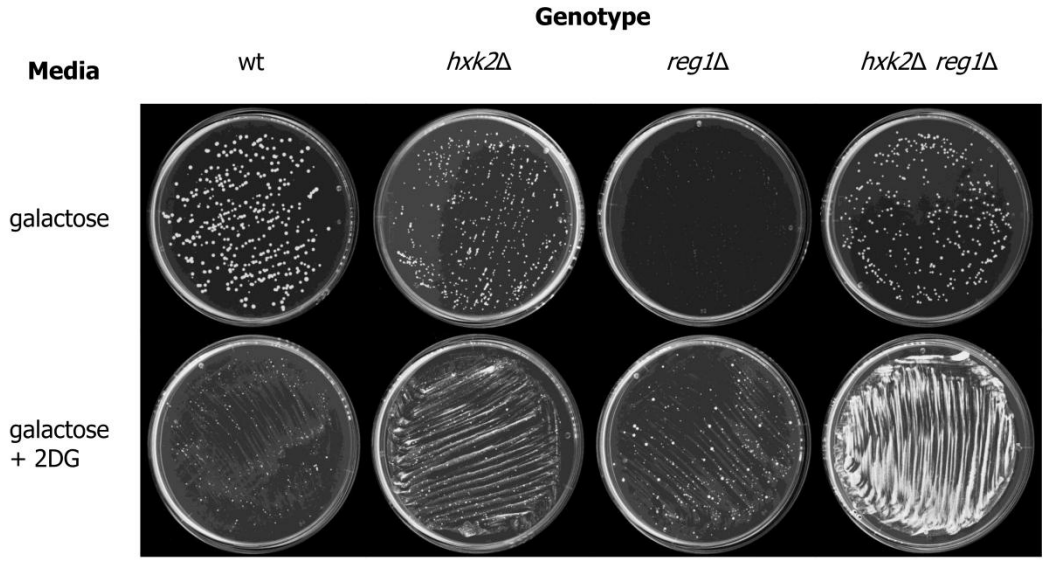


Figure 18. The double deletion genotype *hxk2*Δ *reg1*Δ provides more resistance to 2-deoxyglucose on galactose media than the single deletions. The screen for glucose derepressed mutants was simulated with the wild type and the deletions. Each genotype was grown in YPGal media for 1 day then plated on 20 g/L galactose mineral media with or without 0.3 g/L 2-DG and incubated at 30 °C for a week. The top and bottom rows were plated at 300 and 10^6 cells/plate respectively. The galactose plates in the top row are indicative of the plating efficiency across genotypes. The galactose+2-DG plates in the bottom row are indicative of the proportion of glucose derepressed cells. Shown in this figure are representative plates for 4 biological replicates tested.

Next, these controls were tested for their ability to co-utilize glucose and galactose in liquid media. The extent of the lag phase was used to measure the degree of glucose repression; the stronger the repression the slower the growth at the transition from glucose to galactose metabolism (New et al., 2014). The four genotypes were grown overnight on 20 g/L glucose and a mixture of 5 g/L glucose 20 g/L galactose (Figure 19). As shown in Figure 19A, the deletions have a growth defect on both glucose and the mixed media as evidenced by the lower biomass yields. To better visualize and compare the lag phase, the data was replotted in Figure 19B showing the growth rate as a function of biomass concentration (New et al., 2014). In the wild type, the growth difference at the lag phase was quantified at the point where the maximum difference in growth rate occurred between the single sugar media and mixed sugar media (line 'a' in Figure 19B). The deletions had a similar pattern of growth on both types of media, unlike the wild type, therefore they had a smaller lag phase difference. These plots also illustrate the growth defect in the deletions by their markedly reduced maximum growth rate on glucose (line 'b' in Figure 19B).

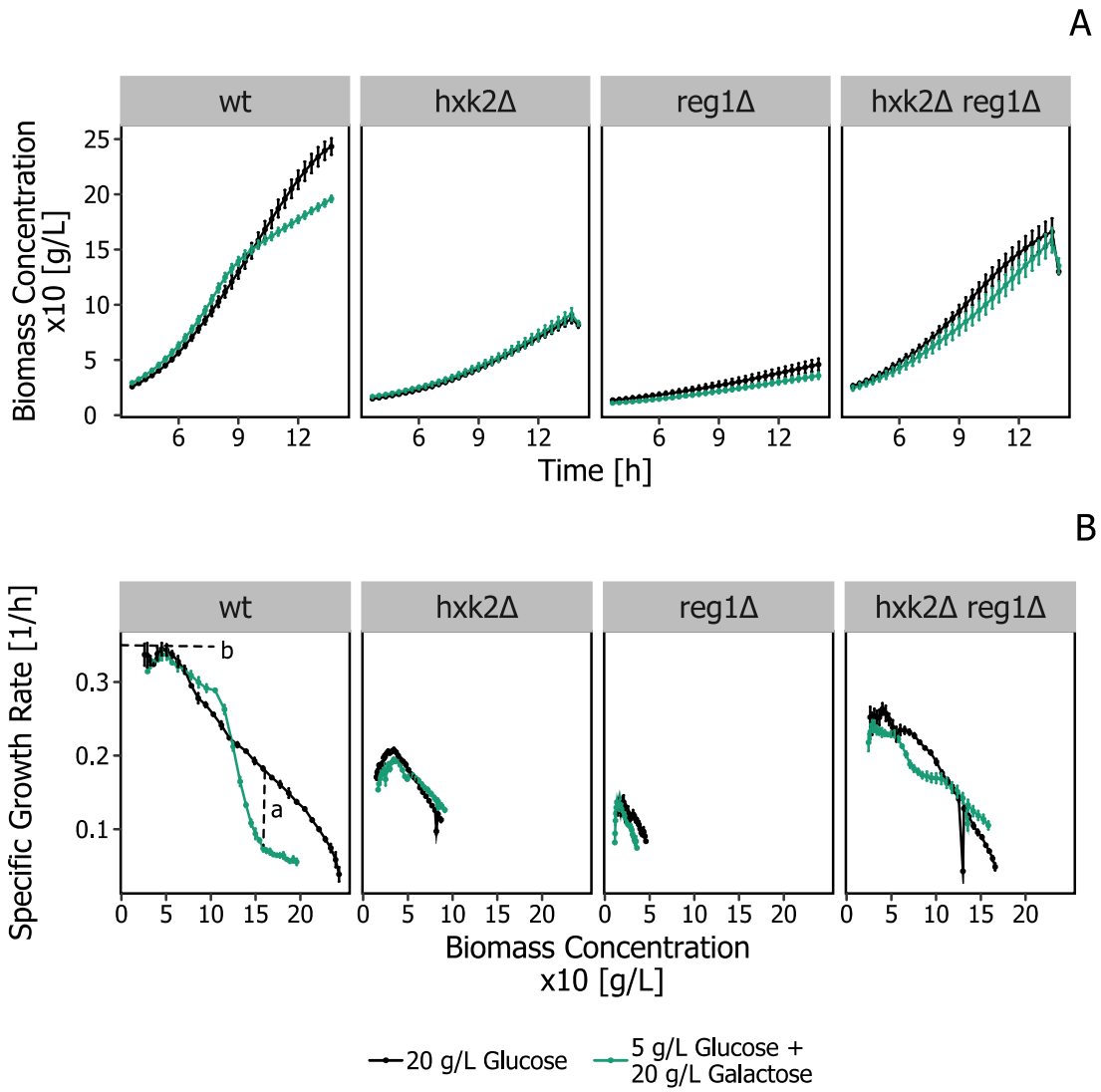


Figure 19. Deletion of *HXK2* and *REG1* results in similar growth patterns on mixed-sugar media and glucose media. Comparing the growth patterns on mixed-sugar media and glucose media can be used to estimate the degree of sugar co-utilization. **(A)** Each genotype was grown in YPR and diluted into either 20 g/L glucose mineral media or 5 g/L glucose 20 g/L galactose mineral media and grown overnight at 30 °C. The biomass concentration was determined from OD measurement at 600 nm and converted to dry cell weight using a calibration curve. Each point is the average of n=2 biological replicates. Error bars represent the SEM and may be obscured by the points. **(B)** The same data from **A**, but plotting growth rate as a function of biomass concentration. Dashed line 'a' represents the maximum growth rate difference between the mixed-sugar media and glucose media. Dashed line 'b' represents the maximum growth rate on glucose media. Symbols: wt, wild type.

3.3.1. Characterizing the growth rates of glucose derepressed mutants in mixed-sugar media.

The strain yBX130DA that carries *HXK2* and *REG1* adjacent to an inducible DSB and carries A3A was grown in the reaction media for up to 2 days and screened on 20 g/L galactose 0.3 g/L 2-DG plates. Mutant colonies that reproducibly grew on 2-DG plates were used for further characterization. The growth rates of 69 mutants were tested in glucose and glucose+galactose media. Representative growth curves of 11 mutants are shown in Figure 20. Similar to the xylose utilization experiment, A3A-carrying backgrounds exhibited highly variable phenotypes in glucose-galactose mixtures. Relative to the wild type (yBX27), mutants generally had lower growth rate differences between the two media. Some mutants such as GRM418 and GRM250 also had greatly reduced growth on glucose media and resemble the single deletions from Figure 19B. From this set of mutants, 5 were identified that appeared to be co-utilizing glucose and galactose. This was based on the fact that their growth profiles on glucose-galactose mixtures were nearly identically to the glucose mixtures; indicative of an absence of lag phase. In Figure 20 these mutants are GRM15, GRM262, GRM416, GRM402, and GRM250. In terms of minimizing lag phase, the top performing mutant was GRM262, which had 26% of the maximum growth rate difference compared to the wild type and 81% of the maximum growth rate on glucose (Table 4).

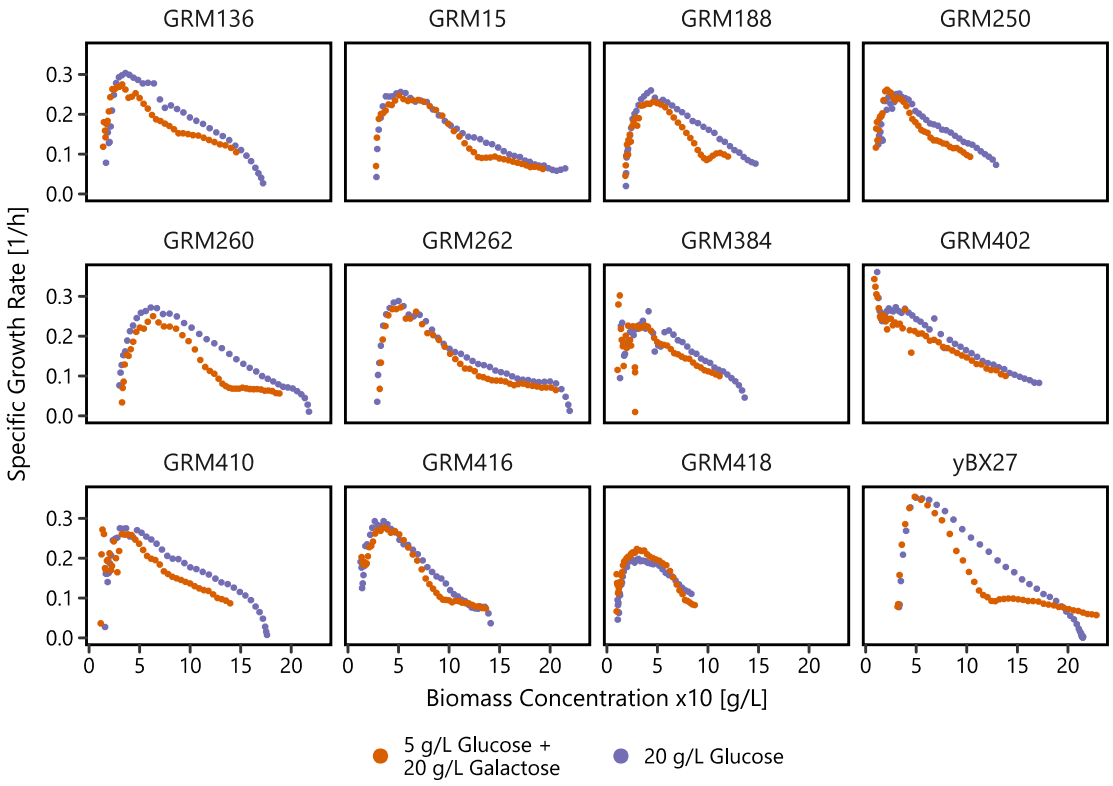


Figure 20. Localized hypermutation with APOBEC3A generated *S. cerevisiae* mutants with improved co-utilization of glucose and galactose. Representative growth curves of 11 mutants growing in either 20 g/L glucose mineral media or 5 g/L glucose 20 g/L galactose mineral media are shown from a sample size of 69 mutants tested. The biomass concentration was determined from OD measurement at 600 nm and converted to cell dry weight using a calibration curve. Each point is the average of n=2 technical culture replicates. Each sample represents an n=1 biological colony replicate.

Table 4. Growth rate parameters for the top 5 glucose and galactose co-utilizing mutants. Local growth rates were calculated between successive days using Equation 3. The largest of these rates is reported as μ_{max}, which was typically on day 1. The value for MaxΔμ was calculated as the greatest difference in growth rate between glucose and glucose+galactose media at the same biomass concentration. The values for μ_{max} and MaxΔμ are representative for n=1 biological replicates.Symbols: wt, wild type.

Strain	μ_{max} [1/h] Glucose	MaxΔμ [1/h] Glucose/Glucose+Galactose
GRM262	0.288	2.99×10^{-2}
GRM416	0.293	4.48×10^{-2}
GRM15	0.256	2.70×10^{-2}
GRM402	0.273	4.02×10^{-2}
GRM250	0.252	3.65×10^{-2}
hxk2Δ	0.208	1.36×10^{-2}
reg1Δ	0.141	3.73×10^{-2}
hxk2Δ reg1Δ	0.263	4.62×10^{-2}
wt	0.352	11.3×10^{-2}

4. DISCUSSION

In this study a system consisting of an inducible double-strand break and the human APOBEC3A was used for selectively mutating up to 40 kb of DNA at higher rates than elsewhere in the genome. This system was tested in *S. cerevisiae* containing a triple-gene reporter system. The effectiveness of localized hypermutation was measured by the inactivation frequency in the reporter genes and by cell survival. Next this technique was applied to improve xylose utilization and co-utilization of glucose and galactose in *S. cerevisiae*.

4.1. Multiple gene inactivation using a triple-reporter system.

4.1.1. Single and double gene inactivation frequency by A3A.

The endogenous inactivation frequency of *CAN1* in other *S. cerevisiae* backgrounds is reported to be between 1.5×10^{-7} to 2.1×10^{-7} (Lis et al., 2008; Shor et al., 2013). In comparison, the background *CAN1* inactivation frequency of ySR128 was 3.7×10^{-6} (Figure 5A, WT). The background rate of canavanine resistance calculated in this study is consistent with reported values in a related background (Chan et al., 2012). This suggests that the ySR128 lineage has an increased endogenous mutation rate. When transferred to the reaction media, the *CAN1* inactivation frequency of the parental strain remained steady over 4 days of incubation. Therefore, the media composition and incubation time did not contribute to DNA damage accumulation nor affect the endogenous mutation

rate. Any increase in the gene inactivation frequency can be attributed to induction of APOBEC3A (A3A).

After 1 day of incubation in the reaction media, the A3A and DSB A3A backgrounds showed a greater increase in canavanine resistance (27-fold and 12-fold respectively) compared to controls that only had a 0.7- to 3-fold increase relative to day 0. A similar effect was seen in *CAN1 URA3* and *CAN1 ADE2* double gene inactivation events (Figure 5BC). Although a greater mutation frequency was found in backgrounds carrying A3A, there was no difference in viability on day 1 (Figure 5D). With further incubation, the inactivation frequency began to reach a plateau with less change over time while viability steadily decreased with each day. This effect was observed in all mutant phenotypes and suggests that localized hypermutation generated the most mutations on day 1 with diminishing effectiveness over time. One explanation for this pattern is that Can^R cells accumulate mutations from A3A activity and can die from inactivation of essential genes. Also, the calculation of mutation frequency depends on viable colony counts. Thus with longer incubation time, nonviable Can^R cells can drop out from the screen and offset the gain of newly acquired canavanine resistance in other cells, which may explain the plateau effect. These results suggest that a longer incubation time is associated with diminishing mutation specificity towards the triple reporter system due to increasing off-target mutations. Instances of possible off-target A3A deamination can occur at times when DNA is temporarily in single-stranded form such as during transcription, replication, or repair. Furthermore, A3A is highly active and

expression of A3A has been shown to mutate most cytosine bases on plasmids within 48 h (Stenglein et al., 2010). This can also explain the bigger increase in gene inactivation events on day 1 compared to other days. Daily cell concentration measurement showed there was little to no growth in the reaction media, therefore the biggest population of cells to acquire canavanine resistance should be on day 1. Combining this with the high activity of A3A lends further support for the observed trends.

In this experiment the mutants were screened for binary traits (either present or not present) and DNA sequencing was not performed on any mutants. This experiment was limited in estimating the number of mutations acquired per gene and it is expected that one benefit of a longer incubation time is the acquisition of more mutations beyond what is necessary for inactivation. Thus in screening of more complex traits, longer incubations may be useful for generating more mutant diversity despite the lower survival.

A3A backgrounds exhibited a similar pattern of mutagenesis as DSB A3A and on average had 2-fold lower inactivation frequency. This was contrary to expectation as the presence of a DSB is known to increase the local mutation frequency of other cytidine deaminases— up to 2 orders of magnitude (Poltoratsky et al., 2010; Taylor et al., 2013). Poltoratsky et al. (2010) used the same I-SceI system and observed a strong reduction in survival by 48 h of DSB induction. In this experiment, a reduction was also observed in survival, but it was not significant at the 5% confidence level until day 3. This observation supports that DSB induction in this study occurred to a lesser degree than what is

observed in the literature and contributed little to gene inactivation relative to A3A alone. One reason for why the enhancement with the DSB was not as high as in the literature could be due to A3A's high activity masking the effect of the DSB. Among the human cytidine deaminases, APOBEC3A is known to have the highest activity (Carpenter et al., 2012; Taylor et al., 2013) and has been shown to cause DSBs while AID cannot (Mussil et al., 2013). The studies that observed an increase in mutation frequency with DSB induction used AID and A3G, and furthermore Taylor et al. (2013) noted that A3G was specifically chosen to test the enhancement with a DSB due A3G generating lower hypermutation relative to other deaminases.

When factoring in the survival drop, 2 days of mutagenesis maximizes the number of viable mutants per plate in DSB A3A backgrounds. This value was about 2.8-fold higher than with 4 days of incubation. For this reason, 2 days of incubation was the maximum time used in later experiments. At this incubation time, the average single gene inactivation frequency, double gene inactivation frequency, and survival were 6.5×10^{-4} , 1.5×10^{-5}, and 39% respectively. These results compare favourably to UV mutagenesis. Hashimoto et al. (2005) generated auxotrophic mutants in diploid brewing strains, which requires simultaneous inactivation of 2 alleles, and reported frequencies in the range of 5.3×10^{-4} to 2.0×10^{-3} and survivals ranging from 8% to 22%. Although the A3A-induced double gene inactivation frequency was 35- to 130-fold lower, the viability in these haploids was between 1.7- to 4-fold higher, which indicates less background mutation. In addition, the reported UV mutation frequencies were based on pooling

different types of nutrient requirements. Therefore, on a per-trait-of-interest basis the values would be lower by a factor of 2 to 3 based on the different phenotypes identified. Localized hypermutation with A3A, despite having up to 2 orders of magnitude lower loss-of-function rate in 2 genes, appears to be comparable to UV mutagenesis in usefulness because of higher survival rates and it is expected to have a higher per-gene mutation rate. A similar *in vivo* targeted mutagenesis approach utilized the DNA glycosylase Mag1 fused to a tet repressor protein to mutate a region of DNA spanning up to 20 kb and reported an 800-fold increase in *URA3* inactivation rate (Finney-Manchester and Maheshri, 2013). The system relies on the binding of the tet repressor to an integrated array of tet operator sequences in order to localize the base excision activity of Mag1, which is known to excise guanine bases from dsDNA (Berdal et al., 1998). A disadvantage of this technique is that two-thirds of the recovered mutants were deleted for the two selection markers used, which is indicative of undesirable chromosomal breaks and rearrangements. Because this study used a triple-reporter system and simultaneous loss of function in all 3 genes was rare, the mutant phenotypes observed in this study were likely not caused by chromosomal rearrangements or breaks since at least one of the three genes was functional at all times.

4.1.2. Comparison of mutation rate under different doxycycline treatments.

Maximizing the mutation frequency while maintaining adequate viability ensures the most diverse sampling of phenotypes. Increasing the doxycycline concentration was

shown to not have a significant effect on mutation frequency (Figure 6). This suggests that either A3A expression is saturated, or that induction of A3A is not limiting for mutagenesis. The starting concentration of 10 µg/mL doxycycline is higher than what is required for expression and is close to the expression saturation limit (Roney et al., 2016). Other studies argue that the persistence of long stretches of ssDNA is limiting for mutagenesis and not the expression levels of A3A (Sakofsky et al., 2019). Higher doxycycline concentrations caused a drop in survival on days 3 and 4, but there was no significant difference in mutation frequency. Thus increasing the amount of doxycycline did not improve mutagenesis and the effect on survival could possibly be attributed to doxycycline toxicity (Bellí et al., 1998).

4.1.3. Effect of rescue with DSB repair oligonucleotides on survival.

Since mutation rates could not be increased any further with additional doxycycline, an alternate route is to increase the number of viable mutants by increasing survival. The use of oligonucleotides as a template for DNA repair has been useful for recovering mutants that were exposed to a DSB (Yang et al., 2008). This experiment was conducted in the *XYL1 XYL2* Ethanol Red background for generating xylose-utilizing mutants. Transformation with repair oligonucleotides resulted in a 2.5-fold drop in survival (Figure 8). Given that the mock transformation without oligonucleotides yielded similar survival rates, any rescuing effect with DSB repair oligonucleotides was negligible compared to the survival drop from the transformation process. Storici et al. (2003) similarly reported

25% survival when transforming with or without repair oligonucleotides. From this data it remains inconclusive whether plating the surviving subpopulation of transformants is more effective in generating mutants than plating untransformed cells. One consideration that the rescue transformation did not improve mutant selection is that the samples from Figure 8 were also plated on xylose plates for mutant screening, and from a qualitative perspective there was no notable difference in the number of colonies formed between the mock and oligonucleotide transformations.

4.2. Generating xylose-utilizing *S. cerevisiae* mutants with LH.

4.2.1. Screening and growth rate determination of xylose-utilizing mutants.

Most if not all engineered *S. cerevisiae* strains in the literature place *XYL1* and *XYL2* under strong constitutive promoters and terminators because of poor expression levels under the *S. stipitis* regulatory elements (Amore et al., 1991; Kötter et al., 1990). Optimization studies revealed the importance of differential expression of *XYL1* and *XYL2* in improving xylose fermentation (Kim et al., 2012b; Zha et al., 2012). This experiment took the opposite approach and aimed to improve expression of *XYL1* and *XYL2* through mutagenesis by starting with low expression and working up. Attfield and Bell (2006) supported the feasibility of this approach by improving xylose utilization in nonrecombinant *S. cerevisiae* through a lengthy adaption procedure exceeding 1,000 days. It was hypothesized that localized hypermutation will accelerate this adaptation process.

In this study the industrial ethanol-producing strain Ethanol Red was used in order to better simulate behaviour in an industrial setting. Ethanol Red is a robust background with higher ethanol tolerance than a typical lab strain (Lam et al., 2014). The innate thermotolerance and inhibitor tolerance of Ethanol Red make it an attractive strain for study in yeast robustness (Demeke et al., 2013b; Wallace-Salinas and Gorwa-Grauslund, 2013). Thus, any emerging xylose-utilizing mutants could be of immediate use to the industry. A representative haploid derived from Ethanol Red was engineered with *XYL1* and *XYL2* from *S. stipitis* to establish a xylose fermentation pathway (Figure 1). Mutagenized cells did not grow above background unless allowed to adapt on xylose media. This adaptation period took around 30 days until colonies were visible above background growth (Figure 10) and this duration was consistent with other studies (Jin et al., 2005; Liu and Hu, 2010; Wahlbom et al., 2003).

As shown in Figure 11, mutants carrying A3A exhibited a spectrum of growth profiles on xylose (n=153). This observation is consistent with the hypothesis that localized hypermutation generates diverse phenotypes. Some of these mutants performed worse than adaptation-only *XYL1 XYL2* controls, but the majority outperformed controls that lacked A3A. The top-performing mutant reached a maximum specific growth rate of 4.74×10^{-2} h^{-1}, which was 1.8-fold higher than the best adapted *XYL1 XYL2* control. Compared to the A3A-carrying backgrounds, the adapted control growth profiles displayed small variability suggesting a limited improvement capacity by adaptation alone; however, the controls also had a much smaller sample size (n=11) and

may not be fully representative of the complete sampling space.

Splitting the A3A-carrying mutants in Figure 11A by the duration spent in the reaction media revealed that mutants without induction of A3A (day 0) had a similar distribution of growth profiles as those with A3A induction (day 1 and 2) as shown in Figure 11B. This finding suggests that the presence of the plasmid harbouring A3A was more important than induction of A3A in the reaction media. One reason for this observation is leaky expression of the doxycycline induction system (Bellí et al., 1998; Roney et al., 2016). This explanation is corroborated by the higher mutation frequency of A3A-carrying backgrounds on day 0 of the triple gene inactivation experiment (Figure 5). The possibility of leaky expression of A3A during the 30 day incubation period is also important to consider. This could mean that the no-induction controls might have had significant A3A expression and could explain why they produced mutants comparable to mutants subjected to A3A induction.

The best mutants obtained in this study grew at rates of up 4-fold less than top-performing engineered strains in the literature (Garcia Sanchez et al., 2010; Kim et al., 2013; Moysés et al., 2016; Scalcinati et al., 2012). Clearly these first-generation mutants are growing on xylose at a fraction of their potential. With a much larger sample size it may be possible to find mutants comparable to the literature that can be used to better distinguish the effect of A3A induction.

4.2.2. Growth on xylose under different oxygen conditions.

Although *S. cerevisiae* can grow fermentatively under aerobic and anaerobic conditions, when growing on xylose the production of biomass and ethanol is strongly decoupled. Oxygen concentration in the media is an important parameter in xylose fermentation as *S. cerevisiae* follows different growth patterns depending on oxygen availability (Wahlbom et al., 2003). Under oxygen-rich conditions, *S. cerevisiae* respires xylose and exhibits CO_2 production and biomass accumulation. Under oxygen-limited conditions, fermentative growth occurs with little biomass production, and under anaerobic conditions fermentative growth is greatly reduced. This oxygen-dependent growth pattern is documented in other yeast species such as natural xylose-consuming *S. stipitis* and *S. passalidarum*, which shows that this effect is an intrinsic part of yeast metabolism (Veras et al., 2017).

Recent research has been focused on improving fermentative growth under strict anaerobic conditions (Hou et al., 2017; Kwak and Jin, 2017). The mechanism behind the inability of *S. cerevisiae* to grow on xylose anaerobically has not yet been completely elucidated and has been attributed to several factors: low expression of the pentose phosphate pathway genes, insufficient ATP production, insufficient CO_2 production, and lack of recognition of xylose as a fermentable carbon source (Bracher et al., 2018; Kuyper et al., 2005; Osiro et al., 2019; Sonderegger et al., 2004). In addition, the cofactor usage of XR and XDH, to some degree, contributes to suboptimal anaerobic growth because the cell relies on respiration for rapid oxidation of excess NADH produced by XR (Kuyper et

al., 2004; Sonderegger et al., 2004). A mutant with altered cofactor preference for XDH that accepts NADH with higher preference over NADPH may alleviate this redox imbalance burden, therefore it was of interest to test how the mutants obtained in this study behaved under different oxygen conditions.

The 10 fastest-growing mutants were tested for their growth rates under hypoxic conditions and some were tested under anaerobic conditions (Figure 12). There was no detectable growth within the course of a week under anaerobic conditions for any of the mutants. In order to test if the apparatus had impacted viability, the tubes were disconnected from the gas lines and allowed to grow in room air for 3 more days (Supplementary Figure 4). All samples except the negative controls quickly resumed growth showing that the lack of oxygen was responsible for the halted growth.

When grown in 1% oxygen, most mutants attained lower biomass than when grown in room air (Figure 12). This result is consistent with the expected growth behavior as described earlier. Moreover, this result not only shows that oxygen was limiting for growth, but also that the mutants were most likely respiring the xylose in room air. A comparison of ethanol titers under the three conditions would provide better information on the performance of these mutants; however, due to a lack of equipment for precise quantification of ethanol this could not be done. Two exceptions to the oxygen dependence on growth rate were mutants yBX191 and yBX192 that grew faster in 1% oxygen, while the top performer grew 10% slower (Table 3).

Future characterization of these mutants should include more biological replicates to perform statistical analysis, and genome sequencing of mutants yBX191 and yBX192 to look for causative mutations that improve growth in 1% oxygen. Finally, this experiment shows how using synthetic gas mixtures with reduced oxygen pressure can help discriminate between xylose-utilizing mutants that otherwise appear similar when grown in room air or nitrogen gas. Typically, studies such as that by Parreiras et al. (2014) improve anaerobic growth in two steps: first screen for aerobic growth, then shift to screening under anaerobic conditions. The selection to go from fully aerobic to anaerobic growth may be too restrictive. For example strain development by Sonderegger and Sauer (2003) required over 100 days of selection in an anaerobic chemostat until anaerobic growth was identified. The findings in this study show that by screening a large mutant pool at progressively lower oxygen concentrations, it may be possible to improve anaerobic growth more efficiently by gradually identifying high-performing mutants and it can also narrow the pool of promising mutants for genetic analysis and further improvement.

4.2.3. Determination of the xylose uptake rate in the top-performing mutant.

The rate of xylose consumption is an important parameter for fermentation as it controls the minimum reaction time, which affects operating costs (heating and cooling) and the ethanol production rate. The cost of producing ethanol also depends to a large degree on capital costs and a lower reaction time provides additional savings in capital costs

through requiring smaller vessels and other equipment (Gubicza et al., 2016). The xylose uptake rate of the fastest-growing mutant was measured in a 200 mL culture and compared to the performance of the *XYL1 XYL2* parental control. Both the mutant and the parental control exhibited detectable biomass yield and xylose depletion in the media as shown in Figure 13. The slight growth of the control strain confirmed the presence of basal activity of either the *S. stipitis* XR and XDH or the endogenous homologs, which appears consistent with the results from Figure 11A. The specific xylose utilization rate of mutant yBX186 was about 5-fold higher than the parental control and decreased linearly from an initial rate of 31.0×10^{-3} h^{-1} to 11.0×10^{-3} h^{-1} (Figure 14).

Consistent with the growth rate comparison, the xylose consumption rates are also lower—by about 10-fold—than strains reported in the literature under similar conditions (Scalcinati et al., 2012; Wei et al., 2018). In addition to the magnitude of the rates, the time dependent profiles were also different between mutant yBX186 and the control, and is suggestive of a difference in xylose utilization. As xylose utilization is very inefficient in *S. cerevisiae* and fails at both the level of transport and metabolism, both processes can affect xylose uptake rate (Hou et al., 2017). Given that the xylose depletion rate for mutant yBX186 declines linearly over time as biomass increases, it is possible that this mutant is limited at the transport level. The drop in uptake rate could be explained by the lower flux of xylose transport per cell as the xylose concentration decreases and the numbers of cells increases. With regard to the control, the xylose uptake profile is suggestive of metabolism-limited uptake. One reason is the low

biomass increase, which means that the cell is not generating sufficient energy from xylose. Second, xylose is transported into the cell through facilitated diffusion (Sedlak and Ho, 2004), therefore it is expected that the transport rate decreases with decreasing extracellular xylose concentration. On the contrary, a constant uptake rate was observed in the control, which suggests a downstream bottleneck. Unfortunately the techniques available for this study only permitted the detection of free reducing sugars. For a more rigorous characterization of xylose utilization, it would be beneficial to study the fate of xylose by monitoring the levels of xylitol, ethanol, and other carbon products.

4.2.4. Whole-genome sequencing of xylose-utilizing mutants.

Whole-genome sequencing of two representative xylose-utilizing mutants revealed no mutations in the vicinity of the DSB site including the *XYL1* and *XYL2* ORFs. Further investigation was conducted since this result was contrary to expectation based on results from the triple gene inactivation experiment. To identify if there was any A3A activity in these strains, the mutations were plotted in a trinucleotide context (Figure 15). The biggest proportion of mutations were at C to T transitions, and among those the frequency of mutations at TCT, TCA, and TCG contexts was significantly higher than the expected value for equal likelihood mutation rates (as shown by the dashed line). The fact that the 3 most preferred motifs of A3A had the highest relative contribution to the mutation spectrum shows that there was A3A activity in these mutants and that it was responsible for up to about 20% of the total acquired mutations. Contrary to

expectation, the C to T transitions appeared to be distributed throughout the genome rather than concentrated near the *XYL1 XYL2* locus (Supplementary Figure 6 and Supplementary Figure 7).

Since the inducible DSB system was intended to localize A3A activity towards the target genes, it is possible that the I-SceI cutting was defective. Both the *SCEI* gene and the I-SceI cut site were correctly integrated in the genome based on sequencing data; however, there was a SBS in the promoter region of *SCEI* that was present in the reference and in the mutants, and it was probably introduced by PCR at the transformation step. It was unclear if this contributed to reduced expression of I-SceI. To test the efficacy of I-SceI induction on generating a DSB, parental strains lacking A3A were grown on galactose media for up to 5 hours. Genomic DNA extracted from these strains was used for PCR amplification across the I-SceI recognition site to test the integrity of the DNA at that locus. There was no significant difference in band intensity at any time point when normalizing with either amplified *HO* gene or gDNA template ($p<0.05$, Supplementary Figure 5). Significant cleavage at the I-SceI cut site should occur within 3 to 6 h of induction (Storici et al., 2003; Yang et al., 2008). In the fraction of cells with damaged DNA, amplification should not occur at the cut site and should result in a reduced band intensity. Studies using a similar galactose inducible endonuclease have shown that PCR is sensitive enough detect these events (Chaurasia et al., 2012). From this data it appears that DSB induction was defective in this background. Despite this problem, it was shown in the triple gene inactivation experiment that A3A DSB

backgrounds had an increase in mutation frequency of the same order of magnitude as A3A only backgrounds, therefore the DSB induction should not necessary to increase mutation frequency (Figure 5).

An explanation for why mutants yBX102 and yBX128 failed to acquire mutations at *XYL1* and *XYL2* despite the success of the triple gene inactivation experiment is likely due to the different nature of the screening process. In the previous experiment the screen identified mutants with loss-of-function mutations, but this experiment expected gain-of-function mutations (increased activity and modified cofactor preference of XR and XDH). Moreover, it is expected that enzyme altering mutations are more likely to disrupt function than enhance it. Thus it is possible that some proportion of cells do acquire hypermutation in *XYL1* and *XYL2*, but more likely than not abolished the pathway for consuming the only available carbon source and dropped out of the screen. To test this hypothesis a much bigger sample size should be sequenced to see if mutants in *XYL1* and *XYL2* can be recovered. Another challenge with this experiment is the possibility of a high false positive rate in the sense that mutants may evade mutation in *XYL1* and *XYL2*, and instead adapt through modifications in other pathways related to pentose metabolism. If further screening fails to yield mutants in *XYL1* and *XYL2*, it would be informative to conduct the triple gene inactivation experiment in the Ethanol Red background to identify any background-specific effects.

Analysis of possible causative mutations for the improved xylose utilization in yBX102 and yBX128 revealed nonsynonymous mutations in about 100 genes

(Supplementary Table 2 and Supplementary Table 3). About 40 mutated genes were common to both mutants and contained identical variants. Since the mutants were isolated from different biological replicates conducted in different experiments, these variants were likely of ancestral origin; likely acquired during construction of the background and were not attributed to improvement in xylose utilization. Of the remaining genes there were none involved in the key pathways for xylose utilization including the pentose phosphate pathway and glycolysis. The processes associated with mutations were diverse and included cell wall organization, osmotic stress, starvation response, ion transport, budding, DNA repair, etc. Only mutations in *HXT7* found in yBX102 stood out as having a direct relation to xylose utilization. *HXT7* encodes for a high-affinity glucose transporter and is used for transport of glucose at low concentrations (Reifenberger et al., 1995). Among the 18 transporters in *S. cerevisiae*, Hxt7 transports xylose the most efficiently (Sedlak and Ho, 2004), and is highly expressed on xylose (Zeng et al., 2016). Furthermore, mutations in *HXT7* have been found that convert it from a glucose transporter into a xylose transporter (Reider Apel et al., 2016; Young et al., 2014). The 2 amino acid changes S316D and S317T found in yBX102 are predicted to be in the cytoplasmic domain (Reider Apel et al., 2016), and because they are away from the sugar binding site in the transmembrane domain, it is expected they contribute little to xylose transport. In the future this hypothesis should be tested by introducing the mutations into the parental strain to see if they improve growth on xylose.

4.3. Improving mixed-sugar co-utilization in *S. cerevisiae* with LH.

4.3.1. Characterizing the growth rates of glucose derepressed mutants in mixed-sugar media.

This study investigated how hypermutation with A3A can be used to generate *S. cerevisiae* mutants with improved capacity to ferment sugars from lignocellulosic hydrolysates. As the main sugar components of lignocellulose are glucose and xylose (Okeke and Obi, 1994), this entails engineering efficient fermentation of glucose and xylose mixtures. The problem was addressed as two separate, simpler problems in parallel: improve xylose utilization in *S. cerevisiae*, and improve co-utilization of mixed-sugar media. In yeast and bacteria, glucose signals for the repression of alternative carbon source utilization pathways (Brückner and Titgemeyer, 2002; Busti et al., 2010). This response known as glucose repression results in a sequential fermentation of glucose first then xylose, which prolongs reaction time and increases ethanol production costs. Studies in *E. coli* have shown that co-utilization of xylose and glucose allows for over 2-fold reduction in reaction time (Chiang et al., 2013; Li et al., 2007). Unfortunately, reaction time reduction in co-utilizing yeast mutants has not been as significant and can even require longer time for full conversion of both sugars (Lane et al., 2018; Papapetridis et al., 2018).

In this study glucose derepressed mutants were screened on glucose and galactose media as a proxy for glucose and xylose co-utilization. This was done to avoid the confounding effect of suboptimal xylose consumption in the Ethanol Red

background. Another advantage of working with galactose is that marine biomass has little to no xylose, but is abundant in galactose (Yanagisawa et al., 2013). Thus galactose is not only a proxy for xylose, but co-utilization of glucose and galactose in itself is of possible interest to bioethanol manufacturers that use marine biomass as feedstock.

To generate glucose depressed mutants, the genes *HXK2* and *REG1* were chosen for hypermutation because they are known key players in glucose repression and arise frequently in genetic screens (McCartney et al., 2014). Despite being well-studied, there is a lack of research exploring mutant allele combinations of these genes (possibly due to the limitations of the mutagenesis techniques used). It is known that mutations in *HXK2* can improve co-utilization in glucose and xylose mixtures (Lane et al., 2018; Papapetridis et al., 2018), therefore it was expected that the phenotypes obtained in this study would extend to xylose mixtures.

Mutants were screened for glucose derepression on plates containing galactose and 2-DG. The principle behind this growth-based screen is that wild-type cells will prioritize consumption of the toxic 2-DG due to glucose repression and fail to grow on galactose. As a control, the selectivity was tested with *hxk2*Δ and *reg1*Δ cells. Hxk2 and Reg1 are positive regulators of glucose repression therefore it was excepted that deletions would have higher survival on the selective plates. The genotypes *hxk2*Δ and *reg1*Δ yielded higher colony counts than the control based on qualitative inspection (Figure 18). Quantitative comparisons of viability could not be made due the tendency of Ethanol Red to grow in clumps and chain-like structures that made accurate counting of

cell titers difficult. It was also apparent from the smaller colony sizes that the deletions had diminished growth, which may explain why mutants in the literature do not exhibit appreciably faster co-fermentation times. The defective growth was also verified in liquid culture with glucose only, and glucose and galactose media (Figure 19A). The resistance to 2-DG by *hxk2*Δ and *reg1*Δ strains in galactose media was confirmed by Curiel et al. (2016), but that study did not look at the *hxk2*Δ *reg1*Δ phenotype. The double deletion genotype *hxk2*Δ *reg1*Δ showed a synergistic effect that produced much higher survival on the 2-DG plates. This was not surprising since Hxk2 and Reg1 are known to act in parallel to regulate glucose repression (Busti et al., 2010). Interestingly, the double deletion also exhibited faster growth in liquid media (Figure 19). This was an encouraging result that suggests mutation in *HXK2* and *REG1* can modulate growth rate while maintaining glucose derepression.

A collection of 69 mutants were evaluated for their growth profile in glucose and galactose media. Figure 20 shows representative growth profiles of 11 mutants grown in 0.5% w/v glucose 2% w/v galactose media. The capacity for mixed-sugar co-utilization was determined indirectly by monitoring the extent of the diauxic shift since instrumentation for precise measurement of glucose and galactose levels in media was not available. A lag phase occurs during the transition from growth on glucose to galactose because of transcriptional reprogramming that happens as a consequence of glucose repression (Johnston et al., 1994; New et al., 2014). The mutants were ranked based on the biggest difference in growth rate between mixed-sugar media and glucose,

which should occur at the lag phase (line 'a' in Figure 19B), and their maximum growth rate on glucose media (line 'b' in Figure 19B). Five exceptional mutants with minimal lag phase and high growth rates on glucose are summarized in Table 4. Mutant GRM262 showed the best performance as its growth profile on mixed and glucose media were nearly superposed, which suggests that this mutant consumed 2 sugars similarly to consuming 1 sugar, i.e. co-utilization. The greatest growth rate discrepancy between mixed-sugar and glucose media in GRM262 was 3.8-fold lower than in wild type and 1.5-fold lower than in *hxk2*Δ *reg1*Δ cells. The maximum growth rate of GRM262 on glucose was 80% of wild type and about 9% higher than *hxk2*Δ *reg1*Δ cells. These results suggest that the mutant does not contain complete loss of function in Hxk2 and Reg1, and may represent a novel allele combination. To confirm this prediction, further replicates are necessary as well as whole-genome sequencing.

A confounding factor in this experiment is the galactose-to-glucose ratio, which is known to cause induction of galactose metabolic genes in between 40% to 60% of the cell population. Another limitation was that a complete set of negative controls including parental backgrounds was not tested. In future work it is crucial that the mutants obtained in this study including controls are tested at different sugar ratios to determine if the phenotypes are reproducible. Other next steps include whole-genome sequencing of these mutants and introducing pertinent mutations into the xylose-utilizing yeasts generated in this study to test if they enable co-utilization of glucose and xylose.

Ultimately, this study showed that the double deletion genotype *hxk2*Δ *reg1*Δ improved the growth defect caused by single deletions in either gene while still exhibiting evidence of sugar co-utilization. Through targeted mutagenesis of these 2 genes the growth rate was improved even further; up to 80% of wild-type. Since this gene combination is not well studied in xylose utilization, these findings may lead to new insights in achieving fast co-fermentation of sugars from lignocellulosic hydrolysates.

5. CONCLUSION

This study aimed to improve the diversity of yeast mutant screening by using localized hypermutation. The focus of this study was strain development for bioethanol production. Bioethanol is an environmentally friendly, carbon-neutral fuel that is a candidate to replace fossil fuels. Its main advantage is that growing crops for biofuel production requires sequestration of carbon dioxide from the atmosphere by photosynthesis. Therefore, unlike burning gasoline, bioethanol reuses the carbon dioxide from burning ethanol to regenerate the fuel. Unfortunately bioethanol production is still limited by prohibitive costs and requires more optimization to compete with fossil fuels. One way to cheapen the cost of producing bioethanol is to improve the efficiency at which microbes ferment the sugars from biomass.

In this study human APOBEC3A (A3A) was used to generate localized hypermutation in *S. cerevisiae*. It was hypothesized that A3A could be targeted to mutate many genes of interest at high rates by placing them near an inducible double-strand break site. Strains expressing A3A for 2 days achieved an increase of 2 orders of magnitude in single gene inactivation and an order of magnitude increase in double gene inactivation. These results confirmed that A3A can mutate at least 3 genes simultaneously at increased rates.

Next, this technique was applied to two important problems in bioethanol production: (1) improving xylose utilization in *S. cerevisiae*, and (2) improving mixed-sugar co-utilization in *S. cerevisiae*. The fastest growing mutant had a specific growth

rate of 4.74×10^{-2} h^{-1}, which was 1.8-fold higher than controls adapted without A3A. Whole–genome sequencing of 2 representative mutants revealed no mutations in *XYL1* and *XYL2* contrary to expectation; however, the limited size provides inconclusive evidence whether the technique failed since false positives can occur. One limitation in this experiment was the inability to measure ethanol production due to a lack of instrumentation for precise measurement. Instead, xylose utilization capacity was evaluated based on growth rate. Despite the underperformance of mutants generated in this study relative to the literature, the sample size in this study was only 159 mutants, which represents first-generation mutants. Given that mutants with better performance than controls were readily attainable even with a small sample size, it is encouraging evidence that better growth rates can be attained with a much larger sample size.

To generate yeast mutants with improved capacity to co-ferment mixed-sugar media, the genes *HXK2* and *REG1* were targeted for hypermutation in *S. cerevisiae*. Co-utilization was determined by how similar growth profiles on mixed-sugar media were compared to glucose media. The top-performing mutant had nearly identical growth profiles on mixed-sugar media and glucose media. One limitation in this experiment was the lack of direct measurement of sugar levels. More rigorous experiments will be necessary to truly know if they were co-utilizing sugars.

Taken together this data shows that even in its first iteration, localized hypermutation was able to improve xylose metabolism and co-utilization of sugar mixtures in *S. cerevisiae*. With further refinement, this approach may be of great use to

the biotechnology industry by providing a faster way of generating mutants in polygenic traits.

6. REFERENCES

Abt, T.D., Souffriau, B., Foulquié-Moreno, M.R., Duitama, J., and Thevelein, J.M. (2016). Genomic saturation mutagenesis and polygenic analysis identify novel yeast genes affecting ethyl acetate production, a non-selectable polygenic trait. Microb. Cell *3*, 159–175.

Ahuatzi, D., Riera, A., Peláez, R., Herrero, P., and Moreno, F. (2007). Hxk2 Regulates the Phosphorylation State of Mig1 and Therefore Its Nucleocytoplasmic Distribution. J. Biol. Chem. *282*, 4485–4493.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Amore, R., Kötter, P., Küster, C., Ciriacy, M., and Hollenberg, C.P. (1991). Cloning and expression in Saccharomyces cerevisiae of the NAD(P)H-dependent xylose reductase-encoding gene (XYL1) from the xylose-assimilating yeast Pichia stipitis. Gene *109*, 89–97.

Ashe, M.P., De Long, S.K., Sachs, A.B., and Wickens, M.P. (2000). Glucose Depletion Rapidly Inhibits Translation Initiation in Yeast. Mol. Biol. Cell *11*, 833–848.

Attfield, P.V., and Bell, P.J.L. (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.

Belinchón, M.M., and Gancedo, J.M. (2007). Glucose controls multiple processes in Saccharomyces cerevisiae through diverse combinations of signaling pathways. FEMS Yeast Res. *7*, 808–818.

Bellí, G., Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1998). An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. Nucleic Acids Res. *26*, 942–947.

Berdal, K.G., Johansen, R.F., and Seeberg, E. (1998). Release of normal bases from intact DNA by a native DNA repair enzyme. EMBO J. *17*, 363–367.

Berger, G., Durand, S., Fargier, G., Nguyen, X.-N., Cordeil, S., Bouaziz, S., Muriaux, D., Darlix, J.-L., and Cimarelli, A. (2011). APOBEC3A Is a Specific Inhibitor of the Early Phases of HIV-1 Infection in Myeloid Cells. PLOS Pathog. *7*, e1002221.

Bracher, J.M., Martinez-Rodriguez, O.A., Dekker, W.J.C., Verhoeven, M.D., van Maris, A.J.A., and Pronk, J.T. (2018). Reassessment of requirements for anaerobic xylose fermentation by engineered, non-evolved Saccharomyces cerevisiae strains. FEMS Yeast Res. *19*.

Broach, J.R. (2012). Nutritional Control of Growth and Development in Yeast. Genetics *192*, 73–105.

Brückner, R., and Titgemeyer, F. (2002). Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. FEMS Microbiol. Lett. *209*, 141–148.

Busti, S., Coccetti, P., Alberghina, L., and Vanoni, M. (2010). Glucose Signaling-Mediated Coordination of Cell Growth and Cell Cycle in Saccharomyces Cerevisiae. Sensors *10*, 6195–6240.

Canelas, A.B., van Gulik, W.M., and Heijnen, J.J. (2008). Determination of the cytosolic free NAD/NADH ratio in Saccharomyces cerevisiae under steady-state and highly dynamic conditions. Biotechnol. Bioeng. *100*, 734–743.

Carpenter, M.A., Li, M., Rathore, A., Lackey, L., Law, E.K., Land, A.M., Leonard, B., Shandilya, S.M.D., Bohn, M.-F., Schiffer, C.A., et al. (2012). Methylcytosine and Normal Cytosine Deamination by the Foreign DNA Restriction Enzyme APOBEC3A. J. Biol. Chem. *287*, 34801–34808.

Chan, K., Sterling, J.F., Roberts, S.A., Bhagwat, A.S., Resnick, M.A., and Gordenin, D.A. (2012). Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. PLoS Genet. *8*, e1003149.

Chan, K., Roberts, S.A., Klimczak, L.J., Sterling, J.F., Saini, N., Malc, E.P., Kim, J., Kwiatkowski, D.J., Fargo, D.C., Mieczkowski, P.A., et al. (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat. Genet. *47*, 1067–1072.

Chaurasia, P., Sen, R., Pandita, T.K., and Bhaumik, S.R. (2012). Preferential Repair of DNA Double-strand Break at the Active Gene in Vivo. J. Biol. Chem. *287*, 36414–36422.

Chiang, C.-J., Lee, H.M., Guo, H.J., Wang, Z.W., Lin, L.-J., and Chao, Y.-P. (2013). Systematic approach to engineer Escherichia coli pathways for co-utilization of a glucose-xylose mixture. J. Agric. Food Chem. *61*, 7583–7590.

Chiang, L.-C., Gong, C.-S., Chen, L.-F., and Tsao, G.T. (1981). d-Xylulose Fermentation to Ethanol by Saccharomyces cerevisiae. Appl. Environ. Microbiol. *42*, 284–289.

Cunha, J.T., Soares, P.O., Romaní, A., Thevelein, J.M., and Domingues, L. (2019). Xylose fermentation efficiency of industrial Saccharomyces cerevisiae yeast with separate or combined xylose reductase/xylitol dehydrogenase and xylose isomerase pathways. Biotechnol. Biofuels *12*, 20.

Curiel, J.A., Salvadó, Z., Tronchoni, J., Morales, P., Rodrigues, A.J., Quirós, M., and Gonzalez, R. (2016). Identification of target genes to control acetate yield during aerobic fermentation with Saccharomyces cerevisiae. Microb. Cell Factories *15*, 156.

Demeke, M.M., Dietz, H., Li, Y., Foulquié-Moreno, M.R., Mutturi, S., Deprez, S., Den Abt, T., Bonini, B.M., Liden, G., Dumortier, F., et al. (2013a). Development of a D-xylose fermenting and inhibitor tolerant industrial Saccharomyces cerevisiae strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. Biotechnol. Biofuels *6*, 89.

Demeke, M.M., Dumortier, F., Li, Y., Broeckx, T., Foulquié-Moreno, M.R., and Thevelein, J.M. (2013b). Combining inhibitor tolerance and D-xylose fermentation in industrial Saccharomyces cerevisiae for efficient lignocellulose-based bioethanol production. Biotechnol. Biofuels *6*, 120.

Escalante-Chong, R., Savir, Y., Carroll, S.M., Ingraham, J.B., Wang, J., Marx, C.J., and Springer, M. (2015). Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. Proc. Natl. Acad. Sci. *112*, 1636–1641.

Farwick, A., Bruder, S., Schadeweg, V., Oreb, M., and Boles, E. (2014). Engineering of yeast hexose transporters to transport d-xylose without inhibition by d-glucose. Proc. Natl. Acad. Sci. U. S. A. *111*, 5159–5164.

Fernández-García, P., Peláez, R., Herrero, P., and Moreno, F. (2012). Phosphorylation of Yeast Hexokinase 2 Regulates Its Nucleocytoplasmic Shuttling. J. Biol. Chem. *287*, 42151–42164.

Finney-Manchester, S.P., and Maheshri, N. (2013). Harnessing mutagenic homologous recombination for targeted mutagenesis in vivo by TaGTEAM. Nucleic Acids Res. *41*, e99.

Gancedo, J.M. (2008). The early steps of glucose signalling in yeast. FEMS Microbiol. Rev. *32*, 673–704.

Garcia Sanchez, R., Karhumaa, K., Fonseca, C., Sànchez Nogué, V., Almeida, J.R., Larsson, C.U., Bengtsson, O., Bettiga, M., Hahn-Hägerdal, B., and Gorwa-Grauslund, M.F. (2010). Improved xylose and arabinose utilization by an industrial recombinant Saccharomyces cerevisiae strain using evolutionary engineering. Biotechnol. Biofuels *3*, 13.

Goldstein, A.L., and McCusker, J.H. (1999). Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast Chichester Engl. *15*, 1541–1553.

Gubicza, K., Nieves, I.U., Sagues, W.J., Barta, Z., Shanmugam, K.T., and Ingram, L.O. (2016). Techno-economic analysis of ethanol production from sugarcane bagasse using a

Liquefaction plus Simultaneous Saccharification and co-Fermentation process. Bioresour. Technol. *208*, 42–48.

Hanaki, K., and Portugal-Pereira, J. (2018). The Effect of Biofuel Production on Greenhouse Gas Emission Reductions. In Biofuels and Sustainability: Holistic Perspectives for Policy-Making, K. Takeuchi, H. Shiroyama, O. Saito, and M. Matsuura, eds. (Tokyo: Springer Japan), pp. 53–71.

Hashimoto, S., Ogura, M., Aritomi, K., Hoshida, H., Nishizawa, Y., and Akada, R. (2005). Isolation of Auxotrophic Mutants of Diploid Industrial Yeast Strains after UV Mutagenesis. Appl Env. Microbiol *71*, 312–319.

Hector, R.E., Dien, B.S., Cotta, M.A., and Mertens, J.A. (2013). Growth and fermentation of D-xylose by Saccharomyces cerevisiae expressing a novel D-xylose isomerase originating from the bacterium Prevotella ruminicola TC2-24. Biotechnol. Biofuels *6*, 84.

Henningsen, B.M., Hon, S., Covalla, S.F., Sonu, C., Argyros, D.A., Barrett, T.F., Wiswall, E., Froehlich, A.C., and Zelle, R.M. (2015). Increasing Anaerobic Acetate Consumption and Ethanol Yields in Saccharomyces cerevisiae with NADPH-Specific Alcohol Dehydrogenase. Appl. Environ. Microbiol. *81*, 8108–8117.

Herwig, C., and von Stockar, U. (2002). Quantitative analysis of the oxidative metabolism in HXK2- and REG1-deletion mutants of Saccharomyces cerevisiae. Enzyme Microb. Technol. *31*, 698–710.

Hoopes, J., Cortez, L., Mertz, T., Malc, E.P., Mieczkowski, P.A., and Roberts, S.A. (2016). APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template during DNA Replication. Cell Rep. *14*, 1273–1282.

Hou, J., Qiu, C., Shen, Y., Li, H., and Bao, X. (2017). Engineering of Saccharomyces cerevisiae for the efficient co-utilization of glucose and xylose. FEMS Yeast Res. *17*.

Huxley, C., Green, E.D., and Dunham, I. (1990). Rapid assessment of S. cerevisiae mating type by PCR. Trends Genet. TIG *6*, 236.

Jin, Y.-S., Laplaza, J.M., and Jeffries, T.W. (2004). Saccharomyces cerevisiae Engineered for Xylose Metabolism Exhibits a Respiratory Response. Appl. Environ. Microbiol. *70*, 6816–6825.

Jin, Y.-S., Alper, H., Yang, Y.-T., and Stephanopoulos, G. (2005). Improvement of Xylose Uptake and Ethanol Production in Recombinant Saccharomyces cerevisiae through an Inverse Metabolic Engineering Approach. Appl. Environ. Microbiol. *71*, 8249–8256.

Johnston, M., Flick, J.S., and Pexton, T. (1994). Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae. Mol. Cell. Biol. *14*, 3834–3841.

Kim, S.R., Ha, S.-J., Wei, N., Oh, E.J., and Jin, Y.-S. (2012a). Simultaneous co-fermentation of mixed sugars: a promising strategy for producing cellulosic ethanol. Trends Biotechnol. *30*, 274–282.

Kim, S.R., Ha, S.-J., Kong, I.I., and Jin, Y.-S. (2012b). High expression of XYL2 coding for xylitol dehydrogenase is necessary for efficient xylose fermentation by engineered Saccharomyces cerevisiae. Metab. Eng. *14*, 336–343.

Kim, S.R., Skerker, J.M., Kang, W., Lesmana, A., Wei, N., Arkin, A.P., and Jin, Y.-S. (2013). Rational and Evolutionary Engineering Approaches Uncover a Small Set of Genetic Changes Efficient for Rapid Xylose Fermentation in Saccharomyces cerevisiae. PLOS ONE *8*, e57048.

Klein, C., and Struhl, K. (1994). Protein kinase A mediates growth-regulated expression of yeast ribosomal protein genes by modulating RAP1 transcriptional activity. Mol. Cell. Biol. *14*, 1920–1928.

Kötter, P., and Ciriacy, M. (1993). Xylose fermentation by Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol. *38*, 776–783.

Kötter, P., Amore, R., Hollenberg, C.P., and Ciriacy, M. (1990). Isolation and characterization of the Pichia stipitis xylitol dehydrogenase gene, XYL2, and construction of a xylose-utilizing Saccharomyces cerevisiae transformant. Curr. Genet. *18*, 493–500.

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

Kuyper, M., Hartog, M.M.P., Toirkens, M.J., Almering, M.J.H., Winkler, A.A., van Dijken, J.P., and Pronk, J.T. (2005). Metabolic engineering of a xylose-isomerase-expressing Saccharomyces cerevisiae strain for rapid anaerobic xylose fermentation. FEMS Yeast Res. *5*, 399–409.

Kvitek, D.J., and Sherlock, G. (2011). Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape. PLoS Genet. *7*.

Kwak, S., and Jin, Y.-S. (2017). Production of fuels and chemicals from xylose by engineered Saccharomyces cerevisiae: a review and perspective. Microb. Cell Factories *16*.

Lam, F.H., Ghaderi, A., Fink, G.R., and Stephanopoulos, G. (2014). Engineering alcohol tolerance in yeast. Science *346*, 71–75.

Lane, S., Xu, H., Oh, E.J., Kim, H., Lesmana, A., Jeong, D., Zhang, G., Tsai, C.-S., Jin, Y.-S., and Kim, S.R. (2018). Glucose repression can be alleviated by reducing glucose phosphorylation rate in Saccharomyces cerevisiae. Sci. Rep. *8*, 2613.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

Li, R., Chen, Q., Wang, P.G., and Qi, Q. (2007). A novel-designed Escherichia coli for the production of various polyhydroxyalkanoates from inexpensive substrate mixture. Appl. Microbiol. Biotechnol. *75*, 1103–1109.

Li, X., Park, A., Estrela, R., Kim, S.-R., Jin, Y.-S., and Cate, J.H.D. (2016). Comparison of xylose fermentation by two high-performance engineered strains of Saccharomyces cerevisiae. Biotechnol. Rep. *9*, 53–56.

Lis, E.T., O'Neill, B.M., Gil-Lamaignere, C., Chin, J.K., and Romesberg, F.E. (2008). Identification of Pathways Controlling DNA Damage Induced Mutation In Saccharomyces cerevisiae. DNA Repair *7*, 801–810.

Liu, E., and Hu, Y. (2010). Construction of a xylose-fermenting Saccharomyces cerevisiae strain by combined approaches of genetic engineering, chemical mutagenesis and evolutionary adaptation. Biochem. Eng. J. *48*, 204–210.

Lorenz, M.C., Pan, X., Harashima, T., Cardenas, M.E., Xue, Y., Hirsch, J.P., and Heitman, J. (2000). The G protein-coupled receptor gpr1 is a nutrient sensor that regulates pseudohyphal differentiation in Saccharomyces cerevisiae. Genetics *154*, 609–622.

Matsushika, A., and Hoshino, T. (2015). Increased ethanol production by deletion of HAP4 in recombinant xylose-assimilating Saccharomyces cerevisiae. J. Ind. Microbiol. Biotechnol. *42*, 1623–1631.

Mazur, X., Eppenberger, H.M., Bailey, J.E., and Fussenegger, M. (1999). A novel autoregulated proliferation-controlled production process using recombinant CHO cells. Biotechnol. Bioeng. *65*, 144–150.

McCartney, R.R., Chandrashekarappa, D.G., Zhang, B.B., and Schmidt, M.C. (2014). Genetic Analysis of Resistance and Sensitivity to 2-Deoxyglucose in Saccharomyces cerevisiae. Genetics *198*, 635–646.

Meijer, M.M.C., Boonstra, J., Verkleij, A.J., and Verrips, C.T. (1998). Glucose Repression in Saccharomyces cerevisiae Is Related to the Glucose Concentration Rather Than the Glucose Flux. J. Biol. Chem. *273*, 24102–24107.

Michael, D.G., Maier, E.J., Brown, H., Gish, S.R., Fiore, C., Brown, R.H., and Brent, M.R. (2016). Model-based transcriptome engineering promotes a fermentative transcriptional state in yeast. Proc. Natl. Acad. Sci. *113*, E7428–E7437.

Miller, G.L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal. Chem. *31*, 426–428.

Mohd Azhar, S.H., Abdulla, R., Jambo, S.A., Marbawi, H., Gansau, J.A., Mohd Faik, A.A., and Rodrigues, K.F. (2017). Yeasts in sustainable bioethanol production: A review. Biochem. Biophys. Rep. *10*, 52–61.

Moysés, D., Reis, V., Almeida, J., Moraes, L., and Torres, F. (2016). Xylose Fermentation by Saccharomyces cerevisiae: Challenges and Prospects. Int. J. Mol. Sci. *17*, 207.

Mukherjee, V., Radecka, D., Aerts, G., Verstrepen, K.J., Lievens, B., and Thevelein, J.M. (2017). Phenotypic landscape of non-conventional yeast species for different stress tolerance traits desirable in bioethanol fermentation. Biotechnol. Biofuels *10*.

Mussil, B., Suspène, R., Aynaud, M.-M., Gauvrit, A., Vartanian, J.-P., and Wain-Hobson, S. (2013). Human APOBEC3A Isoforms Translocate to the Nucleus and Induce DNA Double Strand Breaks Leading to Cell Stress and Death. PLOS ONE *8*, e73641.

New, A.M., Cerulus, B., Govers, S.K., Perez-Samper, G., Zhu, B., Boogmans, S., Xavier, J.B., and Verstrepen, K.J. (2014). Different Levels of Catabolite Repression Optimize Growth in Stable and Variable Environments. PLOS Biol. *12*, e1001764.

Nijland, J.G., Shin, H.Y., de Jong, R.M., de Waal, P.P., Klaassen, P., and Driessen, A.J. (2014). Engineering of an endogenous hexose transporter into a specific D-xylose transporter facilitates glucose-xylose co-consumption in Saccharomyces cerevisiae. Biotechnol. Biofuels *7*.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell *149*, 979–993.

Okeke, B.C., and Obi, S.K.C. (1994). Lignocellulose and sugar compositions of some agro-waste materials. Bioresour. Technol. *47*, 283–284.

Oreb, M., Dietz, H., Farwick, A., and Boles, E. (2012). Novel strategies to improve co-fermentation of pentoses with D-glucose by recombinant yeast strains in lignocellulosic hydrolysates. Bioengineered *3*, 347–351.

Osiro, K.O., Borgström, C., Brink, D.P., Fjölnisdóttir, B.L., and Gorwa-Grauslund, M.F. (2019). Exploring the xylose paradox in Saccharomyces cerevisiae through in vivo sugar signalomics of targeted deletants. Microb. Cell Factories *18*, 88.

Otto, T.D., Dillon, G.P., Degrave, W.S., and Berriman, M. (2011). RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res. *39*, e57.

Ozcan, S., and Johnston, M. (1995). Three different regulatory mechanisms enable yeast hexose transporter (HXT) genes to be induced by different levels of glucose. Mol. Cell. Biol. *15*, 1564–1572.

Ozcan, S., Dover, J., and Johnston, M. (1998). Glucose sensing and signaling by two glucose receptors in the yeast Saccharomyces cerevisiae. EMBO J. *17*, 2566–2573.

Palomino, A., Herrero, P., and Moreno, F. (2006). Tpk3 and Snf1 protein kinases regulate Rgt1 association with Saccharomyces cerevisiae HXK2 promoter. Nucleic Acids Res. *34*, 1427–1438.

Papapetridis, I., Verhoeven, M.D., Wiersma, S.J., Goudriaan, M., van Maris, A.J.A., and Pronk, J.T. (2018). Laboratory evolution for forced glucose-xylose co-consumption enables identification of mutations that improve mixed-sugar fermentation by xylose-fermenting Saccharomyces cerevisiae. FEMS Yeast Res. *18*.

Parreiras, L.S., Breuer, R.J., Narasimhan, R.A., Higbee, A.J., Reau, A.L., Tremaine, M., Qin, L., Willis, L.B., Bice, B.D., Bonfert, B.L., et al. (2014). Engineering and Two-Stage Evolution of a Lignocellulosic Hydrolysate-Tolerant Saccharomyces cerevisiae Strain for Anaerobic Fermentation of Xylose from AFEX Pretreated Corn Stover. PLOS ONE *9*, e107499.

Poltoratsky, V., Heacock, M., Kissling, G.E., Prasad, R., and Wilson, S.H. (2010). Mutagenesis dependent upon the combination of activation-induced deaminase expression and a double-strand break. Mol. Immunol. *48*, 164–170.

Raamsdonk, L.M., Diderich, J.A., Kuiper, A., Gaalen, M. van, Kruckberg, A.L., Berden, J.A., and Dam, K.V. (2001). Co-consumption of sugars or ethanol and glucose in a Saccharomyces cerevisiae strain deleted in the HXK2 gene. Yeast *18*, 1023–1033.

Reider Apel, A., Ouellet, M., Szmidt-Middleton, H., Keasling, J.D., and Mukhopadhyay, A. (2016). Evolved hexose transporter enhances xylose uptake and glucose/xylose co-utilization in Saccharomyces cerevisiae. Sci. Rep. *6*.

Reifenberger, E., Freidel, K., and Ciriacy, M. (1995). Identification of novel HXT genes in Saccharomyces cerevisiae reveals the impact of individual hexose transporters on glycolytic flux. Mol. Microbiol. *16*, 157–167.

Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., et al. (2013). An APOBEC Cytidine Deaminase Mutagenesis Pattern is Widespread in Human Cancers. Nat. Genet. *45*, 970–976.

Roney, I.J., Rudner, A.D., Couture, J.-F., and Kærn, M. (2016). Improvement of the reverse tetracycline transactivator by single amino acid substitutions that reduce leaky target gene expression to undetectable levels. Sci. Rep. *6*, 1–8.

Sakofsky, C.J., Saini, N., Klimczak, L.J., Chan, K., Malc, E.P., Mieczkowski, P.A., Burkholder, A.B., Fargo, D., and Gordenin, D.A. (2019). Repair of multiple simultaneous double-strand breaks causes bursts of genome-wide clustered hypermutation. PLOS Biol. *17*, e3000464.

Salter, J.D., Bennett, R.P., and Smith, H.C. (2016). The APOBEC Protein Family: United by Structure, Divergent in Function. Trends Biochem. Sci. *41*, 578–594.

Salusjärvi, L., Kankainen, M., Soliymani, R., Pitkänen, J.-P., Penttilä, M., and Ruohonen, L. (2008). Regulation of xylose metabolism in recombinant Saccharomyces cerevisiae. Microb. Cell Factories *7*, 18.

Sandell, L.L., and Zakian, V.A. (1993). Loss of a yeast telomere: arrest, recovery, and chromosome loss. Cell *75*, 729–739.

Santangelo, G.M. (2006). Glucose Signaling in Saccharomyces cerevisiae. Microbiol. Mol. Biol. Rev. *70*, 253–282.

Sanz, P., Alms, G.R., Haystead, T.A.J., and Carlson, M. (2000). Regulatory Interactions between the Reg1-Glc7 Protein Phosphatase and the Snf1 Protein Kinase. Mol. Cell. Biol. *20*, 1321–1328.

Scalcinati, G., Otero, J.M., Van Vleet, J.R.H., Jeffries, T.W., Olsson, L., and Nielsen, J. (2012). Evolutionary engineering of Saccharomyces cerevisiae for efficient aerobic xylose consumption. FEMS Yeast Res. *12*, 582–597.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schneper, L., Düvel, K., and Broach, J.R. (2004). Sense and sensibility: nutritional response and signal integration in yeast. Curr. Opin. Microbiol. *7*, 624–630.

Sedlak, M., and Ho, N.W.Y. (2004). Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant Saccharomyces yeast. Yeast *21*, 671–684.

Shi, K., Carpenter, M.A., Banerjee, S., Shaban, N.M., Kurahashi, K., Salamango, D.J., McCann, J.L., Starrett, G.J., Duffy, J.V., Demir, Ö., et al. (2017). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat. Struct. Mol. Biol. *24*, 131–139.

Shor, E., Fox, C.A., and Broach, J.R. (2013). The Yeast Environmental Stress Response Regulates Mutagenesis Induced by Proteotoxic Stress. PLOS Genet. *9*, e1003680.

Snoek, T., Picca Nicolino, M., Van den Bremt, S., Mertens, S., Saels, V., Verplaetse, A., Steensels, J., and Verstrepen, K.J. (2015). Large-scale robot-assisted genome shuffling yields industrial Saccharomyces cerevisiae yeasts with increased ethanol tolerance. Biotechnol. Biofuels *8*.

Sonderegger, M., and Sauer, U. (2003). Evolutionary Engineering of Saccharomyces cerevisiae for Anaerobic Growth on Xylose. Appl. Environ. Microbiol. *69*, 1990–1998.

Sonderegger, M., Jeppsson, M., Hahn-Hägerdal, B., and Sauer, U. (2004). Molecular Basis for Anaerobic Growth of Saccharomyces cerevisiae on Xylose, Investigated by Global Gene Expression and Metabolic Flux Analysis. Appl. Environ. Microbiol. *70*, 2307–2317.

Stavrou, S., and Ross, S.R. (2015). APOBEC3 Proteins in Viral Immunity. J. Immunol. *195*, 4565–4570.

Steensels, J., Snoek, T., Meersman, E., Nicolino, M.P., Voordeckers, K., and Verstrepen, K.J. (2014). Improving industrial yeast strains: exploiting natural and artificial diversity. Fems Microbiol. Rev. *38*, 947–995.

Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J., and Harris, R.S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat. Struct. Mol. Biol. *17*, 222–229.

Storici, F., and Resnick, M.A. (2003). Delitto perfetto targeted mutagenesis in yeast with oligonucleotides. Genet. Eng. (N. Y.) *25*, 189–207.

Storici, F., and Resnick, M.A. (2006). The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. Methods Enzymol. *409*, 329–345.

Storici, F., Durham, C.L., Gordenin, D.A., and Resnick, M.A. (2003). Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. Proc. Natl. Acad. Sci. U. S. A. *100*, 14994–14999.

Storici, F., Snipe, J.R., Chan, G.K., Gordenin, D.A., and Resnick, M.A. (2006). Conservative Repair of a Chromosomal Double-Strand Break by Single-Strand DNA through Two Steps of Annealing. Mol. Cell. Biol. *26*, 7645–7657.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. ELife *2*.

Temer, B., dos Santos, L.V., Negri, V.A., Galhardo, J.P., Magalhães, P.H.M., José, J., Marschalk, C., Corrêa, T.L.R., Carazzolle, M.F., and Pereira, G.A.G. (2017). Conversion of an inactive xylose isomerase into a functional enzyme by co-expression of GroEL-GroES chaperonins in Saccharomyces cerevisiae. BMC Biotechnol. *17*.

Toivari, M.H., Salusjärvi, L., Ruohonen, L., and Penttilä, M. (2004). Endogenous Xylose Pathway in Saccharomyces cerevisiae. Appl. Environ. Microbiol. *70*, 3681–3686.

Tu, J., and Carlson, M. (1995). REG1 binds to protein phosphatase type 1 and regulates glucose repression in Saccharomyces cerevisiae. EMBO J. *14*, 5939–5946.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. Nucleic Acids Res. *40*, e115.

Veras, H.C.T., Parachin, N.S., and Almeida, J.R.M. (2017). Comparative assessment of fermentative capacity of different xylose-consuming yeasts. Microb. Cell Factories *16*.

Verduyn, C., Van Kleef, R., Frank, J., Schreuder, H., Van Dijken, J.P., and Scheffers, W.A. (1985). Properties of the NAD(P)H-dependent xylose reductase from the xylose-fermenting yeast Pichia stipitis. Biochem. J. *226*, 669–677.

Visser, J.A.G.M. de, and Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. Nat. Rev. Genet. *15*, 480–490.

Wahlbom, C.F., van Zyl, W.H., Jönsson, L.J., Hahn-Hägerdal, B., and Otero, R.R.C. (2003). Generation of the improved recombinant xylose-utilizing Saccharomyces cerevisiae TMB 3400 by random mutagenesis and physiological comparison with Pichia stipitis CBS 6054. FEMS Yeast Res. *3*, 319–326.

Wallace-Salinas, V., and Gorwa-Grauslund, M.F. (2013). Adaptive evolution of an industrial strain of Saccharomyces cerevisiae for combined tolerance to inhibitors and temperature. Biotechnol. Biofuels *6*, 151.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164–e164.

Wang, L., Littlewood, J., and Murphy, R.J. (2013). Environmental sustainability of bioethanol production from wheat straw in the UK. Renew. Sustain. Energy Rev. *28*, 715–725.

Wei, S., Liu, Y., Wu, M., Ma, T., Bai, X., Hou, J., Shen, Y., and Bao, X. (2018). Disruption of the transcription factors Thi2p and Nrm1p alleviates the post-glucose effect on xylose utilization in Saccharomyces cerevisiae. Biotechnol. Biofuels *11*, 112.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).

Yanagisawa, M., Kawai, S., and Murata, K. (2013). Strategies for the production of high concentrations of bioethanol from seaweeds. Bioengineered *4*, 224–235.

Yang, Y., Sterling, J., Storici, F., Resnick, M.A., and Gordenin, D.A. (2008). Hypermutability of Damaged Single-Strand DNA Formed at Double-Strand Breaks and Uncapped Telomeres in Yeast Saccharomyces cerevisiae. PLoS Genet. *4*, e1000264.

Ye, L., Kruckeberg, A.L., Berden, J.A., and Dam, K. van (1999). Growth and Glucose Repression Are Controlled by Glucose Transport in Saccharomyces cerevisiae Cells Containing Only One Glucose Transporter. J. Bacteriol. *181*, 4673–4675.

Young, E.M., Tong, A., Bui, H., Spofford, C., and Alper, H.S. (2014). Rewiring yeast sugar transporter preference through modifying a conserved protein motif. Proc. Natl. Acad. Sci. *111*, 131–136.

Young, E.T., Dombek, K.M., Tachibana, C., and Ideker, T. (2003). Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. J. Biol. Chem. *278*, 26146–26158.

Zeng, W.-Y., Tang, Y.-Q., Gou, M., Xia, Z.-Y., and Kida, K. (2016). Transcriptomes of a xylose-utilizing industrial flocculating Saccharomyces cerevisiae strain cultured in media containing different sugar sources. AMB Express *6*.

Zha, J., Hu, M.-L., Shen, M.-H., Li, B.-Z., Wang, J.-Y., and Yuan, Y.-J. (2012). Balance of XYL1 and XYL2 expression in different yeast chassis for improved xylose fermentation. Front. Microbiol. *3*.

Zhang, G.-C., Liu, J.-J., and Ding, W.-T. (2012). Decreased Xylitol Formation during Xylose Fermentation in Saccharomyces cerevisiae Due to Overexpression of Water-Forming NADH Oxidase. Appl. Environ. Microbiol. *78*, 1081–1086.

Zhang, G.-C., Kong, I.I., Wei, N., Peng, D., Turner, T.L., Sung, B.H., Sohn, J.-H., and Jin, Y.-S. (2016). Optimization of an acetate reduction pathway for producing cellulosic ethanol by engineered yeast. Biotechnol. Bioeng. *113*, 2587–2596.

(2014). Yeast Genetics: Methods and Protocols (Humana Press).

7. CONTRIBUTIONS OF COLLABORATORS

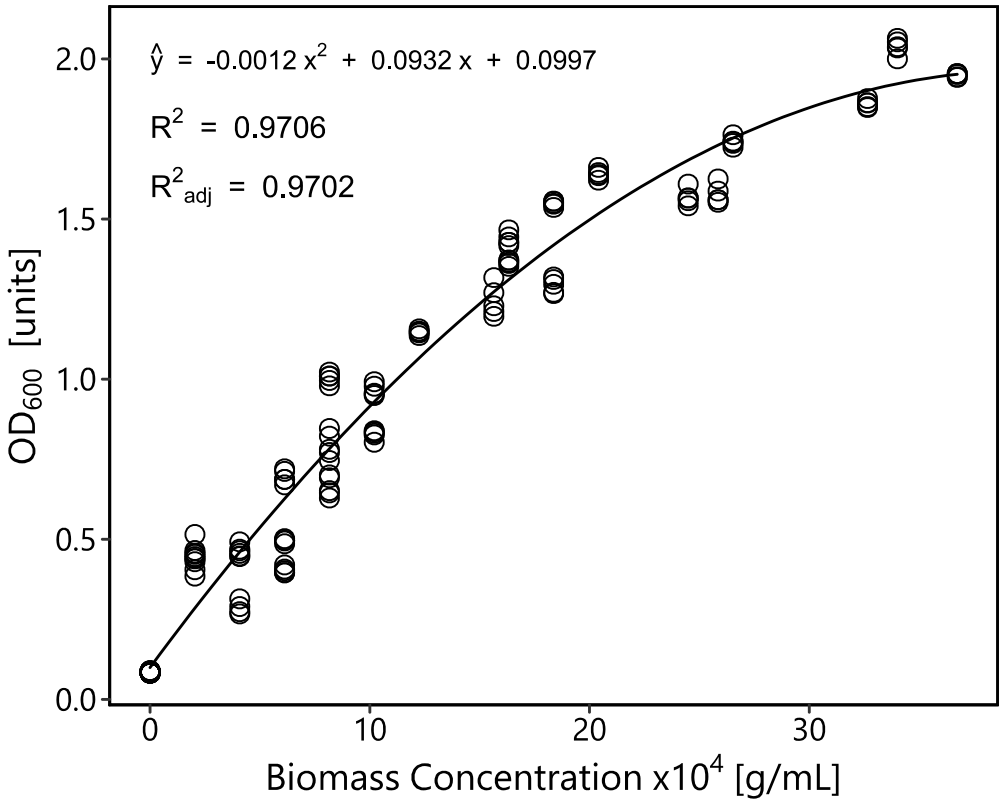
Emma Mcgurk and Ghadir Makki assisted with generating and testing of mutants. Emma Mcgurk also assisted with generation of the biomass concentration calibration curve.

8. APPENDIX

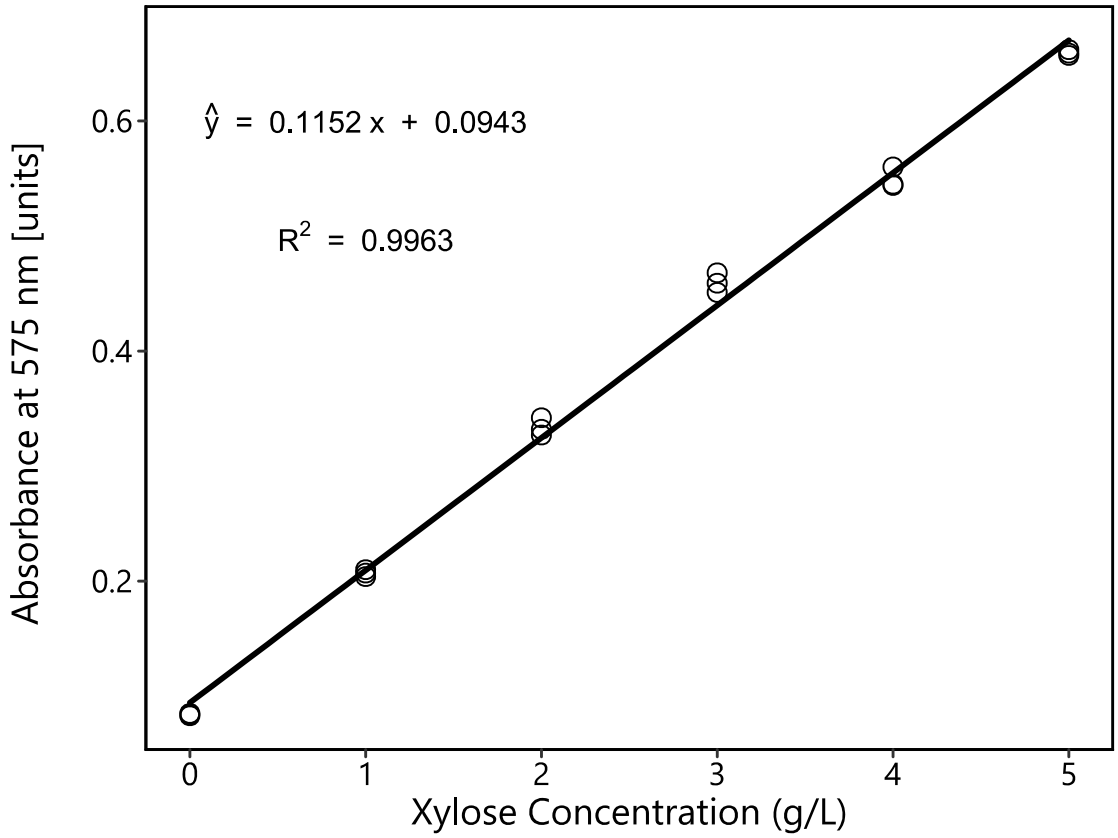
Supplementary Table 1. List of primers and other oligonucleotides used in this study. Underlines represent priming sites. Bold sequences represent the I-SceI recognition site.

Name	Sequence
oBX1	GCGGTATTTCACACCGCATAGGGTAATAACTGATATAATTAAATTGAAGCCCGCGCGTTGGCCGATTCAT
oBX2	TGCCCAGTGTTCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAAACG**TAGGGATAACAGGGTAAT**TTCGTACGCT GCAGGTCGAC
oBX3	GGCAAATTTTTCGTTCCAAG
oBX4	GAACGAAGGAAGGAGCACAG
oBX6	AACGTTTTAATTAAATGCTTTCTGAAAACACGA
oBX7	CTAATAGGCGCGCCCATCCAAAATATTAAATTTTACTTTTATTA
oBX11	TGCCCAGTGTTCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAAACGTTCGTACGCTGCAGGTCGAC
oBX13	ACTATACTTCTATAGACACGC
oBX14	GCGTGAATGTAAGCGTGAC
oBX20	GTAATCTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTTTCGTACGCTGCAGGTCGAC
oBX21	GCTCTAATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTCCGCGCGTTGGCCGATTCAT
oBX22	TTTTTGATTCCGGTTTCTTTG
oBX23	AAATCATTACGACCGAGATTCC
oBX38	TCTCTGTTTCAGGGACATCATGTTGTAAAAGAAAAAGACAGTTTAGGTAATTCGTACGCTGCAGGTCGAC
oBX39	TTAAGCACCGATGATACCAACGGACTTACCTTCAGCAATTCTTTTTTGGGCCGCGCGTTGGCCGATTCAT
oBX40	TCTCGCGGGCAGTTTTTC
oBX41	AAAAGGGCACCTTCTTGTTG
oBX42	TCTCTGTTTCAGGGACATCATGTTGTAAAAGAAAAAGACAGTTTAGGTAACCCAAAAAAGAATTGCTGAAGGTAAGTCCG TTGGTATCATCGGTGCTTAA
oBX43	TTAAGCACCGATGATACCAACGGACTTACCTTCAGCAATTCTTTTTTGGGTTACCTAAACTGTCTTTTTCTTTTACAACATGA TGTCCCTGAAACAGAGA
oBX58	AGTCACATCAAGATCGTTTATGG
oBX59	GCACGGAATATGGGACTACTTCG
oBX60	ACTCCACTTCAAGTAAGAGTTTG
oBX61	CTTTCATTAAATGTGCTCACTGTTTCTCAATCTTTATGCCATATTTTTGCTTCGTACGCTGCAGGTCGAC
oBX62	AATCATCTATCGATTAGTATTCGTACTATTGGCATGTTATCGTATACGGTCCGCGCGTTGGCCGATTCAT
oBX63	AGCTTTCATTAAATGTGCTCACTG
oBX64	CGCGACATGTATCAAATTGC
oBX65	AATGTGCTCACTGTTTCTCAATCTTTATGCCATATTTTTGC**TAGGGATAACAGGGTAAT**ACCGTATACGATAACATGCCA ATAGTACGAATACTAATCGA
oBX66	TCGATTAGTATTCGTACTATTGGCATGTTATCGTATACGGT**ATTACCCTGTTATCCCTA**GCAAAAATATGGCATAAAGATT GAGAAACAGTGAGCACATT
oBX67	ATTCGTACTATTGGCATGTTATCGTATACGGT**ATTACCCTGTTATCCCTA**CCGCGCGTTGGCCGATTCAT
oBX68	CTTTCATTAAATGTGCTCACTGTTTCTCAATCTTTATGCCATATTTTTGCAACATAGTGATTGTCAAGAGCTTAG
oBX69	ATTCGTACTATTGGCATGTTATCGTATACGGT**ATTACCCTGTTATCCCTA**TACTCCCAAGAATTAGAACAAGTAG
oBX70-1	CTTTCATTAAATGTGCTCACTGTTTCTCAATCTTTATGCCATATTTTTGCTAAGTTTGCCAAGCTGGCTG
oBX71	ATTCGTACTATTGGCATGTTATCGTATACGGT**ATTACCCTGTTATCCCTA**GAGTTCACGGGCAAGAGAAG
oBX72	**TAGGGATAACAGGGTAAT**ACCGTATACGATAACATGCCAATAGTACGAATTTCGTACGCTGCAGGTCGAC

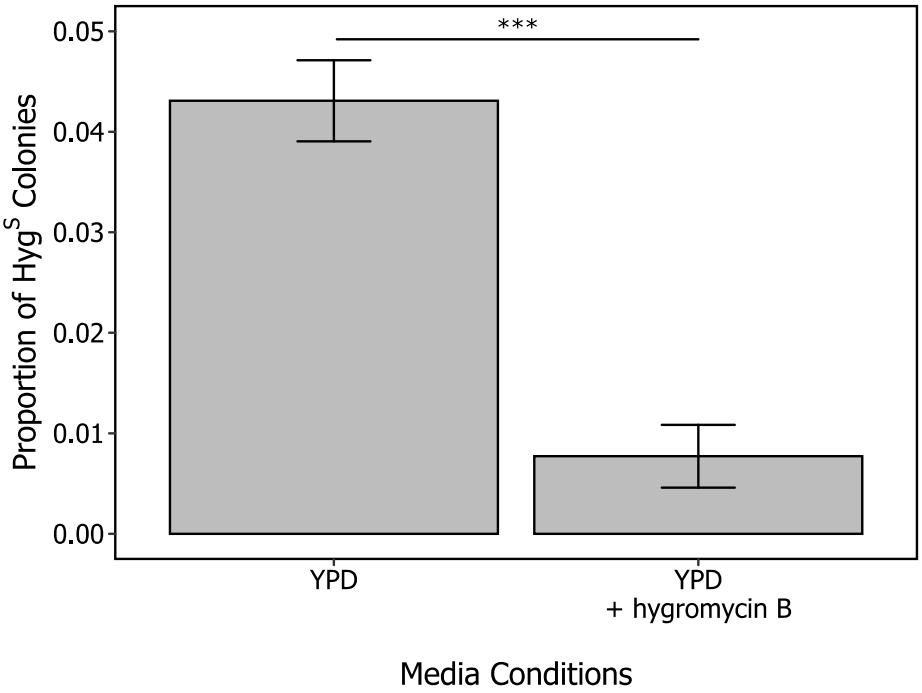
oBX73	AGTTCAAGATGATTGTTGGAATAAGAATCAACAATCATCTATCGATTAGTCCGCGCGTTGGCCGATTCAT
oBX74	ATACAGCTGGTCTGCCCAAG
oBX75	CGATTTTTACGCGTGGATTC
oBX76	**TAGGGATAACAGGGTAAT**ACCGTATACGATAACATGCCAATAGTACGAATTAAAGTAGGATTTTAACGATTCAGT
oBX77	AGTTCAAGATGATTGTTGGAATAAGAATCAACAATCATCTATCGATTAGTTTTGTTTGGCTAAGTATAGTTGGTG
oBX78	**TAGGGATAACAGGGTAAT**ACCGTATACGATAACATGCCAATAGTACGAATGGATCCATCTTTGAATGTTCATAAG
oBX79	AGTTCAAGATGATTGTTGGAATAAGAATCAACAATCATCTATCGATTAGTCGCTTATTGACATTGGCCAG
oBX80	ATGTCAACAAATCTAGCAAATTACTTCGCCGGTAAGAAAGATATTGAAAATTCGTACGCTGCAGGTCGAC
oBX81	CTAACTGCTGTCATTTCCATTTTCTTGTGGCTTGACGTCACTCTTTTTCCCCGCGCGTTGGCCGATTCAT
oBX82	TTAACACCACCTCCTGAAAGAGAAC
oBX83	ATGGAAGCTCAAGAAGTTCTTGTAC
oBX84	AAATCCTAAAGCAAGCATATTGACGAAGACGAGATAAGAAAAATCCAAAAAAGAAAGAATTTTGAAGTCAACATGAAGA GAAAATAAAAATCCAGGTAGT
oBX85	ACTACCTGGATTTTTATTTTCTCTTCATGTTGACTTCAAAATTCTTTCTTTTTTGGATTTTTCTTATCTCGTCTTCGTCAATATG CTTGCTTTAGGATTT
oKC192	AGTGGTTCGTATCGCCTAAAATCATACCAAAATAAAAAGAGTGTCTAGAAGGGTCATATACGTACGCTGCAGGTCGAC
oKC193	ATTAAAAATGACTATATATTTGATATTTATATGCTATAAAGAAATTGTACTCCAGATTTCATCGATGAATTCGAGCTCG
oKC200	ATGGCGAGTGTCACATAATAG
oKC201	CCGTGATAGTTTAATGGTCAG



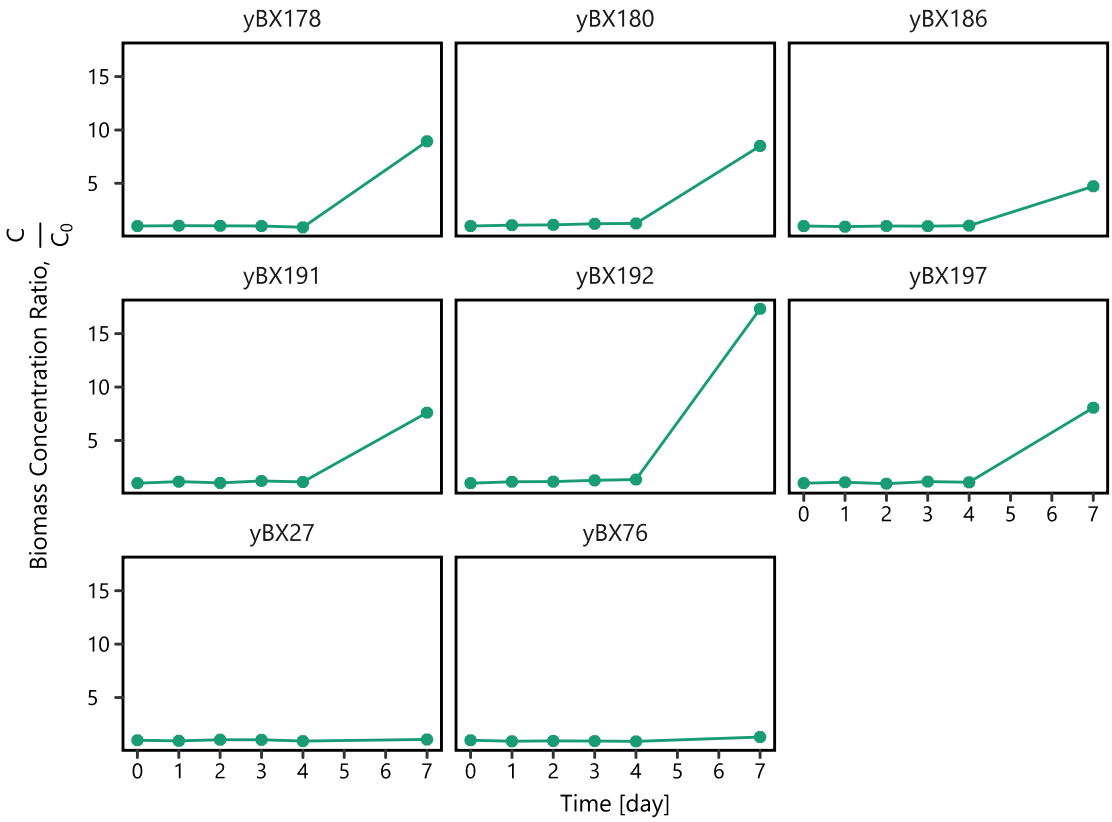
Supplementary Figure 1. Biomass concentration calibration curve. Cultures of 5 different yeast strains were grown in either YPD or 20 g/L raffinose mineral media for 3 days. Various concentrations of these cultures were made (20% v/v to 200% v/v) with sterile media from which cell pellets were collected. The pellets were dried at 105 °C and weighed. The absorbance at 600 nm was measured and plotted against biomass concentration. For this calibration curve 2 fast and 2 slow growing xylose-utilizing mutants grown in mineral media were grouped together along with yBX27 grown in mineral media and YPD for a total of n=6 biological replicates. The absorbance was measured with n=5 technical replicates and the weight with n=1 technical replicates. Each circle is 1 OD reading technical replicate. Symbols: $\hat{y}$, expected response from a quadratic fit; R^2, coefficient of determination; R^2_{adj}, adjusted coefficient of determination; OD_{600}, optical density (absorbance) at 600 nm.



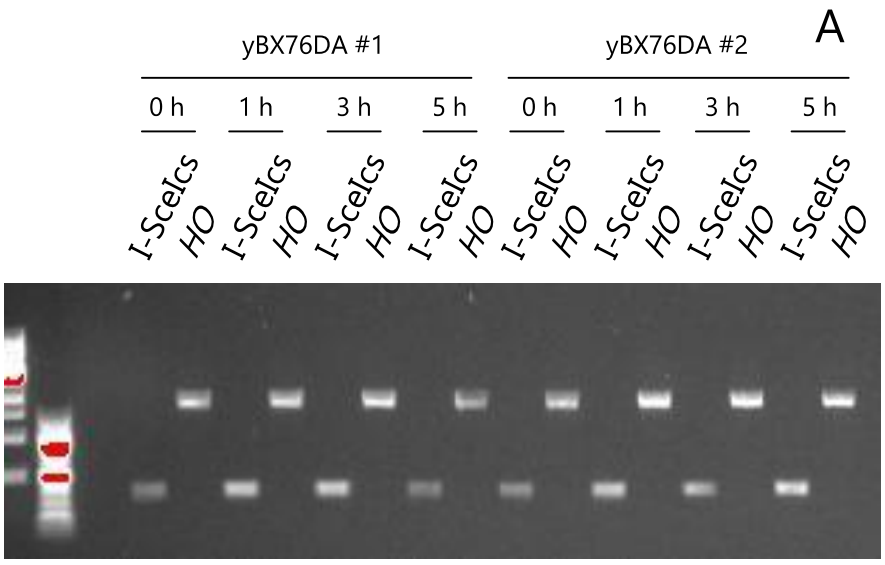
Supplementary Figure 2. DNS assay calibration curve. Various concentrations of xylose mixtures in mineral media were prepared by measuring various amounts of D-xylose using a scale and dissolving into known volumes of mineral media. The xylose concentration was assayed via the DNS method and the absorbance readings at 575 nm were plotted against the known xylose concentration. The experiment was conducted with n=3 OD reading technical replicates and each circle is one replicate. Symbols: $\hat{\boldsymbol{y}}$, expected response from a linear fit; $\boldsymbol{R^2}$, coefficient of determination.

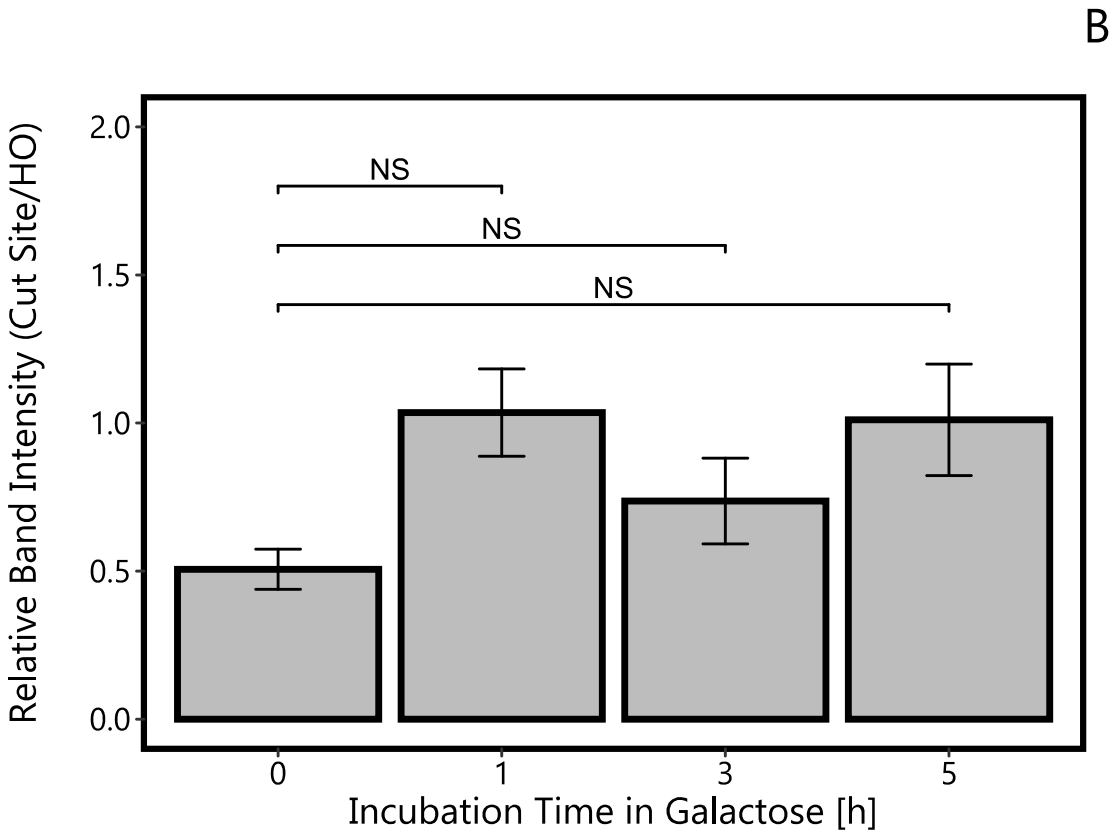


Supplementary Figure 3. Presence of the pSR435 plasmid was detectable by the increase in Hyg^S colonies in media without hygromycin B selection. Cultures of single colony isolates from ySR128DA were either grown in YPD or YPD with 300 µg/mL hygromycin B for 2 days then plated on YPD at 300 cells/plate and incubated at 30°C. Once colonies grew up, the YPD plates were replica plated onto YPD+300 µg/mL hygromycin B and incubated once more. The proportion of colonies that failed to grow on the YPD + hygromycin B replica plates was calculated for each liquid media condition. Each sample is the mean of n=6 independent colony biological replicates. Symbols: Hyg^S, Hygromycin B sensitive; ***, p<0.001.

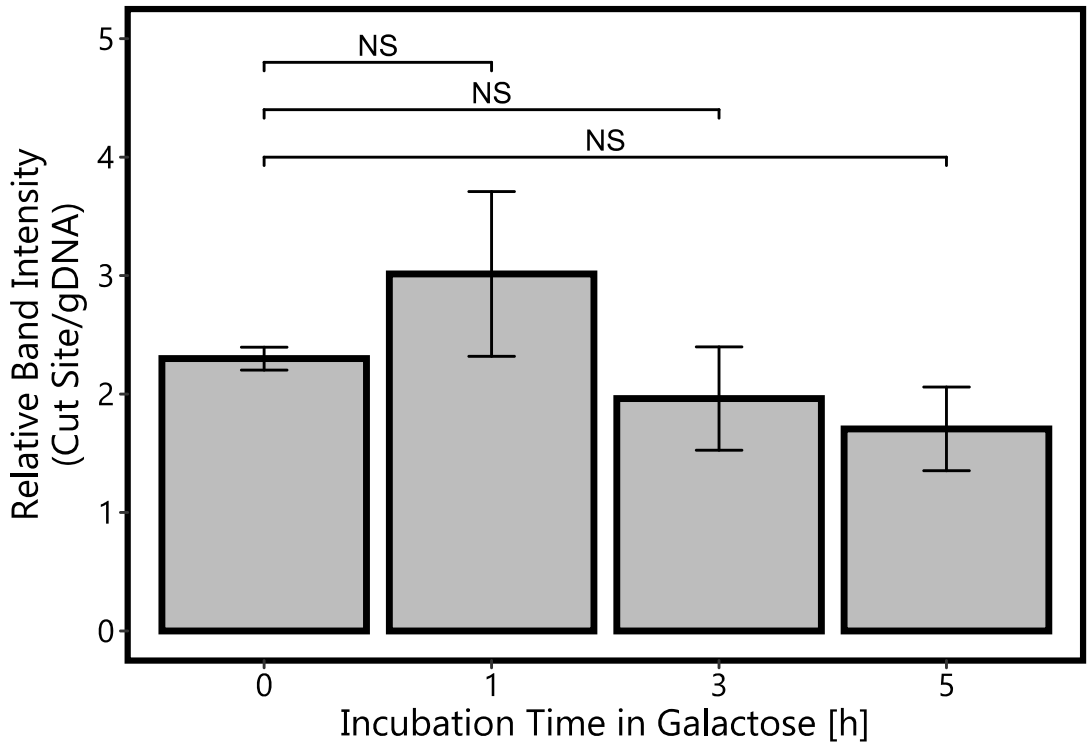


Supplementary Figure 4. Xylose-utilizing mutants resume growth when shifted from anaerobic to aerobic conditions. The same mutants as in Figure 12 that were grown in 0% oxygen were shifted to room air at the end of day 4 and grown for 3 more days.


A
yBX76DA #1
yBX76DA #2
0 h
1 h
3 h
5 h
I-SceIcs
HO


B
Relative Band Intensity (Cut Site/HO)
2.0
1.5
1.0
0.5
0.0
NS
NS
NS
0
1
3
5
Incubation Time in Galactose [h]

C



Supplementary Figure 5. No DSB induction was observed in yBX76. The DSB induction system in yBX76 was tested by PCR amplification across the DSB site to compare the abundance of uncut DNA template before and after galactose induction. Relative amplicon concentration was estimated by band intensity after gel electrophoresis. **(A)** Cultures of yBX76 grown in YPR were diluted 1:37 into YPGal and incubated for up to 5 hours. At each time point gDNA was isolated and used for PCR amplification across the cut site (I-SceIcs) and *HO* (as a reference) with primer pairs oBX74+oBX98 and oBX6+oBX7 respectively. Shown here is a representative gel with the band intensities of the cut site and *HO* for 2 out of n=3 independent colony biological replicates. **(B)** The relative band intensities from **A. (C)** The band intensities of the cut site from **A** relative to the amount of gDNA used as template for PCR. Each bar represents n=3 biological replicates. The error bars represent the SEM. Symbols: NS, not significant (p>0.05).

Supplementary Table 2. List of genes with nonsynonymous mutations in yBX102.

Gene	Base Change	Amino Acid Change
YAL068W-A	c.A245T	p.N82I
YAL068W-A	c.A248C	p.N83T
YAL068W-A	c.T249G	p.N83K
YAL068W-A	c.C251A	p.T84K
FLO9	c.T330G	p.I110M
PSK1	c.C2846T	p.S949F
AVT5	c.G1015A	p.D339N
PHO3	c.C884T	p.S295L
ALG1	c.G610A	p.E204K
TBS1	c.A2935G	p.S979G
ISW1	c.G2986A	p.E996K
HIS7	c.G1270A	p.E424K
SNF5	c.G2170T	p.A724S
POF1	c.G544A	p.E182K
PDI1	c.T1518A	p.D506E
ARE1	c.C179G	p.T60S,YCR049C
ARE1	c.C197G	p.T66S,YCR049C
YCR100C	c.G13A	p.D5N
PAU3	c.A52G	p.I18V
AAD3	c.A601G	p.I201V
COS7	c.C350T	p.S117F
YDL228C	c.C217T	p.Q73X
YDL228C	c.A205G	p.N69D
YDL228C	c.G202A	p.G68S
SEC31	c.G1819A	p.D607N
SEC31	c.G1924A	p.D642N
TSR1	c.C160T	p.R54C
PRM7	c.C634T	p.P212S
PRM7	c.G628A	p.V210I
PRM7	c.T625A	p.S209T
PRM7	c.A604G	p.S202G
PRM7	c.A599C	p.Q200P
BSC1	c.C644G	p.S215C
BSC1	c.T643A	p.S215T
BSC1	c.T598A	p.C200S
ENA5	c.T3175C	p.S1059P
NVJ3	c.C1021T	p.Q341X
PLP1	c.G634A	p.E212K

SIR4	c.C2984T	p.S995L
YDR246W-A	c.A161G	p.D54G
HXT7	c.G950C	p.S317T
HXT7	c.G947A	p.S316N
HXT7	c.A946G	p.S316G
SVF1	c.G1144A	p.E382K
PAL1	c.G473A	p.R158K
UTP5	c.C1583T	p.S528F
CYM1	c.C1897T	p.R633C
TOM1	c.C1955A	p.T652K
FIT1	c.A679G	p.T227A
FIT1	c.T611C	p.V204A
STL1	c.C1646T	p.S549L
PAD1	c.C703T	p.R235C
YDR543C	c.C281T	p.S94F
YIL169C	c.A563T	p.D188V
MLP2	c.C3742T	p.Q1248X
ASG1	c.A2545G	p.N849D
YIR042C	c.G547A	p.V183M
YIR042C	c.T545C	p.V182A
YFL067W	c.T83C	p.I28T
BUD27	c.T1750A	p.Y584N
DCV1	c.A448G	p.I150V
YGL260W	c.T62C	p.I21T
YGL260W	c.A74G	p.H25R
RCK1	c.G995A	p.R332Q
PRM8	c.A25G	p.N9D
PRM8	c.G55A	p.G19S
PRM8	c.T152C	p.V51A
PRM8	c.C157A	p.L53I
PRM8	c.A184G	p.I62V
PRM8	c.C188T	p.S63L
PRM8	c.A190T	p.I64L
PRM8	c.C199T	p.L67F
PRM8	c.C205A	p.H69N
MSB2	c.A1135G	p.T379A
MSB2	c.G1138C	p.A380P
MSB2	c.A1142G	p.N381S
ROM1	c.A2555C	p.H852P
TPO2	c.A511G	p.S171G
YGR174W-A	c.C35T	p.A12V

CPD1	c.C538T	p.Q180X
MTM1	c.G1028C	p.R343P
NPR3	c.T2324C	p.I775T
YHR028W-A	c.T206A	p.F69Y
YHR054C	c.A427G	p.I143V
RSC30	c.G1888A	p.V630I
SSF1	c.C862T	p.P288S
STB5	c.G1738A	p.E580K
OYE2	c.G1004A	p.R335K
GPI16	c.C371T	p.S124L
FLO5	c.G2679A	p.W893X
VTH2	c.C4449G	p.F1483L
GEA1	c.G1023A	p.M341I
DAN4	c.G3010A	p.V1004I
DAN4	c.G814C	p.A272P
YJR162C	c.C55G	p.H19D
TRP3	c.G1043A	p.S348N
TOR2	c.G3429A	p.M1143I
PTK1	c.T1855C	p.S619P
PTK1	c.C1852A	p.P618T
HAP4	c.A512T	p.Y171F
ELM1	c.G877A	p.D293N
YLR001C	c.G2540A	p.R847Q
SLS1	c.G1306T	p.A436S
RRN5	c.C943G	p.Q315E
RRN5	c.C961G	p.Q321E
CSC1	c.G1023A	p.M341I
CDC25	c.C2270T	p.A757V
YLR464W	c.A30C	p.R10S
RSE1	c.T2423A	p.I808K
SPT5	c.C3095T	p.S1032F
YAP1	c.G271A	p.E91K
ALD3	c.G835A	p.E279K
YMR173W-A	c.T71C	p.M24T
CTL1	c.G631A	p.D211N
RPL20A	c.A350G	p.D117G
CUE1	c.G489A	p.M163I
CAT8	c.G2995A	p.D999N
RPD3	c.C860T	p.S287F
PHA2	c.C533T	p.S178L
RPS19B	c.G201A	p.M67I

PIK1	c.C707T	p.S236F
MRPL17	c.C836T	p.A279V
NST1	c.G958A	p.A320T
YNL019C	c.C664T	p.P222S
PHO91	c.G1062A	p.M354I
YNR014W	c.C287T	p.S96L
AGA1	c.T710C	p.L237S
AGA1	c.A746T	p.Q249L
TIR2	c.G518C	p.S173T
YOR034C-A	c.C176T	p.S59F
YOR053W	c.G95T	p.R32L,VHS3
YOR053W	c.C205T	p.H69Y,VHS3
YOR053W	c.C214T	p.R72C,VHS3
VHS3	c.G1678A	p.E560K
VHS3	c.C1082G	p.S361C
NOB1	c.G578A	p.G193E
RPB2	c.G982A	p.E328K
YOR200W	c.A383G	p.N128S
CIN1	c.C311T	p.S104L
PBI1	c.G1072A	p.E358K
GPI2	c.G79A	p.E27K
LCL1	c.T89C	p.V30A
TPO3	c.C130T	p.Q44X
TDA6	c.G780T	p.K260N
TDA6	c.G1021C	p.E341Q
YPR204W	c.G1642A	p.D548N
YPR204W	c.T1712C	p.I571T

Supplementary Table 3. List of genes with nonsynonymous mutations in yBX128.

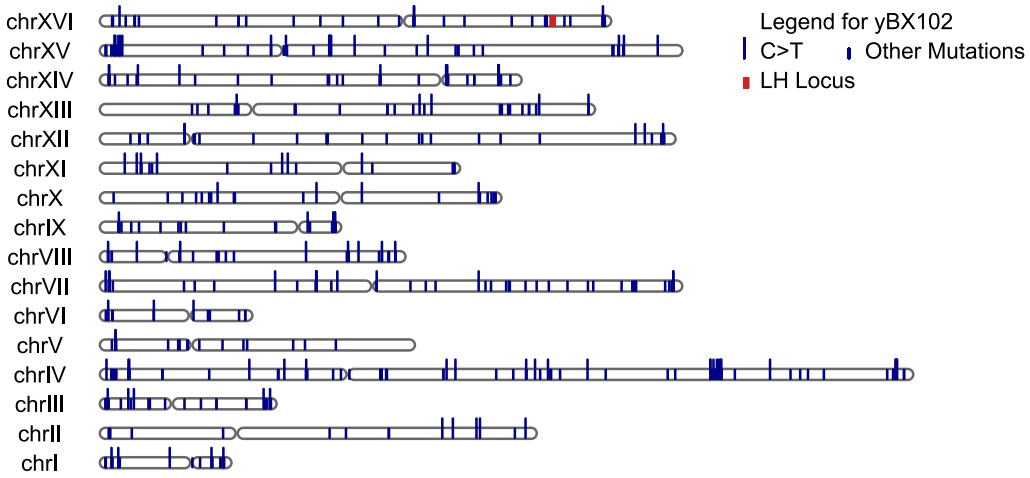
Gene	Base Change	Amino Acid Change
FLO9	c.C2791T	p.P931S
FLO9	c.A2551G	p.I851V
FLO9	c.A2509G	p.I837V
FLO9	c.T330G	p.I110M
YAL004W	c.T271A	p.S91T
PRM9	c.A697G	p.K233E
PRM9	c.A698G	p.K233R
PRM9	c.A702T	p.E234D
PRM9	c.T703C	p.S235P
PRM9	c.T706C	p.S236P
PRM9	c.G709A	p.A237T
PRM9	c.A722C	p.K241T
PRM9	c.A850G	p.N284D
HSL7	c.G1012T	p.D338Y
YBR138C	c.G910A	p.D304N
TBS1	c.A2935G	p.S979G
YBR196C-A	c.C112T	p.R38C
DUG2	c.C1708T	p.Q570X
PDI1	c.T1518A	p.D506E
GID7	c.C1268T	p.S423L
TAF2	c.G974A	p.G325E
ARE1	c.C197G	p.T66S,YCR049C
FIG2	c.T1480C	p.S494P
PAU3	c.A52G	p.I18V
AAD3	c.A601G	p.I201V
COS7	c.C350T	p.S117F
BRE4	c.G2827A	p.E943K
PTP1	c.C442G	p.R148G
YDL228C	c.C217T	p.Q73X
YDL228C	c.A205G	p.N69D
YDL228C	c.G202A	p.G68S
HO	c.G1157A	p.R386K
HO	c.G1018A	p.E340K
CDC13	c.G1988A	p.R663Q
MTF2	c.C259T	p.Q87X
MTF2	c.T252G	p.N84K
PRM7	c.T688G	p.S230A
PRM7	c.C634T	p.P212S

PRM7	c.G628A	p.V210I
PRM7	c.T625A	p.S209T
PRM7	c.A604G	p.S202G
PRM7	c.A599C	p.Q200P
BSC1	c.C644G	p.S215C
BSC1	c.T643A	p.S215T
SLM3	c.G424A	p.E142K
GAL3	c.C1364T	p.S455F
ENA5	c.T3175C	p.S1059P
YDR094W	c.C131G	p.S44X
GIS1	c.C560G	p.S187X
YDR132C	c.G1081A	p.E361K
ENT5	c.C770T	p.S257F
SSY1	c.G1060A	p.V354I
RAD9	c.C1703T	p.S568L
YDR246W-A	c.A161G	p.D54G
OPI7	c.C44T	p.S15L,EAF1
NCB2	c.G127A	p.E43K,YDR396W
RPB7	c.G496A	p.D166N
HKR1	c.T2323C	p.S775P
YDR426C	c.21_24del	p.I7fs
FIT1	c.T611C	p.V204A
VTH1	c.C4449G	p.F1483L
ASG1	c.A2545G	p.N849D
YIR019C	c.C3320T	p.S1107F
YIR019C	c.A1699G	p.T567A
YIR042C	c.C568A	p.R190S
YIR042C	c.G547A	p.V183M
YIR042C	c.T545C	p.V182A
SER3	c.G51C	p.M17I
UTP7	c.C278G	p.S93W
ILV1	c.G150C	p.L50F
DOT6	c.C26G	p.S9X
PTC2	c.C748G	p.H250D
SPT15	c.G388A	p.D130N
YFL067W	c.C488T	p.A163V
YFL067W	c.C518T	p.A173V
MIL1	c.G433A	p.E145K
BUD27	c.T1750A	p.Y584N
DCV1	c.A448G	p.I150V
IRC7	c.G976A	p.V326I

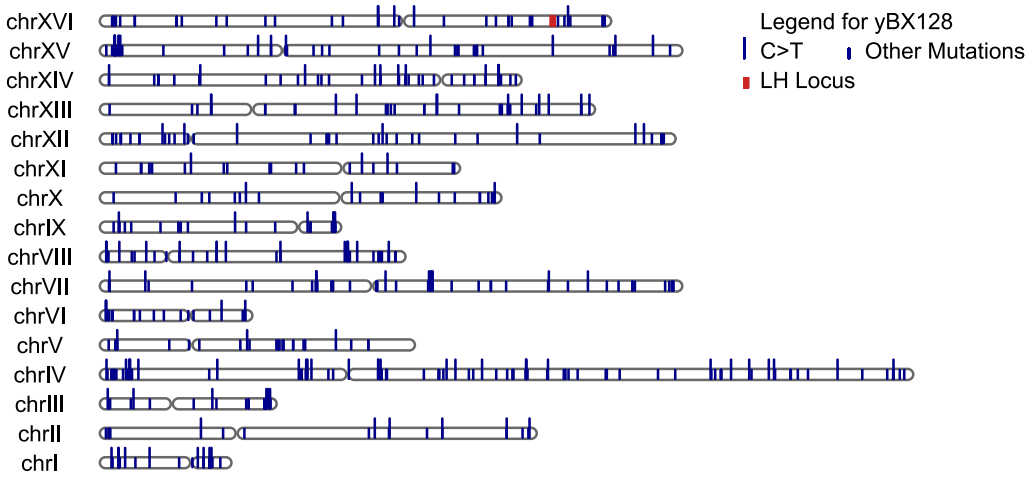
YGL220W	c.G241A	p.G81S
PRM8	c.A25G	p.N9D
PRM8	c.G55A	p.G19S
PRM8	c.T152C	p.V51A
PRM8	c.C157A	p.L53I
PRM8	c.A184G	p.I62V
PRM8	c.C188T	p.S63L
PRM8	c.C199T	p.L67F
PRM8	c.C205A	p.H69N
MSB2	c.A1135G	p.T379A
MSB2	c.G1138C	p.A380P
MSB2	c.A1142G	p.N381S
MSB2	c.G1439A	p.S480N
MSB2	c.G1450A	p.V484I
MSB2	c.A1507T	p.T503S
COX18	c.G950C	p.X317S
SPT4	c.G10A	p.E4K
YGR066C	c.G3A	p.M1I
YGR067C	c.G275C	p.R92T
TPO2	c.A511G	p.S171G
YGR174W-A	c.C35T	p.A12V
MTM1	c.G1028C	p.R343P
MTM1	c.C1027G	p.R343G
YHL050C	c.G782A	p.S261N
MUP3	c.C482T	p.A161V
OTU2	c.G163A	p.D55N
YHR028W-A	c.T206A	p.F69Y
YHR052W-A	c.C47T	p.T16I
SSF1	c.C862T	p.P288S
YHR177W	c.C476T	p.S159L
OYE2	c.C416T	p.S139F
YHR210C	c.T392A	p.L131Q
VTH2	c.C4449G	p.F1483L
MDE1	c.G499A	p.E167K
GRR1	c.G636A	p.M212I
DAN4	c.A723T	p.Q241H
DAN4	c.A722C	p.Q241P
COS5	c.C646A	p.Q216K
COS5	c.T644C	p.V215A
PTK1	c.T1855C	p.S619P
PTK1	c.C1852A	p.P618T

HAP4	c.A512T	p.Y171F
UTP30	c.G268A	p.D90N
YLR124W	c.T247A	p.C83S
SLS1	c.G1306T	p.A436S
RRN5	c.C943G	p.Q315E
RRN5	c.C961G	p.Q321E
TOS4	c.A175G	p.S59G
MDL1	c.C979T	p.H327Y
ENT2	c.A1625C	p.Q542P
COS3	c.A971G	p.K324R
RSE1	c.T2423A	p.I808K
SOK2	c.T251C	p.L84P
SEG1	c.C1060T	p.Q354X
YMR173W-A	c.T41C	p.I14T
YMR173W-A	c.T71C	p.M24T
RPL20A	c.A350G	p.D117G
RPL20A	c.A242T	p.K81M
CAT8	c.G2995A	p.D999N
HSH155	c.C893T	p.S298L
YMR317W	c.T1207G	p.S403A
YMR317W	c.C1208T	p.S403L
YNL337W	c.T52G	p.S18A
SEC21	c.G1312A	p.E438K
BNI1	c.C3721T	p.P1241S
MPA43	c.1617delT	p.F539fs
SRV2	c.G937A	p.E313K
DGR1	c.G86A	p.R29Q
MSK1	c.G1602A	p.M534I
YNL033W	c.C664T	p.P222S
YNL019C	c.C664T	p.P222S
SSK2	c.G247C	p.V83L
AGA1	c.T710C	p.L237S
AGA1	c.A746T	p.Q249L
TLG2	c.G460A	p.E154K
TIR4	c.T538A	p.S180T
TIR4	c.T548C	p.V183A
TIR4	c.G550A	p.A184T
TIR2	c.G518C	p.S173T
YOR034C-A	c.C176T	p.S59F
VHS3	c.C1947A	p.D649E,YOR053W
IAH1	c.G170C	p.R57T

YOR200W	c.A383G	p.N128S
RPA190	c.G1831A	p.E611K
MRX4	c.G457A	p.E153K
TAF14	c.G290A	p.G97D
LCL1	c.T89C	p.V30A
LSP1	c.G3A	p.M1I



Supplementary Figure 6. Genomic distribution of mutations acquired in yBX102. A total of 633 mutations were mapped onto the nuclear chromosomes. The blue lines extending above the chromosomes represent the position of C to T transitions. Blue lines of the same width as the chromosome represent the position all other mutations. The red region represents the position of *XYL1* and *XYL2* (LH locus). The width of the blue lines represents about 3,000 bp.



Supplementary Figure 7. Genomic distribution of mutations acquired in yBX128. A total of 767 mutations were mapped onto the nuclear chromosomes. The blue lines extending above the chromosomes represent the position of C to T transitions. Blue lines of the same width as the chromosome represent the position all other mutations. The red region represents the position of *XYL1* and *XYL2* (LH locus). The width of the blue lines represents about 3,000 bp.