Evaluation of three heterologous xylose isomerase genes for the fermentation of plant biomass to bioethanol in *Saccharomyces cerevisiae*

Evan J. Robinson

Submitted in fulfillment of the requirements for the degree of

Master of Science

at the

Department of Molecular and Cell Biology

Faculty of Science

University of Cape Town

South Africa

February 2008

Evaluation of three heterologous xylose isomerase genes for the fermentation of plant biomass to bioethanol in *Saccharomyces cerevisiae*

Evan J. Robinson, Department of Molecular and Cell Biology at the University of Cape Town, 2008

Abstract: Diminishing oil reserves are causing scientists to explore renewable and sustainable replacement fuel sources. One candidate is plant biomass, although without modification its energy-rich sugars remain unavailable to the standard industrial yeast *Saccharomyces cerevisiae* for fermentation to the fuel ethanol. In this study, three xylose isomerase genes originating from *Haemophilus influenzae* Rd K20, *Arabidopsis thaliana* and *Bacteroides thetaiotaomicron* VPI-5482 were expressed in *S. cerevisiae* to produce a strain capable of metabolising D-xylose (a biomass constituent). Resulting strains were analysed for protein production, growth differences and xylose isomerisation but no introduced characteristics were detected. RT-PCR suggested transcription occurred for all the genes tested but no recombinant protein nor any xylose isomerase activity was detected. An *in silico* bioinformatic analysis showed a putative inhibitory stem-loop structure in the mRNA containing the *B. thetaiotaomicron* gene which may have reduced translation. Otherwise non-functional folding or undetectable activity were concluded as the results of the expression for all genes.

1. INTRODUCTION

1.1 Incorporating high ethanol production into yeast grown on xylose

Dwindling petroleum reserves, concern over their use and increased global warming have caused a surge of interest in the use of sustainable alternative energy sources. Large amounts of renewable energy (estimated at 10^{20} Joules per annum)[1] exist relatively untapped in the form of plant biomass. Biomass has the potential for being fermented to ethanol or chemically converted to its counterpart biodiesel. Ethanol is a readily usable replacement for conventional fuels as it can power vehicles and generators directly or be blended with other fuels[2]. Ethanol is produced industrially by the fermentative yeast *Saccharomyces cerevisiae*. This yeast is chosen for fermentation over bacterial counterparts as it exhibits robust growth on various media, resistance to contamination, better growth at low pH and has a thicker cell wall and larger cell size[3].

The fermentation of biomass to ethanol is, however, currently not economically viable as the hemicellulose fraction of plant carbohydrates cannot be metabolised by *S. cerevisiae*. Wheat biomass, for instance, consists of 25-30% hemicellulose by mass[2], with arabinoxylans

forming up to 81%[4] of that fraction. Arabinoxylan (Fig 2) contains approximately two thirds xylose and one third arabinose[5]. It thus follows that the majority of studies have focused on modifying *S. cerevisiae* to metabolise xylose. Yeasts have the ability to metabolise xylulose, an isomer of xylose, by assimilation into the pentose phosphate pathway (PPP). Both D-xylose and L-arabinose enter the oxidative phase of the PPP via D-xylulose-5-phosphate[3] and are processed into a variety of downstream intermediates by the relevant metabolic pathways (Fig 1).

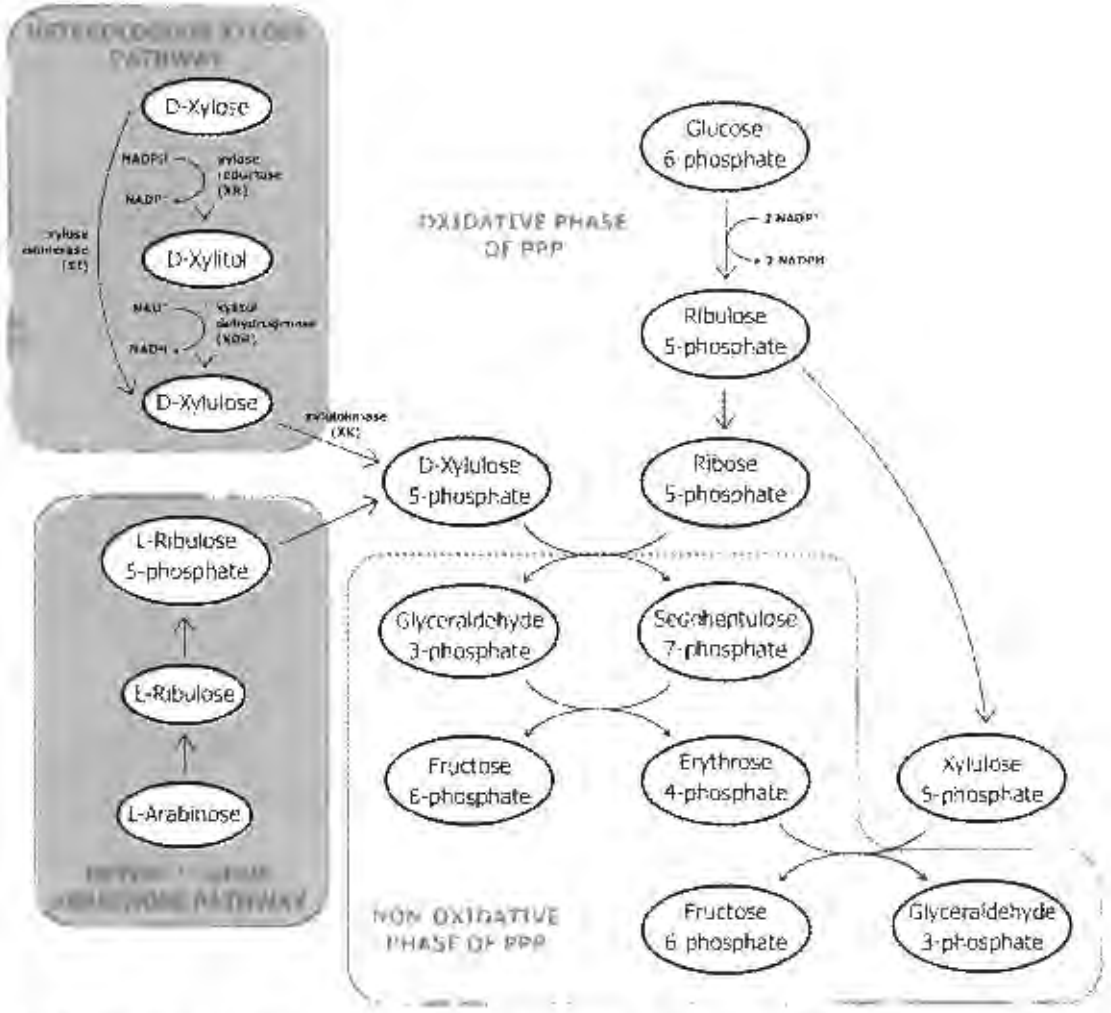


Figure 1: The heterologous xylose and arabinose pathways and their connection to the PPP. Note xylose isomerisation to xylulose occurring via either the XR-XDH enzyme pair or the XI enzyme. Image drawn using Macromedia Flash 4 (Adobe Systems, CA, USA).

The majority of attempts at engineering xylose metabolism for biofuel production have therefore focused on implementing the isomerisation of xylose to xylulose. This can be achieved and optimised in a variety of ways, including metabolic analyses, tuning cellular pathways, evolutionary engineering, transport optimisation and heterologous gene expression, each of which will be discussed.

1.2 Approaches

1.2.1 Modelling cellular metabolism *in silico*

A number of studies have effectively used predictive mathematical models to represent the cellular machinery, confirming hypotheses created *in silico* with data from experiments. Jin and Jeffries[6] tested several environmental extremes on a mathematical model of wild-type yeast metabolism, augmented with heterologous xylose reductase (XR), xylitol dehydrogenase (XDH) and xylulokinase (XK) genes from the xylose fermenting yeast *Pichia stipitis*. A digital flux balance analysis (FBA) showed that ethanol production from xylose would be maximised under oxygen limitation. Fermentation confirmed this result as well as similar predictions relating to ethanol production at different oxygen levels. Unfortunately FBA was unable to predict that greater xylitol production would accompany the increased ethanol production at low oxygen levels.

A study comparing the importance of redox co-factors in two recombinant yeast strains found that mathematical metabolic analysis provided more useful data when augmented with ^{13}C-labelled glucose tracking[7]. Two *S. cerevisiae* strains were used, TMB3001 and CPB.CR4, both expressing XR, XDH and XK from *P. stipitis*, though in the latter strain, *GDH1* was replaced with *GDH2*. While *GDH1* and *GDH2* both encode glutamate dehydrogenases of the ammonia assimilation pathway, Gdh1p is NADPH-dependent whereas Gdh2p is NADH-dependent. Replacing *GDH1* with *GDH2* therefore increases $NADH+H^+$ consumption at the expense of $NADPH+H^+$ consumption. The group was able to explain that the 25% improvement in ethanol production observed for strain CPB.CR4[8] was due to a partial shift in the cofactor use of XR from $NADPH+H^+$ to $NADH+H^+$. This flexibility, as well as expression of *GDH2*, would have increased reoxidation of $NADH+H^+$ (produced by XDH) and increased the flux from xylose to xylulose.

Mathematical models on the genome scale were used by Bro *et al.*[9] to reduce glycerol production and increase ethanol production. Using linear programming to optimise 584 metabolites and 1175 reactions, they predicted that expression of a heterologous non-phosphorylating $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPDH) would increase ethanol production by decreasing glycerol production during growth on glucose. Experiments with the wild-type strain showed a 40% decrease in glycerol production resulting in a 3% increase in ethanol production. The increase rose to 25% when a xylose-fermenting reference strain was grown on xylose and glucose mixed media. The

heterologous GAPDH complemented the activity of endogenous GAPDH and phosphoglycerate kinase (PGK) in the conversion of glyceraldehyde 3-phosphate to 3-phosphoglycerate. With its preference for $NADP^+$, incorporating the additional GAPDH decreased the amount of reduced NAD^+ formed during the production of 3-phosphoglycerate. NAD^+ resulting from the conversion of acetaldehyde to ethanol then accumulated to drive XDH. At the same time, the increased level of available $NADPH+H^+$ would have driven XR, accelerating the production of xylulose.

1.2.2 Experimental tracking of carbon metabolites

A more experimental approach that is effective in combination with computerised models is studying the flux of carbon substrates through the cellular machinery as they become downstream intermediates. Work involving two strains of *S. cerevisiae* engineered to contain *P. stipitis* XR in equal copy numbers, but different copy numbers of XDH, used HPLC to show how aeration was a contributing factor in producing xylulose from xylose without producing xylitol as a byproduct[10]. The study showed that increased aeration favoured the accumulation of xylulose instead of xylitol during xylose fermentation. Increased XDH activity could reduce xylitol formation; however the flux from xylitol to xylulose did not increase in proportion to XDH activity, indicating that the cells' physiological state may have had an influence.

Exact metabolic flux ratios across glycolysis, the PPP, the tricarboxylic acid pathway and one-carbon metabolism were determined with high reliability and high sensitivity[11]. The study made use of fractionally ^{13}C-labelled amino acids (for assimilation and incorporation during metabolism) followed by NMR of dried cell hydrolysates (for post-growth analysis). This method also revealed the entire network of biosynthetic genes required for growth on glucose media as well as the flux ratios between cellular compartments for several compounds. Four of these authors then used this technique to probe the differences in central carbon metabolism between *S. cerevisiae* and *P. stipitis*[12]. Their study showed minimal use of the non-oxidative PPP by *S. cerevisiae* in comparison to *P. stipitis*. This lead to their investigation of the effects of phosphoglucose isomerase (PGI) deletion in *S. cerevisiae*. This enzyme converts glucose-6-phosphate to fructose-6-phosphate and is part of the glycolytic pathway. Knockouts are unable to survive on glucose because metabolism is forced to occur via the PPP. This results in the accumulation of $NADPH+H^+$ that is not reoxidised. Incorporating a soluble transhydrogenase from *Escherichia coli* transferred electrons from $NADPH+H^+$ to NAD^+ resulting in a growth rate half that of the wild-type

strain. Essentially they showed that *S. cerevisiae* can be modified to resemble *P. stipitis* in its metabolism of carbon substrates, this being beneficial for xylose assimilation.

Jin *et al.*[13] showed that by using enzymatic assays to monitor the flux through *P. stipitis* XR and XDH in *S. cerevisiae*, they could optimise the expression of a similarly derived XK for optimum ethanol production and growth on xylose. They showed that uncontrolled high XK activity results in reduced growth and ethanol production on xylose. Using a vector expressing variable copy numbers, they found that an average of four copies of XK per haploid genome increased the efficiency of ethanol production from xylose.

1.2.3 Gene expression analysis

The development of high–throughput analysis of protein and mRNA expression data is allowing a deeper insight into engineering *S. cerevisiae* to ferment xylose. Analysing cDNA microarray data, Sonderegger *et al.*[14] examined differences in the global expression profiles of 6383 genes between *S. cerevisiae* strain TMB3001 and an evolved derivative strain, C1. TMB3001 and C1 both incorporated *P. stipitis* XR, XDH and XK but only the C1 strain showed effective anaerobic growth on xylose alone. The C1 strain showed notably higher transcript levels and rates of flux through central carbon metabolism - especially the xylulose-utilising PPP and glycerol pathways. Increased glycerol production would have consumed greater amounts of cytosolic $NADH+H^+$. Analysing the redox metabolism genes showed that increasing both cytosolic $NADPH+H^+$ formation and $NADH+H^+$ consumption resulted in higher rates of flux from xylose to xylulose.

In a more focused study, Jin *et al.*[15] probed the expression of several metabolic pathway genes in *S. cerevisiae* strain YSX3 expressing *P. stipitis* XR, XDH and XK. They found that on glucose or xylose, with or without aeration, the expression of glycolytic, fermentative and PPP genes did not change significantly. However the expression of genes of the tricarboxylic acid (TCA) cycle, as well as the regulatory genes *HAP4* and *MTH1*, increased significantly in cells grown on xylose. This increase was exaggerated by oxygen limitation, suggesting that *S. cerevisiae* may have trouble recognising xylose as a fermentable carbon source. Increased transcripts of NAD^+/NADH shuttling genes were also noted, presumably to address the redox imbalance caused by XR and XDH. Since the TCA cycle was upregulated in recombinant *S. cerevisiae* grown on xylose, metabolism was assumed to be oxidative. To redirect metabolism from respiration toward ethanol production, a petite mutant was produced. Although the resulting strain produced more ethanol, it was unable to grow on

xylose. An identical comparative study by Salusjarvi *et al.*[16] reached similar conclusions, noting that expression of genes for NADPH+H^+ regeneration, acetyl-coA synthesis and sterol biosynthesis were all upregulated. Interestingly, a number of the genes upregulated during growth on xylose (in place of glucose) matched those usually expressed during the starvation response.

1.2.4 Addressing xylose transport

Xylose can only be used as a substrate if its transport into the cell takes place at a rate sufficient for metabolism. *S. cerevisiae* has been shown to transport xylose across the plasma membrane as effectively as glucose at low glucose concentrations (0.05%)[17]. Xylose transport decreases significantly at higher glucose concentrations (2%). Radio-labelled xylose was used to demonstrate competitive transport by two glucose transporters: one a high-affinity and the other a low-affinity transporter. At low glucose concentrations the V_{max} of xylose transport exceeded the V_{max} previously reported for xylitol production by XR, thus transport of xylose in these conditions may not be a limiting factor.

The next two studies discovered uncharacterised fungal sugar symporters that will likely prove useful in improving xylose transport at low concentrations. Gárdonyi *et al.*[18] evaluated xylose transport in several natural xylose-assimilating yeasts (excludes *S. cerevisiae*), focusing on the strain showing the highest capacity transport: *Candida intermedia* PYCC 4715. Two transport systems were found: a high-affinity xylose/ H^+ symporter that was more repressed by glucose than by xylose and a less specific low-affinity transport system that operated through facilitated-diffusion. Both transporters also accepted glucose. Expression of both transporter genes in *S. cerevisiae* showed that the V_{max} values for glucose and xylose transport by the xylose/H^+ symporter were an order of magnitude lower than that of the facilitated-diffusion transporter[19]. However expressing the transporters was not enough to sustain vigorous growth of *S. cerevisiae* on xylose.

Saloheimo *et al.*[20] recently reported the discovery of a xylose-specific transporter from the filamentous fungus *Trichoderma reesei* after screening a cDNA library expressed in *S. cerevisiae*. The transporter proved as effective (after optimisation over extended growth) as the yeast hexose transporters Hxt1p, Hxt2p, Hxt4p and Hxt7p at transporting xylose, the hexose transporters being tested individually.

1.2.5 Reduction of the redox imbalance

Problems with an incompatibility of redox co-factors arise when XR and XDH genes are expressed simultaneously (Fig 1), leading to the accumulation of xylitol and an impaired production of ethanol[21]. XR oxidises abundant cytoplasmic $NADPH+H^+$ to convert xylose to xylitol with the resulting $NADP^+$ being reduced by the PPP. XDH however reduces NAD^+ to convert xylitol to xylulose – this reaction is energetically unfavourable and reduces the overall flux from xylose to xylulose.

Jeppsson *et al.*[22] mutated *P. stipitis* XR to preferentially use $NADH+H^+$ instead of $NADPH+H^+$. This modification successfully increased anaerobic production of ethanol and CO_2 while reducing xylitol formation and maintaining similar rates of biomass formation. Similarly, Watanabe *et al.* changed the coenzyme specificity of XDH from NAD^+ to $NADP^+$ by site-directed mutagenesis[23], later achieving similar results to Jeppson *et al.* with mutagenesis of XR[24].

Natural xylose fermenting yeast strains have been shown to sustain their redox imbalance with either a non-specific aldose reductase that uses $NADH+H^+$ as well as $NADPH+H^+$ (in the case of *P. stipitis*)[25], or an $NADP^+$-utilising glyceraldehyde-3-phosphate dehydrogenase, in the case of *Kluyveromyces lactis*[26]. A cDNA library of the yeast *Ambrosiozyma monospora* was found to contain a putative NADH-dependent L-xylulose reductase (normally NADPH-dependent)[27]. The reductase was successfully overexpressed in *S. cerevisiae* and shown to be entirely specific for $NADH+H^+$. $NADH+H^+$ oxidation was therefore increased, although the resultant effect on flux through XR and XDH was not explored.

1.2.6 Evolutionary adaptation

Extended growth and the resulting evolution by mutation of cells in xylose media has been shown to positively affect growth rates on xylose. In order to alleviate the redox problem, a xylose isomerase (XI) gene from the anaerobic rumen fungus *Piromyces* sp E2 was expressed in yeast[28]. XI replaces XR and XDH, resulting in a single step conversion of xylose to xylulose without using cofactors and producing xylitol. Initial expression of this gene in *S. cerevisiae* resulted in slow growth ($0.005h^{-1}$ doubling time) on 2% xylose medium and the production of small amounts of ethanol. A strain with an improved doubling time of $0.03h^{-1}$ was created by growth with limited oxygen followed by anaerobic conditions (10 batch cultures each)[29]. This strain however consumed xylose mixed with glucose in a diauxic manner, only consuming xylose at a slower rate after exhausting glucose[30]. This diauxic

growth was not present after prolonged evolutionary adaptation in an anaerobic chemostat. The resultant strain consumed xylose in mixed media at a rate of 0.9g per gram cells per hour.

Similarly, XI from *Thermus thermophilus* was expressed in *S. cerevisiae*[31]. The resulting strain also overexpressed four endogenous enzymes from the non-oxidative phase of the PPP. The ability to grow anaerobically was engineered by repeated cultivation on xylose media. Subsequent replacement of XI with XR and XDH showed four times faster growth, implying that XI was rate limiting.

Sonderegger and Sauer[32] mutagenised *S. cerevisiae* strain TMB3001 with ethyl methane sulfonate (EMS) before continuous cultivation on xylose under aerobic conditions, followed by partially aerobic conditions and finally anaerobic conditions. The resulting cells were mutagenised again with EMS and then grown in both batch and continuous culture. Two different cell populations were derived showing 60% higher xylose consumption and 19% more ethanol production. The genetic mutations underlying these changes were not explored. The two strains were subsequently evaluated against eight other engineered and evolved *S. cerevisiae* strains and one of them declared to be best for ethanol production[33].

Attfield and Bell[34] derived a xylose utilising strain without mutagenesis through breeding and selection of a varied heterogeneous population of *S. cerevisiae* strains. Whereas most studies have reported that this yeast is unable to grow on xylose, selection on solid media, serial dilution and long term batch culture created a strain with a 6 hr doubling time on media containing xylose.

A *S. cerevisae* strain incorporating *P. stipitis* XR, XDH and endogenous XK was optimised for growth on xylose by extended growth in a chemostat and later analysed for changes in the metabolism of xylose, glucose and ethanol[35]. The new strain showed notable improvements in bottlenecks normally associated with xylose metabolism. These included a two-fold improvement in xylose uptake and a two-fold improvement in the expression of key PPP enzymes such as transketolase and transaldolase.

1.2.7 Optimising the PPP

Jin *et al.*[36] cloned fragments from a *P. stipitis* gene library into *S. cerevisiae* expressing *P. stipitis* XR and XDH, applied ten rounds of serial dilution optimisation and subsequently

recovered 16 clones with enhanced growth on xylose. The majority of these clones contained fragments encoding XK. This gene was thus incorporated into the strain and the gene library cloning repeated. The fast growing mutants isolated all contained identical fragments encoding the *P. stipitis* transaldolase gene *PsTAL1*. This gene was consequently expressed under a constitutive promoter, creating a strain that achieved double the rate of growth on xylose and produced 70% more ethanol when compared to the parent strain. The positive effect of transaldolase expression on promoting growth on xylose in XK-expressing strains has since been verified[37]. In contrast, it was previously observed that overexpression of the *S. cerevisiae* transaldolase lead to growth inhibition on glucose[38].

Despite the importance of XK, Jin *et al.*[39] demonstrated that its deletion in *P. stipitis* was not lethal and that growth on xylose occurred via an alternative enzyme. This resulted in a five-fold increase in xylose consumption with increased xylitol production when the yeast was grown with appropriate aeration.

In order to improve ethanol yields, a heterologous phosphoketolase pathway was expressed in *S. cerevisiae* to improve $NADH+H^+$ reoxidation during anaerobic growth[40]. The pathway consisted of an acetaldehyde dehydrogenase from *Entamoeba histolytica* and phosphotransacetylase from *Bacillus subtilis*, as well as the native phosphoketolase. The new pathway resulted in 25% more ethanol being produced on xylose with less xylitol, but large amounts of acetate were also produced. To alleviate this, the pathway was expressed in an *ald6* mutant strain that exhibited reduced acetate formation. This resulted in a 20% increase in ethanol production and a 40% increase in xylose fermentation without acetate production.

Arabinoxylan contains approximately two thirds D-xylose, as well as one third L-arabinose[5] (Fig 2). Becker *et al.*[41] augmented the PPP with an L-arabinose assimilation pathway in order to metabolise this carbon source. Genes for *B. subtilis* L-arabinose isomerase (*AraA*), *E. coli* L-ribulokinase (*AraB*) and L-ribulose-5-phosphate 4-epimerase (*AraD*) were overexpressed with the yeast arabinose transporter galactose permease. The bacterial genes convert L-arabinose to L-ribulose, L-ribulose-5-phosphate and D-xylulose-5-phosphate respectively, the last of which feeds into the PPP (Fig 1). Following repeated serial batch growth and plate selection, a strain with modest ethanol production and a doubling time of 7.9 hr on L-arabinose media was achieved.

1.2.8 Assimilating xylan

Xylose exists within plant biomass predominantly unavailable for metabolism in the form of the polymer xylan. Several studies have addressed the need to digest xylan to xylose to allow fermentation. Crous *et al.*[42] expressed and secreted the β-xylanase of *Aspergillus kawachii* IFO4308 in S. *cerevisiae*. Although effective degradation of a chromogenic xylan substrate occurred, the optimal conditions were at a temperature and pH unsuitable for yeast growth. Co-expression of β-xylanase from *Trichoderma reesei* and β-xylosidase from *Aspergillus niger* resulted in a 57% conversion of birchwood xylan, mostly to xylose monomers but also to xylobiose and xylotriose[43].

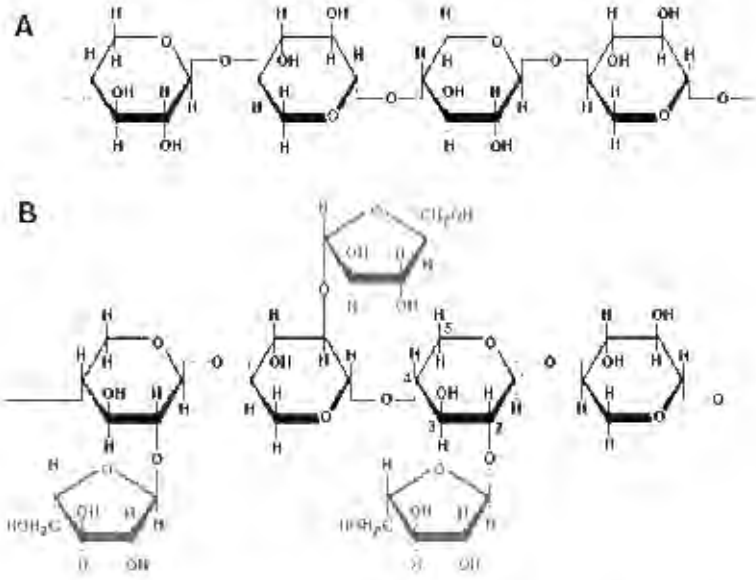


Figure 2: A) The structure of xylan, consisting of β-xylopyranosyl units with β(1→4) linkages. B) Arabinoxylan, the major constituent of hemicellulose, consisting of a xylan backbone with L-arabinose units (grey) attached randomly by α(1→2) and/or α(1→3) linkages. Structures were drawn using Macromedia Flash 4 (Adobe Systems, CA, USA).

To enable direct xylan fermentation by *S. cerevisiae*, Katahira *et al.*[44] co-expressed xylanase II from *T. reesei* and β-xylosidase from *Aspergillus oryzae* on the cell wall, together with *P. stipitis* XR and XDH and an endogenous XK. The resultant recombinant *S. cerevisiae* strain produced ethanol from birchwood xylan (7.1g per litre of culture after 62 hrs of fermentation). Similar targeting, utilising the yeast cell wall protein *PIR4* as a fusion partner, was used to express a *Bacillus* xylanase A either on the cell wall or for secretion into the culture medium[45]. Significant activity was found in the culture medium, however the resulting effect on xylan degradation was not quantified.

1.3 Strain evaluation on plant biomass

Laboratory strains are often tested under defined, ideal conditions. In contrast, industrial use requires significant and robust growth on a variety of substrates. Katahira *et al.*[46] created a recombinant *S. cerevisiae* strain expressing *P. stipitus* XR and XDH together with the endogenous XK and β-glucosidase from *Aspergillus aculeatus*, the latter being targeted to the cell wall. This strain was tested for its ability to produce ethanol from a wood chip lignocellulose prep that had been hydrolysed with concentrated sulphuric acid and subsequently adjusted to pH 7. After 36 h of fermentation, all of the cellooligosaccharides and xylose had been utilised, producing 30.3g ethanol per litre culture. A similar ethanol yield was obtained when a synthetic hydrolysate containing only defined sugars was fermented. This showed that the furfural and hydroxymethylfurfural (HMF) inhibitors present in the wood chip hydrolysate had a negligible effect on fermentation.

S. cerevisiae strain TMB3400 was fermented on finely milled corn stover (waste stalks and leaves) after a pre-treatment with water saturated steam and SO_2[47]. This was supplemented with trace minerals, yeast extract and commercial cellulase and β-glucosidase enzymes. Despite the resulting medium containing high levels of both water insoluble solids and typical inhibitors, a high level of ethanol (almost 40g/l) was produced after 96 h of simultaneous saccharification and fermentation (SSF). The high yields (calculated to be 59% of theoretical) were achieved by prior growth of the yeast beforehand in pre-treatment glucose-supplemented hydrolysate medium.

The presence of fermentation inhibitors in various hydrolysates remains a significant problem. The ability of strain TMB3001 to ferment steam pretreated H_2SO_4-hydrolysed sugarcane bagasse was optimised by lengthy pre-adaptation of the yeast on hydrolysate medium containing phenolic compounds and aliphatic acids, as well as increasing levels of inhibitors[48]. The adapted strain showed more than double the ethanol yield of its parent strain, in addition to more efficient conversion of furfural and HMF to the corresponding alcohols, the latter being beneficial for the cofactor imbalance.

These field trials and the accompanying lab studies show great promise for the direct fermentation of pre-processed plant biomass by recombinant *S. cerevisiae*. Continued studies will likely remove the effect of inhibitory compounds altogether and add capabilities for assimilating additional carbon sources with greater potency and reduced pre-processing.

1.4 Outline of study

Very few studies have focused on the expression of heterologous XI in *S. cerevisiae* for the conversion of xylose to xylulose - most making use of the *P. stipitis* XR and XDH enzyme pair. The benefits of using a single enzyme include simplicity of expression with no change to the balance of redox cofactors. However very few published examples[28,49,31] show the enzyme to be viable. Thus there exists abundant opportunity for the discovery of other XI genes that show greater activity than those demonstrated. Furthermore, it has been shown that strains expressing XI achieve higher final yields of ethanol than strains expressing XR and XDH, although the latter strains achieve comparable high yields faster[50].

The objective of this study was to evaluate other XI genes for functional expression in *S. cerevisiae*. Exploration of the available sequences was to be augmented by *in silico* codon suitability analysis and activity of the expressed protein was to be optimised by expression within, on the surface of and outside of the cell. Overexpression of the endogenous XK and serial batch growth with mutagenesis were also planned to improve growth on xylose media. Lastly, the ability of the resultant recombinant strain to ferment sugarcane bagasse was to be explored.

2. MATERIALS AND METHODS

2.1 Organisms and growth conditions

Wild-type yeast *Saccharomyces cerevisiae* W303 (a/alpha, *ade2-1/ade2-1, trp1-1/trp1-1, leu2-3/leu2-112, his3-11/his3-15, ura3/ura3, canr1-100/CAN*, haploid form) was used throughout and routinely grown at 30°C on a shaker in either YPD (1% w/v yeast extract, 2% w/v peptone, 2% w/v glucose) medium or complete synthetic (CS) minimal medium supplemented with 0.1mg/ml ampicillin (to eliminate bacterial contamination). CS medium contained 0.67% w/v yeast nitrogen base without amino acids (Difco, Sparks, MD, USA), 0.077% w/v complete synthetic medium without uracil (BIO-101 Systems, CA, USA) and the required sugar. The wild-type strain was unable to synthesise uracil, thus allowing successful recombinants to be selected for and grown using CS minimal medium without uracil. Recombinants were routinely grown in CS minimal medium without uracil to prevent reversion to wild-type cells. Cell growth was monitored by measuring the optical density (OD) of cultures at 600nm with an Ultrospec 500 Pro visible wavelength spectrophotometer (Amersham Biosciences, Sweden).

Escherichia coli strain DH5α was used throughout for cloning procedures and was routinely grown at 37°C in Luria medium (1% w/v tryptone, 0.5% w/v yeast extract, 0.5% w/v NaCl) with shaking. 1.5% w/v agar was added to make Luria agar plates.

2.2 General cloning procedures

Restriction enzyme digests, ligations, the polymerase chain reaction (PCR) and nucleotide digestions were all performed using standard methodologies[51]. Restriction enzymes were purchased from Fermentas (Ontario, Canada) and Roche (Mannheim, Germany).

PCR was performed with a high-fidelity *Taq* DNA polymerase from Fermentas (Ontario, Canada). In PCRs requiring two annealing temperatures, the first step amplified DNA from the template using only the portion of the primers homologous to the sequence. In the second step, the entire primer would anneal at a higher temperature to fragments amplified during the first step in order to include flanking sequences containing restriction sites, these not being homologous to the original template. Reactions typically involved (unless otherwise stated) cycles consisting of a denaturation step of 92°C for 30 s, followed by annealing at a specific temperature for 30 s and elongation at 72°C for 30 s. Reactions started with a 3 min denaturation step at 92°C.

Gel extraction of PCR and restriction digest products was performed using a QIAquick Gel Extraction Kit #28704 (QIAGEN, CA, USA). Gene cloning was performed using the TA cloning strategy and the pGEM-T Easy vector cloning kit (Promega, Wisconsin, USA) with positive transformants being identified by colour differences (blue/white colonies) due to alpha-complementation. The plasmids used in this study (other than pGEM-T Easy) were all *E. coli-S. cerevisiae* shuttle vectors. These were chosen so that genetic manipulations and plasmid amplification could be performed in *E. coli* before transformation of the final construct into yeast. These plasmids all made use of highly active promoters from constitutive genes of the yeast glycolytic pathway or the galactose utilisation pathway. All recombinant plasmids were digested with combinations of restriction enzymes to verify their validity before transformation into yeast.

2.3 Bacterial transformation

100µl competent *E. coli* cells were mixed with ~50ng of plasmid DNA and incubated on ice for 10 min. Cells were heat-shocked at 37°C for 5 min before adding 600µl pre-warmed YT (1.6% w/v tryptone, 2% w/v yeast extract, 0.5% w/v NaCl) and incubating at 37°C for 40 min. Cells were pelleted by brief centrifugation at 9000 x *g* and plated onto Luria agar plates supplemented with 0.004% w/v 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-Gal), 0.5mM isopropyl-β-D-thiogalactoside (IPTG) and 0.1mg/ml ampicillin before incubation at 37°C for 24 h.

2.4 Bacterial plasmid extraction

Plasmid DNA amplified in *E. coli* was routinely extracted using a hybrid extraction protocol. 10-12ml of overnight culture grown in Luria broth was pelleted and resuspended in an Eppendorf containing 300µl Buffer P1 (resuspension buffer) from a QIAGEN Plasmid Midi Kit #12143 (QIAGEN, CA, USA). 300µl of Buffer P2 (lysis buffer, from the same kit) was added, the samples were inverted 4-6 times and then left for 5 min at room temperature. 300µl Buffer P3 (neutralisation buffer, from the same kit) was added and samples were mixed immediately before being placed on ice for 15 min. After 5 min of centrifugation at 9000 x *g*, 900µl of supernatant was removed and mixed with 600µl of isopropanol. Samples were left for 2 min at room temperature and then centrifuged for 9 min at 9000 x *g*. The DNA pellet was washed with 70% ethanol, air dried and then resuspended in water or Buffer EB (resuspension buffer) from the kit.

2.5 Oligonucleotide synthesis and sequencing

Primers were designed with DNAMAN software (Lynnon BioSoft, Quebec, Canada) in addition to web-based PrimerPal software designed by the author for this purpose. The latter allows individual primers to be designed by scrolling along a sequence whilst analysing the melting temperature, GC content, fragment length, complementary and reverse-complementary sequences in real-time. Oligonucleotides were synthesized on a Beckman 1000M DNA synthesizer (Beckman Coulter, Inc., CA, USA) running a high purity (above 99%) program and subsequently verified for purity by gel electrophoresis.

The insert in plasmid pAZ4 (see 2.17) was amplified for sequencing in a PCR consisting of 30 cycles annealing at 61°C and elongating for 1 min 15 s at 72°C with the following primers: forward 5`-GTCAATGCAAGAAATACATATTTGG-3` (annealing 124bp upstream of the insertion site), reverse 5`-GCATAAAGGCATTAAAAGAGGAGC-3` (annealing 187bp downstream of the insertion site). The insert in plasmid pYES2 (see 2.20) was amplified for sequencing in a PCR consisting of 30 cycles annealing at 49°C and elongating for 1 min 10 s at 72°C with the following primers: forward 5`-GTAATAAAAGTATCAAC-3` (annealing 181bp upstream of insertion site), reverse 5`-GAAATATAAATAACGTTCTTAATAC-3` (annealing 136bp downstream of the insertion site). Trial amplifications were performed for insert confirmation before samples were sent to the relevant facility for re-amplification before sequencing.

Sequencing was performed either by Macrogen Inc. in Seoul, Korea or by the Core DNA Sequencing Facility at the University of Stellenbosch. Samples that were sequenced at Stellenbosch University were initially prepared by Di James (Molecular and Cell Biology department, University of Cape Town) using a Big Dye terminator v3.1 Cycle Sequencing kit (Applied Biosystems, CA, USA) and amplified with a GeneAmp PCR System 9700 (Applied Biosystems, CA, USA). Subsequent amplification cleanup and electrophoresis was performed by the Core DNA Sequencing Facility itself, making use of a 3130 Genetic Analyser (Applied Biosystems, CA, USA). Sequence information was edited with CHROMAS 2.3 software (Technelysium Pty Ltd) and aligned with DNAMAN.

2.6 *In silico* codon suitability analysis

Evaluation of potential gene sequences for expression in *S. cerevisiae* was performed by computational comparison of the codons present with those favoured by the majority of yeast genes. Codon frequency data were sourced from the Saccharomyces Genome

Database (http://www.yeastgenome.org) from a table compiled by J. Michael Cherry analysing 3222 yeast ORFs with the GCG program CodonFrequency. It was assumed that by finding a gene containing mostly codons that are translated with high efficiency, candidate genes from foreign hosts would be readily translated. Online software was written by the author in the Perl programming language to rate the codons of entire genes both individually and in total. Each codon was scored based on its fractional use converted to a percentage: for example a codon usage of 0.6 was scored as 60%. The total score was listed, total best and worst translated codons displayed as percentages and codon suitability displayed over the length of the gene in a graphical format.

2.7 Bioinformatic evaluation of XI candidates

Several candidate gene sequences were evaluated *in silico* with the goal of finding a XI gene with the potential for high activity in yeast. Since activity had been demonstrated for the XI gene of *Piromyces* sp. E2 (Genbank accession AJ249909) in *S. cerevisiae*[28], this gene's protein sequence was used as a basis for comparison and selection. A BLAST search of this sequence at the National Center for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov) found several potential candidates. The 40 highest matches were evaluated with regard to their sequence similarity and alignment (to the *Piromyces* XI sequence), both codon suitability scores (see 2.6) and the percentage of best and percentage of worst codons. These data, along with the nature of the organism and its availability were taken into account during final candidate selection.

2.8 Yeast transformation

Yeast cells were transformed by electroporation as described previously[52], with minor modifications. Essentially, yeast cells were grown to early exponential phase to an OD_{600} of 0.7. Cells were harvested at 2000 x *g* for 5 min, washed twice in ice-cold sterile water and then once in ice-cold 1M sorbitol. Cell pellets were resuspended in a minimal volume of ice-cold 1M sorbitol and 100µl mixed with 2µg of plasmid DNA. Cells suspensions were transferred to an electroporation cuvette and pulsed once at 1.5kV (200Ω, 25µF) in a Bio-Rad Genepulser (Bio-Rad, CA, USA). 500µl ice-cold 1M sorbitol was immediately added following electroporation. The cells were then pelleted by brief centrifugation at 9000 x *g*, plated onto CS minimal medium plates (selective for transformants) supplemented with 1M sorbitol and incubated at 30°C for 3-4 days.

2.9 Low stringency colony PCR

A low stringency PCR was regularly used to identify positive *E. coli* transformants. Individual colonies were marked and numbered, then picked and mixed directly into 50µl PCR reactions. Appropriate annealing and melting temperatures were used as per the original cloning, though annealing temperatures were often reduced by 5°C to lower the stringency of the PCR. This was done to ensure adequate annealing by long primers affected by chaotropic agents present as a result of the added cells.

Colony PCR was also performed on yeast transformants; however the cells required pre-treatment prior to PCR due to the strength of their cell walls. Colonies of approximately 1mm in diameter were solubilised in 30µl of 0.2% SDS, vortexed at high speed for 15 s and boiled at 90°C for 4 min. Samples were centrifuged for 1 min at 9000 x *g* and the supernatant removed and stored at -20°C until required. 1µl was used for PCR in each case using appropriate annealing and melting temperatures as per the original cloning.

2.10 Yeast genomic DNA extraction

10ml of cells grown to an OD_{600} of 0.7 were harvested at 12000 x *g* for 5 min. The cell pellet was resuspended in the remaining droplets of medium (50µl or less) then centrifuged at 9000 x *g* for 15 s to remove all media. The pellet was transferred into 200µl lysis buffer (2% v/v Triton X-100, 1% w/v SDS, 100mM NaCl, 1mM EDTA, 10mM Tris-HCl pH8.0) and resuspended by vortexing. 200µl of 0.4mm glass beads and 200µl phenol:chloroform:isoamyl alcohol (25:24:1) were added, after which the cells were lysed by vortexing for 4 min. 200µl TE buffer (10mM Tris-HCl, 10mM EDTA pH8.0) was added, followed by 5 min of centrifugation at 9000 x *g*. The top aqueous layer was removed and incubated with 30µl of 1mg/ml RNase A for 10 min at 37°C. One tenth volume of 3M sodium acetate pH5.2 and two volumes of ice-cold 100% ethanol were added before precipitation at -20°C for a minimum of 1 hour. Following a 15 min centrifugation at 9000 x *g* at 4°C, the DNA pellet was rinsed with ice-cold 70% ethanol and dried completely in a rotary vacuum drier for 12 min before resuspension in Millipore water (filtered to 0.22µm, Millipore, MA, USA).

2.11 Total protein extraction and visualisation

Total soluble protein was extracted from 300µl yeast cell pellets by the addition of 300µl phosphate-buffered saline (PBS) pH7.3, 200µl of 0.4mm diameter glass beads, 171.4µl of sample application buffer and 85.7µl 100mM PMSF. Samples were vortexed at high speed

for 4 min before a 1 min centrifugation at 9000 x *g*. The supernatant was loaded onto an SDS polyacrylamide gel (20% acrylamide, 0.1% bis-acrylamide) for electrophoresis in 0.3% w/v Tris-Cl, 1.5% w/v glycine, 0.1% w/v SDS buffer as described[53]. Protein was visualised using Coomassie Blue staining.

2.12 Yeast growth curves

Two methods of monitoring the growth of yeast cells on various substrates were used. The first (standard) approach involved the periodic spectrophotometric analysis at 600nm of cultures grown in either 250ml conical flasks or McCartney bottles. The second approach was devised to reduce contamination and culture volumes, while improving accuracy and increasing throughput. This method made use of 3ml cultures grown in 4ml cylindrical glass vials fixed on a rectangular array and placed on a linear shaker. Despite the reduced head space (one third of the culture volume) cultures grew at the same rate as cultures with 5 culture volumes of head space. A modified spectrophotometer block was created to hold each culture vial for direct analysis of the contents at 600nm (with no absorption occurring in the glass walls), thus preventing contamination that may have occurred due to opening and closing the vials for aliquotting and measurement. The 600nm absorption of samples in vials held by the modified spectrophotometer block showed identical values to samples measured regularly in normal cuvettes. For both methods, cells were inoculated to an OD_{600} of 0.005 before incubation began. Cells were appropriately diluted in the original medium for accurate spectrophotometric analysis when required.

2.13 Xylulose detection assay

Sorbitol dehydrogenase activity was used to detect xylulose, as it reduces xylulose to xylitol with the simultaneous oxidation of $NADH+H^+$. 200µl of exponential phase yeast cells were added to 200µl 0.4mm glass beads and made up to 1.3ml with enzyme extraction buffer (100mM sodium phosphate, 1mM $MgCl_2$, 0.1mM EDTA, 1mM dithiothreitol, 1mM PMSF pH7). After the mixture was vortexed at high speed for 4 min and centrifuged for 30 s at 9000 x *g*, the supernatant was removed and stored on ice. The assay for xylulose was performed as previously described[28] with a few modifications. Each reaction contained: 300µl cell extract, 300µl of 3 x reaction buffer (100mM Tris-HCl, 10mM $MgCl_2$ pH7.5), 300µl of either a water control or 1.5M xylose, 8µl 0.15mM $NADH+H^+$ and 6µl of 0.5U/µl sorbitol dehydrogenase (Roche, Mannheim, Germany). After the reactions were mixed, the absorbance at 340nm was monitored for the consumption of $NADH+H^+$ (extinction coefficient of $6.3mM^{-1}.cm^{-1}$). Readings were recorded automatically every 15 s with a Beckman Spectrophotometer

(Beckman, California, USA) and experiments were performed in quadruplicate. Best-fit linear trendlines were applied to each dataset as a measure of the average consumption rate.

2.14 High performance liquid chromatography

HPLC was used for the detection of xylose, glucose and galactose in the yeast culture medium. Stationary phase cultures were used to inoculate 10ml of CS medium (containing the required mix of 2% xylose with either 1% glucose or 2% galactose) in tightly sealed McCartney bottles to an OD_{600} of 0.005. These bottles were shaken on a linear shaker at 30°C for 48 h. Growth media were removed and filter-sterilised through 0.2µm filters before snap freezing in liquid nitrogen and storage at -70°C. HPLC analysis was performed under the supervision of Annatjie Hugo at the Department of Microbiology, Stellenbosch University. Samples were diluted to 1/625 in Millipore water and run through a Dionex HPAEC-PAD (High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection) (Dionex, California, USA). A CARBOPAC PA-10 column running an optimised NaOH gradient elution was used. This setup was specific for the separation of sugars only. No other components present in the media eluted within the observed range (Fig 3). Peak volume data, rather than peak height data, were used for comparison since these are more accurate. Calibration standards were not used since only comparative data between different modified strains were of interest. Peak identification was achieved by recording the elution times of single sugars present in CS medium. The zero time points shown in Figure 3 contained 2% xylose together with 1% glucose or 2% galactose.

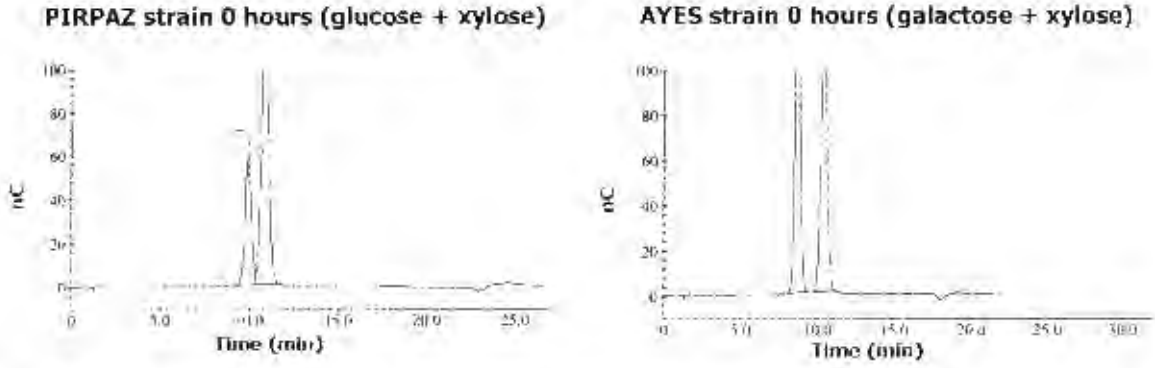


Figure 3: HPLC traces used for peak identification. The first panel shows CS minimal medium containing 1% glucose (1st peak) and 2% xylose (2nd peak). The second panel shows CS minimal medium containing 2% galactose (1st peak) and 2% xylose (2nd peak). The scale of the y-axes was altered to detect any other sugars present in small amounts, and thus the galactose and xylose peaks appear incomplete. The complete peak data were recorded by the HPLC data recorder and used for analyses.

2.15 Yeast RNA isolation

RNA was extracted from yeast as described originally by Chomczynski *et al.*[54] with only minor modifications pertaining to volumes and the preparation of cells. Yeast cells were grown overnight in CS minimal medium (selective for transformants only) to an OD_{600} of 0.7 (early exponential phase) and harvested by centrifugation at 2000 x *g* for 10 min at 4°C. 150µl of cells were added to 850µl Solution D (4M guanidinium thiocyanate, 25mM sodium citrate, 0.5% w/v sarcosyl, 0.1M 2-mercaptoethanol pH7) together with 200µl 0.4mm diameter glass beads and the mixture vortexed at high speed for 4 min. The remaining steps of the protocol were directly followed as described until the final step, when nucleic acid pellets were vacuum dried and resuspended in 50µl DNaseI buffer (50mM Tris-HCl, 10mM $MnCl_2$ pH7.5) at 65°C for 10 min.

Since pure RNA was required for cDNA synthesis and PCR, a DNaseI treatment was incorporated to remove DNA from the extract. The above samples were therefore incubated at 37°C for 40 min with 10U DNaseI (Roche, Mannheim, Germany) and 20U RNase inhibitor (Roche, Mannheim, Germany). RNA was subsequently re-extracted using the latter part of the above protocol (followed from after glass bead disruption) to separate RNA from the DNaseI enzyme.

2.16 cDNA synthesis

2ug of extracted RNA was added to 4µl of 50µM oligo-dT primer stock and made up to 15µl with either Millipore or diethylpyrocarbonate (DEPC) treated sterile water. This was heated to 70°C for 5 min, cooled rapidly on ice and collected by brief centrifugation at 9000 x *g*. 5µl of 5xMMLV reaction buffer (Promega, Wisconsin, USA), 2µl of 6.25mM dNTP stock, 25U RNase inhibitor (Roche, Mannheim, Germany) and 1µl of MMLV reverse transcriptase (Promega, Wisconsin, USA) were added together with Millipore or DEPC treated water to 25µl total volume. Sample tubes were tapped to mix the contents then incubated at 41°C for 1 hour or longer. 1µl of the cDNA and 250ng of the RNA extract was used for RT-PCR.

2.17 Preparation of the targeting plasmid PIRPAZ and the derivative control yeast strain

Initially XI was to be targeted to three locations (the cytoplasm, the cell wall and secreted outside of the cell) to optimise performance. The yeast cell wall protein Pir4p was to be used as the fusion partner for targeting XI to two of these locations: the cell wall and outside of the cell[45]. The plasmid used for expression was pAZ4 (a gift from Willem H. van Zyl of the

Department of Microbiology, Stellenbosch University) featuring the constitutive *S. cerevisiae PGK1* promoter and terminator with *URA3* and Amp^R selectable markers[42]. The plasmid maps for all plasmids are shown in Figure 4. Using the Saccharomyces Genome Database (http://www.yeastgenome.org), primers for the *PIR4* gene (CIS3/YJL158C, Genbank accession AAT93046) were designed to amplify 707bp including the start and stop codons: forward 5`-CAAGAATTCGAAATGCAATTCAAAAACG–3` (includes an *EcoRI* site, underlined), reverse 5`-GATCTCGAGCTTTAACAGTCGACC-3` (includes a *XhoI* site, underlined). Five PCR cycles annealing at 44°C followed by 30 cycles annealing at 68°C with 10µM dNTPs and 1.5mM $MgCl_2$ amplified *PIR4* from yeast genomic DNA. PCR fragments were extracted and ligated to the pGEM-T Easy plasmid for amplification in *E. coli* before extraction, digestion and secondary ligation to an *EcoRI*/*XhoI* digested pAZ4. This plasmid was amplified in *E. coli*, extracted and transformed into wild-type yeast to create the control strain denoted PIRPAZ.

2.18 Amplification of *H. influenzae* XI and creation of yeast strain HPAZ

Haemophilus influenzae Rd KW20 was provided as a gift from Dr Stephen Oliver of the Department of Medical Microbiology, University of Cape Town. Generous scrapings of fresh cells were dissolved in 1ml Millipore water then boiled for 10 min. Genomic DNA was extracted using the yeast genomic DNA extraction protocol. Three pairs of primers (Table 1) were designed to amplify the *H. influenzae* XI (Genbank accession NC_000907) with the required restriction sites to allow targeting of the protein to three separate locations: within the cell, on the cell wall and excreted out of the cell. PCR was performed with 28.75µM dNTPs and 1.5mM $MgCl_2$.

Table 1: Primers used for amplification of *H. influenzae* XI for cloning and fusion with *PIR4*.

Location	Primer sequence	RE site	Annealing temps
Cytoplasm	Fwd: 5`-CAGAATTCTATGACAACCTATTTCGATAAAATTG-3` Rev: 5`-GTCTCGAGATTAGCTATAAATAACTTGGTTTAC-3`	*EcoRI* *XhoI*	5 x 49°C 30 x 60°C
Cell wall (fusion with *PIR4*)	Fwd: 5`-CAGTCGACACAACCTATTTCGATAAAATTG-3` Rev: 5`-GAGTCGACGCTATAAATAACTTGGTTTACC-3`	*SalI* *SalI*	5 x 45°C 30 x 54°C
External to cell (fusion with *PIR4*)	Fwd: 5`-CCAGATCTATACAACCTATTTCGATAAAATTG-3` Rev: 5`-GAGTCGACGCTATAAATAACTTGGTTTACC-3`	*BglII* *SalI*	5 x 45°C 30 x 54°C

Amplified fragments were ligated to pGEM-T Easy and transformed into *E. coli* for amplification. The cytoplasm-targeted gene was digested from plasmid DNA with *EcoRI*/*XhoI*, ligated to similarly digested pAZ4 and amplified in *E. coli*. Creation of the two

PIR4 fusions was delayed at this stage, until proof of XI activity in the cytoplasm was demonstrated. The resulting plasmid was extracted, sequenced (Appendix A1) and transformed into wild-type yeast with successful HPAZ transformants being identified by low-stringency colony PCR.

2.19 Amplification of *A. thaliana* XI and creation of yeast strains AXPAZ and AXENO

The *Arabidopsis thaliana XI* complete cDNA[55] (Genbank accession AY136469, locus At5g57655) was ordered online from the Arabidopsis Biological Resource Center's DNA Stock Center (Ohio, USA) (http://www.arabidopsis.org). The clone (#U25054) was delivered as a stock of *E. coli* strain PIR1 containing the vector pUNI51. This was grown overnight in Luria broth containing 75µg/ml kanamycin before subsequent plasmid extraction. Primers were designed to amplify the complete coding sequence: forward 5`-GGTTCGTGAATTCCTATGAAGAAAGTTGAG-3` (includes an *EcoRI* site, underlined), reverse 5`-GTACTCGAGCTTACATTGCAGATTGGAAAATC-3` (includes a *XhoI* site, underlined). Five cycles annealing at 45°C followed by 30 cycles annealing at 65°C with 1mM $MgCl_2$ and 10µM dNTPs were used to amplify the fragment that was then extracted and ligated to pGEM-T Easy. This was transformed, amplified and subcloned as before into *EcoRI*/*XhoI* digested pAZ4 that was subsequently sequenced (Appendix A2) and transformed into wild-type yeast to create strain AXPAZ.

The same gene was digested from pGEM-T Easy and ligated into plasmid vector YEpENO1 (a gift from Willem H. van Zyl of the Department of Microbiology, Stellenbosch University). This vector is identical to pAZ4 but contains the promoter and terminator of the constitutively expressed yeast *ENO1* gene (ENO1/YGR254W, Genbank accession AAA88712)[56]. The resulting plasmid was transformed into wild-type yeast and the new strain denoted AXENO.

2.20 Amplification of *A. thaliana* XI and creation of yeast strain AYES

Plasmid pUNI51 was used as template for a PCR consisting of 5 cycles annealing at 58°C, followed by 30 cycles annealing at 65°C. 4mM $MgCl_2$ and 400µM dNTPs were used. The following primers were used for amplification: forward 5`-GAAGAATTC**AATATGG**GTAAGAAAGTTGAGTTTTTTATG-3` (includes an *EcoRI* site, underlined), reverse 5`-CGGTCTCGAGCTTACATTGCAGATTGGAAAATC-3` (includes a *XhoI* site, underlined). The commercial yeast protein expression vector pYES2 (Invitrogen, CA, USA) was used in this strain and those following it for its high inducibility with galactose,

with glucose acting as a repressor[57]. The forward primer was designed to incorporate a Kozak translation initiation sequence[58] around the start codon (bold text). Three extra bases encoding a codon for glycine directly after the start codon were incorporated to this end. The reverse primer was designed to incorporate the gene in frame with a backup TAG stop codon 9bp downstream of the *XhoI* site. The amplified fragment was ligated to pGEM-T Easy as before, then subcloned into pYES2. The resulting plasmid was sequenced (Appendix A3) and transformed into wild-type yeast and the new strain denoted AYES.

2.21 Amplification of *B. thetaiotaomicron* XI and creation of yeast strain BYES

Genomic DNA was extracted from cells of *Bacteroides thetaiotaomicron* VPI-5482 (a gift from Dr Ekta Patel, Molecular and Cell Biology department, University of Cape Town) using the yeast colony PCR extraction protocol. This template was used to amplify the XI gene (Genbank accession NC_004663) in a PCR consisting of 5 cycles annealing at 58°C, followed by 30 cycles annealing at 65°C with 4mM $MgCl_2$ and 400µM dNTPs. Primers used were: forward 5`-CTTTTTAAAGCGGCCGC**ATTATGG**CAACAAAAG-3` (includes a *NotI* site, underlined), reverse 5`-GAAAATTCTCGAGATTAGCAATACATATTCAGAATTG-3` (includes a *XhoI* site, underlined). The forward primer was designed to incorporate a Kozak translation initiation sequence[58] (bold text). The reverse primer was designed to incorporate the gene in frame with a backup TAG stop codon 9bp downstream of the *XhoI* site. The gene was ligated to pGEM-T Easy as before and subcloned into pYES2. The resulting plasmid was sequenced (Appendix A4) and transformed into wild-type yeast and the new strain denoted BYES.

2.22 Creation of control yeast strains EMPAZ and EMPYES

The PIRPAZ and AYES plasmids were digested with *EcoRI*/*XhoI*. The restriction sites of the resulting fragments were blunt-ended with DNA polymerase I (Klenow large fragment) (New England Biolabs, United Kingdom) for 15min at 25°C. The resulting DNA was electrophoresed to separate the vectors from the inserts. Both linear vectors were gel extracted individually and circularised with T4 DNA ligase (New England Biolabs, United Kingdom) to form empty vectors devoid of XI genes. The resulting plasmids, denoted EMPAZ and EMPYES respectively, were digested with restriction enzymes and sequenced across the insertion sites (data not shown) to confirm that no XI genes were present and no complete *EcoRI* or *XhoI* restriction sites remained. The plasmids were transformed into wild-type yeast to form two new strains carrying the names of the plasmids.

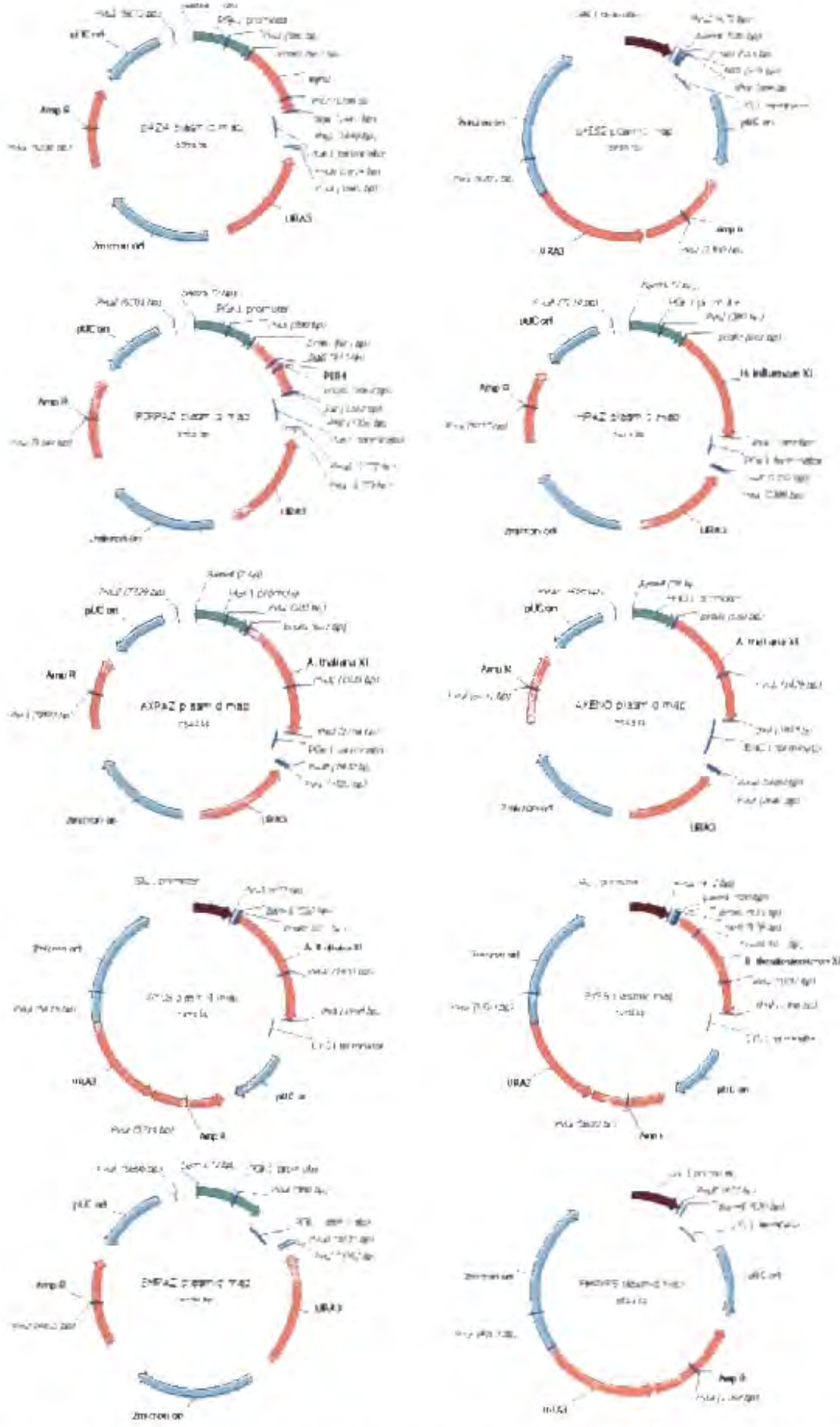


Figure 4: Plasmid maps of all plasmids showing inserted genes, origins of replication, selective markers, promoters, terminators and the restriction sites used for cloning and mapping. Vectors were drawn using Vector NTI 10.1.1 software (Invitrogen, CA, USA).

3.1 Gene candidates showed high codon suitability when compared to native highly expressed genes.

In order to find a XI gene with the potential for efficient translation in *S. cerevisiae*, a variety of XI genes from different organisms were computationally analysed with regard to their codon suitability for expression in yeast. It was expected that selecting genes composed of codons highly expressed in yeast would lead to optimal protein production. Surprisingly, gene candidates from several unrelated organisms were found to contain a high level of codon suitability for expression in *S. cerevisiae* (Table 2). The various scores calculated by the software for some non-yeast genes were comparable and in some cases better than those of some highly expressed *S. cerevisiae* genes.

Table 2: Comparison of codon suitability for *S. cerevisiae* of the XI genes from a variety of organisms and some highly expressed yeast genes [a]

Organism	Codon suitability[b]	Best translated codons[c]	Least translated codons[c]
Clostridium difficile QCD-32g58	45%	54%	18%
Haemophilus influenzae R2866	**45%**	**62%**	**9%**
Flavobacterium johnsoniae UW101	44%	57%	15%
Clostridium phytofermentans ISDg	44%	56%	14%
Photorhabdus luminescens subsp. Laumondii TT01	44%	58%	14%
Piromyces sp. E2	44%	54%	26%
Bacteroides thetaiotaomicron VPI-5482	**42%**	**51%**	**22%**
Alkaliphilus metalliredigenes QYMF	42%	52%	17%
Photobacterium profundum SS9	42%	50%	19%
Flavobacterium sp. MED217	41%	48%	17%
Arabidopsis thaliana (thale cress)	**41%**	**51%**	**24%**
Saccharomyces cerevisiae (*HXK1* gene)	43%	59%	20%
Saccharomyces cerevisiae (*PKC1* gene)	42%	55%	19%
Saccharomyces cerevisiae (*PGM1* gene)	40%	48%	19%
Saccharomyces cerevisiae (*TPS1* gene)	38%	39%	32%

[a] Hosts in bold face text were used as sources for cloning to create the yeast strains evaluated in this study.
[b] Combined percentage, based on fractional use, of each codon for its amino acid
[c] Percentage of codons represented by best and least translated versions

As a result of these calculations, and an evaluation of what genetic stock was most easily available, the first candidate gene to be tested was that of *H. influenzae* Rd KW20. This gene was chosen as it was available locally and had an almost identical sequence to that of *H. influenzae* R2866.

3.2 *H. influenzae* XI was transcribed but not found to be translated.

The first recombinant strain, HPAZ, contained the XI gene from *H. influenzae* Rd KW20 under control of the endogenous *PGK1* promoter. The plasmid construct was sequenced and seven nucleotide differences were found (Appendix A1). Six of these were silent mutations resulting in alternate codon usage, and only one of these resulted in the amino acid substitution, Ser → Pro at position 73. In order to establish whether translation of the XI gene had occurred, exponential phase cells of both the HPAZ and wild-type strains were grown for 4 h in either YPD or YPX (2% w/v xylose in place of glucose) medium before the cells were harvested and total soluble protein was extracted. SDS-PAGE of the extracted protein revealed no differences within the predicted size range between the wild-type and HPAZ strains (data not shown).

To investigate whether small amounts of XI not detectable by SDS-PAGE could be detected as differences in growth between the strains, growth of the HPAZ and wild-type strains was monitored for 400 h. CS medium containing either no carbon source or 2% xylose or 0.1% glucose or 2% xylose, 0.1% glucose was used as the growth medium with uracil supplementation to support growth of the wild-type strain. The extended period of growth was chosen since a similar study[28] involving the *Piromyces* XI gene had only shown significant growth after 400 h. On all media evaluated, the HPAZ strain showed better growth than the wild-type strain in that a higher cell density was observed (Fig 5). The results also showed that the HPAZ strain exhibited improved growth when xylose was added to CS medium containing no carbon source, an effect not replicated by the wild-type strain. As expected, the addition of glucose was noted to significantly increase growth in both strains. Moreover, no increased growth was observed when xylose was added to glucose-containing medium.

In order to investigate whether the wild-type strain was a suitable control for growth experiments, the PIRPAZ strain was created. This strain was identical to the HPAZ strain except that the gene for the cell wall protein Pir4p replaced the XI gene. The above growth experiment was repeated with the PIRPAZ strain. The two growth effects observed previously for the HPAZ strain, namely the general improved growth on all media and enhanced growth due to xylose, were duplicated by the PIRPAZ strain (data not shown). Thus both growth improvements observed previously for the HPAZ strain were not attributable to the XI enzyme. It was thought that the general improved growth observed on media containing xylose might have been due to osmotic stabilisation. Since the wild-type

strain did not need to expend energy replicating a plasmid and because the PIRPAZ strain was able to grow on exactly the same medium as the recombinant strains, the PIRPAZ strain was used as a control strain in all subsequent growth experiments.

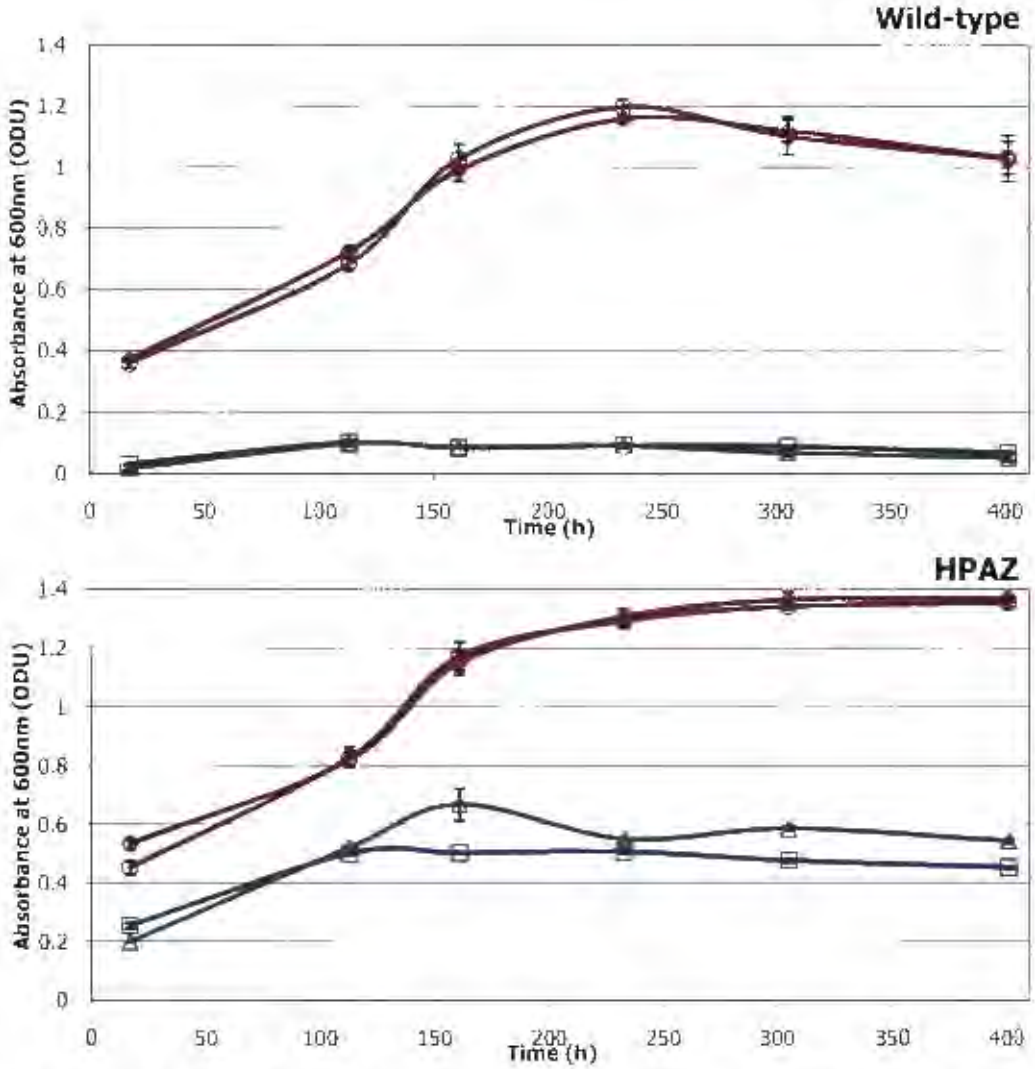


Figure 5: Growth of wild-type and HPAZ strains in CS media containing no carbon source (□—□), 0.1% glucose (◇—◇), 2% xylose (Δ—Δ) and 0.1% glucose combined with 2% xylose (O— O). Data represent the mean ± standard error of three separate experiments. Error bars not visible are within the symbols.

Since no growth differences could be attributed to the XI enzyme, an enzyme-based *in vitro* assay using sorbitol dehydrogenase was used to detect changes in the xylulose concentration that could be attributed to XI activity. Sorbitol dehydrogenase reversibly reduces certain keto sugars to their corresponding alcohols with simultaneous oxidation of $NADH+H^+$; this includes D-xylulose to D-xylitol. This methodology has been used previously[28,49] to detect XI enzyme activity. However, SDH also reversibly reduces four pairs of common cytosolic compounds other than D-xylulose, namely fructose to D-

sorbitol, L-sorbose to L-iditiol, D-ribulose to ribitol and allulose to allitol. The enzyme is also known to utilise NADPH+H$^+$ as well as NADH+H$^+$, though at a reduced rate. It was found that although NADH+H$^+$ consumption occurred due to both strains with or without xylose, the rate of consumption was less for samples containing xylose, contrary to what was expected (Fig 6).

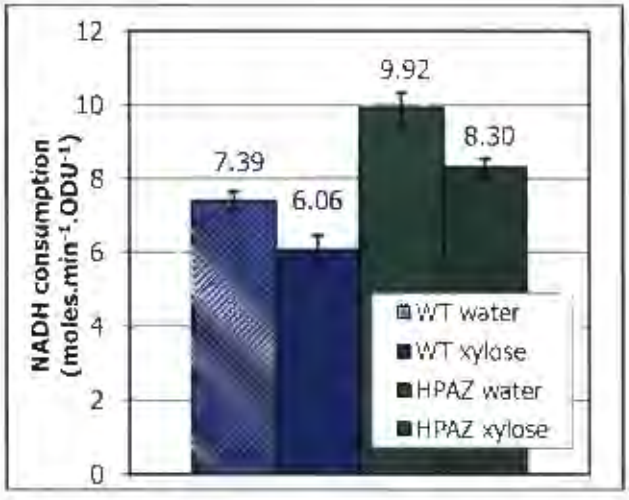


Figure 6: Cell extracts from HPAZ and wild-type cells grown overnight in YPD were used to collect quadruplicate data for NADH+H$^+$ consumption in the presence of sorbitol dehydrogenase. The ratios of consumption with and without xylose for HPAZ and wild-type cells were 0.84 and 0.82 respectively. Data represent the mean ± standard error of four separate experiments.

It was thought that xylose may have acted as a partial inhibitor of the sorbitol dehydrogenase enzyme, thus reducing its ability to oxidise NADH+H$^+$ effectively. Since only reduced rates of consumption were observed, and these rates were almost identical (98%) between strains, the effect could not be attributed to XI.

Since no XI activity was detected by the enzyme-dependent assay, high performance liquid chromatography was used to quantify the relative amounts of both glucose and xylose in the growth medium before and after 48 h of growth. It was expected that glucose would be metabolised completely by both strains as the preferred substrate, and that any decrease in the xylose concentration due to conversion to xylulose by XI would occur subsequently. Previous work indicated that on a mixture of glucose and xylose, recombinant strains initially consumed glucose (while inhibiting xylose consumption) before xylose consumption began[59]. In light of this, the medium used for the experiment contained only 1% glucose to grow cells to a suitable density, after which consumption of the 2% xylose would be expected to commence. All of the glucose was metabolised after 48 h by both strains and xylose

consumptions for the HPAZ strain and the PIRPAZ control strain (Fig 7) were 2.65% and 4.20% of the initial added xylose respectively (Table 3). Since virtually no xylose was consumed by either the HPAZ strain or the control strain, it was concluded that the construct of the putative *H. influenzae* XI gene in the HPAZ plasmid had had no effect on the consumption of xylose.

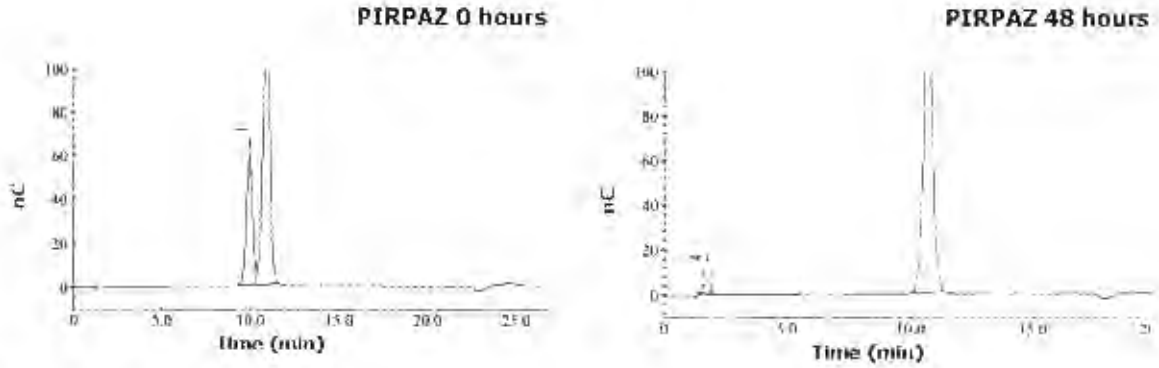


Figure 7: HPLC traces for the PIRPAZ strain in CS medium. The peak eluting first is glucose (1%) and the second peak is xylose (2%). All of the glucose was consumed after 48 hours. The scale of the y-axes was altered to detect any other sugars present in small amounts, and thus the xylose peaks appear incomplete. The complete peak data were recorded by the HPLC data recorder and used for analyses.

Table 3: Consumption of glucose and xylose by HPAZ and PIRPAZ strains over 48 h [a]

Strain	Glucose			Xylose		
	0 h (g/l)	48 h (g/l)	Consumption	0 h (g/l)	48 h (g/l)	Consumption
HPAZ	10	0	100%	20	19.47	2.65%
PIRPAZ	10	0	100%	20	19.16	4.20%

[a] Constituent masses in proportion to peak volumes derived from HPLC data are shown at 0 and 48 h.

Since no data suggesting the activity or the presence of the translated XI protein in the HPAZ strain could be detected, it was investigated whether XI gene transcription had occurred by RT-PCR. The EMPAZ control strain was created for comparison with the HPAZ strain. The EMPAZ plasmid was identical to the HPAZ plasmid but lacked the XI gene insert. Total nucleic acid was extracted from exponential phase cells of both strains, after which DNAseI treatment was used to remove any DNA. cDNA was produced from the remaining RNA and RT-PCR used to detect transcription of the XI gene in the HPAZ and EMPAZ strains. Positive evidence for an mRNA encoding the *H. influenzae* XI was found only in the cDNA of the HPAZ strain (Fig 8). The positive control to demonstrate successful cDNA production was *HSP12* - a constituent of the stress response pathway in yeast. Transcription of *HSP12* was visible in cDNA samples from both strains. The negative controls (indicating no plasmid DNA remaining due to incomplete DNAseI digestion) were the RNA extract, as well as

amplification of the plasmid using the pAZ4 sequencing primers. No PCR products were observed for all negative controls.

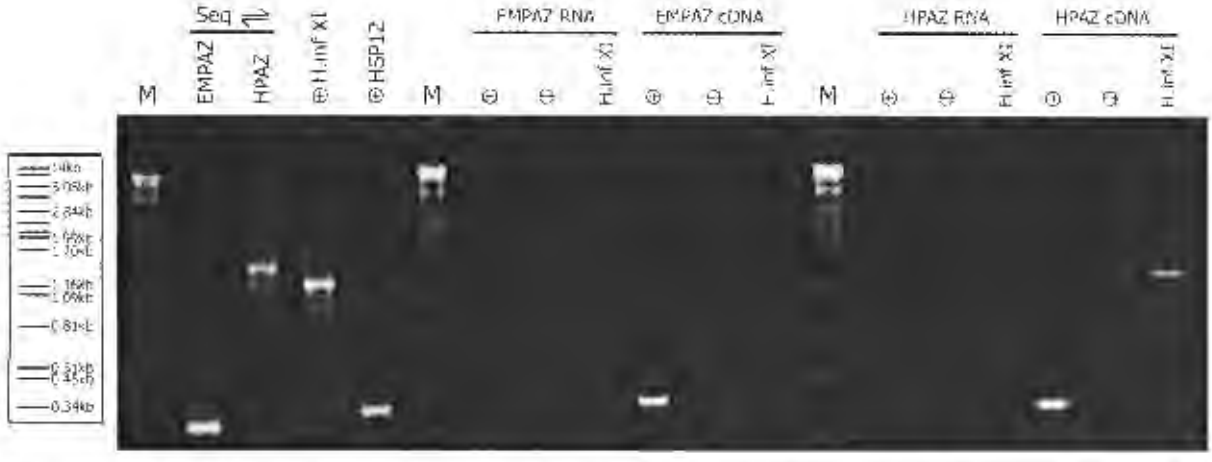


Figure 8: RT-PCR of RNA and cDNA prepared from EMPAZ (empty vector) and HPAZ cells grown to exponential phase. Both strains were tested for transcription of the *H. influenzae* XI gene (1.32kb), *HSP12* (positive control indicating successful cDNA production ⊕, 348bp) and amplification of pAZ4 plasmid DNA with sequencing primers (negative control ⊖, ±310bp for EMPAZ and ±1.63kb for HPAZ). The RNA extract was also used as a negative control to detect background plasmid DNA remaining from an incomplete DNAseI digestion. Transcription of the XI gene occurred only in the cDNA of the HPAZ strain. *HSP12* transcription occurred in the cDNA of both strains and no products were observed for both the RNA and sequencing primer negative controls. The first four lanes demonstrate amplification of the EMPAZ and HPAZ plasmids with pAZ4 primers, followed by amplification of *H. influenzae XI* and *HSP12* to demonstrate the sizes of the expected products. Lambda DNA digested with *PstI* was used as a marker (M) (marker fragment sizes shown in box).

Since the recombinant strain was able to transcribe the XI gene but neither protein nor enzymatic activity were detected, it was concluded that the gene sequence itself was the reason for undetected translation and not the behaviour of the promoter or the vector. Assuming translation had occurred, the protein may have made use of prokaryotic folding motifs not compatible with the eukaryotic host, thus producing an inactive enzyme. A second XI candidate with high codon suitability was therefore chosen from *A. thaliana* (Table 2). It was also hoped that by selecting a gene from a eukaryotic source, translation to a functional protein would be more likely to occur in *S. cerevisiae*.

3.3 *A. thaliana* XI was transcribed but not found to be translated.

The second recombinant strain to be evaluated, AXPAZ, contained the *A. thaliana* XI gene under control of the same endogenous *PGK1* promoter and terminator used previously. The plasmid construct was sequenced and no errors were found (Appendix A2). Protein from exponential and stationary phase cells of the AXPAZ and wild-type strains was extracted and electrophoresed on an SDS gel. No differences between the strains were observed within the predicted size range (45-60kDa), even at high-resolution (Fig 9).

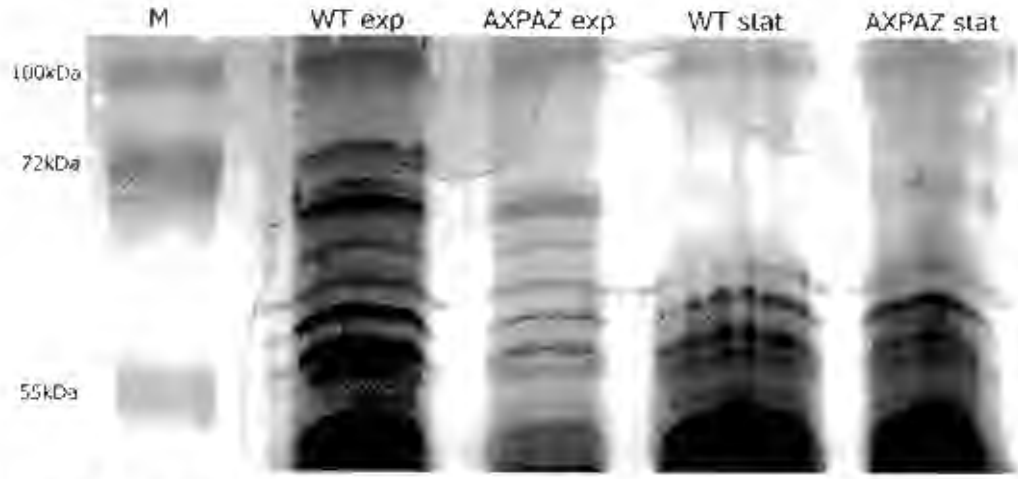


Figure 9: SDS-PAGE of protein extracted from AXPAZ and wild-type cells grown to exponential phase and stationary phase. 150µl of pelleted cells were used for extractions and 120µl of the extracted protein sample was loaded. The marker used was the PageRuler Prestained Protein Ladder (Fermentas, Ontario, Canada).

To investigate whether small amounts of XI not detectable by SDS-PAGE could be detected as differences in growth between the two strains, growth of the AXPAZ strain and the PIRPAZ control strain on xylose or glucose-supplemented xylose media was monitored. A more accurate method of studying growth as a function of time involving more replicate samples (see section 2.12) was used. The experiment was unable to demonstrate improved growth of the AXPAZ strain over that of the control strain and once again all strains grew only to a very low stationary phase density (Fig 10). The faster rate at which the AXPAZ strain grew to stationary phase was possibly due to the AXPAZ strain stock culture used being younger (and thus more viable) than the other stock cultures used. In support of this hypothesis, a definite lag phase was visible for growth of both the HPAZ and PIRPAZ strains (older stocks) on 2% xylose and CS medium containing no carbon source.

Since no data suggesting the activity or the presence of the translated XI protein in the AXPAZ strain could be detected, it was investigated whether XI gene transcription had occurred by RT-PCR. The EMPAZ control strain was used for comparison with the AXPAZ strain. The EMPAZ plasmid was identical to the AXPAZ plasmid but lacked the XI gene insert. Total nucleic acid was extracted from exponential phase cells of both the AXPAZ and the EMPAZ strains, followed by a DNAseI treatment to remove DNA. cDNA was produced from the remaining RNA and RT-PCR used to detect transcription of the XI gene in the AXPAZ and EMPAZ strains. Positive evidence for an mRNA encoding the *A. thaliana* XI was found only

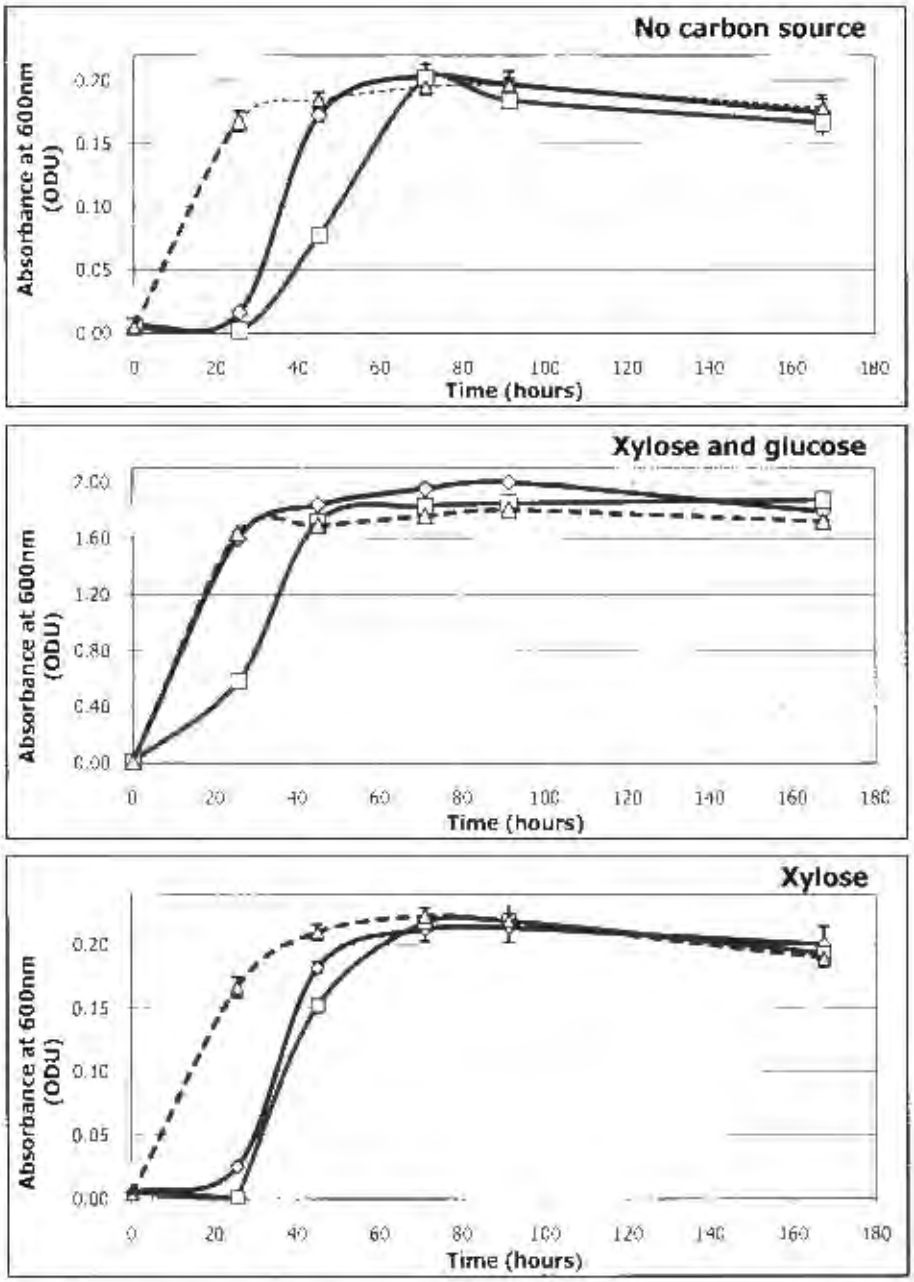


Figure 10: Growth of the AXPAZ (Δ– – Δ), PIRPAZ (◇—◇) and HPAZ (□—□) strains in CS medium supplemented with no carbon source or 1% glucose, 2% xylose or 2% xylose. Data represent the mean ± standard error of four separate experiments. The HPAZ strain was included to validate previous findings. Note that the density of cells grown in the presence of glucose is 10-fold greater than the density achieved without glucose. Error bars not visible are within the symbols.

in the cDNA of the AXPAZ strain (Fig 11). The positive control to demonstrate successful cDNA production was *HSP12* - a constituent of the stress response pathway in yeast. Transcription of *HSP12* was visible in cDNA samples from both strains. The negative controls (indicating no plasmid DNA remaining due to incomplete DNAseI digestion) were the RNA extract, as well as amplification of the plasmid using the pAZ4 sequencing primers. No PCR products were observed for all negative controls.

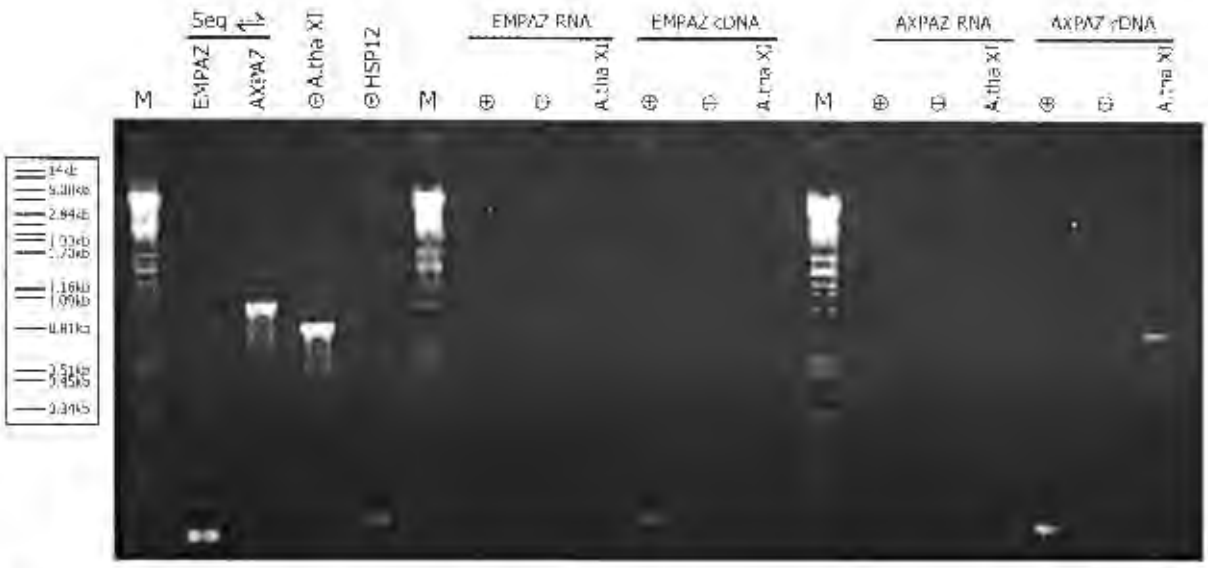


Figure 11: RT-PCR of RNA and cDNA prepared from EMPAZ (empty vector) and AXPAZ cells grown to exponential phase. Both strains were tested for transcription of the *A. thaliana* XI gene (1.43kb), *HSP12* (positive control indicating successful cDNA production ⊕, 348bp) and amplification of pAZ4 plasmid DNA with sequencing primers (negative control ⊖, ±310bp for EMPAZ and ±1.75kb for AXPAZ). The RNA extract was also used as a negative control to detect background plasmid DNA remaining from an incomplete DNaseI digestion. Transcription of the XI gene occurred only in the cDNA of the AXPAZ strain. *HSP12* transcription occurred in the cDNA of both strains and no products were observed for both the RNA and sequencing primer negative controls. The first four lanes demonstrate amplification of the EMPAZ and AXPAZ plasmids with pAZ4 primers, followed by amplification of *A. thaliana XI* and *HSP12* to demonstrate the sizes of the expected products. Lambda DNA digested with *PstI* was used as a marker (marker fragment sizes shown in box).

Since PCR provides no indication of the relative abundance of specific mRNAs within cells, one possibility was that the transcript levels were too low for adequate translation resulting in undetectable activity. Despite previous work[60,61] identifying *PGK1* as a strong promoter, it was thought prudent to check whether expression of the same gene under a different promoter could make a difference.

The same *A. thaliana* XI gene was ligated to the plasmid YEpENO1 – a vector identical to pAZ4 except that the *ENO1* promoter and terminator replaced those of *PGK1*. Eno1p or enolase is responsible for the conversion of 2-phosphoglycerate to phosphoenolpyruvate during glycolysis and the activity of its promoter has been reported to be comparable to that of *PGK1*[56]. The resulting plasmid was transformed to produce the AXENO strain. This strain's growth was compared with that of the PIRPAZ control strain in quadruplicate for 165 h in CS medium with or without 2% xylose. No significant difference in growth was observed between the strains even after 8 days of incubation on CS plates containing 2% xylose (data not shown). Thus no effect was detected for the *A. thaliana* XI gene under expression of the *ENO1* promoter.

With evidence only for transcription, but not translation of either of the two gene candidates tested, it was assumed that active protein production was not occurring for one of two possible reasons: either the genes had been translated and not folded correctly, or the flanking base pairs on the mRNA molecules themselves were unsuitable for translation. To address both of these concerns, a new XI gene with high amino acid sequence identity (83%) to that of the previously proven *Piromyces* XI gene[28] was chosen as a final candidate. In addition, a commercially constructed plasmid vector created for high-level protein production was used.

3.4 *B. thetaiotaomicron* VPI-5482 XI was also transcribed but again not found to be translated.

The widely used expression vector pYES2 was chosen for subsequent experiments. This vector has been designed for high-level inducible expression using the yeast *GAL1* promoter and the *CYC1* transcription terminator. A Kozak protein translation initiation sequence[58] was included in the PCR primers. This sequence (ANNAUGG) occurs with high regularity flanking the start codon of most vertebrate mRNAs. Approximately 2-3 fold improved expression has been attributed to the presence of this sequence in *S. cerevisiae*.[62]

Two new recombinant strains denoted AYES and BYES, carrying the *A. thaliana* and *B. thetaiotaomicron* XI genes respectively, were constructed by ligating these genes into pYES2. The constructs were sequenced (Appendix A3 and A4) and only BYES showed sequence differences that may have been mutations. These differences included amino acid substitutions Val → Ala at position 165, Thr → Ala at position 249 and Ala → Thr at position 386. To verify that galactose-induced protein expression did occur, exponential phase cells were harvested from CS medium containing 2% glucose before being resuspended and grown in CS medium containing 2% galactose for a further 10 hours. Samples were removed at 2 hour intervals and total soluble protein was extracted for SDS-PAGE. No differences were observed at the predicted size over the 10 hour growth period (data not shown). This suggested that either the expressed proteins had only been translated to a low level or that the expression period selected was too short.

To address this latter concern, both strains were grown for 48 h on medium containing 2% galactose and 2% xylose after which HPLC was used to determine the consumption of xylose. Galactose was used as the primary carbon source to grow recombinant cells to a

high density whilst activating the promoter driving XI expression. Use of galactose also avoids catabolite repression and any potential inhibition of xylose transport[63]. No significant consumption was observed for either strain; in fact negative consumption of xylose was observed using the BYES strain (Table 4). This might have been due to the accuracy of the serial dilutions used or the reproducibility of the HPLC data. Interestingly, the AYES strain showed a ratio of xylose to galactose consumption the same as that of the control strain (2.44%).

Table 4: Consumption of galactose and xylose by AYES, BYES and PIRPAZ strains over 48 h [a]

Strain	Galactose			Xylose			$\Delta Xyl / \Delta Gal$
	0 h (g/l)	48 h (g/l)	Consumption	0 h (g/l)	48 h (g/l)	Consumption	
AYES	20	6.69	66.56%	20	19.67	1.63%	2.44%
BYES	20	16.49	17.53%	20	20.52	-2.60%	-14.84%
PIRPAZ	20	15.27	23.65%	20	19.88	0.58%	2.44%

[a] Constituent masses in proportion to peak volumes derived from HPLC data are shown at 0 and 48 h.

Since no data suggesting the activity or the presence of the translated XI proteins in the AYES and BYES strains could be detected, it was investigated whether XI gene transcription had occurred by RT-PCR. The EMPYES control strain was created for comparison with the AYES and BYES strains. The EMPYES plasmid was identical to both the AYES and BYES plasmids but lacked a XI gene insert. Exponential phase cells of all three strains were grown in CS medium containing 2% galactose at 30°C for 10 h. Total nucleic acid was then extracted from the cells, followed by a DNAseI treatment to remove DNA. cDNA was produced from the remaining RNA and RT-PCR used to detect transcription of the XI gene in the AYES, BYES and EMPYES strains. Positive evidence for mRNAs encoding the *A. thaliana* and *B. thetaiotaomicron* XI genes was found only in the cDNA of the relevant strains (Fig 12). The positive control to demonstrate successful cDNA production was *HSP12* - a constituent of the stress response pathway in yeast. Transcription of *HSP12* was visible in cDNA samples from all three strains. The negative controls (indicating no plasmid DNA remaining due to incomplete DNAseI digestion) were the RNA extract, as well as amplification of the plasmid using the pYES2 sequencing primers. No PCR products were observed for all negative controls.

All three attempts to express the XI gene were unsuccessful. Although transcription was shown to occur for all three XI genes used, no data suggesting XI translation or enzyme

activity was obtained. A bioinformatic analysis of the sequences was therefore performed to see whether this might provide an explanation.

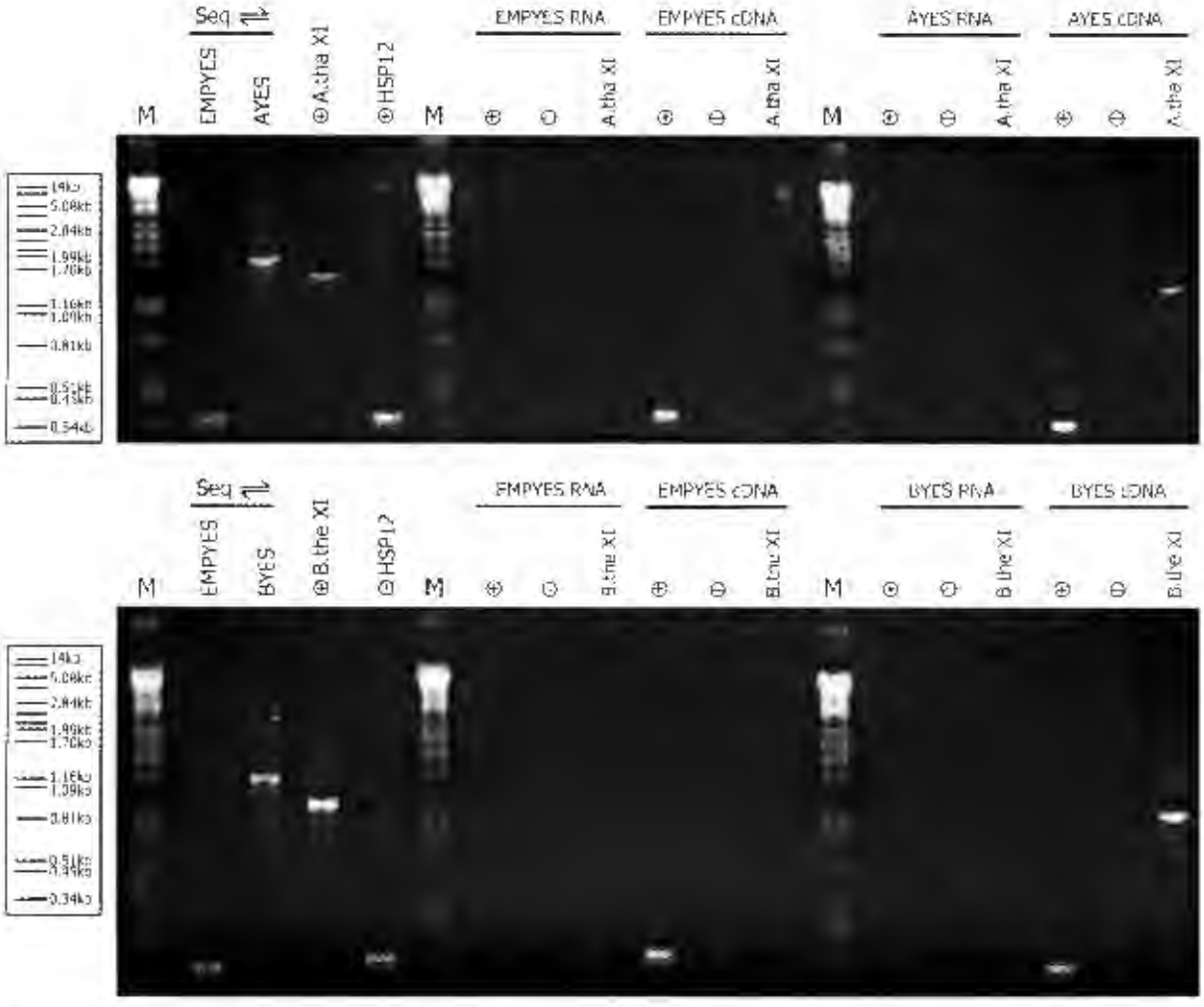


Figure 12: RT-PCRs of RNA and cDNA prepared from EMPYES (empty vector), AYES and BYES cells grown in galactose medium for 10 hours from exponential phase. All strains were tested for transcription of the XI gene (1.43kb for *A. thaliana* and 1.31kb for *B. thetaiotaomicron*), *HSP12* (positive control indicating successful cDNA production ⊕, 348bp) and amplification of pYES2 plasmid DNA with sequencing primers (negative control ⊖, ±317bp for EMPYES, ±1.75kb for AYES and ±1.63kb for BYES). The RNA extract was also used as a negative control to detect background plasmid DNA remaining from an incomplete DNAseI digestion. Transcription of the XI gene occurred only in the cDNA of the AYES and BYES strains. *HSP12* transcription occurred in the cDNA of all strains and no products were observed for both the RNA and sequencing primer negative controls. The first four lanes demonstrate amplification of the EMPYES and AYES/BYES plasmids with pYES2 primers, followed by amplification of *A. thaliana*/*B. thetaiotaomicron XI* and *HSP12* to demonstrate the sizes of the expected products. Extra bands visible at the top of the lanes for the control amplification of *HSP12* and the EMPYES cDNA *XI* in the AYES data are due to marker overloading. Lambda DNA digested with *PstI* was used as a marker (marker fragment sizes shown in box).

3.5 A translation-inhibiting stem-loop was found on the BYES mRNA

A variety of *in silico* analyses were performed on the sequence of each transformed plasmid. Previous studies in *S. cerevisiae* indicated that the presence of small ORFs upstream of the start codon could inhibit translation of the intended gene by between 60% and 90% and

decrease mRNA stability[64]. No small ORFs were detected within 50bp upstream of the start codons of four of the five plasmid sequences used (Appendix B). The AXENO plasmid, however, showed a single ORF of six amino acids 16bp upstream of the start codon which may have contributed to reduced XI translation in this strain. Sequences were searched for stop codons that may have caused premature termination. Stop codons were identified within 50bp upstream of the start codon in four of the five plasmids (not BYES), though these were unlikely to have affected translation since no start codons were found upstream of these stop codons. Although multiple alternative ORFs were found in each sequence's alternate frames, sequencing results for plasmids HPAZ, AXPAZ, AYES and BYES (Appendix A) showed no point insertions or deletions that would have made these relevant. mRNA stability has been reported to be reduced by the presence of nonsense codons in the sequence[65]. No such codons were detected.

Several authors[65,66,67] have described the inhibitory effect on translation of stem-loop structures in the mRNA leader region. Vega Laso *et al*[67] showed that a stem-loop of 15bp in length (stability of –30 kcal/mol) reduced translation by 98% and a smaller loop of 6bp (stability –14 kcal/mol) reduced translation by 65%. They concluded that hairpins with a stability of approximately –20 kcal/mol inhibit translation by up to 90% and promote mRNA degradation. An *in silico* analysis of 200bp on either side of the start codons of each plasmid revealed the presence of a significant stem-loop in one of the strains, BYES (Table 5).

Table 5: Thermodynamic stability of hairpin loops in the various mRNA sequences [a]

mRNA	Highest stability (kcal/mol)	2nd highest stability (kcal/mol) [b]	Position	Stem length (bp)	Loop length (bp)
HPAZ	-8.0	-4.0	+31 to +47	5	6
AXPAZ	-5.2	-4.0	+32 to +46	4	6
AXENO	-5.2	-4.2	+32 to +46	4	6
AYES	-5.2	-5.0	+35 to +49	4	6
BYES	-29.4	-5.0	-54 to -4	14	22

[a] Calculated with Vector NTI 10.1.1 software (Invitrogen, CA, USA)
[b] Included for comparison

The stability of this loop is high enough (-29.4 kcal/mol) for almost complete translational inhibition and its position is close enough to the start codon to strongly disrupt initiation of the translational machinery. This loop contains 33 base matches, 22 of which comprise the pairing of bases on the pYES2 vector sequence only (Fig 13). Only a third of the overall stability is due to 11 discontinuous base matches occurring between the insert and the vector. It is surprising that this inhibitory feature occurs on a commercial vector that is

widely used. The same stem-loop does not occur on the AYES mRNA because the 5` restriction site (*EcoRI*) exists 26bp upstream of the *NotI* site (Fig 13).

Figure 13: Structure of putative RNA stem-loops on the BYES and AYES sequences, as predicted by Vector NTI 10.1.1 software (Invitrogen, CA, USA). Purple bases: pYES2 vector; blue bases: *NotI* and *EcoRI* restriction sites respectively; green bases: XI gene; red bases: start codon. Vertical symbols (|) mark the predicted stem structure with plus symbols (+) marking contributing homology. Note that the majority of the complementarity on the BYES stem-loop occurred within the vector sequence and restriction site. The stem-loop shown for the AYES strain is of insignificant stability.

4. CONCLUSION

In this study three heterologous XI genes were evaluated by expression in *S. cerevisiae* followed by assays for the translated proteins, functionality of those proteins and transcription of the relevant genes. Despite evidence suggesting gene transcription to mRNA, no protein was found to be translated, nor was any activity detected for any of the genes.

It is possible that the SDS gels and Coomassie Blue stain used for direct protein visualisation were of insufficient sensitivity and that the activity of all translated protein products was too low for detection. More sensitive protein detection may have been achieved using MALDI-TOF analysis. The *H. influenzae* XI gene sequence showed seven substitution mutations (Appendix A1), though only one of these resulted in an amino acid change. These sequence aberrations may not have been PCR-induced mutations, but rather differences between the NCBI database sequence for the strain (used as the template for sequence alignment and comparison) and the actual sequencing result obtained from the strain used. The single amino acid substitution (Ser → Pro) may have reduced the activity of a translated protein to an undetectable level as proline is often found in β-bends resulting in a sharp turn in the peptide backbone. In the case of the *B. thetaiotaomicron* gene, the presence of a high-stability stem-loop before the start codon possibly inhibited translation to a high degree. Sequencing of the BYES plasmid (Appendix A4) revealed three mutations, each resulting in amino acid substitutions that may have reduced activity of the translated protein. As with the HPAZ strain however, these aberrations may have been differences between the strain sequenced and the online sequence records, rather than mutations due to the PCR. It is also possible that had both this gene and that of *H. influenzae* been translated, they would not have folded effectively since these were both prokaryotic sequences expressed in a eukaryotic environment. The eukaryotic *A. thaliana* gene may have been translated and even folded correctly, but exhibited activity too low for detection. Since no studies have reported a functional XI from *A. thaliana*, it is possible that this is either an inactive enzyme or a pseudo-gene.

Even with so few studies[28,49,31] showing effective heterologous XI expression in *S. cerevisiae*, the fermentation of plant biomass remains exceedingly viable via implementation of the XR and XDH enzyme pair and other modifications to the cellular machinery involved. New developments including sequencing of the *P. stipitis* genome[68] and the development of novel

screening techniques[69] will undoubtedly bring the effective and viable fermentation of biomass to biofuels closer to reality.

ACKNOWLEDGEMENTS

The author would like to thank his lab mates Vanessa Shamrock, Andrew Kessler, Grant Theron, Yvette Le Roux and Leela Kleinschmidt for the advice, support, materials and solutions shared, George Lindsey for his assistance as supervisor, as well as his friends and family for their valued moral support. The financial support of the NRF and UCT is also graciously acknowledged, as well as the commendable efficiency of both the MCB Department and Science faculty head office.

REFERENCES

1. Parikka M: **Global biomass fuel resources.** *Biomass Bioenergy* 2004, **27**: 613-620.

2. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ Jr, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T: **The path forward for Biofuels and Biomaterials.** *Science* 2006, **311**: 484-489.

3. Jeffries T: **Engineering yeasts for xylose metabolism.** *Curr. Opin. Biotechnol.* 2006, **17**: 320-326.

4. Thygesen A, Thomsen AB, Schmidt AS, Jørgensen H, Ahring BK, Olsson L: **Production of cellulose and hemicellulose-degrading enzymes by filamentous fungi cultivated on wet-oxidised wheat straw.** *Enzyme Microb. Technol.* 2003, **32**: 606-615.

5. Dudkin MS, Sorochan DV, Kozlov GF: **The structure of the water-soluble arabinoxylan of wheat endosperm.** *Chem. Nat. Compd.* 1977, **12**: 10-12.

6. Jin YS, Jeffries TW: **Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae* strains.** *Metab. Eng.* 2004, **6**: 229-238.

7. Grotkjaer T, Christakopoulos P, Nielsen J, Olsson L: **Comparative metabolic network analysis of two xylose fermenting recombinant *Saccharomyces cerevisiae* strains.** *Metab. Eng.* 2005, **7**: 437-444.

8. Roca C, Nielsen J, Olsson L: **Metabolic engineering of ammonium assimilation in xylose-fermenting *Saccharomyces cerevisiae* improves ethanol production.** *Appl. Environ. Microbiol.* 2003, **69**: 4732-4736.

9. Bro C, Regenberg B, Forster J, Nielsen J: ***In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production.** *Metab. Eng.* 2005, **8**: 102-111.

10. Jin YS, Jeffries TW: **Changing flux of xylose metabolites by altering expression of xylose reductase and xylitol dehydrogenase in recombinant *Saccharomyces cerevisiae*.** *Appl. Biochem. Biotechnol.* 2003, **105**: 277-285.

11. Maaheimo H, Fiaux J, Cakar ZP, Bailey JE, Sauer U, Szyperski T: **Central carbon metabolism of *Saccharomyces cerevisiae* explored by biosynthetic fractional ^{13}C labelling of common amino acids.** *Eur. J. Biochem.* 2001, **268**: 2464-2479.

12. Fiaux J, Cakar ZP, Sonderegger M, Wüthrich K, Szyperski T, Sauer U: **Metabolic-flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*.** *Eukaryotic Cell* 2003, **2**: 170-180.

13. Jin YS, Ni H, Laplaza JM, Jeffries TW: **Optimal growth and ethanol production from xylose by recombinant *Saccharomyces cerevisiae* require moderate D-xylulokinase activity.** *Appl. Environ. Microbiol.* 2003, **69**: 495-503.

14. Sonderegger M, Jeppsson M, Hahn-Hägerdal B, Sauer U: **Molecular basis for anaerobic growth of *Saccharomyces cerevisiae* on xylose, investigated by global gene expression and metabolic flux analysis.** *Appl. Environ. Microbiol.* 2004, **70**: 2307-2317.

15. Jin YS, Laplaza JM, Jeffries TW: ***Saccharomyces cerevisiae* engineered for xylose metabolism exhibits a respiratory response.** *Appl. Environ. Microbiol.* 2004, **70**: 6816-6825.

16. Salusjärvi L, Pitkänen JP, Aristidou A, Ruohonen L, Penttilä M: **Transcription analysis of recombinant *Saccharomyces cerevisiae* reveals novel responses to xylose.** *Appl. Biochem. Biotechnol.* 2006, **128(3)**: 237-261.

17. Lee WJ, Kim MD, Ryu YW, Bisson LF, Seo JH: **Kinetic studies on glucose and xylose transport in *Saccharomyces cerevisiae*.** *Appl. Microbiol. Biotechnol.* 2002, **60**: 186-191.

18. Gárdonyi M, Osterberg M, Rodrigues C, Spencer-Martins I, Hahn-Hägerdal B: **High capacity xylose transport in *Candida intermedia* PYCC 4715.** *FEMS Yeast Res.* 2003, **3**: 45-52.

19. Leandro MJ, Goncalves P, Spencer-Martins I: **Two glucose/xylose transporter genes from the yeast *Candida intermedia*: first molecular characterisation of a yeast xylose/H+ symporter.** *Biochem. J.* 2006, **395**: 543-549.

20. Saloheimo A, Rauta J, Stasyk OV, Sibirny AA, Penttilä M, Ruohonen L: **Xylose transport studies with xylose-utilizing *Saccharomyces cerevisiae* strains expressing heterologous and homologous permeases.** *Appl. Microbiol. Biotechnol.* 2007, **74**: 1041-1052.

21. Kötter P, Ciriacy M: **Xylose fermentation by *Saccharomyces cerevisiae*.** *Appl. Microbiol. Biotechnol.* 1993, **38**: 776-783.

22. Jeppsson M, Bengtsson O, Franke K, Lee H, Hahn-Hägerdal B, Gorwa-Grauslund MF: **The expression of a *Pichia stipitis* xylose reductase mutant with higher K_m for NADPH increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*.** *Biotechnol. Bioeng.* 2005, **93(4)**: 665-673.

23. Watanabe S, Kodaki T, Makino K: **Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc.** *J. Biol. Chem.* 2005, **280**: 10340-10349.

24. Watanabe S, Pack SP, Saleh AA, Annaluru N, Kodaki T, Makino K: **The positive effect of the decreased NADPH-preferring activity of xylose reductase from *Pichia stipitis* on ethanol production using xylose-fermenting recombinant *Saccharomyces cerevisiae*.** *Biosci. Biotechnol. Biochem.* 2007, **71(5)**: 1365-1369.

25. Rizzi M, Erleman P, Bui-Tanh NA, Dellweg H: **Purification and kinetic studies of xylose reductase from *Pichia stipitis*.** *Appl. Biochem. Biotechnol.* 1988, **29**: 148-154.

26. Verho R, Richard P, Jonson PH, Sundqvist L, Londesborough J, Penttilä M: **Identification of the first fungal NADP-GAPDH from Kluyveromyces lactis.** *Biochemistry* 2002, **41(46)**: 13833-13838.

27. Verho R, Putkonen M, Londesborough J, Penttilä M, Richard P: **A novel NADH-linked L-xylulose reductase in the L-arabinose catabolic pathway of yeast.** *J. Biol. Chem.* 2004, **279**: 14746-14751.

28. Kuyper M, Harhangi HR, Stave AK, Winkler AA, Jetten MS, de Laat WT, den Ridder JJ, Op den Camp HJ, van Dijken JP, Pronk JT: **High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*?** *FEMS Yeast Res.* 2003, **4**: 69-78.

29. Kuyper M, Winkler AA, van Dijken JP, Pronk JT: **Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: a proof of principle.** *FEMS Yeast Res.* 2004, **4(6)**: 655-664.

30. Kuyper M, Toirkens MJ, Diderich JA, Winkler AA, van Dijken JP, Pronk JT: **Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting *Saccharomyces cerevisiae* strain.** *FEMS Yeast Res.* 2005, **5(10)**: 925-934.

31. Karhuuma K, Hahn-Hägerdal B, Gorwa-Grauslund MF: **Investigation of limiting metabolic steps in the utilization of xylose by recombinant *Saccharomyces cerevisiae* using metabolic engineering.** *Yeast* 2005, **22(5)**: 359-68.

32. Sonderegger M, Sauer U: **Evolutionary engineering of *Saccharomyces cerevisiae* for anaerobic growth on xylose.** *Appl. Environ. Microbiol.* 2003, **69(4)**: 1990-1998.

33. Sonderegger M, Jeppsson M, Larsson C, Gorwa-Grauslund MF, Boles E, Olsson L, Spencer-Martins I, Hahn-Hägerdal B, Sauer U: **Fermentation performance of engineered and evolved xylose-fermenting *Saccharomyces cerevisiae* strains.** *Biotechnol. Bioeng.* 2004, **87(1)**: 90-98.

34. Attfield PV, Bell PJ: **Use of population genetics to derive nonrecombinant *Saccharomyces cerevisiae* strains that grow using xylose as a sole carbon source.** *FEMS Yeast Res.* 2006, **6(6)**: 862-868.

35. Pitkänen JP, Rintala E, Aristidou A, Ruohonen L, Penttilä M: **Xylose chemostat isolates of *Saccharomyces cerevisiae* show altered metabolite and enzyme levels compared with xylose, glucose, and ethanol metabolism of the original strain.** *Appl. Microbiol. Biotechnol.* 2005, **67(6)**: 827-837.

36. Jin YS, Alper H, Yang YT, Stephanopoulos G: **Improvement of xylose uptake and ethanol production in recombinant *Saccharomyces cerevisiae* through an inverse metabolic engineering approach.** *Appl. Environ. Microbiol.* 2005, **71(12)**: 8249-8256.

37. Ni H, Laplaza JM, Jeffries TW: **Transposon mutagenesis to improve the growth of recombinant *Saccharomyces cerevisiae* on D-xylose.** *Appl. Environ. Microbiol.* 2007, **73(7)**: 2061-2066.

38. Sundström M, Lindqvist Y, Schneider G, Hellman U, Ronne H: **Yeast *TKL1* gene encodes a transketolase that is required for efficient glycolysis and biosynthesis of aromatic amino acids.** *J. Biol. Chem.* 1993, **268**: 24346-24352.

39. Jin YS, Cruz, J, Jeffries TW: **Xylitol production by a *Pichia stipitis* D-xylulokinase mutant.** *Appl. Microbiol. Biotechnol.* 2005, **68**: 42-45.

40. Sonderegger M, Schümperli M, Sauer U: **Metabolic engineering of a phosphoketolase pathway for pentose catabolism in *Saccharomyces cerevisae*.** *Appl. Environ. Microbiol.* 2004, **70(5)**: 2892-2897.

41. Becker J, Boles E: **A modified *Saccharomyces cerevisiae* strain that consumes L-arabinose and produces ethanol.** *Appl. Environ. Microbiol.* 2003, **69(7)**: 4144-4150.

42. Crous JM, Pretorius IS, van Zyl WH: **Cloning and expression of an *Aspergillus kawachii* endo-1,4-β-xylanase gene in *Saccharomyces cerevisiae*.** *Curr. Genet.* 1995, **28**: 467-473.

43. La Grange DC, Pretorius IS, Claeyssens M, van Zyl WH: **Degradation of xylan to D-xylose by recombinant *Saccharomyces cerevisiae* coexpressing the *Aspergillus niger* β−xylosidase (xlnD) and the *Trichoderma reesei* xylanase II (xyn2) genes.** *Appl. Environ. Microbiol.* 2001, **67(12)**: 5512-5519.

44. Katahira S, Fujita Y, Mizuike A, Fukuda H, Kondo A: **Construction of a xylan-fermenting yeast strain through codisplay of xylanolytic enzymes on the surface of xylose-utilizing *Saccharomyces cerevisiae* cells.** *Appl. Environ. Microbiol.* 2004, **70(9)**: 5407-5414.

45. Andrés I, Gallardo O, Parascandola P, Pastor FI, Zeuco J: **Use of the cell wall protein Pir4 as a fusion partner for the expression of *Bacillus* sp. BP-7 xylanase A in *Saccharomyces cerevisiae*.** *Biotechnol. Bioeng.* 2005, **89(6)**: 690-697.

46. Katahira S, Mizuike A, Fukuda H, Kondo A: **Ethanol fermentation from lignocellulosic hydrolysate by a recombinant xylose- and cellooligosaccharide-assimilating yeast strain.** *Appl. Microbiol. Biotechnol.* 2006, **72**: 1136-1143.

47. Ohgren K, Bengtsson O, Gorwa-Grauslund MF, Galbe M, Hahn-Hägerdal B, Zacchi G: **Simultaneous saccharification and co-fermentation of glucose and xylose in steam-pretreated corn stover at high fiber content with *Saccharomyces cerevisiae* TMB3400.** *J. Biotechnol.* 2006, **126(4)**:488-498.

48. Martín C, Marcet M, Almazán O, Jönsson LJ: **Adaptation of a recombinant xylose-utilizing *Saccharomyces cerevisiae* strain to a sugarcane bagasse hydrolysate with high content of fermentation inhibitors.** *Bioresour. Technol.* 2007, **98(9)**: 1767-1773.

49. Voronovsky AY, Ryabova OB, Verba OV, Ishchuk OP, Dmytruk KV, Sibirny AA: **Expression of *xylA* genes encoding xylose isomerases from *Escherichia coli* and *Streptomyces coelicolor* in the methylotrophic yeast *Hansenula polymorpha*.** *FEMS Yeast Res.* 2005, **5**: 1055-1062.

50. Karhumaa K, Garcia Sanchez R, Hahn-Hägerdal B, Gorwa-Grauslund MF: **Comparison of the xylose reductase-xylitol dehydrogenase and the xylose isomerase pathways for xylose fermentation by recombinant *Saccharomyces cerevisiae*.** *Microb. Cell Fact.* 2007, **6**:5 Published online.

51. Sambrook J, Fritsch EF, Maniatis T: **Molecular cloning: A laboratory manual, 2nd ed.** Cold Spring Harbour Laboratory Press (1989), Cold Spring Harbour, New York.

52. Delorme E: **Transformation of *Saccharomyces cerevisiae* by electroporation.** *Appl. Environ. Microbiol.* 1989, **55**: 2242-2246.

53. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, **227**: 680-685.

54. Chomczynski P, Sacchi N: **Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.** *Anal. Biochem.* 1987, **162(1)**: 156-159.

55. Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR: **Empirical analysis of transcriptional activity in the Arabidopsis genome.** *Science* 2003, **302**: 842-846.

56. Den Haan R, Rose SH, Lynd LR, van Zyl WH: **Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*.** *Metab. Eng.* 2007, **9**: 87-94.

57. Giniger E, Varnum SM, Ptashne M: **Specific DNA binding of GAL4, a positive regulatory protein of yeast.** *Cell* 1985, **40**: 767-774.

58. Kozak M: **An analysis of vertebrate mRNA sequences: intimations of translational control.** *J. Cell Biol.* 1991, **115(4)**: 887-903.

59. Govindaswamy S, Vane LM: **Kinetics of growth and ethanol production on different carbon substrates using genetically engineered xylose-fermenting yeast.** *Bioresour. Technol.* 2007, **98(3)**: 677-685.

60. Walfridsson M, Anderlund M, Bao X, Hahn-Hägerdal: **Expression of different levels of enzymes from the *Pichia stipitis XYL1* and *XYL2* genes in *Saccharomyces cerevisiae* and its effects on product formation during xylose utilisation.** *Appl Microbiol Biotechnol* 1997, **48**: 218-224.

61. Chambers A, Tsang JS, Stanway C, Kingsman AJ, Kingsman SM: **Transcriptional control of the *Saccharomyces cerevisiae PGK* gene by *RAP1*.** *Mol. Cell. Biol.* 1989, **9(12)**: 5516-5524.

62. Baim SB, Sherman F: **mRNA structures influencing translation in the yeast *Saccharomyces cerevisiae*.** *Mol. Cel. Biol.* 1988, **8(4)**: 1591-1601.

63. Hahn-Hägerdal B, Wahlbom CF, Gárdonyi M, van Zyl WH, Cordero Otero RR, Jönsson LJ: **Metabolic engineering of *Saccharomyces cerevisiae* for xylose utilization.** *Adv. Biochem. Eng. Biotechnol.* 2001, **73:** 53–84.

64. Oliveira CC, McCarthy JEG: **The relationship between eukaryotic translation and mRNA stability.** *J. Biol. Chem.* 1995, **270(15)**: 8936-8943.

65. McCarthy JEG: **Posttranscriptional control of gene expression in yeast.** *Microbiol. Mol. Biol. Rev.* 1998, **62(4)**: 1492-1553.

66. Oliveira CC, van den Heuvel JJ, McCarthy JEG: **Inhibition of translational initiation in *Saccharomyces cerevisiae* by secondary structure: the roles of the stability and position of stem-loops in the mRNA leader.** *Mol. Microbiol.* 1993, **9(3)**: 521-532.

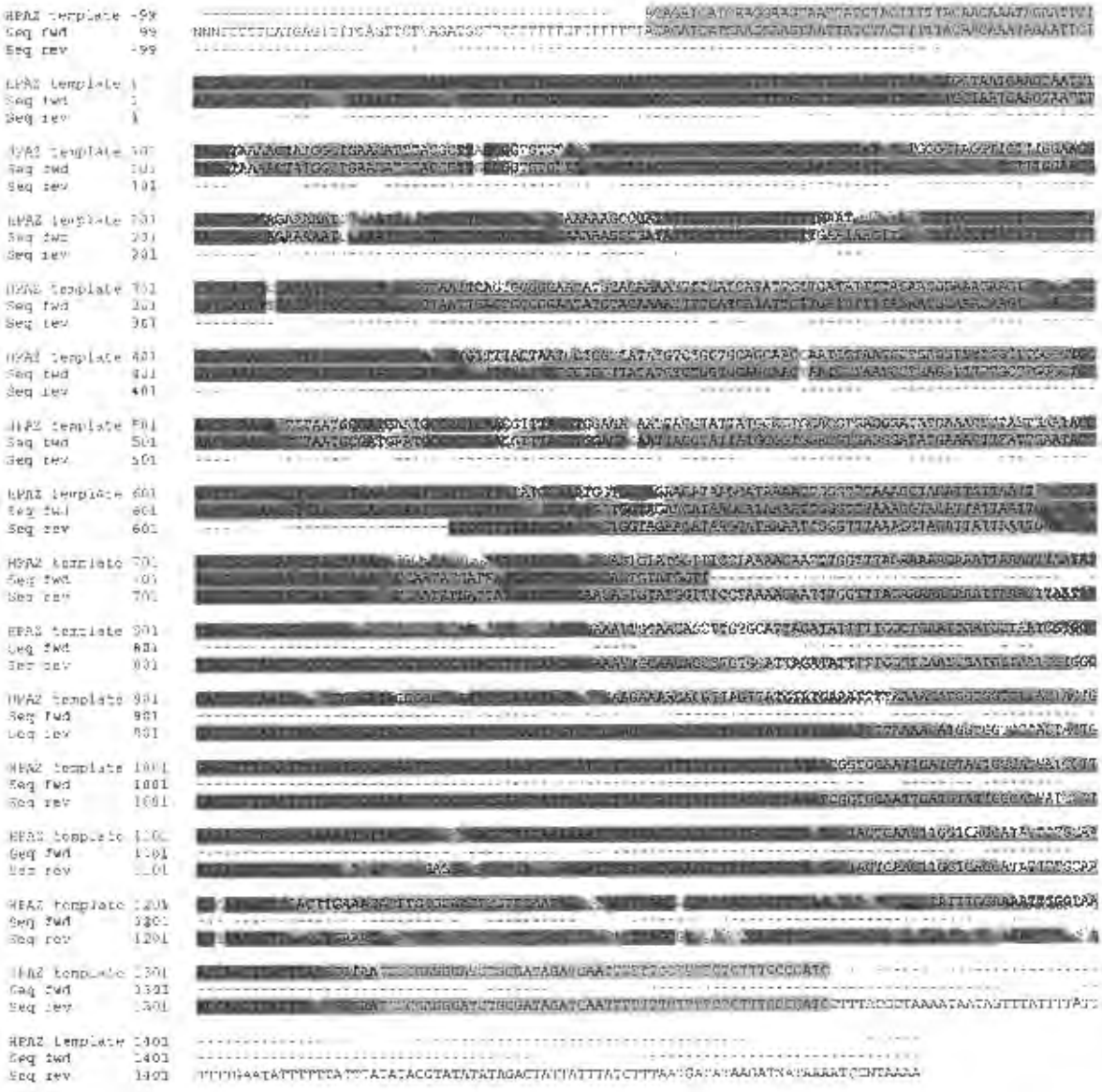
67. Vega Laso MR, Zhu D, Sagliocco F, Brown AJ, Tuite MF, McCarthy JE: **Inhibition of translational initiation in the yeast *Saccharomyces cerevisiae* as a function of the stability and position of hairpin structures in the mRNA leader.** *J. Biol. Chem.* 1993, **268(9)**: 6453-6462.

68. Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, Salamov A, Schmutz J, Lindquist E, Dehal P, Shapiro H, Jin YS, Passoth V, Richardson PM: **Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*.** *Nat. Biotechnol.* 2007, **25(3)**: 319-326.

69. Grabek-Lejko D, Ryabova OB, Oklejewicz B, Voronovsky AY, Sibirny AA: **Plate ethanol-screening assay for selection of the *Pichia stipitis* and *Hansenula polymorpha* yeast mutants with altered capability for xylose alcoholic fermentation.** *J. Ind. Microbiol. Biotechnol.* 2006, **33**: 934-940.

APPENDIX A1

Comparison of the HPAZ plasmid sequence data obtained with published sequence data (Genbank accession NC_000907). Seq fwd: sequence obtained using the forward sequencing primer; Seq rev: sequence obtained using the reverse sequencing primer.

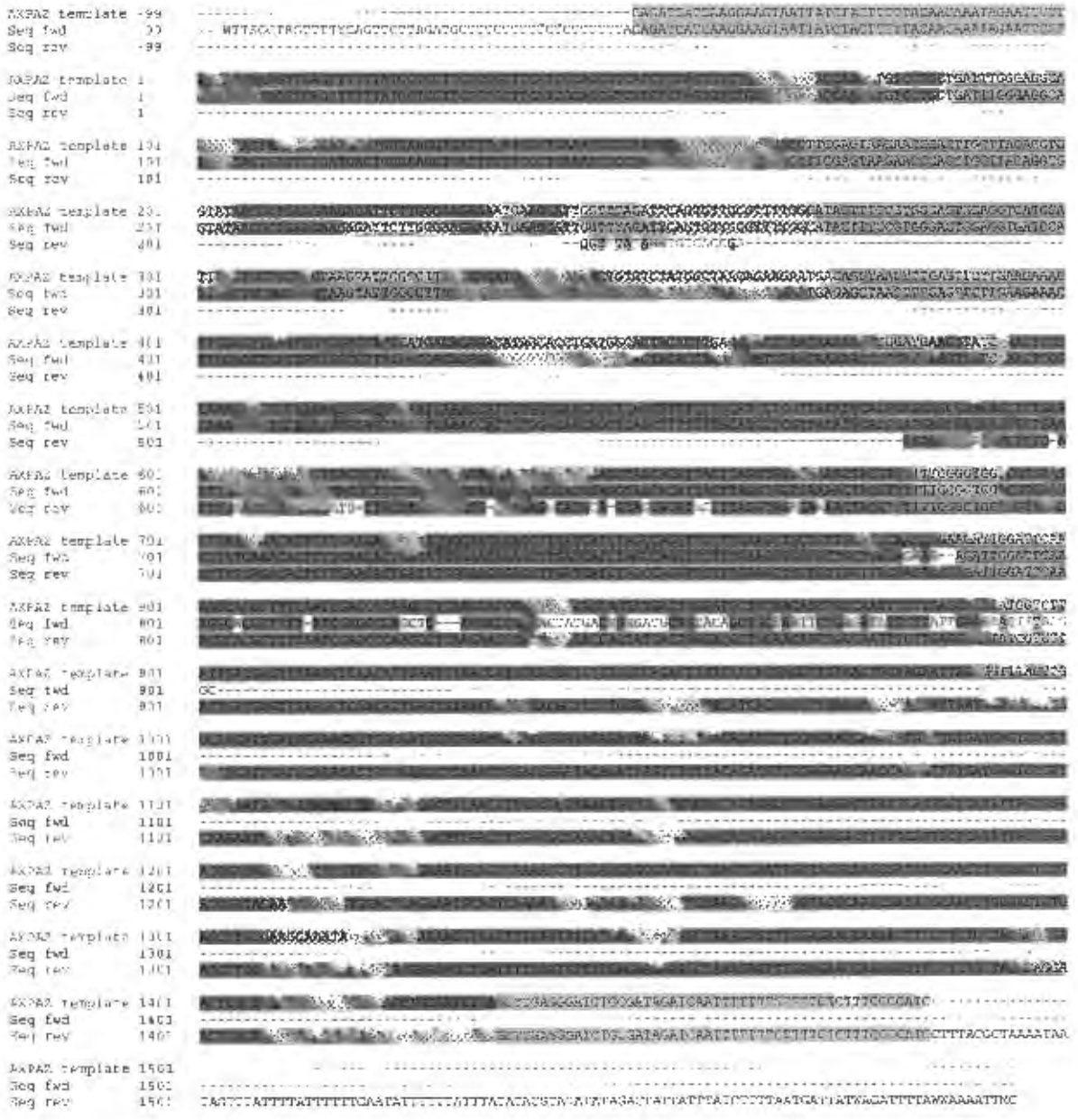


Summary of differences:

132bp	Fwd primer only	G in place of A	Leu in place of Leu
217bp	Fwd primer only	C in place of T	Pro in place of Ser
309bp	Fwd primer only	T in place of C	Val in place of Val
468bp	Fwd primer only	T in place of C	Thr in place of Thr
549bp	Fwd primer only	G in place of A	Glu in place of Glu
1203bp	Rev primer only	T in place of C	Gly in place of Gly
1254bp	Rev primer only	G in place of A	Ala in place of Ala

APPENDIX A2

Comparison of the AXPAZ plasmid sequence data obtained with published sequence data (Genbank accession AY136469). Seq fwd: sequence obtained using the forward sequencing primer; Seq rev: sequence obtained using the reverse sequencing primer.

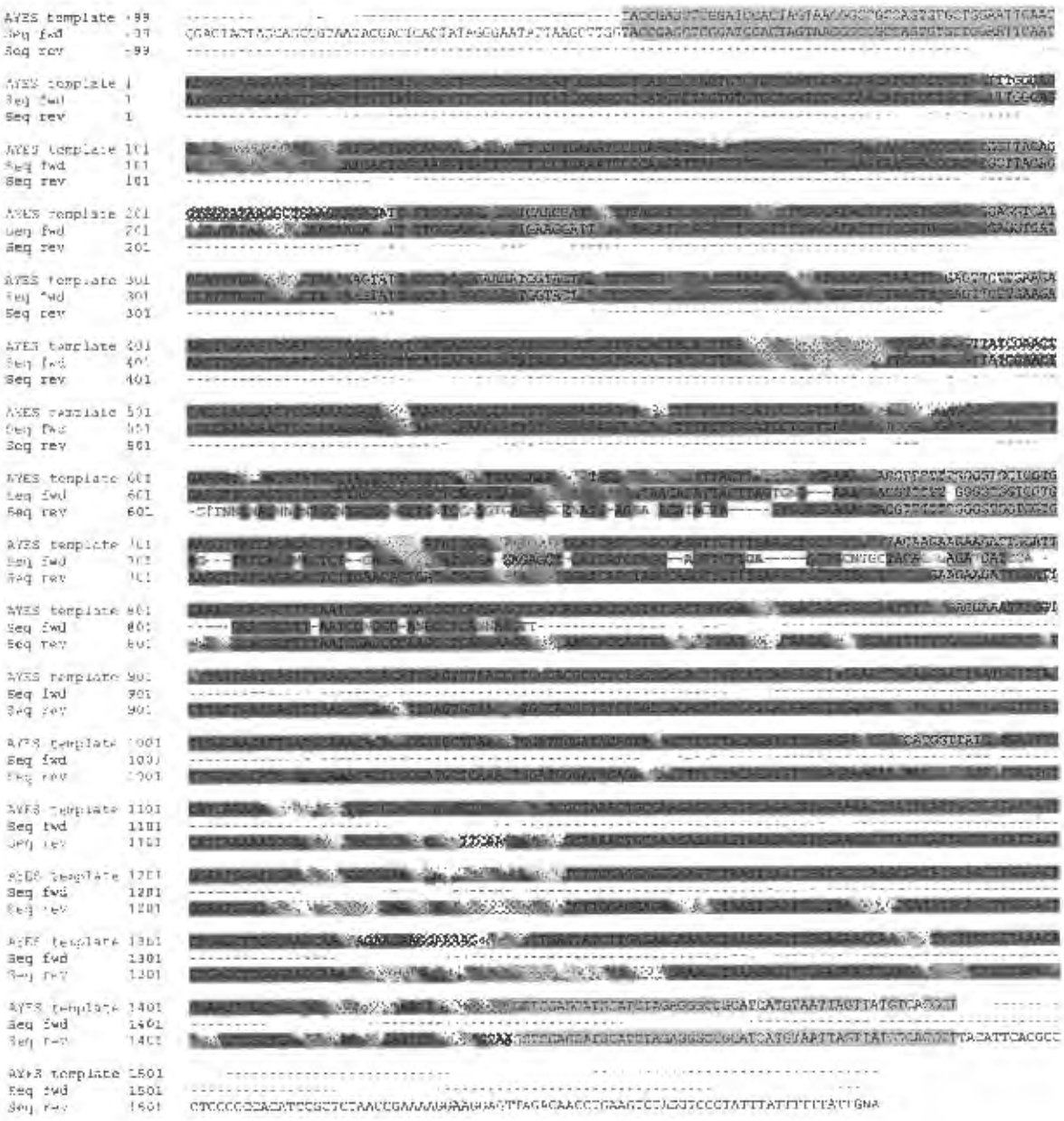


Summary of differences:

None

APPENDIX A3

Comparison of the AYES plasmid sequence data obtained with published sequence data (Genbank accession AY136469). Seq fwd: sequence obtained using the forward sequencing primer; Seq rev: sequence obtained using the reverse sequencing primer.

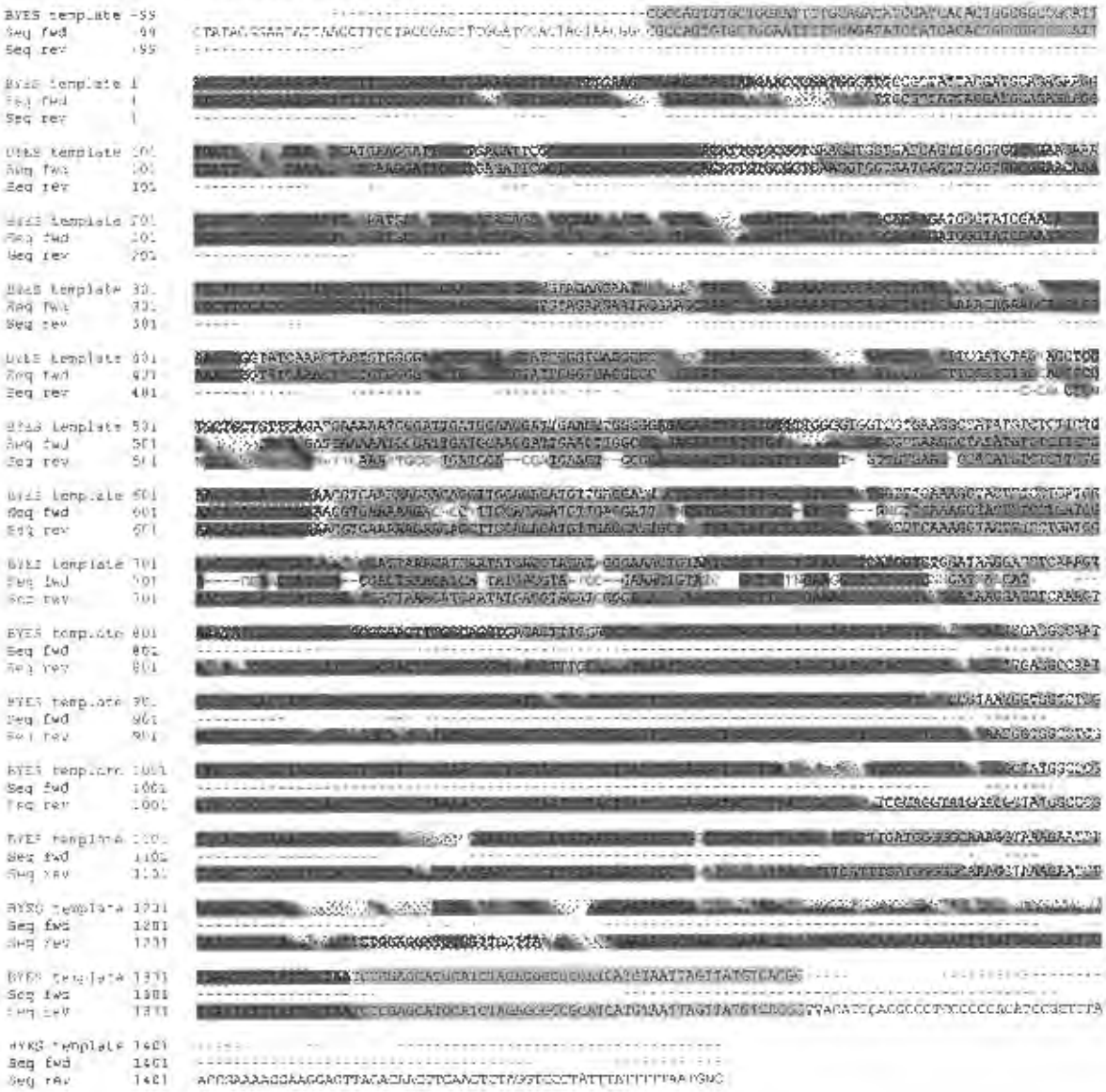


Summary of differences:

None

APPENDIX A4

Comparison of the BYES plasmid sequence data obtained with published sequence data (Genbank accession NC_004663). Seq fwd: sequence obtained using the forward sequencing primer; Seq rev: sequence obtained using the reverse sequencing primer.



Summary of differences:

494bp Fwd primer only	C in place of T	Ala in place of Val
745bp Rev primer only	G in place of A	Ala in place of Thr
1156bp Rev primer only	A in place of G	Thr in place of Ala

APPENDIX B

Open reading frames detected within 50bp upstream of the start of translation for all plasmids transformed into yeast.

Possible Open Reading Frame in HPAZ template(1-1370)
Universal code

Strand	RF	AA Num	Position	Sequence
Plus	3	439	51-1370	atagaattctATGacaac...atagcTAA
Plus	1	39	1033-1152	attttaaaacATGgtggt...atcctTAAaacgt
Plus	1	33	1201-1302	aaaattgttaATGaacgc...agattTAGcacca
Plus	1	25	352-429	tattgttttcATGatgtc...tatttTAGaacgc
Plus	1	11	778-813	aaacatcaatATGattat...tttccTAAaacaa
Plus	2	8	464-490	taaaactcttATGgggaa...tttacTAAtcctc
Plus	1	8	892-918	acctttcaacATGaaatt...tgcatTAGatatt
Plus	1	8	937-963	ggctcaattgATGctaat...acaatTAGgctgg
Plus	1	7	214-237	ttttgctggaATGggaat...tgggcTAGgttct
Minus	1	35	1243-1350	tagcatcataATGtttga...gtagaTAAttact
Minus	3	28	228-314	tttaaggataATGccaat...aaaatTAAaacct
Minus	1	24	658-732	acccaattttATGcttat...ttcaaTAAagttt
Minus	1	22	478-546	ttgcaatttcATGttgaa...ttttcTAAaccaa
Minus	3	19	456-515	aaaatatctaATGcacaa...gtggtTAGcttca
Minus	1	18	130-186	cttgcaaaatATGctgac...ttttgTAAggttt
Minus	1	11	598-633	aatcatattgATGtttag...tcaatTAAtaatg
Minus	1	10	958-990	tatcaacaatATGatgaa...cgcacTGAattac
Minus	1	8	1018-1044	tatcgacatcATGaaaac...acgccTAActtat
Minus	1	7	1198-1221	ctaagcgtaaATGttcag...ttaccTAAaatta
Minus	1	6	1171-1191	agcaaaaggtATGccaat...accgcTAAgcgta

Possible Open Reading Frame in AYES template(1-1487)
Universal code

Strand	RF	AA Num	Position	Sequence
Plus	3	478	51-1487	ggaattcaatATGggtaa...caatgTAA
Plus	1	65	946-1143	ttgaggaaatATGgtctt...ggttaTGAtgagt
Plus	1	62	499-687	atagcacctgATGgcact...tcaggTGAagaaa
Plus	1	40	169-291	agtgattctgATGactgg...gaaaaTGAaggat
Plus	1	39	1159-1278	gtcatcaaaaATGgtggg...tggacTGAgaaat
Plus	2	33	1091-1192	ctcaaactggATGggata...aacttTGAcgcta
Plus	1	12	388-426	ccttgggaagATGgtact...aagaaTGAgagct
Plus	1	11	904-939	aagcaccagtATGactgg...tttctTGAggaaa
Plus	1	11	1336-1371	cgcaagcgatATGcaact...gcaaaTAGaagaa
Plus	1	10	1282-1314	ggactgagaaATGcagtc...aagtcTAAgtgaa
Minus	1	200	778-1380	cacctaagtaATGtgtta...gacacTAGagatg
Minus	1	50	472-624	caagctcgtgATGacaag...tcgatTAAaagcg
Minus	2	44	251-385	aatatgagcaATGaataa...gaaacTGAtctgt
Minus	2	31	1394-1489	tgacgctgcgATGaagca...
Minus	2	31	512-607	gttacactcaATGttgag...gttctTGAggctt
Minus	1	28	688-774	acctggctagATGatcaa...ccaccTAAgtaat
Minus	3	15	1134-1181	gcagcaccaaATGgatca...aacacTGAatcta
Minus	2	14	428-472	gtttgcatcaATGttgcc...gctcgTGAtgaca
Minus	3	13	1356-1397	tcagcaggacATGttggt...tgcgaTGAagcag

Possible Open Reading Frame in AXPAZ template(1-1484)
Universal code

Strand	RF	AA Num	Position	Sequence
Plus	3	477	51-1484	tagaattcctATGaagaa...caatgTAA
Plus	1	65	943-1140	ttgaggaaatATGgtctt...ggttaTGAtgagt
Plus	1	62	496-684	atagcacctgATGgcact...tcaggTGAagaaa
Plus	1	40	166-288	agtgattctgATGactgg...gaaaaTGAaggat
Plus	1	39	1156-1275	gtcatcaaaaATGgtggg...tggacTGAgaaat
Plus	2	33	1088-1189	ctcaaactggATGggata...aacttTGAcgcta
Plus	1	12	385-423	ccttgggaagATGgtact...aagaaTGAgagct
Plus	1	11	901-936	aagcaccagtATGactgg...tttctTGAggaaa
Plus	1	11	1333-1368	cgcaagcgatATGcaact...gcaaaTAGaagaa
Plus	1	10	1279-1311	ggactgagaaATGcagtc...aagtcTAAgtgaa
Minus	1	200	778-1380	cacctaagtaATGtgtta...gacacTAGagatg
Minus	1	50	472-624	caagctcgtgATGacaag...tcgatTAAaagcg
Minus	2	44	251-385	aatatgagcaATGaataa...gaaacTGAtctgt
Minus	2	31	512-607	gttacactcaATGttgag...gttctTGAggctt
Minus	1	28	688-774	acctggctagATGatcaa...ccaccTAAgtaat
Minus	2	23	1394-1465	tgacgctgcgATGaagca...gtagaTAAttact
Minus	3	15	1134-1181	gcagcaccaaATGgatca...aacacTGAatcta
Minus	2	14	428-472	gtttgcatcaATGttgcc...gctcgTGAtgaca
Minus	3	13	1356-1397	tcagcaggacATGttggt...tgcgaTGAagcag

Possible Open Reading Frame in BYES template(1-1367)
Universal code

Strand	RF	AA Num	Position	Sequence
Plus	3	438	51-1367	cggccgcattATGgcaac...attgcTAA
Plus	1	75	1039-1266	attatccgtaATGgtggt...caagcTGActctg
Plus	1	33	265-366	ttcccatggaATGgtaat...tgacgTAGacttg
Plus	1	29	922-1011	gctgtagacaATGgtatg...tgaacTGActcag
Plus	1	26	430-510	atcgtagcttATGcaaaa...ctataTGAacggt
Plus	1	26	571-651	cagatcaaaaATGcgatt...tctttcTGAacaca
Plus	1	25	1276-1353	actctggaggATGtggtt...aattcTGAatatg
Plus	1	11	712-747	gctcgtgactATGcccgt...tttccTGAtcgaa
Plus	1	8	826-852	ctgaaagctcATGgtctg...caaagTAAatatc
Plus	2	7	197-220	tcgctatggcATGgtggc...tgcgcTGAaggtg
Minus	1	144	469-903	cagccaattcATGctcga...cacagTAGtttga
Minus	1	119	1009-1368	agtctacgtcATGgaagc...
Minus	2	31	1043-1138	catcttctgcATGaattc...cgaacTGAtcacc
Minus	3	27	144-227	cccccatcaaATGaagca...ggccaTAGcgtcc
Minus	3	23	1107-1178	ttaccattccATGggaat...tgccaTAGcgaat
Minus	3	12	1242-1280	taataacggaATGccatc...aaatttAAtcttt

Possible Open Reading Frame in AXENO template(1-1484)
Universal code

Strand	RF	AA Num	Position	Sequence
Plus	3	477	51-1484	ccgaattcctATGaagaa...caatgTAA
Plus	1	65	943-1140	ttgaggaaatATGgtctt...ggttaTGAtgagt
Plus	1	62	496-684	atagcacctgATGgcact...tcaggTGAagaaa
Plus	1	40	166-288	agtgattctgATGactgg...gaaaaTGAaggat
Plus	1	39	1156-1275	gtcatcaaaaATGgtggg...tggacTGAgaaat
Plus	2	33	1088-1189	ctcaaactggATGggata...aacttTGAcgcta
Plus	1	12	385-423	ccttgggaagATGgtact...aagaaTGAgagct
Plus	1	11	1333-1368	cgcaagcgatATGcaact...gcaaaTAGaagaa
Plus	1	11	901-936	aagcaccagtATGactgg...tttctTGAggaaa
Plus	1	10	1279-1311	ggactgagaaATGcagtc...aagtcTAAgtgaa
Plus	1	6	34-54	tcaaaactgtATGgaaac...tcctaTGAagaaa
Minus	1	200	778-1380	cacctaagtaATGtgtta...gacacTAGagatg
Minus	1	50	472-624	caagctcgtgATGacaag...tcgatTAAaagcg
Minus	2	44	251-385	aatatgagcaATGaataa...gaaacTGAtctgt
Minus	2	31	512-607	gttacactcaATGttgag...gttctTGAggctt
Minus	2	30	1394-1486	tgacgctgcgATGaagca...
Minus	1	28	688-774	acctggctagATGatcaa...ccaccTAAgtaat
Minus	3	15	1134-1181	gcagcaccaaATGgatca...aacacTGAatcta
Minus	2	14	428-472	gtttgcatcaATGttgcc...gctcgTGAtgaca
Minus	3	13	1356-1397	tcagcaggacATGttggt...tgcgaTGAagcag