Linking single nucleotide polymorphisms to hardwood spent sulfite liquor tolerance in evolutionary engineered *Saccharomyces cerevisiae*

David Colatriano

A Thesis

in

The Department

of

Biology

Presented in partial fulfillment of the requirements

for the Degree of Master of Science (Biology) at

Concordia University

Montréal, Québec, Canada

April 2013

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: David Colatriano

Entitled: Linking single nucleotide polymorphisms to hardwood spent sulfite liquor tolerance in evolutionary engineered *Saccharomyces cerevisiae*

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Selvadurai Dayanandan Chair

Dr. David Walsh External Examiner

Dr. Christopher Brett Examiner

Dr. Luc Varin Examiner

Dr. Vincent Martin Supervisor

Approved by Dr. Selvadurai Dayanandan

Chair of Department or Graduate Program Director

April 2013 _____

Dean of Faculty

Abstract

Linking single nucleotide polymorphisms to hardwood spent sulfite liquor tolerance in evolutionary engineered *Saccharomyces cerevisiae*

David Colatriano

A strain of *Saccharomyces cerevisiae* (R57) highly tolerant to inhibitors found in the lignocellulosic substrate hardwood spent sulfite liquor (HWSSL) was created through genome shuffling. The goal of my research was to identify the mutations responsible for the tolerance phenotype. The genomes of R57 and wild type strain (CEN.PK 113-7D) were previously sequenced and compared for mutations (Dominic Pinel unpublished results). Forty-seven single nucleotide polymorphisms (SNPs) predicted by genome sequence comparison were PCR amplified from R57 genomic DNA for confirmation by Sanger sequencing. Of the 47, 20 were confirmed affecting 17 genes. These SNP-containing regions were amplified from mixed populations of cells from the UV mutagenesis treatment and rounds 1, 3 and 5 of the evolutionary engineered pool of cells. These amplicons were sequenced using the Ion Torrent sequencer revealing the frequencies of each SNP from each pool of cells as they progressed through the evolution. Mutated regions amplified from isolated colonies taken from the growth frontier of HWSSL gradient plates for the UV mutagenized and round 5 pools were sequenced using Sanger sequencing to reveal specific combinations of SNPs. The homozygous *UBP7* c.2466T>A mutation was reconstituted in a diploid wild type background strain and tested against the diploid wild type and R57 strains. The *UBP7* mutant showed increased tolerance to HWSSL. The combination of this data led to the

hypothesis that of the 17 directly affected genes, 6 are more likely to play a role in the tolerance phenotype observed in R57 (*UBP7, NRG1, MAL11, GDH1, GSH1* and *ART5*).

Acknowledgements

I would like to thank my supervisor Dr. Martin for his constant support, guidance and invaluable advice on this project.

I would also like to thank Dr. Brett and Dr. Varin for sitting on my committee.

I would like to thank the entire Martin Lab for their friendship and willingness to lend their expertise whenever needed. I would like to give special thanks to Dominic Pinel who helped me learn what I needed to get started on this project as well as his invaluable advice and expertise on the project, and to Nicholas Gold who was always willing to help troubleshoot problems and to discuss possible new ideas.

Lastly, I would also like to thank my family and friends for their endless love, support and patience.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5-HMF	5-hydroxymethylfurfural
AA	Amino acid
ABC	ATP-binding cassette
ADH	Alcohol dehydrogenase
ALDH	Aldehyde dehydrogenase
ATP	Adenosine triphosphate
Blast	Basic local alignment search tool
bp	Base pair
C1	Common region 1
C6	Common region 6
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
HOG	High osmolarity glycerol
HSP	Heat shock protein
HWSSL	Hardwood spent sulfite liquor
kb	Kilobase
LB	Luria broth
MAPK	Mitogen-activated protein kinase
MAQ	Mapping and Assembly with Qualities
MDR	Multidrug response
MVB	Multivesicular body
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
ORF	Open reading frame
PCR	Polymerase chain reaction
PDH	Pyruvate dehydrogenase
PDR	Pleiotropic drug response
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SD	Synthetic defined
SGD	*Saccharomyces cerevisiae* genome database
SNP	Single nucleotide polymorphism
SSL	Spent sulfite liquor
SWSSL	Softwood spent sulfite liquor
UTR	Untranslated region
WT	Wild type
wt/vol	Weight per volume
YNB	Yeast nitrogen base
YPD	Yeast-peptone-dextrose

1. Introduction

1.1 Introduction to the problem, goal and objectives

Hardwood spent sulfite liquor (HWSSL) is a readily available lignocellulosic hydrolysate with the potential to be fermented to ethanol. Unfortunately, it contains many inhibitors to the growth of one of the most readily used fermentative organisms, *Saccharomyces cerevisiae*. In previous work, *S. cerevisiae* strain R57 was engineered through a process of genome shuffling to be resistant to the inhibitors found in HWSSL [1].

The process of genome shuffling used to create this strain utilized UV mutagenesis and as such, the mutations conferring the tolerance phenotype were unknown. Although the genome of R57 was sequenced and SNPs were identified, the contribution of each SNP with respect to the tolerance phenotype was still unknown.

My hypothesis was that there were a few important SNPs that when combined through recursive mating would act synergistically to account for the tolerance phenotype observed in R57. Therefore, the goal of this work was to narrow down the number of SNPs responsible for the phenotype and to identify which of the mutations were real and most relevant to the observed tolerance phenotype. This information would be very useful for identifying new gene targets to study for increasing tolerance to many of the inhibitors common amongst most lignocellulosic hydrolysates.

Two methods were attempted in this study. The first method utilized was to track the frequency of these same SNPs through several rounds of the genome shuffling that led to the R57 strain. This was achieved by growing up and plating frozen stocks of the

shuffled rounds on HWSSL gradient plates, scraping the growth that grew above the wild type frontier of confident growth and sequencing that population for the SNPs on the Ion Torrent PGM sequencer. This method, would allow one to observe which SNPs showed up in which round, and how the frequencies of these SNPs change throughout the rounds. I hypothesized that if a SNP's frequency was enriched (i.e. selected for) throughout the rounds, it would be most important for the tolerance phenotype. If a SNP's frequency were to be selected against, it would suggest it to be less important for the tolerance phenotype.

The second method was to create 4 plasmids containing subsets of the 19 SNPs confirmed in R57. These plasmids were to be transformed into wild type haploid CEN.PK strains in different permutations and plated on HWSSL gradient plates. This was expected to show differences in tolerance with different combinations of genes and thus give a better understanding as to which genes were important and in what combinations.

1.2 Background

1.2.1 Cellulosic Ethanol

With most of the planet currently dependant on fossil fuels and the growing concerns over its long-term availability as well as its contribution to greenhouse gas emissions, it is no surprise that new forms of fuel are being sought after. In recent years, much attention has been drawn to the possibility of ethanol-based biofuels as an alternate renewable form of fuel. At the moment, most bioethanol is produced from sugar containing crops such as sugar cane and beetroot, and starch containing crops such as corn, wheat, rice and sweet sorgum, with an emphasis on corn and sugar cane use. This

type of ethanol production is known as 1st generation biofuel production [2]. However, concerns about green house gas emission savings being offset by land use changes and deforestation, the cost effectiveness of production and the main concern of land being used in competition with food crops have arisen [2-5].

The concerns and shortcomings of 1st generation biofuels, have led to the creation of a 2nd generation in biofuel production. Second generation biofuels can be produced from cheap, abundant, non-food lignocellulosic biomass such as straw, forestry residues, organic components of municipal waste and energy crops like purpose-grown vegetative grasses [6]. Due to the recalcitrant nature of these feedstocks, a pretreatment step is necessary to separate cellulose and hemicellulose from lignin in the lignocellulosic matrix, making the cellulose and hemicellulose available for hydrolysis. Various pretreatment methods can be utilized, these include physical pretreatment such as pyrolysis, physicochemical pretreatment such as steam explosion (the most commonly used form of pretreatment), ammonia fibre explosion, carbon dioxide explosion, and chemical pretreatment which includes ozonolysis, acid hydrolysis, alkaline hydrolysis and oxidative delignification [7]. Biological pretreatment, which uses various types of microorganisms can also be used [7]. In the majority of cases a hydrolysis step to break the recently liberated cellulose and hemicellulose into their corresponding sugar monomers is required. This is usually accomplished through dilute acid, concentrated acid or enzymatic hydrolysis [7, 8]. However, these methods do not come without their shortcomings. The pretreatment processes used to turn lignocellulosic material into a fermentable substrate generates several compounds that are inhibitory to microorganism growth and productivity [8].

1.2.2 Hydrolysate Inhibitors

Pretreatment of lignocellulosic biomass usually results in the formation of inhibitory compounds. The most common compounds are furans, weak organic acids and phenolic compounds. Furans, 5-hydroxymethyl-2-furaldehyde (HMF) and 2-furaldehyde (furfural) are formed from the degradation of hexose and pentose sugars respectively, while weak organic acids such as acetic acid, levulinic acid and formic acid are formed from the degradation of hemicellulose, the breakdown of both furfural and HMF and from the breakdown of HMF alone, respectively [8]. Phenolic compounds are formed from the partial breakdown of lignin [8]. Other stresses can include dissolved solids, wood extractives, lignosulfonates, ammonia, salts conferring osmotic stress and heavy metals from corrosion of the process machinery [1, 9-13]. Depending on the biomass being used, other acids may also be present. These include vanillic acid, syringic acid, 4-hydroxybenzoic acid, and caproic, caprylic, pelargonic and palmitic acids [8].

Furans are considered to be amongst the most potent growth inhibitors in lignocellulosic hydrolysates because they diminish overall biological activities by diminishing enzyme activity, breaking down DNA, inhibiting the synthesis of both RNA and proteins, creating vacuole and mitochondrial membrane damage, and by inducing the accumulation of reactive oxygen species (ROS) [14-16]. HMF and furfural can act synergistically to inhibit cell growth [15], and can directly inhibit the metabolic enzymes alcohol dehydrogenase (ADH), pyruvate dehydrogenase (PDH), aldehyde dehydrogenase (ALDH), hexokinase and glyceraldehyde-3-phosphate dehydrogenase [16-18]. To detoxify these substances, the yeast cell will reduce 5-HMF and furfural into less

inhibitory alcohols. These reductions are nicotinamide adenine dinucleotide (phosphate) (NAD(P)H) dependent and may lead to NAD(P)H depletion of the cell [19].

Weak acids such as acetic acid are thought to inhibit the growth of *S. cerevisiae* in two ways. The undissociated form of the acid crosses the plasma membrane where the neutral intracellular pH causes it to dissociate. This leads to the decrease of cytosolic pH, which must be compensated for by the removal of protons from the cell by the plasma membrane ATPase instead of being used for ATP hydrolysis. In extreme cases, the proton pumping capacity of the cell is exhausted and ATP is depleted as well as the proton motive force [8, 19-23]. The second method weak acids can inhibit yeast growth is through anion accumulation leading to toxicity [8, 19-23]. It would seem the smaller the molecular size and the more hydrophobic, the more toxic the anion [19].

Phenolic compounds like 4-hydrozybenzoic acid and vanillin are thought to contribute a significant inhibitory effect. Although the exact mechanisms that make phenolic compounds toxic have not been completely elucidated, it is thought that they might cause a loss of integrity of biological membranes [8]. Another possible explanation is that slightly acidic phenolic compounds may interfere with the proton gradient by transporting protons back across the mitochondrial membranes [19, 24].

Lignocellulosic hydrolysates also contribute osmotic stress in the form of hyperosmolarity, where the osmolarity surrounding the cell is higher than inside the cell. This situation causes water efflux which increases cytosolic ion concentration as well as cell shrinkage, both of which are detrimental to the organism [25].

Reactive oxygen species (ROS) oxidize lipids, proteins and nucleic acids thereby damaging them [26-28]. ROS transform proteins to protein carbonyls, which when aggregated disrupt cell homeostasis [28]. Major proteins targeted for inactivation by ROS are proteins involved in the citric acid cycle, such as α-ketoglutarate dehydrogenase, pyruvate dehydrogenase and succinate dehydrogenase as well as actin in the cells cytoskeleton [28] . ROS can cause various types of mutational damage to nucleic acids, including single or double strand breaks, base modifications and protein-DNA cross-linkages [28]. These are all forms of mutational damage and compromise the nucleic acid's functionality and integrity. The accumulation of ROS leads to programmed cell death [28].

Although each of the stressors described above inhibit cell growth on their own, it is also suggested that the synergistic effect of these compounds results in even greater inhibition. For example, it has been shown that acetic acid and furfural interact antagonistically on growth, causing a synergistic decrease in growth rate [8, 29].

1.2.3 Spent Sulfite Liquor

Spent sulfite liquor (SSL) is the effluent from pulp and paper mills that use the acid sulfite process, where wood chips are delignified by cooking in acid bisulfite to form purified cellulose for further processing. The high temperatures of 125-145°C and acidic conditions of pH 1-2 cause partial hydrolysis of hemicellulose and cellulose to their respective monomeric sugars [30]. These sugars are the targets of fermentation to bioethanol in SSL. SSL can come in two forms depending on the type of woods used for pulping. Softwood SSL (SWSSL), from pulping coniferous wood, contains a high proportion of hexose sugars (~74% of total sugars), while hardwood SSL (HWSSL) from

"harder" deciduous woods contains a very low amount of hexose sugars and a higher percentage of pentose sugars (~50-70% of total sugars), predominantly xylose [31]. Due to the high biological oxygen demand SSL creates, it is fermented before being disposed of in order to lessen its burden on the environment. Because of the high hexose concentrations found in SWSSL it is less of a problem to use as a substrate for yeast fermentation than HWSSL [32]. HWSSL also contains many of the inhibitors common to most other lignocellulosic hydrolysates. These inhibitory compounds include: dissolved solids, wood extractives, phenolic compounds, sugar degradation products such as HMF and furfural, weak organic acids (primarily acetic acid), lignosulfonates, ammonia and salts conferring osmotic stress [1, 9-13]. It is likely that the synergistic effect of these inhibitors is the main reason for the growth inhibition observed, while each inhibitor alone does not significantly reduce growth [13, 33].

1.2.4. Yeast Stress Response

Yeasts are usually found in non-ideal conditions and as such have evolved many mechanisms in the way of stress tolerance. The main stressors usually found in lignocellulosic hydrolysates are osmotic stress, weak organic acids, phenolic compounds, oxidative stress and furans. This section will discuss the various ways *S. cerevisiae* has evolved to deal with these stresses. These tolerance mechanisms include the high osmolarity glycerol (HOG) pathway, specific ATP-binding cassette (ABC) transporters as well as transcription factors and proteins involved in the detoxification of toxic compounds and misfolded proteins.

1.2.4.1 Osmotic stress response

With high dissolved solids and inorganic salts, lignocellulosic hydrolysates usually create hyperosmotic stress conditions for yeast. In conditions of hyperosmolarity, *S. cerevisiae* will simultaneously internalize the plasma membrane glycerol channel Fps1p, ensuring the retention and eventual accumulation of intracellular glycerol, as well as induce the high osmolarity glycerol (HOG) mitogen activated protein (MAP) kinase pathway [34]. This leads to a phosphorylation cascade resulting in the phosphorylation of the MAPK Hog1p and its translocation to the nucleus were it regulates gene expression of many genes [34]. Two regulated genes of interest are *GPD1*, which encodes a glycerol 3-phosphate dehydrogenase, which is transcriptionally activated by Hog1p resulting in the increased production of intracellular glycerol, and *ENA1*, a gene encoding a sodium-pump which is also upregulated by Hop1p [34]. The HOG pathway can also play a role in maintaining cell wall integrity under hyperosmotic conditions as well as stunting cell growth and repressing protein synthesis [34, 35]. Microarray analysis shows that under osmotic stress, yeast upregulate the expression of *GPD1, RHR2*, both involved in glycerol metabolism, and *TPS1, TPS1, TPS3, ATH1* and *NTH1*, which are all involved in trehalose metabolism [36].

1.2.4.2 Weak organic acid stress response

As previously mentioned, weak organic acids are one of the most common and important inhibitors found in lignocellulosic hydrolysates. This being the case, it is important to understand the yeast's native response to weak acid stress. Low internal pH can activate plasma membrane ATPase in order to increase the proton pumping action of the cell [8]. Cell volume is also reduced in order to increase the concentration of cellular

compounds, therefore increasing the buffering capacity of the cell [8]. Low pH has also been shown to induce the expression of heat shock genes [8]. The ATP-binding cassette (ABC) transporter Pdr12p, which is part of the yeast pleiotropic drug response is shown to transport weak organic acids, namely acetic acid, across the plasma membrane and out of the cell. This reduces intracellular acetic acid concentrations and confers resistance [37, 38]. Recent work has also shown an upregulation in the transcription factor Haa1p in response to weak acids including, benzoic, octanoic, propinoic and acetic acid. This transcription factor regulates expression of many genes and as a result leads to a reduction of the weak acid-induced lag phase of the cell by reducing the internal accumulation of the acid. It is not yet known how this reduction is achieved, but it has been postulated that restricting influx of the acid while increasing efflux of counter-ions is the cause [39].

1.2.4.3 Furan stress response

To detoxify the cell of the sugar degradation products 5HMF and furfural, *S. cerevisiae* reduces both to their less inhibitory alcohols, 2,5-bis-hydroxymethylfuran and furfuryl alcohol, respectively. These reduction reactions are coupled with nicotinamide adenine dinucleotide phosphate (NADPH) and nicotinamide adenine dinucleotide (NADH) oxidations. One of the enzymes responsible for this reduction has been identified as NADPH-dependent alcohol dehydrogenase (ADH6p) [15, 19, 40, 41]. It has been shown that under furfural conditions many heat-shock proteins (HSPs) are upregulated. These HSPs include Hsp104p, Hsp31p, Ssa4p and Hsc82p. Also upregulated is the expression of Gre2p, a 3-methylbutanal reductase and NADPH-dependent methylglyoxal reductase, which is regulated by the HOG pathway and increase in the

presence of various stressors including oxidative, osmotic, ionic, heat shock and heavy metal stress [42]. Other genes upregulated in response to furfural are *SRX1, CTA1,* and *GRX5,* all of which are known oxidative stress response genes [43].

1.2.4.4 Stress response to phenolic compounds

Upon exposure to phenolic compounds, *S. cerevisiae* will upregulate the expression of the transcription factor Yap1p and the Yap1p inducible *ATR1* and *FLR1* genes [44]. Yap1p has been linked to oxidative stress while Atr1p and Flr1p are both drug/ proton antiporters [44]. A deletion experiment performed to find genes required for vanillin tolerance showed 76 deletion strains that were vanillin-sensitive. These genes could be broken up into 3 main categories, chromatin remodeling, vesicle transport and ergosterol biosynthesis [45]. Ergosterol synthesis was later proven to be very relevant for phenolic tolerance when microarray data of a vanillin tolerant strain of *S. cerevisiae* showed upregulation of five ergosterol biosynthesis genes (*ERG28, HMG1, MCR1, ERG5* and *ERG7*) compared to wild type [46].

1.2.4.5 Oxidative stress

Another common stress caused by lignocellulosic hydrolysates is oxidative stress. *S. cerevisiae* has evolved two main defense mechanisms against oxidative stress. The first consists of the non-enzymatic defense system, which consists of small molecules that act as radical scavengers. They become oxidized, and therefore remove the oxidants before they can cause damage. These molecules include glutathione, phytochelatins, polyamines, ascorbic acid, metallothioneins, thioredoxin and glutaredoxin [47]. Other molecules that have been suggested to play a role in oxidative stress include lipid-soluble

antioxidants, trehalose, flavohaemoglobin and the maintenance of metal ion homeostasis [47]. The second general oxidative stress defense mechanism is enzyme-based. It consists of enzymes that can remove reactive oxygen species, and/or repair the damage caused by them. These enzymes include catalase, which breaks H_2O_2 down into O_2 and H_2O, superoxide dismutase, which catalyzes the dismutation of superoxide to H_2O_2 and O_2 and glutathione reductase, which is responsible for reducing oxidized glutathione so it can be once again reduced by an oxidant. Glutathione reductase, thioredoxin peroxidase and thioredoxin reductase all catalyze the reduction of hydroperoxides, as well as a proposed role of methionine sulphoxide reductase [47]. Enzymes of the pentose phosphate pathway (which include glucose 6-phosphate [Zwf1p], transketolase [Tkl1p] and ribulose 5-phosphate epimerase [Rpe1p]) are very important for the cellular production of NADPH, which is needed for the reduction of oxidized glutathione and thioredoxin by glutathione reductase and thioredoxin reductase [47].

Upon oxidative stress exposure many stress response transcription factors are upregulated, including Yap1p, which is known to regulate many oxidative stress related genes including *GSH1* [47]. Mdl1p, a mitochondrial inner membrane ABC transporter also plays a role in mitigating oxidative stress. [48, 49]

1.2.4.6 Environmental stress

As well as stress responses to specific stressors, there are several more general stress response systems utilized by *S. cerevisiae*. These systems include the so-called environmental stress response, which is the accumulation of about 900 genes that are differentially regulated in response to many stressors [50]. This system can regulate a diverse array of cellular processes, including energy generation and storage, defense

against reactive oxygen species, synthesis of internal osmolytes, protein folding and turnover, DNA repair, protein synthesis and cellular growth [50]. The pleiotropic and multi-drug response is a large system of ABC transporters that are utilized for the detoxification of a large array of compounds. These include the transporter Ycf1p, which is shown to participate in the vacuole detoxification of heavy metals, Mdl1p and Pdr12p [48, 49]. Under stress conditions, it is not uncommon for protein synthesis to be affected. Misfolded or unfolded proteins can cause aggregates, which are thought to be cytotoxic to the cell. Eukaryotic cells use the chaperone family of proteins, namely the heat shock proteins (HSP) as well as the ubiquitin-dependent proteasome pathway, to target and degrade damaged or misfolded proteins or refold them to their proper configuration in order to mediate this stress [51-53].

1.2.5. Engineering *S. cerevisiae* for resistance to biomass hydrolysate inhibitors

1.2.5.1. Engineering strains resistant to furans

TMB3000 is a robust strain of *S. cerevisiae* isolated from spent sulfite liquor with known tolerance to inhibitors found in lignocellulosic hydrolysates [41]. Through microarray measurements of mRNA levels it was determined that TMB3000 had deferential expression of at least fifteen aldehyde reductases (alcohol dehydrogenases), compared to a non-tolerant laboratory strain, CBS8066, when exposed to HMF. Alcohol dehydrogenases are responsible for reducing furfural and HMF to their less inhibitory alcohols [41]. The alcohol dehydrogenase gene *ADH6* from TMB3000 was overexpressed in both TMB3000 and CEN.PK 113-5D yeast and increases in HMF reduction were observed in both strains [19, 41].

In an effort to produce a strain of yeast tolerant to furfural, Heer et al. performed a targeted evolution experiment on the industrial *S. cerevisiae* strain TMB3400 [54]. The wild type strain was transferred to minimal medium supplemented with increasing concentrations of furfural. After ~300 generations a strain of yeast that could survive on furfural supplemented medium, as well as medium containing hydrolysate concentrations was identified. Surprisingly, the evolved strain's furfural reduction rate did not increase, which suggests that this strain utilizes another mechanism to remain viable under furfural toxicity [54]. Using global microarray-based transcript analysis on the parental strain, TMB3400, and the evolved strain, they were able to determine genes that were differentially regulated in cells exposed to an increase in furfural from 15 mM to 25 mM. The two most highly differentially expressed genes, *ADH7* (an alcohol dehydrogenase) and *YKL071W* (a putative protein of unknown function), from the evolved strain were sequenced and compared to the parental genes showing no difference. This led to the hypothesis that the increase in transcription was due to changes to a transcription factor of *ADH7* or another regulatory mechanism. These genes were then overexpressed in the parental strain and tested for furfural tolerance. The strains overexpressing the genes showed significant increases in tolerance to furfural compared to the parent strain [55].

In another study, Liu et al. created strains of *S. cerevisiae* with increased ability to reduce HMF and furfural to their less inhibitory alcohols [56]. This was done by transferring cultures of cells into progressively higher concentrations of either HMF or furfural. In the end, the parental strain NRRL Y-12632 was adapted into strains 307 12H60 and 307 12H120, capable of converting 100% of HMF at 60 mM and 307-12-F40 capable of converting 100% furfural at 30mM [56].

Eleven *S. cerevisiae* strains were tested for tolerance in a cocktail of hydrolysate inhibitors including formic and acetic acid, furfural, HMF cinnamic acid and coniferyl aldehyde. It was found that 2 strains, CEN.PK and SSL3 were able to convert furfural, HMF and coniferyl aldehyde into less inhibitory compounds. This is thought to be linked to their resistance to the compounds' inhibitory effects or could be due to general betterment in growth and metabolic rates [57].

Liu et al. used a pathway-based (glucose metabolism) qRT-PCR array to compare expression between a strain of *S. cerevisiae* tolerant to HMF and furfural (NRRL Y-50049) and the non tolerant parental strain (NRRL Y-12632) [58]. In this study, they discovered 16 genes with increased expression in NRRL Y-50049. These included *ZWF1, GND1, GND2, TDH1* and *ALD4*. From these results, it was hypothesized that in NRRL-50049 glucose metabolism favored the pentose phosphate pathway instead of glycolysis. The pentose phosphate pathway is a major source of NADPH for the cell. NADPH is used in the reduction of HMF and furfural to their less inhibitory alcohols, which then regenerates $NADP^+$. It is thought that this reprogramming of the yeast cellular metabolism to favor the pentose phosphate pathway is what led to the increased tolerance of the Y-50049 strain [58].

1.2.5.2. Engineering strains resistant to acetic acid

Another growth limiting compound found in hydrolysate substrates is weak organic acids, mostly in the form of acetic acid. In an attempt to increase *S. cerevisiae* resistance to acetic acid Wright et al. conducted two experiments. In one, the parent xylose-fermenting strain RWB218 was grown in a sequential anaerobic batch cultivation with increasing concentrations of acetic acid, leading to an adapted strain capable of

growing in acetic acid concentrations of up to 6 gL^{-1}, substantially higher than the parental strain which is inhibited at $1gL^{-1}$ [59]. The second method utilized a continuous anaerobic growth where acetic acid was increased by ammonium assimilation. After 400 generations, the culture was able to grow at acetic acid concentration of 5 gL^{-1} [59].

In an experiment to further understand acetic acid tolerance in *S. cerevisiae*, Mira et al. used the yeast deletion library in a comparative growth assay on 70 mM - 110 mM acetic acid plates. Although improvements to acetic acid tolerance due to any one mutation was not found, 650 genes affecting acetic acid tolerance were found [60]. These 650 genes associated with a vast number of cellular functions are all potential targets for improvement of acetic acid tolerance [60].

Wei et al. were able to create a strain of *Candida krusei,* with improved acetic acid tolerance. To achieve this, genome shuffling of the parental strain GL560 was performed and after 4 rounds of shuffling strain S4-3, demonstrating improved acetic acid tolerance, was created [61]. Although the genetic changes that enabled this improvement were not ascertained, it was demonstrated that strain S4-3 demonstrated greater membrane integrity, as well as higher intracellular catalase activity compared to the parental GL560 strain [61].

1.2.5.3. Engineering strains resistant to phenolics

Increased tolerance to phenolic compounds was attempted by engineering a strain of *S. cerevisiae* with the laccase gene from *Trametes versicolor*. Laccase is an enzyme that has been shown to remove low-molecular-mass phenolic compounds from hydrolysates and has been previously used in enzymatic detoxification procedures [19].

The addition of this gene allowed the strain to grow in medium contain 1.25mmol L^{-1} coniferly aldehyde, which inhibited the growth of the control strain [19].

1.2.5.4. Engineering strains resistant to oxidative stress

Yoo et al. overexpressed human superoxide dismutase in *S. cerevisiae* strain SHY4 and tested its tolerance to the oxidative stressors paraquat, menadione, H_2O_2 and heat shock. The results showed a marked improvement in oxidative stress tolerance to 5mM paraquat, 5mM menadione and 48°C heat shock with no improvement in tolerance to 5 mM H_2O_2 based on a 60 minute survivability assay [62].

1.2.5.5. Engineering strains resistant to complex mixtures

Recently, Pinel et al. developed a strain of *S. cerevisiae* capable of growing on undiluted HWSSL producing a theoretical ethanol yield of 80% (calculated from grams of ethanol per grams of glucose and mannose) from the sugars found in the SSL [1]. This strain, CEN.PK R57, was created using the technique known as genome shuffling. Two wild type haploid CEN.PK strains, CEN.PK 113-1A (*MAT*α) and CEN.PK 113-7D (*MAT*a), were UV mutagenized and plated on HWSSL gradient plates alongside their corresponding wild type strains. All growth of the mutagenized strains that grew above the wild type growth frontier was scraped and mated together, creating a pool of diploid strains with random mutations. These diploids were then put through 5 rounds of re-iterative mating with enrichment on HWSSL gradient plates before each mating step [1]. As well as being able to grow on and ferment sugars from undiluted HWSSL, the R57 strain showed a marked increase in resistance to individual lignocellulosic inhibitors that include acetic acid (pH5.5), HMF (0.5% wt/vol), osmotic stress from NaCl (3% and 7% wt/vol) and sorbitol (2M) and oxidative stress caused by H_2O_2 (1mM). Although many

colonies in the 5th and final round of genome shuffling were comparable to R57 with respect to tolerance, R57 was chosen due to its higher ability to ferment the hexose sugars in HWSSL to ethanol [1]. Genetic characterization of this strain was the focus of my work.

Due to the random mutagenesis employed in this method, the genetic reasons for this tolerance phenotype were unknown. The genomic DNA from the diploid R57 strain was then sent for Illumina genome sequencing along with the genome from the *MAT*a haploid CEN.PK strain, 113-7D. The two genomes were then compared using MAQ and many SNPs were found (over 40). Of these, 20 SNPs in 17 different genes were shown to be non-silent mutations. These mutations will be the focus of this thesis work.

2. Materials and Method

2.1 Strains and plasmids

S. cerevisiae strain CEN.PK is a model strain used in industrial biotechnology research and was therefore chosen for these studies. The prototrophic wild type (WT) diploid strain CEN.PK 122 and haploid strains CEN.PK 113-1A (*MAT*α) and CEN.PK 113-7D (*MAT*a) were used. Auxotrophic strains CEN.PK 113-7A (Δ*HIS3, MAT*a), CEN.PK 113-3C (Δ*TRP1, MAT*a), CEN.PK 113-32D (Δ*LEU2, MAT*a) and CEN.PK 113-5D (Δ*URA3, MAT*a) were also used. All yeast strains used in this body of work were supplied by EUROSCARF collection (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/index.html).

Plasmids used in this study were the yeast centromeric pGreg transformation vectors, which afforded a low copy number (i.e. 1-2 copies) in yeast and a very high copy

number in *E. coli* [63]. The specific vectors used in these experiments were pGreg 503, 504, 505 and 506. These offered the auxotrophic selection markers *HIS3*, *TRP1*, *LEU2* and *URA3* respectively. All contained the kanamycin resistance gene, *KamMX*, as well as an ampicillin resistance gene (*AmpR)* and the yeast GAL1 promoter. The plasmids used in these experiments were modified to only contain one KpnI restriction site. The pGreg plasmids were all supplied by EUROSCARF. Also used in this body of work was the centromeric YCp50::HO plasmid provided by Dr. Beverly Errede from the University of North Carolina and modified by Damien Biot-Pelletier (Martin lab) to contain a hygromycin resistance gene instead of a uracil auxotrophic marker. This plasmid was used to switch mating types of haploid strains. It contains CEN4, a hygromycin resistance gene as well as an ampicillin resistance gene and the HO gene used for mating type switching. Plasmids were maintained in *Escherichia coli* DH5α frozen at -80°C.

All yeast cultures were grown in either yeast-peptone-dextrose medium (YPD) (1% yeast extract, 2% peptone, 2% glucose [wt/vol]) or minimal synthetic defined (SD) medium (0.67% yeast nitrogen base without amino acids [YNB], 2% glucose [wt/vol]). Solid media were prepared by adding 2% (wt/vol) agar to the liquid media described above with the addition of 200 μg/ml G418 to YPD if selection for kanamycin resistance was necessary. For selection using hygromycin, 200 μg/ml hygromycin was added to the molten YPD agar before solidification. All *E. coli* cultures were grown in Luria Broth (LB), with the addition of 2% (wt/vol) agar for solid media. The addition of 100 μg/ml ampicillin to the media was used for selection by ampicillin resistance.

Yeast cultures were incubated at 30°C while *E. coli* cultures were incubated at 37°C. All liquid cultures were shaken at 200 rpm.

2.2 Sanger sequencing predicted SNPs

PCR amplification and Sanger sequencing of regions of the R57 genome thought to contain a SNP (primers Table 1) was done in order to confirm the SNPs identified by Illumina sequencing of the entire R57 genome. These primers were ordered through Integrated DNA Technologies and designed to produce an amplicon of about 500 base pair (bp) where the SNP in question would be around base pair 250 from the 5' end of the PCR product. Gel electrophoresis was performed on each PCR product and bands of the correct length were cut out and gel purified using the Qiagen gel extraction kit (Qiagen; Hilden, Germany) in accordance with its protocol. The PCR products were Sanger sequenced in both the forward and reverse directions using the Genome Québec Innovation Center sequencing facilities. Sequencing information was then analyzed to determine which of the SNPs were real and which were false positives. This was accomplished by using NCBI's basic local alignment search tool (blast) to compare the sequencing results to the same gene on the reference genome, CEN.PK 113-7D. I was also able to confirm that a mutation was heterozygous or homozygous by viewing the peaks in the chromatograph. Mutations that were homozygous showed one strong peak corresponding to the nucleotide while heterozygous mutations showed two weaker overlapping peaks, one of the wild type and one of the mutant nucleotide. In addition, each confirmed gene containing a SNP was analysed using Pfam (Sanger Institute) to determine if the SNP fell within a characterized protein domain.

Table 1. List of primers used to PCR amplify gene regions where Illumina sequencing predicted SNPs to be.

Gene name/ location		Primer
YLL056C	Forwards	AAATCCTAAATGAATAACGCCATCAGATTC
	Reverse	ATTGTGCTGGAAAAAGTTTAAAGGGAAAG
STE5	Forwards	CCTATTTCCTCCTCTACATTTTCG
	Reverse	GGACTGAGGTGTGATTGATAAC
SSA1	Forwards	GGGTAAAGTTCTTGGTTTCACCC
	Reverse	GGGGTGGATTGATGTATCTTCG
PHO3	Forwards	ATGAGTAACGTACCAATTCATCTTCG
	Reverse	AGACAATGTTTTGAATCCATACACTGG
YCR108C	Forwards	GTGAAAAAGAGGTATAACGTATGTATTAAGG
	Reverse	CCATCCATCTCTCTACTTACTACC
DOP1	Forwards	CGATAATGAGGATGTCAATCTTCG
	Reverse	GGAATATTTAGAGTCCGATTACCG
PHO5	Forwards	ATCGTCATCACGGATGAAAAACTCG
	Reverse	GCATATACCCATTTGGGATAAGGG
NUM1	Forwards	CGAAACAATACCTGTCAGATAAGGC
	Reverse	GCCTTGGCGTGTTCAACTAAGTAT
UBP7	Forwards	AGATGCCCTAAAGGTTTCAAACAACG
	Reverse	GTCTCCAAAAAAGTCGAGATGACG
MAL11 (104)	Forwards	CCAAGACAGTACTGAATTAGAGAAAAAGTAA
	Reverse	CCCTTTAGCATTAAGATATTACATGACC
MAL11 (161)	Forwards	GCTAAAGGGCAAACTTCCGAAG
	Reverse	AAAGCATGACTTTGAAGCAGGCG
DAN4 (381)	Forwards	AGGTGAATTCACTGACAGTAGTAGG
	Reverse	TCTGCATCCTCTGTCATTTCTACC
DAN4 (532)	Forwards	GGTTCAACTGATGAACTGATCCC
	Reverse	AAGTCACTTCATCCGCTGAACC
TOF2	Forwards	CCAGTCTAAAAAATTACTACAATTCGCC
	Reverse	TTGATAATAGTAAAGACAAAGTGGAGGC
AVL9	Forwards	ATCTTTATCTGCAACGGTTTCGTCC
	Reverse	TCCTGGAAAGAATACAGACAAGCC
ZDS1	Forwards	TAGTTTCTGCGCGGAACGGC
	Reverse	CCGGTTCAATTCACAGACAGTGCCTTTGG
YNL058c	Forwards	TGTATACAGTTCCATTTGGTTTATCAGC
	Reverse	AAACCCTTAATATGATTTCCTAAAGCGG
HPF1	Forwards	GCTAGGAACCGAAGAAGCGG
	Reverse	CGCTAGTAACAGCTCCACCG
SGO1	Forwards	TAAGCCAGCTGGTACAAGAAAACG
	Reverse	CTATTTTCCTCTTCAGGGTTTAAGGC
GDH1	Forwards	CCTTCAAGTCCACACATAGACC
	Reverse	GGCCGGCTTCTTGTTAAAGC
282bp5' of YBR013C	Forwards	ACACTATAATGAGAGAGAGTTACCCC
	Reverse	GGTATAAAAATCAACTATCGTCTATCAACC
219bp5' of *ATC1*	Forwards	CGATGCAACGTATGTTCTATATTGGC
	Reverse	TACATCTTTTCTCAAGCATTTTTTCTCAAGC
23bp5' of *CCT6*	Forwards	GATTCTGACGAAACGAGATATGTTGG
	Reverse	TTGGTCAAAGATACTTAGCAACTGGC
43bp5' of *BCS1*	Forwards	TACCAAGGATCATAAGACCACCGCCG

	Reverse	GAATGAATGAGGCGTTGTGGCCAAG
3bp5' of *PAU10*	Forwards	TGACGAATTCACGAATCGTTCAAGG
	Reverse	GTGAAGTCACCGTAGTTGAAAACG
645bp3' of *PAU10*	Forwards	GTCATGGTACCAAACTCTCATTTGG
	Reverse	TATTGCAATTTGCTTGAACGGATGCC
161bp5' of *Pbp1*	Forwards	AGTTGGTGGAGCTATCTCTTTTCC
120bp3' of *Okp1*	Reverse	ATTGTTTCAGAGTTTACGATTACGATGC
1362bp5' of *Flo5*	Forwards	GATTAGGGCTAGATTCTCTTCAAGG
	Reverse	AAATGCAAAAGTTGATAATGTAATATGAAAATTAAGG
100bp5' of *MNI1*	Forwards	AAAAAATGCCAATCCTTGGGGAGG
	Reverse	CTGCACAACTCCAGCTTGCG
TYA Gag and	Forwards	GTTACTTCCCTAAGACTGTTTATATTAGG
TYB Pol genes	Reverse	AGTTCAGCTATAATAGGGACGACTTCC
NSP1	Forwards	GCGCTAAACCGGATGAAAATAAAGC
	Reverse	CTTTCTTTCTTTTCGTTAGCCTTTGCTCC
rRNA-NTS1-2	Forwards	AAATTGTCCTCCACCCATAACACC
	Reverse	GGAAAGTATTAACCGTAGAAGAGAGG
217bp5' of *PRP39*	Forward	TGATTCAAAAGCCCATCAGTTTGCC
	Reverse	TATAACCATCAAAAAACACAGAAAAGGCC
141bp5' of *TVP18*	Forward	GTGCAATACACAAGATTATGTTGATATATCC
	Reverse	GTGCTAGCCAAATTGAAAAGAGGC
205bp5' of *RPB11*	Forward	AATCTGTCTGGAGCATTCATGATGC
379bp of *SIN3*	Reverse	CTGAAGAACTGCAAAATGAAGGCG
25bp3' of *NOP58*	Forwards	AGTTCAATTCAATATGGGCGTACAGCACT
	Reverse	CTGATACCGCAAAGGCTGCTTCTGA
43bp5' of *FIT3*	Forwards	GAGACTGCAGCATTACTTTGAGAAGCTCGGTTTCAAACCAGAAAAAAGC
	Reverse	ACAACTCATGGTGATGTGATTGCCCCTATCTAGAGAGGCTTTATTGG
82bp5' of *YOR387C*	Forwards	TAGATTGACCCAAAAATGTACCATTTTGG
	Reverse	TGCTTCTATTCCTGGAATGGAAGG
404bp5' of *YOR387C*	Forward	TGTTTCTATATTTCCAACCTCAGCGG
	Reverse	ATTTCTGAATTCAACAGAACCAAATTGTCC
226bp5' of *YPL279C*	Forward	AAAATCATCAGAGCATTACACGCTCG
310bp5' of *YPL279C*	Reverse	TTACCAAAAAAAAAAAGATATGAAATCAAGATAAAACG
NRG1	Forward	GGCTGTTGTATTATTGATGAATTACAAGT
	Reverse	CGTATACGATTCTGCAAAGTATTATTG
YNR065C	Forward	GTGACGTGTGCTCTTGTTCGGGC
	Reverse	TCTTCTCGTTATTTTCTTTGCCGCATGGTT
GSH1	Forward	CCATGTACTCAAGCTCGTCTCCCCA
	Reverse	CATGCTCTGTGCCTCCCTTGACAAA
ART5	Forward	TTGGAGTTTTTGCAAGTTGGCCGAT
	Reverse	ACGAACTTCCAGACCCGTCTCTGTT
ARO1	Forward	TGGTAAGTGCTATGGTGACTCCGCT
	Reverse	GACCAAGGCAGCCAAGGAAGTCAAA
IGD1	Forward	GGACTGGCTTGTCTACGCTTTGAGC
	Reverse	AGTCACTACGCTCCCGTAACCAGTG
ROG3	Forward	TCCTTCCGGCTTAGCCCTGTTGATT
	Reverse	CTAGCGTCCTGGGGCCTCTCTAATG

2.3 Finding and isolating mutated genes from heterozygous mutations

Seventeen of the twenty verified non-silent SNPs (contained in 14 genes) were heterozygous mutations. In order to create 4 plasmids containing the mutated genes, I needed to isolate the 17 mutated alleles from their wild type counterparts for those genes with heterozygous mutations. In order to achieve this, PCR amplification of the 14 genes with their native promoter and terminator regions was performed. The priming regions were about 1 kilobase (kb) upstream from the open reading frame (ORF), ensuring that the promoter region was included, as well as about 500 bp downstream from the ORF stop codon. The primers used in this amplification also added a common region to either end of the amplicon (Table 2). Common region 1 (C1) was added to the 5' end and common region 6 (C6) was added to the 3' end. These PCR reactions were then verified via gel electrophoresis and bands of the predicted size were gel purified.

Table 2. Primers for PCR amplification of the entire gene including promoter and terminator region, with the addition of common region 1 to the 5' end and common region 6 to the 3' end. Included in this list are primers to amplify the 17 heterozygously mutated genes and the 3 homozygously mutated genes. In bold is the C1 region for the forwards primers and the C6 region for the reverse primers.

Genes		Primer (5' to 3')
SSA1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**TCATATCGGGGGTCTTACTTCC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**CTGACCAATTTGTATATATATACATCTATCC
STE5	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**GCAGTTTGTGCAATTTGTCCAGC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**CGCGATGCTTTACGCTTTGGTTTGG
UBP7	Forwards	G**AGACTGCAGCATTACTTTGAGAAG**GCAACCAGTCTAACGGTTCC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GGGTGACAAAGATAACATTCACAAGAG
NRG1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**GCTATGTACGAATAGTTTCAGACG
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GCCACAGTTTCTAAAAGAAAAGCAATG
DOP1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**GGAACTGTGTATACCAAGCACAG
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GTGACAGTTGCAAAACCTTTCACG
TOF2	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**GGAAGGTCTCACTTTAACTCAGG
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GTATACTTGAATTCTAAAATTTTCAACAAATAAGTG
MAL11	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**CCTCATTGGTACCATATGTGGG
	Reverse	**ACAACTCATGGTGATGTGATTGCC**CCAATAAAGCATTGAATTATGAATGTCG
AVL9	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**CCTAAGTTGATAACAGGAAGAGTTGC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GCAAGATTACTATTAATATTAAAGAGGGTC
YNL058c	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**GGTTTCGAACATTGTGATAAACTCC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GCTGTGTCGGTGTTTCATGAG
SGO1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**GGAAGAGGATAGTATGCCCG
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GGGAGTACCATTTAACATTGCATC
GDH1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**CTTCATATTCTACCAAGAATGACAGC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GGATAAACACTTATAACGATTTCCACC
BCS1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**CGTATGCCATACTTTTCTAACTCCC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**CGGAAGCATTGGCCAATTATTTGC
RPB11	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**GGGAGGATAACCTAATATGTGAGC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GCATCAGCAATAGTCATTTTTACGCG
Pbp1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**CCAAATTTCCCTCTAGGATATCGG
	Reverse	**ACAACTCATGGTGATGTGATTGCC**CGTTTAGGTGTATGAAAAAATATAGAATACC
ARO1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**CTGATCGTGCCTATAATTATACTACC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**TGATGCAACCACCTTGGAGG
ART5	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**CAATTACTGTATATCAACAATTGTCTTCTCC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**GGTTTGTTGAAATATGTAGGACAGG
GSH1	Forwards	**GAGACTGCAGCATTACTTTGAGAAG**CTTTACAACTATCGAATAGTGTGAACC
	Reverse	**ACAACTCATGGTGATGTGATTGCC**CGGAACTGTAACCAATCGTAGC

To clone these genomic regions in pGreg plasmids, the PCR products were then used as templates for a second round of PCR amplification, this time with primers complimentary to the common regions with the addition of regions of homology to a linearized pGreg505 plasmid vector (restriction digested with KpnI and NotI restriction enzymes). Again, these PCR products were gel purified prior to the cloning step.

For cloning by homologous recombination, the PCR products along with the linearized pGreg505 vector were transformed into electrocompetent haploid *S. cerevisiae* CEN.PK 111-32D cells (protocol is described in Appendix A). The transformed cells were then plated on SD agar and incubated at 30°C for 48 hours. Four to six colonies were picked and grown in SD broth overnight. Plasmid DNA was then isolated with the use of a Qiagen mini-prep kit from overnight cultures. The plasmid DNA was suspended in 30µl of autoclaved distilled water and transformed into electrocompetent DH5α *E. coli* (Appendix B). Transformants were plated on LB agar containing 100 µg/ml ampicillin and grown overnight at 37°C. To identify the mutated allele, plasmid DNA from 5 colonies was purified from *E. coli* overnight cultures and sequenced by the Genome Québec Innovation Center using the forward primer for each specific mutated gene region (Table1). Sequencing results were then compared to the reference CEN.PK 113-7D gene to identify the plasmid containing the mutant gene allele.

2.4 Making plasmids

Four plasmids containing the mutated version of the genes were constructed using *in vivo* homologous recombination in *S. cerevisiae*. The pGreg 503, 504, 505 and 506 series of plasmids containing the *HIS3*, *TRP1*, *LEU2* and *URA3* auxotrophic markers were used for cloning and expression of the mutated genes. The 17 genes were separated

into 4 groups based on their predicted function as determined by their annotation found in the *S. cerevisiae* genome database (SGD). One plasmid was to contain genes related to either intracellular transport or cell cycle, *ART5, MAL11, FIT3, SGO1* and *AVL9* (plasmid 1). A second plasmid was to contain genes related to protein processing and transcriptional activation, *NRG1, STE5, BCS1, SSA1* and *UBP7* (plasmid 2). A third and forth plasmid were to contain genes related to biosynthesis as well as RNA related genes *RPB11, PBP1, ARO1, GSH1* and *GDH1* (plasmid 3), and uncharacterized genes or genes with a function that could not be categorized with the rest, *DOP1, TOF2* and *YNL058c* (plasmid 4) respectively.

Each gene for each insert cassette was PCR amplified with their respective native promoter and terminator regions while also adding a common region C1 to the 5' end and region C6 to the 3' end (Table 2). The templates for these PCR reactions were the R57 genome for the genes harbouring homozygous mutations and the pGreg plasmids described in section 2.3 for those genes harbouring heterozygous mutations. The common regions acted as priming regions for the next PCR amplification to add the appropriate homology regions (H regions) to each part of the cassettes (Table 3). These H regions are what allow the *in vitro* homologous recombination event to occur in *S. cerevisiae.* The unique sequences of the H regions also allow for the proper ordering of each of the genes in the cassette (Fig. 1, Table 4).

Table 3. Primers used to attach homology regions to common regions on the gene. Genes with matching homology regions will recombine in *S. cerevisiae* ensuring proper order of recombination. The C region priming site is depicted in bold and the H region in italics.

Primer Name	Primer (5' to 3')
C1:50X	*TAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCGTTTAAACGGCGCGCC***GAGACTGCAGCATTACTTTGAGAAG**
C1:H1 F	*CTCATGGCGGGGGTCGGAATGATTAAAGAAAGGGGCTGTGGGCGAGATTG***GAGACTGCAGCATTACTTTGAGAAG**
C1:H2 F	*CCAGTTAATAAACCGTGGCAAACATGATGGTGGCCTAATGGAGGTCACCA***GAGACTGCAGCATTACTTTGAGAAG**
C1:H3 F	*ATTTTACAACCAGAACACAAAAGTGCGAAGTTTGAGCAACGGCGACGGAT***GAGACTGCAGCATTACTTTGAGAAG**
C1:H4 F	*ATCCAAGGCAGTAGCAATCACGGTATAGATCAGGCCTGATACCGGCGAGC***GAGACTGCAGCATTACTTTGAGAAG**
C1:H5 F	*TATGCTGACAAAACGCTGAGGATCTACTACGCATTAATAAGTGAGGCGAG***GAGACTGCAGCATTACTTTGAGAAG**
C6:50X	*ATAACTTCGTATAATGTATGCTATACGAAGTTATTAGGTACCGCGGCCGC***ACAACTCATGGTGATGTGATTGCC**
C6:H1 R	C*AATCTCGCCCACAGCCCCTTTCTTTAATCATTCCGACCCCCGCCATGAG***ACAACTCATGGTGATGTGATTGCC**
C6:H2 R	*TGGTGACCTCCATTAGGCCACCATCATGTTTGCCACGGTTTATTAACTGG***ACAACTCATGGTGATGTGATTGCC**
C6:H3 R	*ATCCGTCGCCGTTGCTCAAACTTCGCACTTTTGTGTTCTGGTTGTAAAAT***ACAACTCATGGTGATGTGATTGCC**
C6:H4 R	*GCTCGCCGGTATCAGGCCTGATCTATACCGTGATTGCTACTGCCTTGGAT***ACAACTCATGGTGATGTGATTGCC**
C6:H5 R	*CTCGCCTCACTTATTAATGCGTAGTAGATCCTCAGCGTTTTGTCAGCATA***ACAACTCATGGTGATGTGATTGCC**

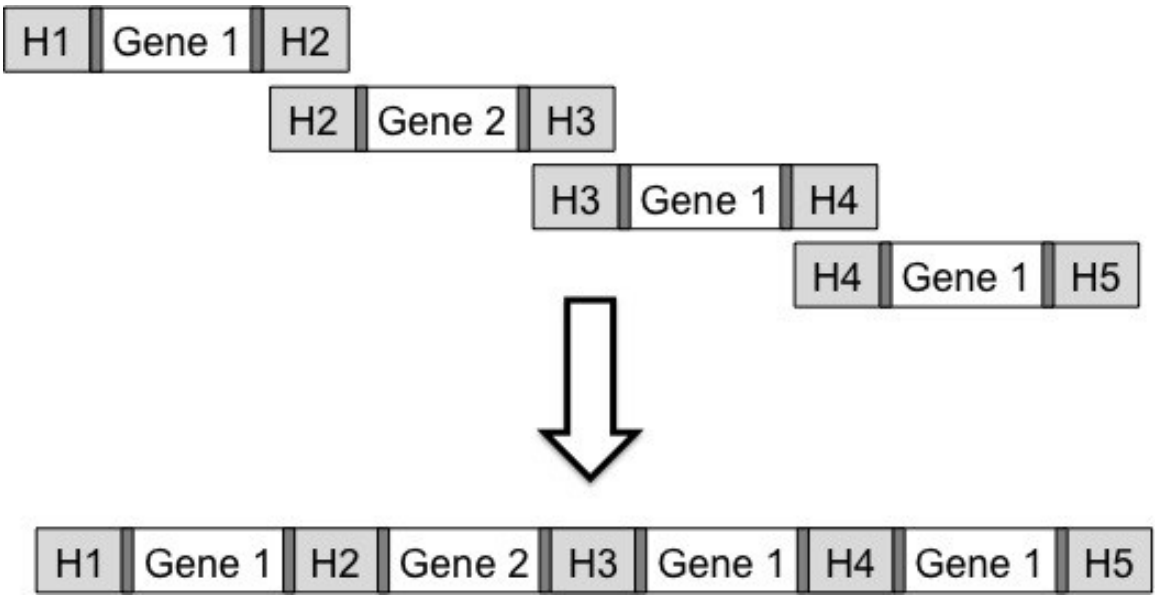

Fig 1. Example of how the homology regions work to produce and ordered assembly of genes within the yeast cell, using *in vivo* homologous recombination. A homology region (H) is attached to the common regions (dark rectangle a the 5' and 3' end of a gene insert). The homology regions are attached in such a way that and insert that is supposed to follow another will have a 5' homology region identical to the 3' homology region of the insert before it. All the inserts are transformed into yeast and through a process of homologous recombination, all the homology regions that are the same will recombine, linking the inserts they are attached to.

Table 4. List of genes for each plasmid, with appropriate 5' and 3' homology regions.

Gene	Plasmid	5' Homology region	3' Homology region
ART5	1	H:50X	H1
MAL11	1	H1	H2
FIT3	1	H2	H3
SGO1	1	H3	H4
AVL9	1	H4	H:50X
NRG1	2	H:50X	H1
STE5	2	H1	H2
BCS1	2	H2	H3
SSA1	2	H3	H4
UBP7	2	H4	H:50X
RPB11	3	H:50X	H1
PBP1	3	H1	H2
ARO1	3	H2	H3
GSH1	3	H3	H4
GDH1	3	H4	H:50X
DOP1	4	H:50X	H1
TOF2	4	H1	H2
YNL058c	4	H2	H:50X

Once each promoter-gene-terminator region had been amplified with the appropriate H regions. They were transformed, along with a linearized pGreg plasmid into electrocompetent CEN.PK yeast containing the appropriate auxotrophy to select for the assembled pGreg vector (plasmid 1, pGreg503, CEN.PK 113-7A; plasmid 2, pGreg505, CEN.PK 111-32D; plasmid 3, pGreg504, CEN.PK 113-3C; plasmid 4, pGreg506, CEN.PK 113-5D).

Twenty microliters of each gene product, along with 10ul of the linearized pGreg plasmid, were transformed via electroporation into the CEN.PK *S. cerevisiae* strains. The transformed cells were suspended in 1ml of YPD and allowed to recover for at least 1

hour. The cells were washed 3 times in 1x PBS and plated on either YPD agar with 200 µl/ml G418 or SD plates (2% wt/vol agar). The plates were incubated for 48 hours at 30°C (until colonies formed). Cells from several colonies (4-10) were inoculated in 10ml of SD medium overnight. Plasmid DNA isolated from these cultures (Qiagen) was dissolved in 30µl of dH_2O and transformed into electrocompetent *E. coli* DH5α where the pGreg plasmids are found in high copy. The transformed DH5α cells were then suspended in 1ml of LB and incubated for 1 hour at 37°C to recover. The cells were then plated on LBamp (100µg/ul ampicillin) and incubated for 24 hours. Plasmid DNA was isolated from several colonies and used as PCR template to confirm the presence of each promoter-gene-terminator fragment. The primers used to identify the SNPs (Table 1) were used to identify each gene in the cassette as well as new primers to check the junctions between cassettes were used.

Two version of each of the plasmids were constructed, one containing the mutated genes and one containing the wild type sequence of the genes. The plasmids that were confirmed by PCR to containing the proper fragments were transformed back into haploid wild type *S. cerevisiae* for phenotypic testing.

2.5 Testing Plasmids

2.5.1 SSL gradient plate

Gradient plates of HWSSL were used to assess the inhibitor tolerance phenotype of the *S. cerevisiae* strains transformed with the plasmids described in section 2.4. The plates were made by propping 22cm × 22 cm square petri plates up 7mm at one end. Two-hundred millilitres of undiluted HWSSL adjusted to a pH of 5.5 with NaOH was boiled with 2% wt/vol agar. This HWSSL agar was poured into the petri plate and left to

solidify creating a wedge of 100% HWSSL. The plate was then laid flat and 200ml of SD medium supplemented with 0.87g/100ml glucose and 2% (wt/vol) agar poured on top of the HWSSL wedge. The gradient plate was left to solidify and equilibrate for approximately 12 hours. This gave a gradient of ~ 40%-70% HWSSL. Plastic dividers were inserted into the medium separating it into lanes. Four centimeters at either end of the plate were neglected and then each lane was measured to be 2.5cm.

Strains to be tested on HWSSL gradient plates were grown in 5ml SD broth overnight at 30°C and 200rpm. One hundred microliters from these cultures were used to seed new 5ml overnight cultures. After a second overnight incubation in SD, the cells were adapted to HWSSL prior to plating on the gradient plates. To achieve this, the cells were harvested by centrifugation and the pellet was suspended in 5ml of liquid HWSSL and incubated overnight at 30°C and 200rpm. The following day the cultures were harvested by centrifugation and the cells were washed with 1x PBS. Methylene blue stain and a hemocytometer were used to count the number of viable cells in each culture (Appendix C). Each culture was then diluted to the same cell density using 1x PBS.

One hundred microliters of each diluted culture was pipetted into their corresponding lanes on the gradient plate. Each strain was plated twice on each plate to ensure the gradient was consistent and 3 replicates of each plate were made. The cells were spread in each lane using sterile glass beads. The plate was then incubated at 30°C until growth was observed (~6 days).

2.5.2 H_2O_2 spot plates

The strain harbouring plasmid 3 was tested on H_2O_2 plates because it contains the mutated *GSH1* gene, which is involved in glutathione synthesis and therefore thought to be involved in oxidative stress tolerance. Thirty-five percent wt/vol H_2O_2 was added to molten SD medium with 1% wt/vol agar with added amino acids to make final concentrations of 0.5mM, 1mM, 2mM, 3mM and 5mM H_2O_2. CEN.PK 113-3C strains harbouring the mutated and wild type versions of plasmid 3 as well as the same wild type haploid with an empty pGreg504 were grown overnight in 5ml SD medium at 30°C and 200rpm (done in biological triplicates). These cultures were used to seed new 5ml SD medium cultures for a second overnight incubation. Before testing on H2O2 plates, cells were adapted to HWSSL overnight as described in section 2.5.1 above. HWSSL adapted cells were harvested by centrifugation and the pellet washed in 1ml of 1x PBS. Cells were then counted using methylene blue staining as described in Appendix C. Ten cells/5µl, 100cells/5µl, 1000cells/5µl and 10000cells/5µl dilutions were then made. Five microliters of each dilution was then spotted onto each of the 5 plates of different concentrations. Plates were then wrapped in saran wrap to avoid H_2O_2 from diffusing out and placed at 30°C to incubate until sufficient growth was seen.

2.6 Ion Torrent pilot experiment

As previously described, two approaches were taken to evaluate the contribution of the SNPs identified in the R57 strain to the phenotype observed. The second approach required tracking the abundance of any given SNP in a population as well as tracking the appearance (or disappearance) of the SNPs through the genome evolution process. To achieve this, I first ran a pilot experiment. Frozen stocks of round 5 genome shuffling

scraping (same lane R57 was pulled from) was used to see if it was possible to observe the frequency of specific SNPs in a mixed population of mutants. A sample of 100 µl of the frozen stock was used to make serial dilution plates in order to quantify the number of live cells per ml in the frozen stock. Ten fold dilutions starting from 1/1000, were plated on YPD 2% (wt/vol) agar and incubated at 30°C for 48 hours. Plates for dilutions with a countable number of colonies were used to calculate the cfu/ml. The stock tube contained 10^10 cells/ml.

Genomic DNA from one hundred µl or 10^9 cells from the frozen stock was purified using a Qiagen DNeasy Blood and Tissue kit. The genomic DNA was then used as template to PCR amplify all 20 SNP regions. The primers were designed so that the SNPs would be located approximately at bp position 50 in the PCR amplified segment of ~100bp. The primers were also designed to attach the adapter P1, key sequence and adapter A region needed by the Ion Torrent sequencer (Table 5). Gel electrophoresis was then performed on the PCR products to ensure that the PCR products were of the appropriate size and gel purification was performed to eliminate any non-specific amplicons from being sequenced.

Table 5. Primers used for PGM sequencing. Primers used to amplify mutated regions and add the adapter regions and key sequence for Ion Torrent sequencing. Adapter region P1 is depicted in bold in the forward primers while adapter region A is depicted in bold in the reverse primers. The key sequence is italicized in the reverse primers.

Primer Name	Primer
SSA1_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GCATCACCAATCAATCTTTCAGTGTC
SSA1_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CGTGTGTTGCTCACTTTGCTAATG
TOF2_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CAGCTGGTCCATGACTTGTTTTG
TOF2_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CTTCCATGTCACTTGAATCTTCATTG
UBP7_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CCATGAAGTTGAGTAAACTTGGTAGG
UBP7_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GGCAATCCCATGCATTTTCACC
YNL058c_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GGAAGGTCTCTTCTGACCAGC
YNL058c_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CTAGACTCATACCTTGTCAAAAGTTC
MAL11a_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CCTTCCATAACCAGGGTAGTAG
MAL11a_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GCTAACAGCGAGGAAAAAAGCATG
MAL11b_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CCATATAAGTCGTGATTTGCAAACC
MAL11b_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GGAGGGTTCTTACGAAATTACTTCC
AVL9_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CCATCATCATCATCCTCAGAGTC
AVL9_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CTGATGCGGAGACAGATCAAC
GDH1a_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CCTCTTTGGAAGACTCTACTCTTTTC
GDH1a_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CCCAGGTGACTCTGAATTGTATG
GDH1b_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GTCAGAGCCAGAATTTCAACAAGC
GDH1b_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GCAAAACCTTTCTGTATTCTGGGTG
GSH1_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GTATATTTTCGATACTCTAAACCACCC
GSH1_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GCTGGAGTAGTTGGATCTTTCC
DOP1_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GCTCGACATTAGCACGAAACTTC
DOP1_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GGATCGACAAAAAATGTCCTTACCAC
SGO1_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GCATGAATCAAGTTTTAACAAGGACG
SGO1_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CGGTTTCGTCTTCAGGTTCTAAAC
BCS1_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CGACATGGTAGATTGAGGGC
BCS1_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GTAAAGATTTCCACTCTTCTATATTTGC
RPB11_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CGGATAGCAGCATTATCCATGAC
RPB11_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GAATTTACAGACTACTGGAGAGGG
FIT3_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GTGTGTTATTAAAATTTTTTTATTCTAACATAACTTCG
FIT3_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CCGTCATGTTATTGTAAATGATATGTG
ARO1_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**CCCTGCTGATCAACAGAAAGTTG
ARO1_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GATTTTACATTGACCTTCACCGAGG
ART5_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GCAGTACAAGCAACCAAGATATGG
ART5_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CTTCGAATTATTTTGTTGAAAACAGGG
NRG1_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GCAGTCTTATTTAATTTGTGTTTTAGTTCATTC
NRG1_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*GGAAACGTTGAAATAAGCCCGG
PBP1_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GGAAAGAACAAAAAGAAAGATAGAAGAAAAC
PBP1_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CGTTTTTGTAAAGCAGTTCTTAAATCG
STE5_F_PGM	**CCTCTCTATGGGCAGTCGGTGAT**GGTCGCTCCATTTGGCTATC

STE5_R_PGM	**CCATCTCATCCCTGCGTGTCTCCGAC***TCAG*CCCTTTCTTCTGTTAGAAATAGGC

The PCR products were then quantified using the Promega Quantifluor dsDNA system (Promega; Fitchburg, Wisconsin, USA) in triplicate, diluted to 16 pmol and combined to make a multiplexed sequencing sample.

The Ion Torrent Personal Genome Machine (Life Technologies, Carlsbad, California, USA) was used to perform sequencing of the PCR products. The Ion Torrent is a semiconductor chip based sequencer. The template DNA to be sequenced is adhered to ion sphere beads at a ratio of one template per bead. These beads are then put through an emulsion PCR that amplifies the number of identical templates on the bead. The beads are then loaded onto a chip containing 6 million machined microwells at a ratio of one bead per well. The chip is then loaded into the Ion Torrent where dNTPs are passed over the chip one at a time. If a base in complementary to the template strand it is added, releasing a hydrogen ion. Because there are several copies of the same templates per well, many hydrogen ions are released. The change in ion concentration is recognized by the semiconductor and a base is called. Sequencing was performed with the PGM Ion Torrent 200bp PGM sequencing kit and a 316 chip. The 200bp kit protocol was used to prepare the sample and chip for sequencing (Life Technologies, Carlsbad, California, USA).

The resulting sequencing data was analysed using the CLC genomics workbench program (CLC bio, Aarhus, Denmark). The sequence reads were aligned to the CEN.PK 113-7D reference genome and SNP calling was performed with a cut-off frequency of 1 base change/ 100 reads.

2.7 Tracking SNP frequency through strain evolution.

The following heterogeneous populations of cells were used to track SNP frequency through the strain's evolution process: wild type CEN.PK 122, a pooled UV mutant population (3 pools of CEN.PK 113-1A UV mutants and 2 pools CEN.PK 113-7D UV mutants) and cells scraped from HWSSL gradient plates from round 1, round 3 and round 5. Cells from the UV mutant and genome shuffling rounds were scraped from above the frontier of wild type growth observed on the HWSSL gradient plate (Appendix C). A sample of 10^9 cells from each population was plated on 40% - 90% HWSSL gradient plates. Gradient plates were prepared by plating 258ml of HWSSL 2% (wt/vol) agar overlaid with 142ml of SD medium with 0.87 g/ 100ml of glucose and 2% (wt/vol) agar. The gradient plates were separated into 4 lanes of 3.5cm with 4cm spaces at either end using plastic separators. Two lanes were for wild type CEN.PK 122 and two were for one of the 4 other populations (UV, round 1, round 3, round 5). Each plate was made in duplicate. This made for 8 plates. One hundred microliters (or 10^9 cells) from each strain was plated in each lane, spread using sterile glass beads and incubated for 6 days at 30°C.

After 6 days of growth, lanes containing strains other than wild type were scraped above the frontier of confident wild type growth (where individual colony formation could start to be observed). Scraped cells were suspended in 1x PBS and counted with the aid of a hemocytometer. Aliquots of 4.15×10^7 cells of each round, suspended in 100µl of 1x PBS were used for genome extraction using a Qiagen DNeasy Blood and Tissue kit.

Five microliters of each genome preparation was then used as a template for a polymerase chain reaction with primers specific for each of the 19 SNP regions (*ARO1* mutations were close enough for one primer set) discovered in R57 (Table 5). Gel

purification was then performed on all 76 PCR products in order to eliminate any non-specific amplicons. Purified products were quantified using the Promega Quantifluor dsDNA system in triplicate. Products were then diluted to a concentration of 16 pmol. Twenty microliters of each of the diluted samples from each population were then pooled to make 4 pools, one for each round. Each pool was then sequenced on the Ion Torrent PGM on a 316 chip with the 200bp kit (following the Ion Torrent protocols).

Each round of sequencing was analyzed separately on the CLC genomics workbench 5 program where the sequences were aligned to the CEN.PK 113-7D reference genome and SNPs were called with a cut-off frequency of 1 base change/ 100 reads.

2.8 Checking CEN.PK 113-1A for SNPs

Since I was looking at SNPs in diploid populations and my reference sequence came from the haploid CEN.PK 113-7D, I wanted to make sure that the opposite mating type haploid CEN.PK 113-1A was isogenic at least for the SNP regions I was interested in. Genomic DNA isolated from cells originating from a single colony of CEN.PK 113-1A was purified using the Qiagen DNeasy Blood and Tissue kit and used as a template for PCR amplification of each SNP region. The primers used were select primers (those for *NRG1, STE5, BCS1, SSA1, UBP7, ART5, MAL11, FIT3, SGO1, AVL9, PBP1, ARO1, GSH1, GDH1, DOP1, TOF2* and *YNL058c*) from the ones used to identify the SNPs in section 2.2 (Table1). PCR products were then gel purified using the Qiagen gel extraction kit and protocol. Purified PCR products with corresponding primers were then sent to the Genome Québec Innovation center for Sanger sequencing. Sequences were compared to

the CEN.PK 113-7D genome sequences for any differences, specifically the SNPs found in R57.

2.9 Sequencing SNPs from individual colonies isolated from the UV mutant and round 5 populations

Seven isolated colonies were picked from the top of the growth frontier (i.e. highest HWSSL concentration on the gradient plate) for the UV population and 20 colonies from the round 5 HWSSL gradient plate. These colonies were then struck for isolation on 50% HWSSL plates to ensure that no other cells were carried through. Individual colonies from each plate were picked and grown in 5ml YPD broth overnight at 30°C and 200rpm. A Qiagen DNeasy Blood and Tissue kit was then used to extract genomic DNA from each of the strains. The genomic DNA from each strain was used as a template in the amplification of the mutated regions of the following genes: *STE5, ARO1, NRG1* (not amplified for the UV strains)*, BCS1, MAL11, PBP1, ART5, UBP7, GSH1, YNL058c, SGO1* and *GDH1*. The amplification products from each of the above reactions were sent for Sanger sequencing at the Genome Québec Innovation Center. Sequences were then analysed for presence or absence of the SNP and if a mutation was present, whether it was heterozygous or homozygous.

2.10 Testing the UBP7 mutation

As a proof of concept, one of the mutations predicted to be important based on the data accumulated in this work was introduced into CEN.PK 113-7D and this strain was tested for tolerance to HWSSL. The mutation in *UBP7* was the mutation chosen for this experiment. This mutation is homozygous in R57 and as such I decided to create a homozygous diploid strain with the *UBP7* mutation. To do this I first attempted to introduce the SNP into both haploid strains CEN.PK 113-7D and CEN.PK 113-1A, but

was unable to get integrants in CEN.PK 113-1A. As a result I used the mate type switch plasmid YCp50::HO in order to create a diploid strain harbouring a homozygous *UBP7* mutation. A wild type diploid was also created in the same way using the haploid CEN.PK 113-7D.

To insert the mutation into the wild type *UBP7* six sets of primers were designed (Table 6). The first set, UBP7mut_1 and UBP7mut_2 where used to amplify a region just 5' of the mutation up to the end of *UBP7*'s 3'UTR as well as add a region of homology to the loxP-kanamycin resistance cassette from the PUG6 plasmid. The second set of primers (UBP7mut_3 and UBP7mut_4) were designed to amplify half of the loxP-kanamycin resistance cassette as well as add homology to end of the 3'UTR (homology to the 3' end of the previously described amplicon) to the 5' end of the amplicon. The third set of primers (UBP7_5 and UBP7_6) were used to amplify the second half of the loxP-kanamycin with 5' homology to the first half of the cassette and to add homology at the 3'end to a region just 3' of *UBP7'*s 3'UTR. The first and second amplicons described were then combined using a PCR sewing reaction with primers UBP7mut_1 and UBP7mut_4. This combined product, as well as the second half of the loxP-kanamycin resistance cassette were transformed into CEN.PK 113-7D and CEN.PK113-1A (Fig. 2) using the lithium acetate method (Appendix D). Transformants were then plated on YPD kanamycin plates and incubated at 30°C for 48hours.

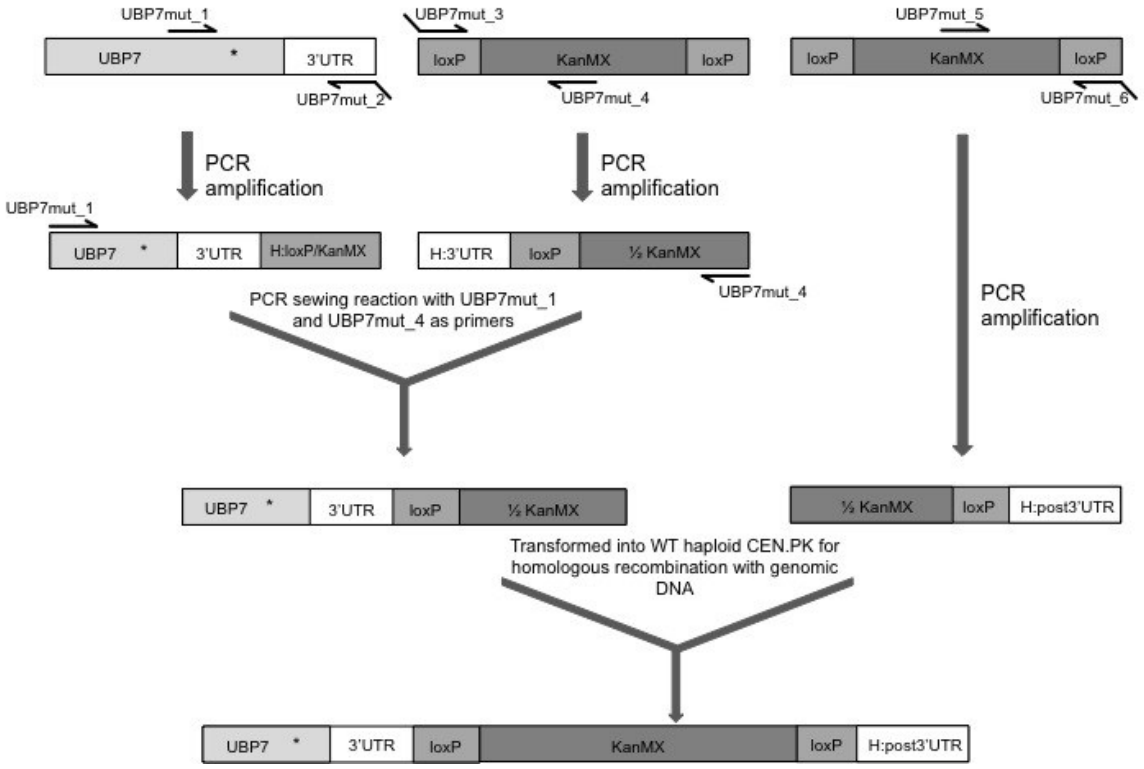

Figure 2. Integration of the mutation found in *UBP7* into the wild type CEN.PK background.

Table 6. List of primers used to integrate the mutation found in *UBP7* into the chromosome**.**

Primer names	**Primer sequence**
UBP7mut_1	CCATGAAGTTGAGTAAACTTGGTAGGTCTACTGAGAAAAGAGTTAAGTTAGAGG
UBP7mut_2	ACGAAGTTATATTAAGGGTTGTCGACCTGCAGCGTACGAAGCTTCAGCTGGGTGACAAA GATAACATTCACAAGAG
UBP7mut_3	CTCTTGTGAATGTATCTTTGTACCCCAGCTGAAGCTTCGTACGCTG
UBP7mut_4	GGAACACTGCCAGCGCATC
UBP7mut_5	GATGCGCTGGCAGTGTTCC
UBP7mut_6	CATCAAGACGTTTGGTGTCTAAATCGGCCGCATAGGCCACTAG

Colonies from the transformation plates were picked and genomic DNA isolated. Genomic DNA was used as a template for a PCR amplification reaction to amplify the region of *UBP7* with the mutation using the previously described primers (Table 1). PCR products were gel purified and sent to the Genome Québec Innovation Center for Sanger

sequencing. Sequencing results were analysed by comparing them to the CEN.PK 113-7D and R57 *UBP7* genomic sequence.

In order to create a homozygous diploid strain, a strain that was found to contain the mutant *UBP7* allele, as well as a wild type CEN.PK 113-7D (to make a control) were transformed with the centromeric mate type switch plasmid YCp50::HO using the lithium acetate method (Appendix D). Transformants were plated on YPD hygromycin plates and incubated for 48 hours at 30°C. After 48 hours 5 colonies from each plate were inoculated in YPD hygromycin broth and grown overnight. To confirm whether the strains were in fact diploid or not, they were patched with both wild type haploids on YPD plates and incubated overnight. Cells from each patch were then examined under the microscope for shmooing. If no shmooing was observed with either mating types, the strain was plated on sporulation medium (1% wt/vol potassium acetate, 0.1% wt/vol bacto-yeast extract, 0.05% wt/vol dextrose, 2% wt/vol agar) to confirm that they were in fact diploid strains. Spore formation was observed after 5 days.

The *UBP7* mutant and wild type strains were grown overnight in YPD to cure them of their YCp50:HO plasmid. Cells from overnight cultures were grown on YPD and replicated on YPD hygromycin plates to test for hygromycin sensitivity.

Hygromycin sensitive strains of *UBP7*, CEN.PK 113-7D diploid, wild type CEN.PK113-7D diploid as well as R57 were tested for their tolerance in parallel. Seed cultures for testing on HWSSL gradient plates were prepared as described in section 2.5.1. Each culture was diluted to the cell concentration of the least concentrated culture and 100µl of each culture was plated on HWSSL gradient (40%-70%) plates and grown

at 30°C for 72 hours. Each plate was separated into six 2.8cm lanes with 3cm of space on either side. Each biological triplicate was duplicated on each plate to ensure a consistent gradient on each plate. Each plate contained 2 lanes for the diploid *UBP7* mutant, 2 lanes for the diploid wild type and 2 lanes for strain R57.

3 Results

3.1 Sanger sequencing of predicted SNPs in R57

Sanger sequencing of all 47 predicted SNPs in R57 revealed 20 SNPs from 17 different genes (Table 7) to be true mutations while 27 were false positives called by the SNP predicting software. Three of these mutations are located within 190 bp 5'of the ORF and believed to be in promoter regions. One mutation is located 42 bp downstream of the *FIT3* ORF, possibly in the terminator region. Three of the mutations are homozygous mutations in the diploid R57 strain while the remaining 17 are heterozygous.

Table 7. List of SNPs confirmed by Sanger sequencing. Shows the 20 mutations along with the 17 affected genes that were found to be real through Sanger sequencing. The affected region, amino acid change, genotype and nucleotide change are also displayed.

Affected Gene	Nucleotide Change	Affected region	Amino Acid change	Genotype
YAL005C/ SSA1	A to G	ORF	Q31K of 642	Heterozygous
YDR043C/ NRG1	G to T	ORF	P46Q of 231	Homozygous
YDR103W/ STE5	C to T	ORF	S171F of 917	Heterozygous
YDR127W/ ARO1	C to T	ORF	S428F of 1589	Heterozygous
YDR127W/ ARO1	C to T	ORF		Heterozygous
YDR141C/ DOP1	T to A	ORF	N14Y of 1698	Heterozygous
YDR375C/ BCS1	T to A	Promoter - 43 bp upstream		Heterozygous
YGR068C/ ART5	G to T	ORF	L152K of 586	Heterozygous
YGR178C/ PBP1	T to C	Promoter- 190 bp upstream		Heterozygous
YGR289C/ MAL11	G to A	ORF	P104S of 616	Heterozygous
YGR289C/ MAL11	A to T	ORF	M161K of 616	Heterozygous
YIL156W/ UBP7	T to A	ORF	N822K of 1071	Homozygous
YJL101C/ GSH1	A to T	Promoter- 72bp upstream		Heterozygous
YKR010C/ TOF2	G to A	ORF	S714L of 771	Heterozygous
YLR114C/ AVL9	C to G	ORF	E602D of 764	Homozygous
YNL058C	T to C	ORF	K3E of 316	Heterozygous
YOR073W/ SGO1	C to G	ORF	S192Y of 590	Heterozygous
YOR375C/ GDH1	G to A	ORF	S16F of 454	Heterozygous
YOR375C/ GDH1	A to G	ORF	F23S of 454	Heterozygous
YOR383C/ FIT3	C to T	Terminator - 42 bp downstream of Stop codon		Heterozygous

3.2 Ion torrent pilot experiment

I wanted to see if I could track the frequency of each of the 20 SNPs through the evolution of the strain from the UV mutagenized populations to the 5th round of genome shuffling. The hypothesis was that if a SNP was enriched in the population as the strain evolved from a sensitive to a resistant phenotype, it would indicate that the SNP is most likely involved in the tolerance phenotype. I decided to determine if this would be at all possible by sequencing the SNPs from the heterogeneous round 5 (R5) population, where we would be most likely to have larger (i.e. detectable) frequencies of the mutations. Regions containing the 20 SNPs were amplified from a genomic DNA preparation of the heterogeneous population and sequenced with the Ion Torrent sequencer. There were a total of 3.76×10^6 reads with an average read length of 104 bp and an average fold coverage of 1.35×10^5 reads per nucleotide sequenced. Ion Torrent data was analysed with CLC genomics workbench's SNP calling software. The software called a SNP if a particular nucleotide was changed from the reference CEN.PK genome to another nucleotide at a frequency greater than or equal to 1%. A total of 31 SNPs were called. Twelve of the 31 were discarded due to the fact they had a fold coverage orders of magnitudes smaller than the rest. Nineteen were accepted as being real SNPs (Table 8).

Table 8. SNPs and their frequencies from the heterogeneous population of cells resulting from 5 rounds of genome shuffling. SNPs were called if they appeared at a frequency of more than 1% of the reads covering a specific nucleotide. Shows the genes the SNPs were called in. An asterix (*) indicates a mutation that was not found in R57 but is within 50 bp of the R57 mutation.

Genes	Nucleotide position in ORF	Position on Reference genome	Allele variations	Fold coverage	Frequency of mutation in %
ART5	454	626459	G/T	116989	49.6
AVL9	1806	360439	G/C	88643	46.4
PBP1	190bp upstream of ORF	853220	T/C	48102	35.1
YNL059c	7	510912	T/C	165388	33.9
SGO1	575	465488	C/A	153261	30.7
BCS1	43bp upstream of ORF	1225616	T/A	55448	30.4
MAL11	310	1075321	G/A	123859	29.7
GDH1	68	1042948	A/G	271444	29.2
UBP7	2466	50554	A/T	131873	24.1
ARO1	1284	698281	C/T	139828	19.9
STE5	512	651696	C/T	95126	16.4
ARO1	40	698282	C/T	143183	10
GSH1	72bp upstream of ORF	218979	A/T	162767	9.8
GDH1	47	1042969	G/A	278645	7.4
NRG1	137	542392	G/T	286520	6.8
SSA1	91	141252	G/T	111406	5.7
MAL11	482	1075149	A/T	165650	3.6
*STE5**	169	651689	A/T	10673	1.6
*SSA1**	65	141226	A/G	44099	1.4
*SSA1**	92	141253	A/G	109361	1.1

Half of the SNPs had a frequency of over 19%: *ARO1* c.1284C>T with a frequency of 19.9%, the mutation found 43 bp upstream of the *BCS1* ORF with 30.4%, *ART5* c.454G>T with 49.6%, the mutation found 190bp upstream of the *PBP1* ORF with 35.1%, *MAL11* c.310G>A with 29.7%, *UBP7* c.2466T>A with 24%, *AVL9* c.1806C>G with 46.4%, *YNL058c* c.7T>C with 33.9%, *SGO1* c.575C>G with 30.7% and *GDH1*

c.68A>G with 29.2%. The R57 SNPs found in *DOP1, FIT3* and *TOF2* were not found in this Ion Torrent dataset and thus had frequencies of less than 1%. Another interesting result was the fact that 2 additional mutations were discovered within 50bp of the R57 *SSA1* mutation, one of them having a frequency of 1.4% and the other having a frequency of 1.1%, both of which were lower than the R57 mutation at 5.7%. A second mutation in *STE5* was also discovered, it too had a lower frequency than the R57 mutation with a frequency of 1.6% compared to the R57 mutation at 16.4%.

This experiment indicated that it is possible to detect most of the SNPs at a reasonably high frequency in a heterogeneous population and that tracking the SNPs through the evolution of the strain would be possible.

3.3 Tracking SNP frequency through strain evolution by genome shuffling

Because the pilot experiment described in 3.2 was successful, I set out to track the frequency of the 20 SNPs throughout the rounds of shuffling. Heterogeneous population of cells from each round was grown on HWSSL gradient plates alongside the wild type diploid CEN.PK 122. Cells that grew above the wild type frontier of confident growth (Fig. 3) were scraped and their genomic DNA was extracted to be used as a template for the amplification of the mutated regions. Ion Torrent sequencing was performed on amplified regions of 100bp, with the SNP located at around base pair 50. The average total number of reads for the 4 chips used in this experiment was 1.5×10^6 at an average read length of 100 bp.

To determine what percent of the unique genotypes in the population was sequenced, I calculated the percentage based on different amounts of doubling and

various assumptions. The number of cells scraped from each lane was about 4×10^7. If we assume that each unique genotype in this 4×10^7 is represented 100 times (7 doublings, which is a conservative scenario) and at an equal frequency then we theoretically end up with 4×10^5 unique genotypes in the population. The average number of total reads for the 4 chips used was 1.5×10^6 reads. If we divide this total number of reads by the number of gene regions sequenced (19) we get 78947. Assuming that the 4×10^5 unique sequences have an equal chance of being sequenced, only 19.7% of them will have the possibility of being sequenced $((78947/4 \times 10^5) \times 100)$. However, if we assume that each unique genotype is represented 1000 times (10 doublings) in the lane scraping and we apply the same logic, 197% of them will have the possibility to be sequenced, meaning all unique sequences will have been sequenced, some more than once. This is all assuming that each unique sequence was used as a template in the PCR amplification step. Because the plates were left to incubate for 6 days, more than 10 doubling events took place and thus we could accept that every unique genotype was sequenced at least once. Sequences were then analyzed using CLC genomics workbench 5 and frequencies are reported in Fig. 4.

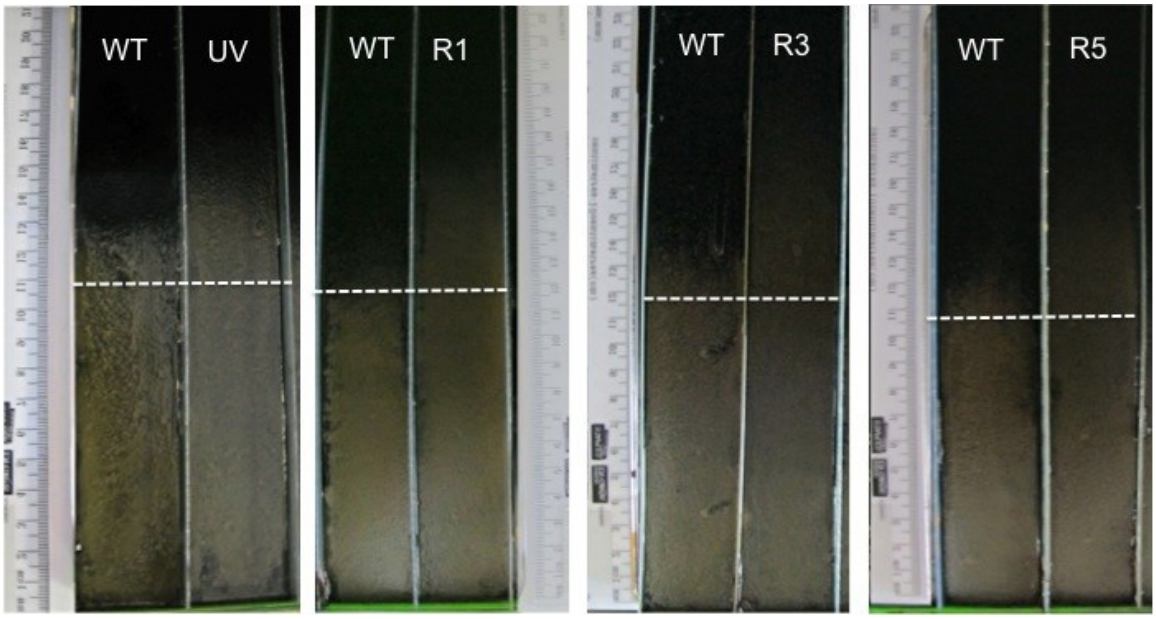

Figure 3. HWSSL gradient plates used to select the population for SNP analysis. Shows the 40%- 90% HWSSL gradient plates plated with the pooled UV mutagenized populations (UV), round 1 (R1), round 3 (R3) and round 5 (R5) of genome shuffling. The frontier of confident wild type growth is shown by the dotted line. Cells growing above the dotted line for each lane were scraped and used for genomic DNA preparation.

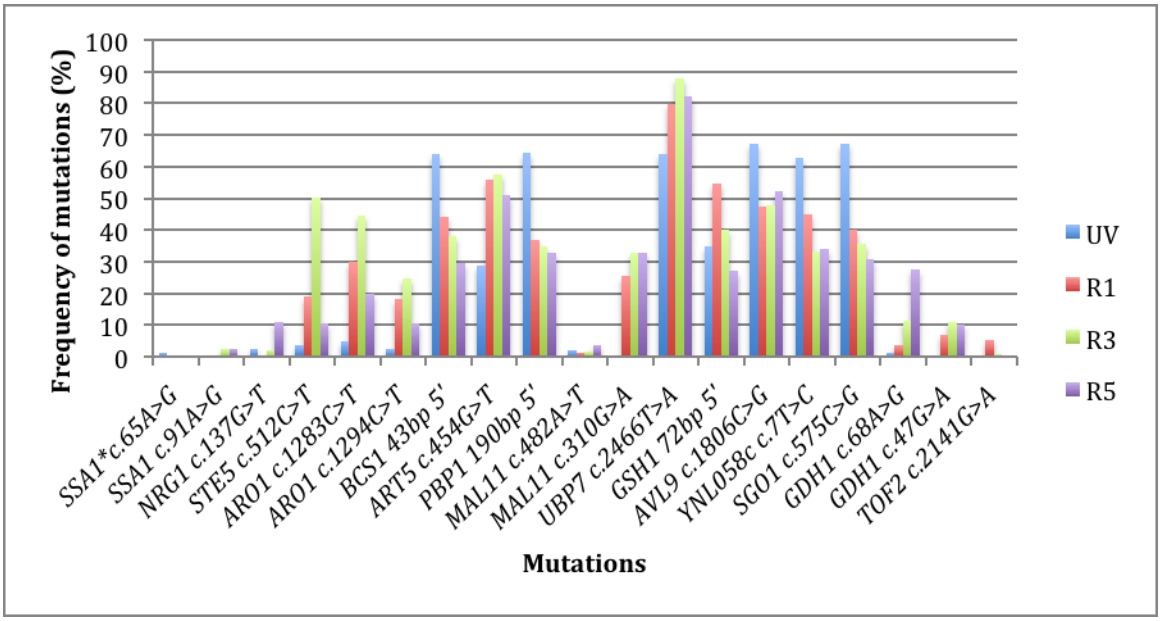

Figure 4. Graphical representation of the frequencies of each mutation in all 4 shuffling rounds. The frequency of the SNPs in the mixed population of cells scraped from each of the 4 gradient plates above the wild type growth frontier are represented in percentage. An asterix (*) is used to show genes found during Ion Torrent sequencing that are not present in the R57 strain. Mutations not found within an ORF are shown with the number of nucleotide 5' of the ORF start site. *FIT3* and *DOP1* were not found at a frequency higher than 1% in any round.

3.3.1 *SSA1*

The frequency of the SNP found in the R57 version of *SSA1* started under 1% in the UV mutagenesis and round 1 populations, and then appeared at a frequency of 2.6% for the round 3 and 5 populations. Another SNP in *SSA1* was discovered 26 bp away from the one found in R57. However, this SNP only shows up in the UV mutagenized population and at a low frequency of 1.1%.

Initial sequencing of R57 revealed a G to T mutation at nucleotide 91 in the *SSA1* ORF. This mutation leads to a glutamine to lysine amino acid change at the 31st amino acid in the polypeptide. This position falls within the HSP70 domain of the protein. Ion

torrent sequencing of the rounds of shuffling revealed another mutation, this time an A to G mutation leading to a silent mutation at amino acid 39 of the Saa1p.

3.3.2 *ART5*

R57 was found to have a c.454G>T mutation in the ORF of *ART5* leading to a leucine to isoleucine amino acid change at position 152 of Art5p.

The frequency of the SNP found in *ART5* increases from 28% in the UV population to 56.1% in round 1. From there it stays in the 50s, having a frequency of 57.6% and 51.1% in rounds 3 and 5.

3.3.4 *UBP7*

Similarly, the frequency of the SNP found in *UBP7* goes from 63.9% in the UV population, up to 80% in round 1. From there it remains in the 80s going to 87.8% in round 3 and 82.1% in round 5.

R57 contains a homozygous c.2466T>A mutation in the ORF of *UBP7* resulting in an asparagine to lysine amino acid change at position 822 in the Ubp7p. Asparagine is an amino acid containing a polar uncharged side chain while lysine has a charged side chain. This mutation is located in the ubiquitin carboxyl-terminal hydrolase domain of the protein.

3.3.5 *GDH1*

There were 2 mutations found in the *GDH1* gene in the R57 strain. *GDH1* c.47G>A shows an increase from a frequency less than 1% to 7.1% from the UV to round 1, then up to 11.2% and down again to 10.3%. The frequency changes for the *GDH1* c.68A>G SNP shows an increasing trend. It goes from a frequency of 1.4% in the UV

round to 3.6% in round 1, then increases again to 11.3% for round 3 and 27.6% in round 5.

These 2 mutations *GDH1* c.47G>A and *GDH1* c.68A>G, lead to a serine to phenylalanine change at amino acid 16 and a phenylalanine to serine amino acid change at position 23 respectively. Phenylalanine contains a hydrophobic side chain with a ring structure while serine has a small polar uncharged side chain with no ring structure.

3.3.6 *MAL11*

The *MAL11* c.310G>A SNP rises from the UV round to round 1, going from a frequency below 1% to 25.4% of the population. It then jumps up to 32.8% at round 3 and remains there for round 5. The *MAL11* c.482A>T mutation shows an increasing trend through the evolution as well. It starts off at 2.1% and goes down to 1.4% in round 1, but picks up to 1.5% in round 3 and 3.9% in round 5.

The two mutations found in the R57 version of this gene, c.310G>A and a c.482A>T resulted in amino acid changes of proline to serine at position 104 and methionine to lysine change at the 161th amino acid in the Mal11 protein respectively. Proline has a side chain with a ring structure while serine is an amino acid with a polar uncharged side chain with no ring structure. Both of these mutations fall within the sugar transport domain of Mal11p.

3.3.7 *NRG1*

NRG1 shows a similar but slightly different trend, going from a frequency of 2.3% in the UV population down to below 1% in round 1, then back up to 2.1% in round 3 and 11.2% in round 5.

The mutation found in *NRG1* was a c.137G >T mutation within the ORF resulting in an amino acid change of proline to glutamine at position 46 of the polypeptide. This mutation is homozygous in R57, suggesting that the mutation was sufficiently enriched through the evolution for two haploids harbouring the mutation to mate creating a diploid homozygously mutated cell.

3.3.8 *STE5, GSH1* and *ARO1*

The SNPs found in *STE5*, *GSH1* and both of the SNPs in *ARO1* show steady increases until round 5 where the results show a large drop in frequency. Sequencing revealed a c.512C>T mutation in the ORF of *STE5* leading to a serine to phenylalanine amino acid change at position 171 of the 917 amino acid protein. The frequency of the mutation starts at 3.8% in the UV population and then increases to 18.9% in round 1. It then dramatically increases to 50.1% in round three and plummets to 10.7% in round 5.

The *GSH1* mutations starts at 34.7% then increases to 54.7%, decreases to 40.1% in round 3 and 27.2% in round 5. The mutation was found 72bp upstream of the ORF in the 5' UTR and thus any effects due to this mutation are likely to be due to changes in expression level.

R57 was found to have two C to T mutations directly next to each other, at nucleotide positions 1283 and 1284 within the ORF of *ARO1*. These nucleotide changes result in a serine to phenylalanine amino acid change at amino acid 428 of the Aro1 protein. This amino acid falls within the EPSP (3-phosphoshikimate 1-carboxyvinyltransferase) synthase domain of the protein (PF00275). Both mutations display the same type of behaviour with one of the SNP frequencies starting at 5% going

up to 30.2% then increasing again to 44.6% then dropping down to 19.9%. The second mutations is almost half of the first, with the SNP frequency starting at 2.5%, rising to 18.2%, and 24.6% then dropping to 10.4%.

3.3.9 *PBP1, BCS1, YNL058c* and *SGO1*

The mutations found in *PBP1*, *BCS1*, *YNL058c* and *SGO1* show a steady decreasing trend through the rounds of shuffling. R57 contains a T to A mutation 43bp upstream of the *BCS1* start site, in the 5' UTR. This means any phenotypic changes found related to this mutation are likely to be due to differences of expression levels of Bcs1p. The mutation in *BCS1* starts off at a high frequency of 64.2% in the UV mutagenized round and decreases to 44.2 % then 38.3% and finally 29.5% in rounds 1, 3 and 5 respectively.

R57 contains a T to C mutation in the 5'UTR, 190bp 5' of the *PBP1* ORF. This means that any changes observed due to this mutation are likely to be due to changes of gene expression levels. This mutation also starts at a high frequency of 64.5% but lowers to 36.8% in round 1, 35% in round 3 and 32.8% in round 5.

YNL058c encodes a putative protein of unknown function. In R57 there is a c.7T>C mutation in the ORF resulting in a lysine to glutamic acid amino acid change for the 3rd amino acid in the protein. The SNP in *YNL058c* follows the same trend as the previous two, going from 67.2% to 45% then to 33.4% and remaining in the 30s at 34.1% in round 5.

Similarly, the mutation in *SGO1* starts of at 67.2% then decreases to 45% then to 35.5% and finally to 30.9%. R57 contains a c.575C>G mutation in the ORF of *SGO1*

resulting in a serine to tyrosine amino acid change at the 192nd amino acid in the Sgo1 protein.

3.3.10 *AVL9*

R57 contains a c.1806C>G mutation in the ORF resulting in a glutamic acid to aspartic acid amino acid change at position 602 of the Avl9 protein. Evolution of the SNP in *AVL9* does something slightly different than the rest of the SNPs, it starts off quite high with a frequency of 67.3% at the UV round, then drops to 47.4% in round 1, then increases to 48% and 52.4% in rounds 3 and 5.

3.3.11 *TOF2, DOP1, FIT3*

The mutation in *TOF2* was not visible at a frequency higher than 1% for the UV round, but jumped to a frequency of 5.3% for round 1, then dropped to 1 and then less than one percent for rounds 3 and 5, respectively. The SNPs found in the genes *DOP1* and *FIT3* of strain R57 were not represented at a frequency of 1% or higher in the heterogeneous populations suggesting that they are also not important for the tolerance phenotype.

By changing the minimum frequency (i.e. the percentage a nucleotide at any one position needs to be different than the reference nucleotide) needed to be called SNP on the CLC genomics workbench program to the highest setting were every base pair sequenced was called as an SNP, I was able to determine the background frequency of this experiment. This was determined to be 0.001%. Since I used a minimum frequency of 1% as my limit I can be confident that the SNPs called for every round, as well as their frequencies are correct plus or minus 0.001%.

3.4 Checking CEN.PK 113-1A for SNPs

Due to the results of the previous experiment, specifically the fact that there were 6 SNPs (*BCS1, PBP1, UBP7, AVL9, YNL058c* and *SGO1)* found in higher than sixty percent of the population in the UV population, I hypothesized that perhaps some of these SNPs were actually found in the wild type CEN.PK 113-1A haploid strain. Because CEN.PK 113-7D was the only wild type strain sequenced to compare to R57 it was possible that there was a difference between the wild type haploids and that these differences were being picked up as SNPs caused by the UV mutagenesis. The results of sequencing CEN.PK 113-1A showed that the SNP found in *AVL9* was indeed found in the wild type haploid strain but the remaining 5 were true mutations in R57 and not an allelic difference between the haploids.

In a previous study, it was shown that CEN.PK 113-1A survived longer than CEN.PK 113-7D in undiluted HWSSL [1]. It is possible that this allelic difference, along with many possible others, could be responsible for the observed difference in death rates of the two strains.

3.5 Mutation combination in single colonies from the UV population and round 5

Because tracking the frequency of the SNPs through the evolution showed the frequency of a mutation from the wild type growth frontier to the top of its own rounds growth frontier, and the fact that I was sequencing mixed populations and not individual cells, I wanted to see if I could detect linkages of these mutations in cells that grew at the top end of the growth frontier.

To do this, 7 colonies from the UV round populations were picked from the top of the growth frontier (only 7 colonies were chosen because only 7 could reliably be picked

as single colonies), and 20 colonies from the round 5 population's growth frontier. Each of these strains was sequenced for the mutations thought to be important based on the frequency data. For the UV strains, this meant the mutations found in *STE5, UBP7, MAL11* (c.482A>T)*, YNL058c, SGO1, GDH1* (c.68A>G)*, BCS1, PBP1, GSH1, ART5* and *ARO1*. The mutations sequenced for the round 5 strains were the ones found in *STE5, UBP7, MAL11* (c.310G>A)*, MAL11* (c.482A>T)*, YNL058c, SGO1, GDH1* (c.47G>A)*, GDH1* (c.68A>G)*, BCS1, PBP1, GSH1, ART5, ARO1, NRG1*.

Results showed that the 7 UV mutagenized strains were all clonal and contained 6 mutations, including all of the mutations that were found above 60% in the frequency experiment (Table 9). This supports the hypothesis that one cell was mutated in several locations on the chromosome and was selected for in the UV mutagenized population.

Table 9. Results from sequencing 11 SNP regions from 7 colonies picked from the top of the UV mutants growth frontier. (red) represents no mutation found, while (green) represents that the SNP was present. Both mutations in *ARO1* are represented in the *ARO1* column because both acted identically. All 7 strains appear to be clonal.

	UV_1	UV_2	UV_3	UV_4	UV_5	UV_6	UV_7
STE5 c.512C>T							
UBP7 c.2466T>A							
MAL11 c.482A>T							
YNL058c c.7T>C							
SGO1 c.575C>G							
GDH1 c.68A>G							
BCS1 43bp 5'							
PBP1 190bp 5'							
GSH1 72bp 5'							
ART5 c.454G>T							
ARO1							

Of the 20 round 5 strains, not one was alike (Table 10). Of the 20 strains, only one did not contain the *UBP7* mutation (R5_2), while all 19 other strains did, 12 as homozygous mutations and 7 with heterozygous mutations. There did not seem to be any genetic linkage between any of the mutations except for the mutations in *ARO1,* which were always mutated or not mutated at the same time as one another. Strain R5_13 only contained two (including *UBP7* c.2466T>A) of the 14 sequenced mutations, supporting the hypothesis that there are likely other mutations caused by the UV mutagenesis that can contribute to the phenotype, that are not seen in R57. However, the fact that this particular strain does contain the mutation in *UBP7* further supports the hypothesis that *UBP7* could be important to stress tolerance.

Table 10. Sequencing results of SNP regions from 14 genes from twenty individual strains picked from the top of the round 5 growth frontier. (green) indicates the mutation was homozygous, (yellow) indicates a heterozygous mutation and (red) indicates no mutation was observed.

	STE5 **c.512C>T**	***ARO1***	***NRG1*** **c.137G>T**	***BCS1*** **43bp 5'**	***MAL11*** **c.310G>A**	***MAL11*** **c.482A>T**	***PBP1*** **190bp5'**	***ART5*** **c.454G>T**	***UBP7*** **c.2466T>A**	***GSH1*** **72bp 5'**	***YNL058c*** **c.7T>C**	***SGO1*** **c.575C>G**	***GDH1*** **c.47G>A**	***GDH1*** **c.68A>G**
Chromosome	4	4	4	4	7	7	7	7	9	10	14	15	15	15
R5_1														
R5_2														
R5_3														
R5_4														
R5_5														
R5_6														
R5_7														
R5_8														
R5_9														
R5_10														
R5_11														
R5_12														
R5_13														
R5_14														
R5_15														
R5_16														
R5_17														
R5_18														
R5_19														
R5_20														

3.5 Testing the mutation in *UBP7*

After narrowing down the list of most important mutations I decided to test one of them to see if the mutation did in fact confer some form of resistance phenotype to wild type CEN.PK. From the results obtained from the previous experiments I hypothesized that the mutation found in *UBP7* was one of the more important mutations for the tolerance phenotype. As such, I chose this mutation to test. After inserting the SNP into the haploid CEN.PK 113-7D chromosome and creating a mutant homozygous diploid strain by switching its mating type, I plated the strain on 40%-70% HWSSL gradient plates alongside a wild type CEN.PK 113-7D diploid and R57 (Fig. 5). As predicted, the strain containing the *UBP7* mutation did better (grew to a higher concentration of HWSSL on the gradient plate) than the wild type, but not as well as R57. This experiment was done in triplicate and all plates showed the same pattern of growth.

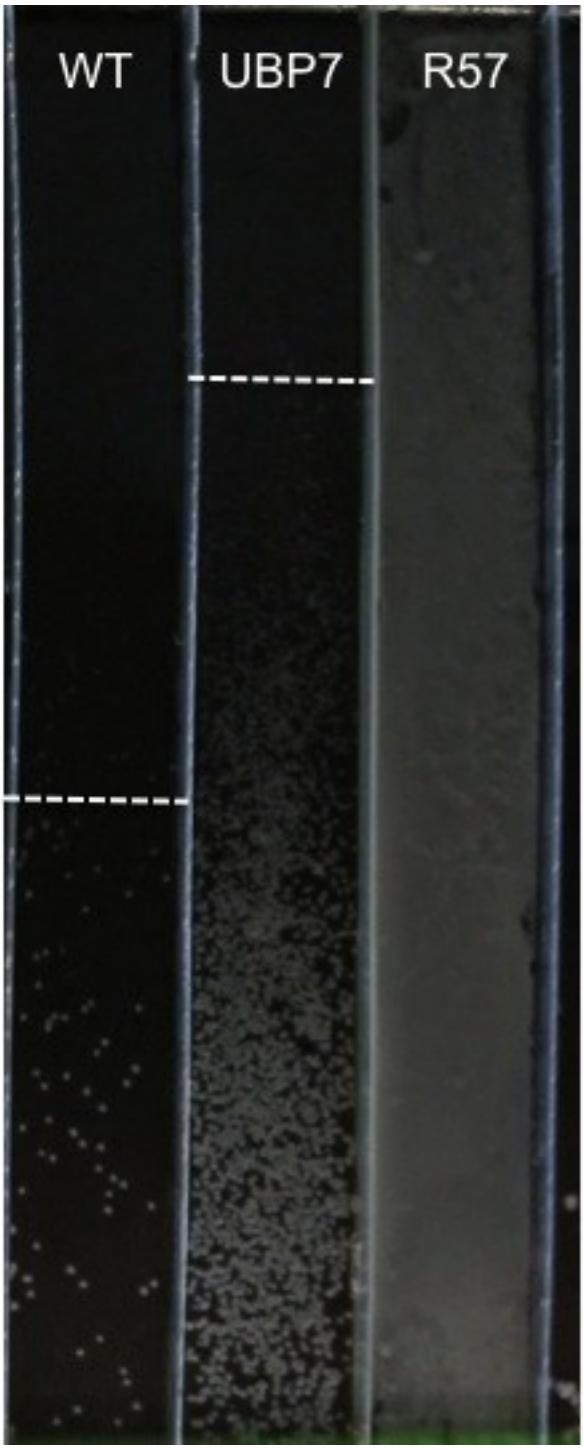

Figure 5. HWSSL gradient plate (40%-70% HWSSL from the bottom of the plate to the top) showing the phenotypic effect of the *UBP7* c.2466T>A. From left to right the lanes are plated with wild type CEN.PK 113-7D diploid, CEN.PK 113-7D diploid with the R57 homozygous mutation in *UBP7* and R57. The CEN.PK 113 7D containing the mutation grows to a higher HWSSL concentration than the wild type but not as high as R57. Dotted white lines indicate frontiers of growth.

3.6 Plasmid 3 and Plasmid 4

Plasmid 3 (P3) and 4 (P4), containing *RPB11, PBP1, ARO1, GSH1*and *GDH1* (*RPB11* was found not to be a mutation after the plasmid was already constructed and tested), and *DOP1, TOF2,* and *YNL058c* respectively were constructed using *in vivo* homologous recombination if CEN.PK 113-3C (P3) and 113-5D (P4). The plasmids consisted of the pGreg504 and pGreg506 backbone vectors and the aforementioned genes. Two versions of each plasmid was constructed, one containing the wild type CEN.PK 122 genes and one containing the R57, point mutant version of the genes. The plasmid assemblies were verified by PCR amplification with primers specific for each gene as well as the junctions between genes (Fig. 6).

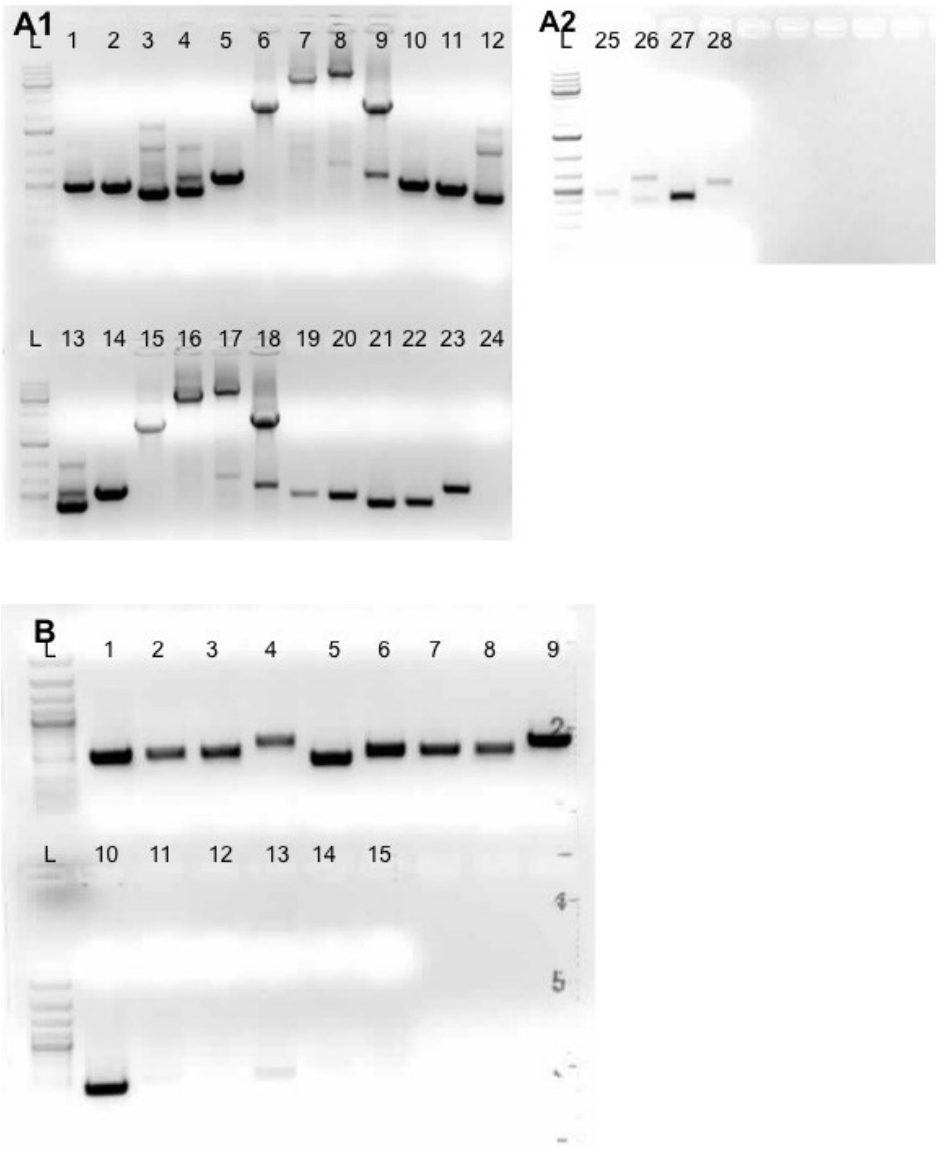

Figure 6. Agarose gel electrophoresis of PCR amplification for the confirmation of plasmid assembly. A) The ladder used in 1kb plus, then *RPB11, PBP1, ARO1, GSH1 and GDH1* and the junctions between the genes from the R57 version of plasmid are represented in lanes 1-9. Lanes 10-16 represent the same thing but with the wild type version of plasmid 3. Lanes 17-21 are products form a positive control for the primers used to amplify the gene region with wild type genomic DNA template. Lanes 22-28 lanes are the negative controls for the gene regions and the junctions between genes, in the same order using an empty vector as a template. B) The ladder lanes are 1kb plus ladder. Lanes 1-5 show the PCR products from *DOP1, TOF2, YNL058c* and the two junctions between. Lanes 6-10 show the same thing for the wild type version of the plasmid. Lanes 11-13 show the negative controls for the gene region primers on an empty vector and lanes 14 and 15 show the negative controls for the junctions between genes.

In order to test whether the plasmids affected yeast stress tolerance to HWSSL, wild type yeast strains harbouring the plasmids were plated on HWSSL gradient plates with a gradient of ~30 to 70% HWSSL. The results showed that expression of the R57 mutant version of the genes on plasmid 3 in a wild type haploid grew to the same point as the same strain harbouring an empty vector. In contrast, overexpression of the wild type gene in a wild type haploid background resulted in growth to a lower HWSSL concentration (Fig. 7). No differences in growth on HWSSL gradient plates was observed for plasmid 4 harbouring the mutant or wild type version of the genes (Fig. 8). Although the constructed plasmid assemblies were verified via PCR amplification, the assemblies were never sent for sequencing to verify that all the gene inserts had been assembled fully and in the correct order.

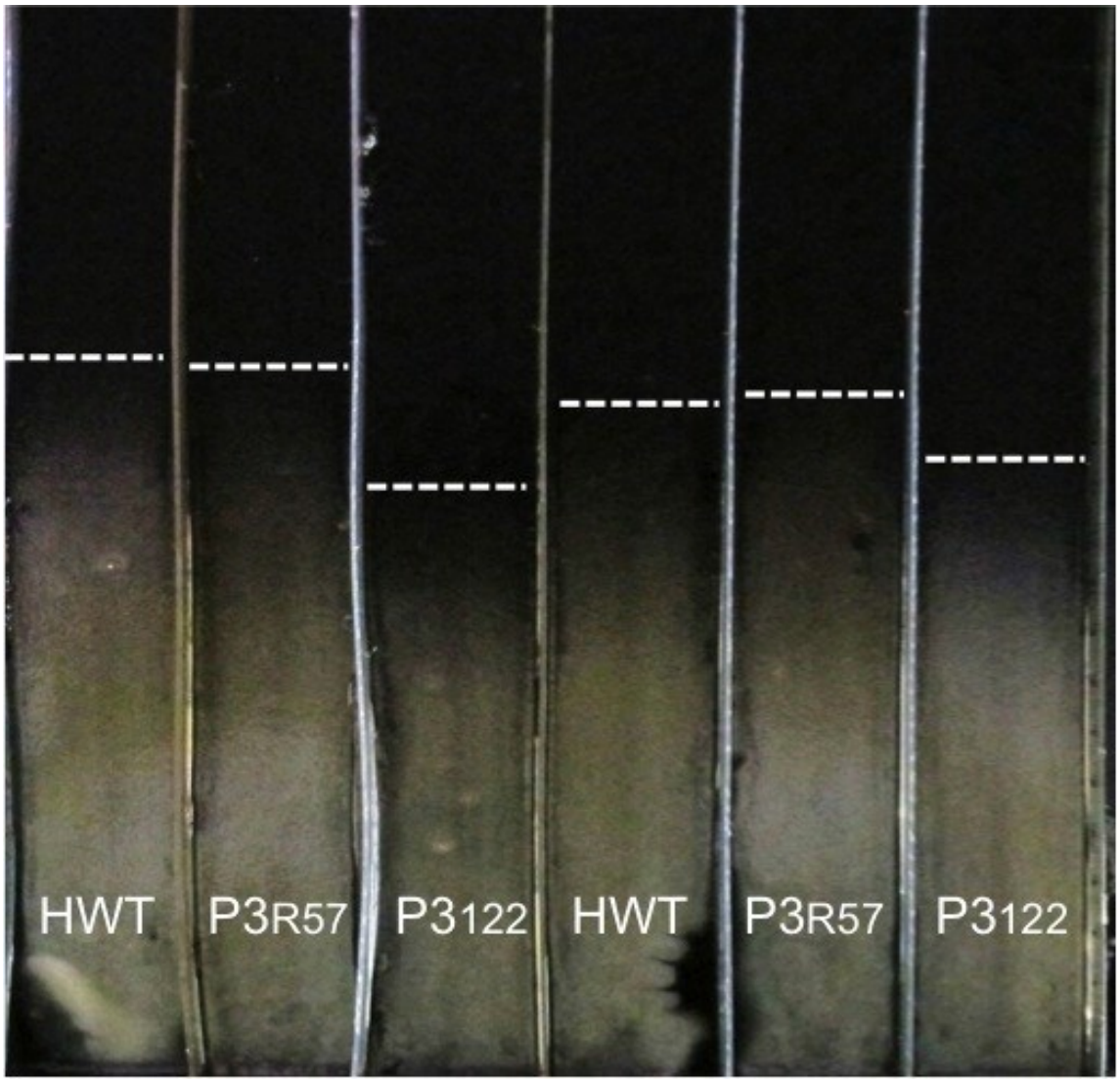

Figure 7. Growth of plasmid 3 harbouring strains on HWSSL gradient plate. HWT represents haploid wild type CEN.PK 113-3C transformed with an empty pGreg 504 vector. P3R57 represents CEN.PK 113-3C transformed plasmid 3 containing the mutated genes and P3122 represents CEN.PK 113-3C transformed with plasmid 3 containing the wild type version of the genes. Each strain was plated twice to ensure the gradient was level. The genes contained in plasmid 3 are *RPB11, PBP1, ARO1, GSH1* and *GDH1*.

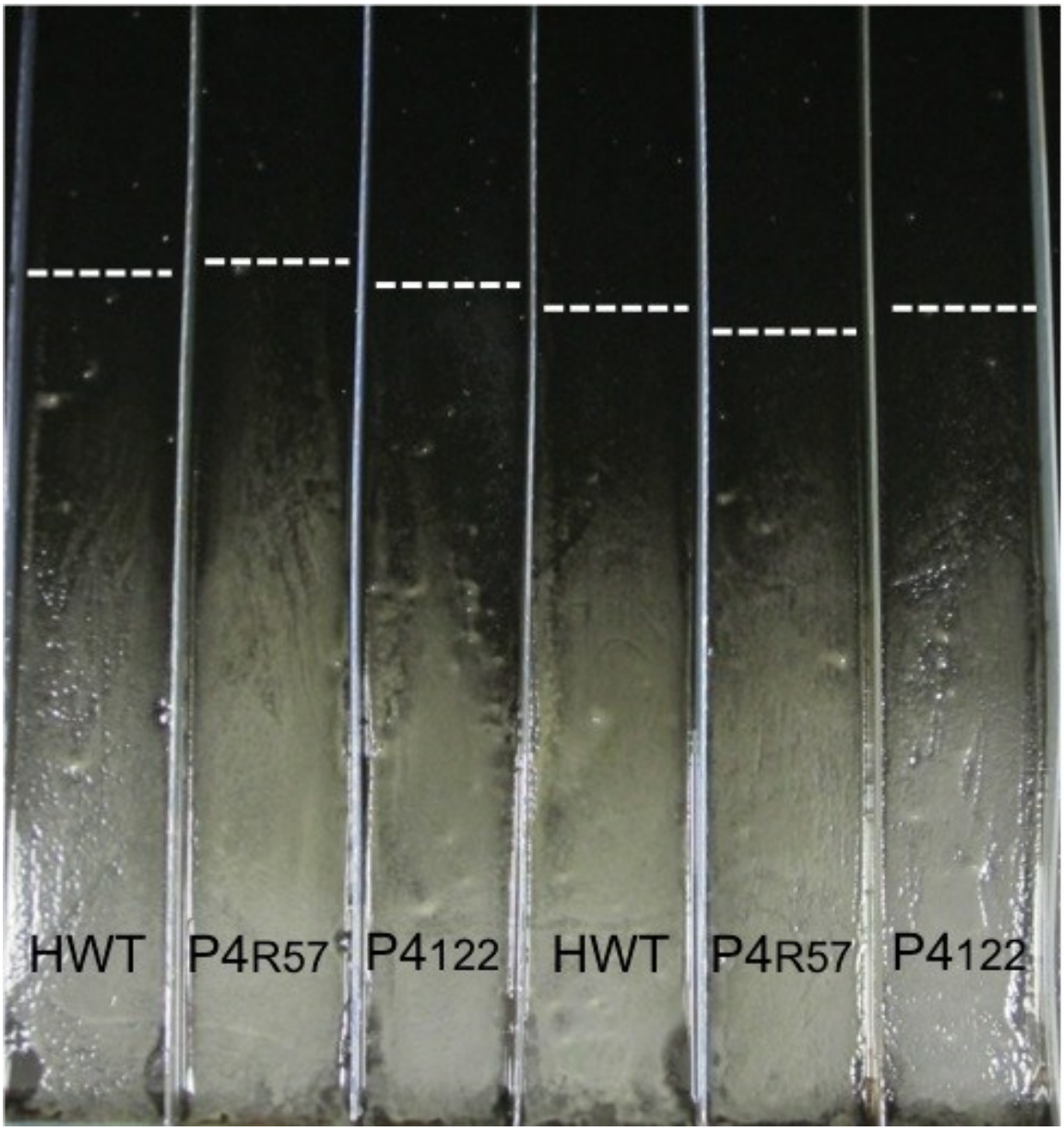

Figure 8. Growth of plasmid 4 harbouring strains on HWSSL gradient plate. HWT represents haploid wild type CEN.PK 113-5D transformed with an empty pGreg 506 vector. P4R57 represents CEN.PK 113-5D transformed plasmid 4 containing the mutated genes and P4122 represents CEN.PK 113-5D transformed with plasmid 4 containing the wild type version of the genes. Each strain was plated twice to ensure the gradient was level. The genes contained in plasmid 4 are *DOP1, TOF2* and *YNL058c.*

3.7 Testing plasmid 3 for oxidative stress

To test for the effects mutant gene expression would have on resistance to oxidative stress, the strain harbouring plasmid 3, described in section 3.6, was plated on H_2O_2 containing YNB. The reasoning behind this was that plasmid 3 contains *GSH1,* which encodes the first enzyme involved in glutathione biosynthesis. Glutathione is known to be an important antioxidant that protects yeast from ROS [64]. Transciptional activation of *GSH1* has also been linked to H_2O_2 [65]. Each plate was spotted with haploid CEN.PK containing the 122 version of plasmid 3, the R57 version of plasmid 3 and the wild type haploid with an empty pGreg504 vector. The results indicated that the over-expressed genes had no effect on resistance to oxidative stress as shown by the absence of difference in growth between any of the strains at any of the concentrations of H_2O_2 (Fig. 9).

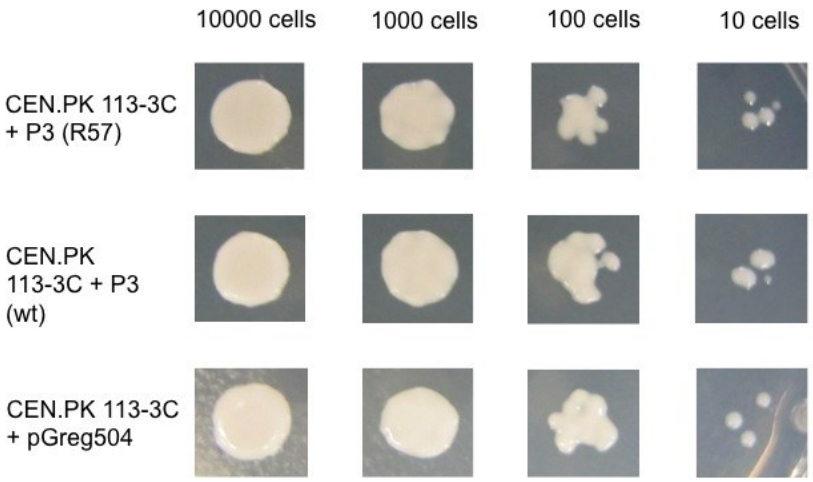

Figure 9. Testing plasmid 3 for oxidative stress tolerance. CEN.PK 113-3C harbouring plasmid 3 with mutated genes, CEN.PK 113-3C harbouring plasmid 3 containing the wild type genes and CEN.PK 113-3C harbouring a pGreg504 plasmid were plated on YNB (with amino acids) plates containing 0.5mM, 1mM, 2mM, 3mM and 5mM H_2O_2. This figure depicts a 2mM H_2O_2. 10000, 1000, 100 and 10 cells of each strain were plated and incubated (2 day incubation). No differences between the strains were observed in on plates of any concentration.

4. Discussion

In a previous body of work [1] we were able to identify a strain of yeast capable of surviving the various stresses found in the harsh lignocellulosic substrate HWSSL. This strain, R57, was subsequently sequenced and we were able to identify point mutations within the R57 genome (D. Pinel PhD thesis and this work). Using this information I tracked the frequency of these specific point mutations through the evolution of the wild type strain to the tolerant R57 strain, and I was able to determine how the frequency of the point mutations changed through the evolution. I was then able to pick several colonies from the starting UV mutant pool and several from round 5 of the genome shuffling evolution to determine if there were any specific and consistent combinations of these mutations that allowed for increased tolerance. Unfortunately this was not the case, instead this data gave us insight into the frequency of the mutations at the top of growth frontier of round 5 which is the region of growth subjected to the highest concentration of inhibitors. Finally, I was able to test one mutation, *UBP7* c.2466T>A, that I hypothesized contributed significantly to tolerance based on its abundance and evolution during the genome shuffling process.

4.1 *UBP7*

UBP7 is a de-ubiquitin protease that cleaves ubiquitin-protein fusions. It is therefore implicated in the ubiquitin-dependent protease pathway which clears misfolded or damaged proteins. It was shown to antagonize the multivesicular body (MVB) sorting of Csp1p (a regulator of meiosis and mitosis) [66]. De-ubiquitinating proteins play an essential roll in the ubiquitin pathway as they liberate ubiquitin from ubiquitin-protein conjugates so as to be conjugated again [67, 68]. Altering this function would alter the

efficiency of the ubiquitin pathway. The deubiquitinating protein family are linked to a variety of cellular processes including chromatin remodelling (Ubp8p) and gene silencing (Ubp10p) [68]. *UBP3* is also upregulated under osmotic stress [68]. Because protein (de)ubiquitination plays a large regulatory role to a vast number of cellular processes, it is easy to imagine how changes in the function of *UBP7* could impact stress tolerance on a general scale. This mutation is also homozygous which means that it must have been selected for enough for two separate cells with this mutation to mate.

Six of the 20 SNPs sequenced started at a frequency of over 60% in the UV population (*BCS1, PBP1, UBP7, YNL058c* and *SGO1*). Of these five, only the mutation in *UBP7* increased in frequency through the rounds of shuffling (Fig. 3). It has a frequency of 63.9% in the UV population and increases to 82.1% by round 5. The other 5 mutations show a decrease through the rounds of evolution. These 6 SNPs were present in all 7 clonal strains sequenced from the UV mutagenized population, indicating that they were likely clonal (i.e. originated from the same mutated cell). The mutation in *UBP7* was present in 19 of the 20 strains sequenced in round 5 and was found as a homozygous mutation in 12 of those 19 strains (63%), while mutations in *BCS1, PBP1, YNL058c* and *SGO1* were homozygous only 20%, 13%, 38% and 38% of the time, respectively (Table 10). The increase in frequency of the *UBP7* mutation from a high 60% to 82.1% (the highest frequency of any of the mutations) through the rounds, coupled with the higher incidences of homozygosity of the mutation and the steady decrease of the other 4 mutations through the rounds leads one to believe that the mutation in *UBP7* was likely the mutation which provided the mutated cell some form of tolerance to HWSSL and was therefore selected for. Because the other SNPs (*BCS1,*

PBP1, YNL058c and *SGO1*) were initially within the same cell as the *UBP7* mutation in the UV mutagenized population (Table 9), they would also be selected for and would also show up at a high frequency. After genome shuffling by recursive mating the mutations would have had a chance to segregate from one another allowing us to see a decline in the frequency of the other 4 SNPs, which were not contributing to the tolerance phenotype and therefore not being selected for. Although these 4 SNPs would gradually decrease, I believe that they would not reach zero and would eventually reach a plateau frequency due to the fact that while shuffling, the shuffled population were never being back-crossed with the wild type genome. Due to their high starting frequencies, they would decrease for a time and then reach a point where there would not be enough wild type gene influx to decrease their frequencies much further. The fact that they start at such a high frequency increases the chance that these mutations will segregate and combine with beneficial mutations. Due to this, even if these mutations are not being selected for they would remain in the population. The fact that the *UBP7* mutation is enriched through the evolution while the others are not corroborates the theory that the mutation in *UBP7* was the gene being enriched for.

Although there has not been much research on *UBP7* specifically, it is known to encode a deubiquitinating enzyme [69] and therefore contributes to the ubiquitin-dependent protease pathway, which affects many cellular processes. Due to the large frequency of this allele in the UV population, the fact that the frequency of this mutation is enriched throughout the evolution, the fact it is part of the deubiquitinating protein family which can play roles in a vast number of the cellular processes [68] and the fact that the mutation confers a polar uncharged to a charged amino acid change in the

catalytic domain of the protein it is not hard to envision it playing an important role in stress tolerance to HWSSL. Because I hypothesized this mutation in *UBP7* to be one of the more important mutations for the tolerance phenotype, I decided to introduce this mutant homozygous allele in a wild type diploid strain of CEN.PK 113-7D and tested it on HWSSL gradient plates. As a result, the strain containing the mutation grew to a higher HWSSL concentration compared to the wild type control, showing that it is in fact important for the phenotype (Fig. 4).

4.2 *BCS1, PBP1, YNL058c, SGO1*

BCS1 plays a crucial role in the expression of two of the cytochrome bc_1 complex subunits; Qcr10p and Rieske FeS. Bcs1p is also intimately involved in the assembly of the cytochrome bc_1 complex by associating with the cytochrome bc_1 pre-complex during its assembly, letting it remain in a state competent for the addition of Qcr10p and the Rieske FeS subunits. Prior to the addition of these subunits, Bcs1p is released in an ATP-dependent manner. It is thought that this binding to the pre-complex is to prevent adverse folding and interactions of the pre-complex subunits [70, 71].

Pab1p-binding protein (Pbp1p) interacts with the C-terminus of Pab1p, which is involved in mRNA polyadenylation by regulating poly(A) tail length and translational activation [72, 73]. It is thought that Pbp1p negatively regulates Pab1p. It can also negatively regulate the poly(A)-nuclease by interacting with Pan2p of the Pan2-Pan3 complex [74, 75]. *PBP1,* along with *PUB1* is also required for yeast stress granule assembly under glucose deprivation. Stress granules have been shown to form due to defects in translation initiation [76, 77] and accumulate under various stresses, including, heat shock [77, 78] and severe ethanol stress [77], but are thought to accumulate under a

large number of other stressors [76]. Overexpression of *PBP1* has been linked to an increased sequestration of Torc1p, which has been linked to regulation of permease sorting and turnover, actin organization, cell cycling and genes induced by nutrient deprivation, to stress granules (131). A component of the Torc1p complex, Tor1p has been linked, through a deletion experiment, to acetic acid, peroxide and osmotic stress tolerance (134, 133). Pbp1p also complexes with Mkt1p, a positive regulator of the HO endonuclease gene that initiates mating-type switching in *Saccharomyces cerevisiae*, to act as a regulator of HO gene expression at the post-transcriptional level. The deletion of *PBP1* results in a decrease in the amount of HO endonuclease produced [79].

This suggests that mutations leading to a different expression level of *PBP1* can in fact affect many cellular processes including some directly related to stress tolerance to inhibitors found in HWSSL. These include, peroxide tolerance, acetic acid tolerance and osmotic stress tolerance. Changes in the turnover rate of permeases can also aid in stress tolerance.

YNL058c is a gene of unknown function. However, it has been shown to localize to the vacuole in *S. cerevisiae* in a GFP-fusion protein experiment [80]. It has also been shown to be a non-essential gene in yeast [81]. In a deletion study, it was shown to be a regulator of cell wall integrity through the cell wall integrity mitogen-activated protein kinase (CWI-MAPK) pathway. In this study, deleting *YNL058c* resulted in increased phosphorylation of the CWI-MAPK which is known to increase cell wall integrity [82]. Cell wall integrity could play a crucial role in mitigating the effects of many of the stressors found in HWSSL, including acetic acid and osmotic stress.

Sgo1p (also known as shugoshin) protects centromeric cohesion during meiosis 1. It also functions as a means to sense lack of tension on mitotic chromosomes, acting as a spindle checkpoint. If *SGO1* is knocked out of budding yeast, the strain looses the ability to arrest the cell cycle when chromosomes are not under tension, leading to yeast death after microtubule depolymerisation [83]. A homologue of budding yeast *SGO1* found in fission yeast has been shown to function as a protector of centromeric cohesion Rec8p from degradation during meiosis 1 thereby protecting sister chromatid cohesion during meiosis 1. [84-86].

Unfortunately, the frequency of the mutations in *BCS1, PBP1, YNL058c* and *SGO1* all decrease through the rounds, essentially either being selected against or not selected for, thereby being diluted in the population. They show up at relatively high frequencies in the 20 strains picked from round 5, but this is due to the fact that they started at a high frequency of over 60% in the UV round. As stated in section 3.1 it is my hypothesis that due to the lack of wild type back-crossing these mutations, although not being selected for, will reach a plateau frequency. This plateau can be observed in figure 3, where there frequencies decrease dramatically from the UV mutagenized population to round 1 and then decrease at a much lower rate from then on. Moreover, *BSC1* and *SGO1* are not shown to be involved in yeast stress tolerance. For these reasons I do not think that these genes should be considered good gene targets for more lignocellulosic inhibitor tolerance.

4.3 *NRG1*

NRG1 encodes a DNA-binding, glucose dependent transcriptional repressor. Nrg1p recruits the general co-repressor Ssn6p (Cyc8p) -Tup1p complex to promoters in

order to repress expression. It was originally implicated in glucose repression, repressing such genes as *STA1* (encoding glucoamylase isozyme GA1)[87], *SUC2* and *GAL* [88]. It was later discovered that Nrg1p plays a role in stress response, regulating a large set of stress response genes [89]. Through the use of deletion experiments, it was shown to play a role in osmotic stress as well as oxidative stress responses. When an *nrg1Δ* strain is plated on NaCl spot plates it does significantly better than the wild type strain [89, 90]. *NRG1* was shown to be a negative regulator of ion tolerance as well as a negative regulator of two pH-induced genes, *ENA1* and *ZPS1* [91]. It therefore plays a major role in pH-responsive gene regulation and excess ion (Na^+ and Li^+) tolerance. *NRG1* can act as a negative regulator of *FLO11*, a gene in the Sn1p-pathway, which mediates invasive and filamentous growth [92]. The transcription of Nrg1p has been suggested to be regulated by Haalp in times of acetic acid stress [93].

The mutation in *NRG1* increases through the rounds of shuffling starting with a frequency of 2.3% in the UV population, a frequency under 1% in round one and then increasing to 2.1% and 11.2% in rounds 3 and 5 respectively. This jump from 2.3% to 11.2% is quite large and shows a selection for this mutation in the population. The fact that we see a frequency of less than 1% in round 1 could be explained by the fact that in R57 this mutation is homozygous. I hypothesize that this mutation must be homozygous to confer a tolerance phenotype. I detect it in the UV population because the cells in the UV population are haploids and therefore any mutations are automatically "homozygous" because of the singular chromosome set. Round 1 is the first round where the vast majority of cells are diploid and therefore it is less likely that two haploids with the *NRG1* mutation have had a chance to mate. The more instances of recursive mating that

occur the higher the chance of mating two haploids with this mutation. Once a homozygous diploid has been created it is enriched, further increasing the probability the haploids from that round mating with another cell with the mutation. It is seen at a low frequency in the 20 strains picked from the round 5 growth frontier, but I believe this to be due to the fact that the overall frequency of this mutation is still relatively low. Although found in low abundance at the growth frontier, the allelic frequency of this mutation is enriched. With more rounds of shuffling I would expect the frequency to rise.

NRG1 encodes a repressor of genes that are linked to osmotic, oxidative and acetic acid stresses, which are all found in a HWSSL environment. This coupled with the increasing frequency of this mutation, the fact that it is enriched at the round 5 growth frontier, that the mutation causes an amino acid change from one with a ring structure to one without and the fact that is found as a homozygous mutation in R57 leads to the hypothesis that *NRG1* should be considered a possible target of further lignocellulosic inhibitor tolerance research.

4.4 *GDH1*

GDH1 encodes one of 2 NADP(+)-dependent glutamate dehydrogenase isoenzymes in *S. cerevisiae*. These isoenzymes catalyze the synthesis of glutamate from ammonium and α-ketoglutarate in an NADPH-dependent manner. Glutamate is responsible for 80% of cellular nitrogen in yeast [94]. Of the two genes responsible for glutamate catabolism (*GDH1* and *GDH3)*, *GDH1* is the primary one. It is NADPH-dependent and thus knocking *GDH1* out creates an increase in available cytosolic NADPH [95], which can be used in other NADPH-dependent reactions such as the reduction of 5-HMF and furfural into their less toxic alcohols. Nitrogen metabolism

genes, including *GDH1* were shown to be down-regulated in a strain of yeast engineered to be tolerant to many of the inhibitory compounds found in HWSSL [96]. This led to the belief that decreases in nitrogen metabolism rate conserves energy to be applied towards stress tolerance [96]. The expression of *GDH1*, regardless of carbon source, is activated by several chromatin remodelling factors, including Spt3p, Ada3p and Snf2p of the Swi-SNF complex, while Spt3 and Ada2p up-regulate expression under ethanol-grown conditions[97]. Under glucose rich conditions *GDH1* is the primary glutamate dehydrogenase expressed, while under glucose poor, ethanol rich conditions *GDH3* is primarily expressed [98, 99]. *GDH1* expression is repressed in the presence of amino acids. When *S. cerevisiae* is grown with glucose as its sole carbon source, Gdh1p is expressed normally when given any nitrogen source. This differs when ethanol is used as a sole carbon source. In this case, maximal transcriptional activity was only found when ammonium was given as a nitrogen source [97]. Deletion of *GDH1* leads to increased ethanol production compared to wild type strains in anaerobic growth on glucose [100].

Interestingly, two separate mutations in the *GDH1* gene were enriched enough during the evolution of the strain to come together in the final R57 strain and that these 2 non-silent mutations both increase in frequency with of one of them increasing dramatically through the rounds of shuffling, beginning at 1.4% in the UV population and ending at 27.6% in round 5 and the other beginning at less than 1% in the UV population and ends at 10.3%. Of the 20 individual round 5 strains that were picked from the top of the growth frontier, over 50% (11/20) contained the *GDH1* c.68A>G mutation and 6 contained the *GDH1* c.47G>A mutation. Due to the fact that there are two mutations in this gene in R57, that both increase in frequency through the rounds (Fig. 3), and that the

mutations seem to be enriched at round 5 growth frontier (Table 10), it would seem that mutations in this gene are selected for and enriched. Moreover, this gene codes for an NADPH dependent enzyme, which when knocked out increases cytosolic NADPH which can be used for 5HMF reduction [95] or for the reduction of oxidized radical scavengers, and a previous study has shown that *GDH1* is down-regulated in another inhibitor tolerant strain of yeast. For these reasons, it is conceivable that these mutations play some role in the tolerance phenotype observed in R57 and as such, this gene should be a target for further research into tolerance to lignocellulosic hydrolysates.

4.5 *MAL11*

MAL11, also known as *AGT1*, encodes a high-affinity α-glucoside transporter. Mal11p is a highly hydrophobic integral membrane protein with high sequence homology to members of the 12 transmembrane domain superfamily of sugar transporters. It is a high affinity maltose/proton symporter, but has also been shown to transport isomaltose, α-methylglucoside, maltotriose, palatinose, trehalose, sucrose and melezitose [101]. Moreover, this gene is also essential for efficient maltotriose fermentation in *S. cerevisiae*. Because Mal11p is an H^+ symporter it is conceivable that it could affect cytosolic pH levels. Trehalose can aid in mediating osmotic stress, which is a stress associated with HWSSL [102]. The expression of this gene is maltose inducible and is mediated by the Mal-activator as well as subject to glucose induced inactivation [101, 103-107]. Threonine 505 (in the transmembrane segment) and Serine 557 (in the intracellular segment) are essential for efficient transport of maltotriose in *S. cerevisiae* [108].

Similarly to *GDH1*, the R57 *MAL11* gene contains 2 mutations, one of which increases from less than 1% in the UV population to 32.8% in round 5 (Fig 3). *MAL11* is a confirmed transporter of trehalose [101], a mediator of osmotic stress in yeast [102], which is one of the main inhibitors found in HWSSL. Of the 20 strains picked from the top of the round 5 growth frontier, 11 contained the *MAL11* c.310G>A mutation and 2 contained the *MAL11* c.482A>T mutation indicating that the *MAL11* c.310G>A mutation is enriched at the round 5 growth frontier. This gene should be regarded as a possible gene target for increased tolerance to the inhibitors found in lignocellulosic hydrolysates due to the fact that two mutations are found within this gene in R57, that the frequency of one of them is so drastically increased from the UV population to round 5 and the fact that the *MAL11* gene has been linked to osmotic stress and cytosolic pH. Moreover, the mutations found in this gene are located in the sugar transport domain of the protein and one leads to an amino acid change from an amino acid containing a ring structure to one that does not.

4.6 *GSH1*

GSH1 encodes the first enzyme involved in glutathione biosynthesis, γ-glutamylcystein synthetase, in yeast. Glutathione (g-L-glutamyl-L-cysteinylglycine) is an important antioxidant in yeast, protecting against ROS and hydrogen peroxide in the organism [64]. Hydrogen peroxide can induce *GSH1* transcription in the presence of glutamate, glutamine and lysine [65]. Glutathione also plays a role in protecting the cell against heavy metals and pesticides [109]. Glutathione, with its redux-active sulphydryl group, acts as a free-radical scavenger. It interacts with oxidants to produce oxidised glutathione then recycles back to its reduced form using NAD(P)H. *GSH1* mutants have,

in the past, created yeast strains which are hypersensitive to oxidants, in particular H_2O_2, contributing further proof that *GSH1* plays a role in oxidative stress tolerance [65, 110, 111]. The absence of glutathione activates HSPs, proteins that are important in cell signalling and response to proteotoxic stress [112].

Although the mutation in *GSH1* seems to increase from 34.7% in the UV round to 54.7% in round 1, then diminish to 40.1% and then 27.2% in rounds 3 and 5 respectively (Fig. 3), it shows up in 13 of the 20 strains picked in round 5, where 9 of the 13 strains were homozygous mutations, resulting in the mutant allele representing 55% for the 20 strains (Table 10). Therefore, although it might diminish in frequency over the whole population from the wild type cut-off to the growth frontier, the mutation is still highly represented at the growth frontier, which is the area of growth that is exposed to the highest concentration of inhibitors. A *GSH1* mutation in R57 could potentially contribute to the proteotoxic stress tolerance of the strain, as well as tolerance to the oxidative stress caused by HWSSL. These include hydrogen peroxide and other ROS found in HWSSL or produced by toxic compounds such as 5-HMF or furfural. Furthermore, the mutation is found in the *GSH1* promoter in R57. This means that expression levels of the gene could be affected. If the mutation strengthens the promoter there would be an increase in glutathione biosynthesis and therefore increased stress tolerance. For these reasons I would consider *GSH1* as a possible gene target for stress tolerance.

4.7 *ART5*

ART5, or arrestin-related trafficking adapter 5, is necessary for the inositol-induced internalization of the plasma membrane transporter Itr1p [113]. Knocking down,

or knocking out the *TAM41* gene product leads to impaired mitochondrial protein transport via the Tim23 complex due to a destabilization of the Tim23 complex [114]. This leads to growth defects as well as accumulation of precursor forms of mitochondrial proteins such as HSP60p and Mdj1p. Overexpression of *ART5* in *TAM41* knockout strains seems to re-stabilize the Tim23 complex and leads to lesser accumulation of mitochondrial proteins suggesting that *ART5* plays a role in mitochondrial protein import. Creating a double knockout strain, *tam41Δ art5Δ,* leads to more sever growth defects as well as more mitochondrial protein accumulation [114].

The arrestin family of proteins have been found to bind to Rsp5p, acting as adaptors. Rsp5p is a ubiquitin ligase that promotes endocytosis of plasma membrane proteins for vacuolar degradation. *ART5* has homology to *ART2,8,4,7,6,3* and *9* and all have the potential to act as Rsp5p adaptors, but Art5p and Art9p were not shown to co-purify with Rsp5p in a large proteomic study [115].

The mutation found in *ART5* increases in frequency from 28.7% in the UV population to 56.1% in round 1 and remains in the 50s thereafter (Fig. 3). Due to the fact that *ART5* is a member of the arrestin family of proteins, it is possible that it plays a role in ubiquitination and thus has the potential to play a large regulatory role in many cellular processes, which can help confer HWSSL tolerance. However, the mutation observed confers a leucine to isoleucine amino acid change. These amino acids are both very similar and within the same hydrophobic side chain group.

The fact that the mutation starts off at such a high frequency above the wild type cut off suggests that the cells containing this mutation were selected for. This frequency

then increases and plateaus in the 50s, suggesting enrichment of the mutation in the population. Because it plateaus in the 50s and does not really increase or decrease in frequency to a great degree it suggests that this mutation was possibly important in the initial enrichment from the UV mutagenized population to round 1, but then became less important for the increase in tolerance in subsequent rounds. Because of its large frequency in round 1 and the fact that no back-crossing with wild type strains took place, the frequency plateaued. Because it is still useful for strains growing to a certain concentration of HWSSL above the wild type growth cut-off, the frequency does not decrease greatly and remains in the 50s. Additionally, of the 20 strains collected from the growth frontier of round 5, 14 contained the *ART5* mutation. This is a frequency of 70% in the area most concentrated with HWSSL. Again, this would suggest some selection. Additionally, this gene is also associated with the ubiquitin-dependent pathway [115], which can affect many pathways in the yeast cell. For these reasons I believe *ART5* to be a good candidate fro further stress tolerance research. However, due to the similarity of the mutated protein's amino acid to the wild type and the data suggesting that this mutation plays a role in the stress tolerance phenotype, I believe further exploration of this gene and this mutation should be considered to better understand if and how this leucine to isoleucine change is conferring a phenotypic change to inhibitor tolerance.

4.8 *STE5*

STE5 encodes a pheromone-responsive mitogen-activated protein kinase (MAPK) scaffold protein. Mating pheromones are a trigger for Ste5p plasma membrane recruitment by a pheromone activated Gβγ (STE4p-STE18p) dimer as well as binding to the plasma membrane using the protein's N terminal plasma membrane/ nuclear

localization signal domain. Correct recruitment of Ste5p requires both [116]. Interaction of the Gβγ with Ste5p produces a conformational change allowing Ste20p to interact with Ste11p [117]. It also interacts with Ste7p (a MEK) and Ste11p (a MEKK). Ste5p forms a multiprotein complex with Ste7p, Ste11p and FUS3, thereby allowing the hyperphosphorylation of Ste7p by Ste11p, which in turn phosphorylates Fus3p, activating it and releasing it from the complex to phosphorylate several targets involved in mating (including transcription factors, a cyclin-dependent kinase inhibitor, and upstream signalling components) [118-121]. Mating pheromones act as a signal for recruitment of Ste5p to the membrane. This allows the activation of the MAPK cascade by bringing Ste11p in proximity of its kinase, Ste20p. The membrane recruitment is also thought to relive auto-inhibition of Ste5p from activating Fus3p [122].

The switch-type response to mating pheromones is thought to be modulated by the dephosphorylation of 4 sites on the Fus3p-Ste5p complex. These dephosphorylations, thought to be mediated by Ptc1p, release Fus3p from Ste5p allowing it to be completely active [123].

Due to the mutation's low frequency in the 5th round of shuffling as well as its low abundance in the 20 coloines picked from the round 5 growth frontier, and that fact that *STE5* has not yet been shown to be related to stress tolerance, one would expect that Ste5p is not likely to participate in the stress response phenotype observed in R57. However, *STE5* is an integral part of the yeast response to mating pheromones, and as such, it is possible that the mutation found in R57 could have somehow helped with genome shuffling by increasing the rate of mating during the mating steps.

4.9 *ARO1*

The *ARO1* gene produces a large pentafunctional arom protein which catalyzes five consecutive steps in the production of chorismate from erythrose-4-phosphate and phosphoenolpyruvate [124]. The protein product of *ARO1* catalyzes the steps 2 through 6 in the aromatic amino acid biosynthesis pathway. The enzymatic activities of the protein in order are that of a 3-dehydroquinate synthase, a 3-dehydroquinase, a shikimate dehydrogenase, shikimate kinase and 5-enolpyruvylshikimate 3-phosphate (EPSP) synthase [125]. It was shown that weak acids inhibit the ability of yeast to uptake aromatic amino acids from the environment [126]. It has also been shown that the deletion of *ARO1* leads to ethanol and osmotic stress sensitivities. It is thought that the ethanol sensitivity is due to the lack of tryptophan production and the osmotic stress sensitivity is due to the simultaneous lack of tyrosine and phenylalanine production[127].

Due to the fact that *ARO1* has been linked to acetic acid, ethanol and osmotic stress response, and the fact that acetic acid along with osmotic stress are found in HWSSL, it is conceivable that the mutations in *ARO1* could play a role in the observed tolerance phenotype.

Unfortunately, as with the mutation found in *STE5,* the mutations found in *ARO1* increase until the third round of shuffling and then decrease to 19.9% and 10.4% for the fifth round of shuffling. Only 3 of the 20 colonies picked at the round 5 growth frontier contained these mutations, but all 3 contained both mutations. There are two mutations directly adjacent to each other, c.1283C>T and c.1284C>T, but they are genetically linked suggesting they occurred at the same time and because of their proximity never had a good chance to segregate. For the reasons listed above I would not consider this

gene to be an important gene target for further research on lignocellulosic inhibitor tolerance. I believe it might play a minor role in increasing tolerance but not a major one.

4.10 *SSA1, TOF2*

There was one mutation found in the *SSA1* gene of R57, but while analyzing for the frequency of this mutation through the rounds of shuffling, a second mutation was discovered in the population. Both of the mutations maintain a very low frequency through the rounds, with one starting at 1.1% and then lowering to and staying at a frequency of less than 1%, and the other mutation going from less than one percent to 2.6%.

SSA1 encodes a member of the heat shock 70 proteins (HSP70). It is one of the two constitutively active HSP70s in yeast and can be a regulator of *HSF1* expression. Its deletion leads to the expression of Hsf1p at normal growth temperatures [128, 129]. Hsf1p is a transcription factor that responds to various stresses. It has been linked to the maintenance of cell wall integrity, oxidative stress, glucose starvation, carbohydrate metabolism, energy generation and protein folding [130-132]. *SSA1* also participates in protein ubiquitination and ultimately degradation, playing an essential role in the ubiquitin-proteasome system [133, 134]. Another study suggests that Ssa1p plays a role in protein translocation into the endoplasmic reticulum (ER) [135]. Ssa1p is also a mediator in the degradation of the gluconeogenic enzyme fructose-1,6-bisphosphatase [136] and was shown to diminish prion propagation [137] and play a role in translation initiation by interactions with Sis1p and Pab1p [138]. Mutations in the same amino acid position as our mutation were shown to decrease prion propagation [138]

Although 2 mutations were found within *SSA1*, their low frequencies suggest that they do not contribute much, if any, to the tolerance phenotype.

Similarly, the mutation found in *TOF2* stays at a very low frequency throughout the rounds of shuffling. *TOF2*, encoding a topoisomerase interacting factor that has been demonstrated to promote rDNA silencing and mitotic rDNA condensation as well as condensing recruitment to the replication fork [139-141]. *TOF2* has also not been associated with any sort of stress response in the past. Therefore, I do not think it likely that this mutation confers any phenotypic changes in regards to stress tolerance.

4.11 *AVL9*

AVL9 is a gene that when mutated is synthetically lethal in *vps1Δ apl2Δ* strain backgrounds [142]. VPS1p is a GTPase required for vacuolar sorting [143], but is involved in cytoskeleton organization [144], endocytosis [145], late Golgi retention of some proteins [146] and the regulation of peroxisome biogenesis [147]. Apl2p is a subunit of the clatherin-associated protein complex which binds clatherin and is involved in clatherin dependent Golgi sorting [148]. Depletion of Avl9p in the same double knockout strain results in secretory defects, the accumulation of Golgi-like membranes, as well as a depolarization of actin cytoskeleton and defects in polarized secretion. Starting at amino acid position 601 *AVL9* is suggested to have a variable region that is not conserved between *S. cerevisiae* species, but is translated non-the-less. Overexpressing Avl9p is toxic in wild type *S. cerevisiae*. This overexpression leads to vesicle accumulation and a post-Golgi defect in secretion [142]. Although the specific biochemistry involved in Avl9p's activity is still unknown, it is clear that it has a role in the late secretory pathway.

By sequencing the SNPs from the CEN.PK 113-1A genome I discovered that the *AVL9* gene of CEN.PK 113-1A and CEN.PK 113-7D, the wild type haploids of the CEN.PK, are not isogenic. The genome of CEN.PK 113-1A contains the c.1806C>G difference we predicted to be an SNP. The fact that CEN.PK 113-1A and CEN.PK 113-7D are not actually isogenic could explain the reason CEN.PK 113-1A seems to survive longer in HWSSL compared to CEN.PK 113-7D [1]. This also explains why this allele was found in greater than 60% of the UV population. The fact that this percentage decreases would indicate that it is either being selected against of just not selected for from our rounds of shuffling. Because the polymorphism in *AVL9* initially decreases from 67.3% to 47.4%, then increases to 48% and 52.4%, I would hypothesize that the mutation is not being selected for or against, but is actually at the mercy of the other mutations it segregates with. This SNP is present in over 60% of the population initially because it is present in one of the mating pair, the CEN.PK 113-1A strain, which is what 3 of the 5 pooled UV mutagenized populations were before the UV mutagenesis (60%). I hypothesize that the initial large drop in frequency is due to the fact that this mutation is not very important for the tolerance phenotype and it is therefore not being selected for. The reason it stays in the population at around the same frequency from round 1 on is due to the fact that it started with a high frequency of 67.3% and because there was no back-crossing with the wild type strain, the mutation will plateau and vary slightly depending on which other mutations it segregates with. If the other mutations the *AVL9* mutation segregated with are selected for then we would see an enrichment in the *AVL9* mutation as well. If the *AVL9* mutation did not segregate with beneficial mutations then the cell would not be selected for and would therefore go down in frequency. Moreover, the fact

that the allelic variation is in the variable region of this protein leads one to believe that this variation is not extremely important for tolerance.

4.12 *DOP1* and *FIT3*

Due to the fact that the overall frequency of the mutations in *DOP1* and *FIT3* in the population of each round was less than 1% (Fig. 3) it is unlikely that these two mutations are very important for the tolerance phenotype of R57. Moreover, *DOP1* is a gene of unknown function that has not been implicated in any stress response pathway [149-153]. *FIT3* on the other hand has been implicated in acetic acid and furfural tolerance. However, the mutation in question is located 43bp 3' of the ORF and is therefore not likely to cause a major phenotypic effect.

4.12 Creating plasmids to test mutation phenotypes

Although I attempted to create 4 plasmids containing the divided mutant genes in the hopes of simulating heterozygous mutations of the genes, I later decided to switch approaches. The reasons for this were 5-fold. I was able to create 2 of the 4 plasmids, but the remaining two were proving difficult to assemble. It is very possible that some of the homozygous mutations would only confer their phenotype if there were two sets of the mutated gene and my strategy did not take this into account. Another reason for not pursuing this was the discovery that the 2 *MAL11* mutations were not present on the same chromosome in R57. To recreate this we would have needed to assemble two separate *MAL11* mutant genes into the plasmids, which would give a total of 3 *MAL11* genes in the cell (two on the plasmid and one on the chromosome) instead of 2, possibly confounding the results. I also realized that putting plasmids containing the mutated

genes into a haploid, resulting in all the mutated genes having two copies, one on the chromosome and one on the plasmid, and all the non-mutated genes only having one copy did not truly mimic, or represent, having mutations in a diploid cell. After testing 2 of the plasmids, I also realized that the method I was using would only be able to test for gain of function mutations because I was adding a mutated gene to a cell that already had a functional non-mutated version of the gene on its chromosome. This method could not take into account the impact of loss of function mutations because if the mutated gene on the plasmid was non-functional, the cell would just use the non-mutated gene on the chromosome and behave normally. For these reasons I discontinued my efforts on this experiment and opted for the approaches above.

5. Conclusions and Future Work

This body of work has been able to take an initial list of over 40 predicted SNPs in a strain of yeast tolerant to HWSSL and narrow it down to a few gene targets that are more likely to affect tolerance.

I was able to eliminate the vast majority of them as errors in SNP calling using Sanger sequencing. Of the 20 that remained the mutations found in *UBP7, NRG1, MAL11, GDH1, GSH1* and *ART5* all show enrichment through shuffling and all but *NRG1* are represented to a high degree at the growth frontier of round 5, which is the area of growth subjected to the highest concentration of HWSSL. They have all also been associated to stress tolerance pathways in yeast. The accumulation of this data leads me to believe that there are likely a few genes that confer a large change in tolerance, these would include *UBP7, MAL11, GDH1* and *ART5*, and some that confer smaller changes

that when combined with the others give even greater resistance. These would include *NRG1* and *GSH1*. To take this one step further I decided to test my hypothesis and chose one of the mutations I thought largely contributed to the phenotype to be tested. In doing this I confirmed that the mutation found in *UBP7* did in fact contribute to the tolerance phenotype, thereby confirming that *UBP7* should be researched as a gene target for increased tolerance to lignocellulosic hydrolysate inhibitors.

There are likely to be more mutations that confer a major change in resistance that we do not see here (that were not present in the R57 strain) that could be revealed through the genome sequencing of other round 5 strains that grew at the growth frontier, strains like R5_2 (which does not contain the *UBP7* mutation), R5_13 (which only has two of the R57 mutations) and R5_18 (which only has 4 of the R57 mutations, 3 of which are heterozygous).

We were also able to discover an allelic difference between the two wild type haploids of CEN.PK yeast in the *AVL9* gene.

Future work would have to be completed to substantiate these hypotheses. One would have to create diploid strains of CEN.PK yeast with the mutations of interest inserted into the chromosome and tested for improved tolerance on HWSSL gradient plates, as was done for the mutation in *UBP7*. This should be done for each of the 5 mutations individually, in both homozygous and heterozygous forms, and then different combinations of these mutations should be made as well.

Most of the inhibitors found in HWSSL are common to all lignocellulosic hydrolysates [1, 9-13], and discovering new gene targets for conferring tolerance would

be a huge step towards building strains of yeast capable of high throughput biofuel production from lignocellulosic waste.

6. References

1. Pinel, D., et al., *Saccharomyces cerevisiae Genome Shuffling through Recursive Population Mating Leads to Improved Tolerance to Spent Sulfite Liquor.* Applied and Environmental Microbiology. **77**(14): p. 4736-4743.
2. Naik, S.N., et al., *Production of first and second generation biofuels: A comprehensive review.* Renewable and Sustainable Energy Reviews, 2010. **14**(2): p. 578-597.
3. Havlik, P., et al., *Global land-use implications of first and second generation biofuel targets.* Energy Policy, 2011. **39**(10): p. 5690-5702.
4. Hill, J., et al., *Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels.* Proceedings of the National Academy of Sciences, 2006. **103**(30): p. 11206-11210.
5. Martin, M.A., *First generation biofuels compete.* New Biotechnology, 2010. **27**(5): p. 596-608.
6. Sims, R.E.H., et al., *An overview of second generation biofuel technologies.* Bioresource Technology, 2010. **101**(6): p. 1570.
7. Kumar, P., et al., *Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production.* Industrial & Engineering Chemistry Research, 2009. **48**(8): p. 3713-3729.
8. Palmqvist, E. and B. Hahn-Hägerdal, *Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition.* Bioresource Technology, 2000. **74**(1): p. 25-33.
9. Helle, S.S., T. Lin, and S.J.B. Duff, *Optimization of spent sulfite liquor fermentation.* Enzyme and Microbial Technology, 2008. **42**(3): p. 259-264.
10. Nigam, J.N., *Ethanol production from hardwood spent sulfite liquor using an adapted strain of Pichia stipitis.* Journal of Industrial Microbiology & Biotechnology, 2001. **26**(3): p. 145-150.
11. Smith, M.T., D.R. Cameron, and S.J.B. Duff, *Comparison of industrial yeast strains for fermentation of spent sulphite pulping liquor fortified with wood hydrolysate.* Journal of Industrial Microbiology & Biotechnology, 1997. **18**(1): p. 18-21.
12. Keating, J.D., C. Panganiban, and S.D. Mansfield, *Tolerance and adaptation of ethanologenic yeasts to lignocellulosic inhibitory compounds.* Biotechnology and Bioengineering, 2006. **93**(6): p. 1196-1206.
13. Olsson, L. and B.r. Hahn-HÃ¤gerdal, *Fermentative performance of bacteria and yeasts in lignocellulose hydrolysates.* Process Biochemistry, 1993. **28**(4): p. 249-257.
14. Allen, S.A., et al., *Furfural induces reactive oxygen species accumulation and cellular damage in Saccharomyces cerevisiae.* Biotechnol Biofuels, 2010. **3**(2): p. 1-10.
15. Liu, Z., et al., *Adaptive response of yeasts to furfural and 5-hydroxymethylfurfural and new chemical evidence for HMF conversion to 2, 5-bis-hydroxymethylfuran.* Journal of Industrial Microbiology & Biotechnology, 2004. **31**(8): p. 345-352.
16. Modig, T., G. Lidén, and M.J. Taherzadeh, *Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase.* Biochemical Journal, 2002. **363**(Pt 3): p. 769.
17. Taherzadeh, M., et al., *Physiological effects of 5-hydroxymethylfurfural on Saccharomyces cerevisiae.* Applied microbiology and biotechnology, 2000. **53**(6): p. 701-708.

18. Banerjee, N., R. Bhatnagar, and L. Viswanathan, *Inhibition of glycolysis by furfural in Saccharomyces cerevisiae.* Applied microbiology and biotechnology, 1981. **11**(4): p. 226-228.
19. Almeida, J.R.M., et al., *Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae.* Journal of Chemical Technology & Biotechnology, 2007. **82**(4): p. 340-349.
20. Pampulha, M. and M. Loureiro-Dias, *Combined effect of acetic acid, pH and ethanol on intracellular pH of fermenting yeast.* Applied microbiology and biotechnology, 1989. **31**(5): p. 547-550.
21. Palmqvist, E., et al., *Main and interaction effects of acetic acid, furfural, and p-hydroxybenzoic acid on growth and ethanol productivity of yeasts.* Biotechnology and Bioengineering, 2000. **63**(1): p. 46-55.
22. Verduyn, C., *Physiology of yeasts in relation to biomass yields.* Antonie van Leeuwenhoek, 1991. **60**(3): p. 325-353.
23. Verduyn, C., et al., *Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation.* Yeast, 1992. **8**(7): p. 501-517.
24. Terada, H., *Uncouplers of oxidative phosphorylation.* Environmental Health Perspectives, 1990. **87**: p. 213.
25. Saito, H. and F. Posas, *Response to Hyperosmotic Stress.* Genetics, 2012. **192**(2): p. 289-318.
26. Halliwell, B. and J.M.C. Gutteridge, *Free radicals in biology and medicine*. Vol. 3. 1999: Oxford university press Oxford.
27. Moradas-Ferreira, P., et al., *The molecular defences against reactive oxygen species in yeast.* Molecular Microbiology, 1996. **19**(4): p. 651-658.
28. Farrugia, G. and R. Balzan, *Oxidative stress and programmed cell death in yeast.* Frontiers in Oncology, 2012. **2**.
29. Myers, R.H., D.C. Montgomery, and C.M. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*. Vol. 705. 2009: Wiley.
30. Xavier, A.M.R.B., et al., *Second-generation bioethanol from eucalypt sulphite spent liquor.* Bioresource Technology, 2010. **101**(8): p. 2755-2761.
31. Helle, S.S., et al., *Xylose fermentation by genetically modified Saccharomyces cerevisiae 259ST in spent sulfite liquor.* Bioresource Technology, 2004. **92**(2): p. 163-171.
32. Pereira, S.R., et al., *Advances in ethanol production from hardwood spent sulphite liquors.* Process Biochemistry, (0).
33. Palmqvist, E. and B.r. Hahn-HÃ¤gerdal, *Fermentation of lignocellulosic hydrolysates. I: inhibition and detoxification.* Bioresource Technology, 2000. **74**(1): p. 17-24.
34. Mager, W.H. and M. Siderius, *Novel insights into the osmotic stress response of yeast.* FEMS Yeast Research, 2002. **2**(3): p. 251-257.
35. Dihazi, H., R. Kessler, and K. Eschrich, *High Osmolarity Glycerol (HOG) Pathway-induced Phosphorylation and Activation of 6-Phosphofructo-2-kinase Are Essential for Glycerol Accumulation and Yeast Cell Proliferation under Hyperosmotic Stress.* Journal of Biological Chemistry, 2004. **279**(23): p. 23961-23968.
36. Hirasawa, T., et al., *Comparison of transcriptional responses to osmotic stresses induced by NaCl and sorbitol additions in< i> Saccharomyces cerevisiae</i> using DNA microarray.* Journal of Bioscience and Bioengineering, 2006. **102**(6): p. 568-571.

37. Piper, P., et al., *The Pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast.* EMBO J, 1998. **17**(15): p. 4257-4265.
38. Casey, E., et al., *Effect of acetic acid and pH on the cofermentation of glucose and xylose to ethanol by a genetically engineered strain of Saccharomyces cerevisiae.* FEMS Yeast Research, 2010. **10**(4): p. 385-393.
39. Fernandes, A.R., et al., *Saccharomyces cerevisiae adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes.* Biochemical and Biophysical Research Communications, 2005. **337**(1): p. 95-103.
40. Taherzadeh, M.J., et al., *Conversion of furfural in aerobic and anaerobic batch fermentation of glucose by Saccharomyces cerevisiae.* Journal of Bioscience and Bioengineering, 1999. **87**(2): p. 169-174.
41. Petersson, A., et al., *A 5-hydroxymethyl furfural reducing enzyme encoded by the Saccharomyces cerevisiae ADH6 gene conveys HMF tolerance.* Yeast, 2006. **23**(6): p. 455-464.
42. Lin, F.M., Y. Tan, and Y.J. Yuan, *Temporal quantitative proteomics of Saccharomyces cerevisiae in response to a nonlethal concentration of furfural.* Proteomics, 2009. **9**(24): p. 5471-5483.
43. Li, B.Z. and Y.J. Yuan, *Transcriptome shifts in response to furfural and acetic acid in Saccharomyces cerevisiae.* Applied microbiology and biotechnology, 2010. **86**(6): p. 1915-1924.
44. Sundström, L., S. Larsson, and L.J. Jönsson, *Identification of Saccharomyces cerevisiae genes involved in the resistance to phenolic fermentation inhibitors.* Applied biochemistry and biotechnology, 2010. **161**(1): p. 106-115.
45. Endo, A., et al., *Genome-wide screening of the genes required for tolerance to vanillin, which is a potential inhibitor of bioethanol fermentation, in Saccharomyces cerevisiae.* Biotechnology for Biofuels, 2008. **1**(1): p. 3.
46. Endo, A., T. Nakamura, and J. Shima, *Involvement of ergosterol in tolerance to vanillin, a potential inhibitor of bioethanol fermentation, in Saccharomyces cerevisiae.* FEMS microbiology letters, 2009. **299**(1): p. 95-99.
47. Jamieson, D.J., *Oxidative stress responses of the yeast Saccharomyces cerevisiae.* Yeast, 1998. **14**(16): p. 1511-1527.
48. Jungwirth, H. and K. Kuchler, *Yeast ABC transporters – A tale of sex, stress, drugs and aging.* FEBS Letters, 2006. **580**(4): p. 1131-1138.
49. Chloupková, M., L.S. LeBard, and D.M. Koeller, *MDL1 is a High Copy Suppressor of ATM1: Evidence for a Role in Resistance to Oxidative Stress.* Journal of Molecular Biology, 2003. **331**(1): p. 155-165.
50. Gasch, A.P., et al., *Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes.* Molecular Biology of the Cell, 2000. **11**(12): p. 4241-4257.
51. Goldberg, A.L., *Protein degradation and protection against misfolded or damaged proteins.* Nature, 2003. **426**(6968): p. 895-899.
52. Nillegoda, N.B., et al., *Ubr1 and Ubr2 Function in a Quality Control Pathway for Degradation of Unfolded Cytosolic Proteins.* Molecular Biology of the Cell, 2010. **21**(13): p. 2102-2116.
53. Wickner, S., M.R. Maurizi, and S. Gottesman, *Posttranslational Quality Control: Folding, Refolding, and Degrading Proteins.* Science, 1999. **286**(5446): p. 1888-1893.
54. Heer, D. and U. Sauer, *Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain.* Microbial biotechnology, 2008. **1**(6): p. 497-506.

55. Heer, D., D. Heine, and U. Sauer, *Resistance of Saccharomyces cerevisiae to High Concentrations of Furfural Is Based on NADPH-Dependent Reduction by at Least Two Oxireductases.* Applied and Environmental Microbiology, 2009. **75**(24): p. 7631-7638.
56. Liu, Z.L., P.J. Slininger, and S.W. Gorsich. *Enhanced biotransformation of furfural and hydroxymethylfurfural by newly developed ethanologenic yeast strains*. in *Twenty-Sixth Symposium on Biotechnology for Fuels and Chemicals*. 2005. Springer.
57. Martín, C. and L.J. Jönsson, *Comparison of the resistance of industrial and laboratory strains of< i> Saccharomyces</i> and< i> Zygosaccharomyces</i> to lignocellulose-derived fermentation inhibitors.* Enzyme and Microbial Technology, 2003. **32**(3): p. 386-395.
58. Lewis Liu, Z., M. Ma, and M. Song, *Evolutionarily engineered ethanologenic yeast detoxifies lignocellulosic biomass conversion inhibitors by reprogrammed pathways.* Molecular Genetics and Genomics, 2009. **282**(3): p. 233-244.
59. Wright, J., et al., *Batch and continuous culture-based selection strategies for acetic acid tolerance in xylose-fermenting Saccharomyces cerevisiae.* FEMS Yeast Research, 2011. **11**(3): p. 299-306.
60. Mira, N.P., et al., *Genome-wide identification of Saccharomyces cerevisiae genes required for tolerance to acetic acid.* Microbial cell factories, 2010. **9**: p. 79-79.
61. Wei, P., et al., *Genome shuffling in the ethanologenic yeast Candida krusei to improve acetic acid tolerance.* Biotechnology and applied biochemistry, 2008. **49**(2): p. 113-120.
62. Yoo, H.Y., S.S. Kim, and H.M. Rho, *Overexpression and simple purification of human superoxide dismutase (SOD1) in yeast and its resistance to oxidative stress.* Journal of Biotechnology, 1999. **68**(1): p. 29-35.
63. Sikorski, R.S. and P. Hieter, *A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae.* Genetics, 1989. **122**(1): p. 19-27.
64. Grant, C.M., *Role of the glutathione/glutaredoxin and thioredoxin systems in yeast growth and response to stress conditions.* Molecular Microbiology, 2001. **39**(3): p. 533-541.
65. Stephen, D.W.S. and D.J. Jamieson, *Amino acid-dependent regulation of the Saccharomyces cerevisiaeGSH1 gene by hydrogen peroxide.* Molecular Microbiology, 1997. **23**(2): p. 203-210.
66. Ren, J., et al., *Hse1, a Component of the Yeast Hrs-STAM Ubiquitin-sortinf Complex, Associates with Ubiquitin Peptidases and a Ligase to Control Sorting Efficiency into Multivesicular Bodies.* Molecular Biology of the Cell, 2007. **18**: p. 324-335.
67. Mukhopadhyay, D. and H. Riezman, *Proteasome-independent functions of ubiquitin in endocytosis and signaling.* Science Signalling, 2007. **315**(5809): p. 201.
68. Finley, D., et al., *The Ubiquitin–Proteasome System of Saccharomyces cerevisiae.* Genetics, 2012. **192**(2): p. 319-360.
69. Ren, J., et al., *Hse1, a Component of the Yeast Hrs-STAM Ubiquitin-sorting Complex, Associates with Ubiquitin Peptidases and a Ligase to Control Sorting Efficiency into Multivesicular Bodies.* Molecular Biology of the Cell, 2007. **18**(1): p. 324-335.
70. Cruciat, C.-M., et al., *Bcs1p, an AAA-family member, is a chaperone for the assembly of the cytochrome bc1 complex.* EMBO J, 1999. **18**(19): p. 5226-5233.
71. Nobrega, F.G., M.P. Nobrega, and A. Tzagoloff, *BCS1, a novel gene required for the expression of functional Rieske iron-sulfur protein in Saccharomyces cerevisiae.* the EMBO Journal, 1992. **11**(11): p. 3821-3829.

72. Wyers, F., et al., *Deletion of the PAT1 gene affects translation initiation and suppresses a PAB1 gene deletion in yeast.* Molecular and Cellular Biology, 2000. **20**(10): p. 3538-3549.
73. Amrani, N., et al., *Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro.* Molecular and Cellular Biology, 1997. **17**(7): p. 3694-701.
74. Mangus, D.A., N. Amrani, and A. Jacobson, *Pbp1p, a Factor Interacting withSaccharomyces cerevisiae Poly(A)-Binding Protein, Regulates Polyadenylation.* Molecular and Cellular Biology, 1998. **18**(12): p. 7383-7396.
75. Mangus, D.A., et al., *Identification of Factors Regulating Poly(A) Tail Synthesis and Maturation.* Molecular and Cellular Biology, 2004. **24**(10): p. 4196-4206.
76. Buchan, J.R., D. Muhlrad, and R. Parker, *P bodies promote stress granule assembly in Saccharomyces cerevisiae.* The Journal of Cell Biology, 2008. **183**(3): p. 441-455.
77. Kato, K., Y. Yamamoto, and S. Izawa, *Severe ethanol stress induces assembly of stress granules in Saccharomyces cerevisiae.* Yeast, 2011. **28**(5): p. 339-347.
78. Buchan, J.R. and R. Parker, *Eukaryotic stress granules: the ins and outs of translation.* Molecular Cell, 2009. **36**(6): p. 932-941.
79. Tadauchi, T., et al., *Posttranscriptional Regulation of HO Expression by the Mkt1-Pbp1 Complex.* Molecular and Cellular Biology, 2004. **24**(9): p. 3670-3681.
80. Huh, W.-K., et al., *Global analysis of protein localization in budding yeast.* Nature, 2003. **425**: p. 686-691.
81. Giaever, G., et al., *Functional profiling of the Saccharomyces cerevisiae genome.* Nature, 2002. **418**: p. 387-391.
82. Arias, P., et al., *Genome-wide survey of yeast mutations leading to activation of the yeast cell integrity MAPK pathway: Novel insights into diverse MAPK outcomes.* BMC genomics, 2011. **12**(1): p. 390.
83. Indjeian, V.B., B.M. Stern, and A.W. Murray, *The Centromeric Protein Sgo1 Is Required to Sense Lack of Tension on Mitotic Chromosomes.* Science, 2005. **307**(5706): p. 130-133.
84. Kitajima, T.S., S.A. Kawashima, and Y. Watanabe, *The conserved kinetochore protein shugoshin protects centromeric cohesion durin meiosis.* Nature, 2004. **427**: p. 510-517.
85. Kiburz, B.M., et al., *The core centromere and Sgo1 establish a 50-kb cohesin-protected domain around centromeres during meiosis I.* Genes & Development, 2005. **19**(24): p. 3017-3030.
86. Dudas, A., S. Ahmad, and J. Gregan, *Sgo1 is required for co-segregation of sister chromatids during achiasmate meiosis I.* Cell Cycle, 2011. **10**(6): p. 951-955.
87. Park, S.H., et al., *Nrg1 Is a Transcriptional Repressor for Glucose Repression of STA1 Gene Expression inSaccharomyces cerevisiae.* Molecular and Cellular Biology, 1999. **19**(3): p. 2044-2050.
88. Zhou, H. and F. Winston, *NRG1I is required for glucose repression of SUC2 and Gal genes of Saccharomyces cerevisiae.* BMC Genetics, 2001. **2**(5).
89. Vyas, V.K., et al., *Repressors Nrg1 and Nrg2 Regulate a Set of Stress-Responsive Genes in Saccharomyces cerevisiae.* Eukaryotic Cell, 2005. **4**(11): p. 1882-1891.
90. Ye, T., K. Elbing, and S. Hohmann, *The pathway by which the yeast protein kinase Snf1p controls acquisition of sodium tolerance is different from that mediating glucose regulation.* Microbiology, 2008. **154**(9): p. 2814-2826.
91. Lamb, T.M. and A.P. Mitchell, *The Transcription Factor Rim101p Governs Ion Tolerance and Cell Differentiation by Direct Repression of the Regulatory Genes NRG1 and SMP1 in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 2003. **23**(2): p. 677-686.

92. Kuchin, S., V.K. Vyas, and M. Carlson, *Snf1 Protein Kinase and the Repressors Nrg1 and Nrg2 Regulate FLO11, Haploid Invasive Growth, and Diploid Pseudohyphal Differentiation.* Molecular and Cellular Biology, 2002. **22**(12): p. 3994-4000.
93. Mira, N.P., J.D. Becker, and I. Sá-Correia, *Genomic expression program involving the Haa1p-regulon in Saccharomyces cerevisiae response to acetic acid.* Omics: a journal of integrative biology, 2010. **14**(5): p. 587-601.
94. Magasanik, B., *Ammonia Assimilation by Saccharomyces cerevisiae.* Eukaryotic Cell, 2003. **2**(5): p. 827-829.
95. Asadollahi, M.A., et al., *Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering.* Metabolic engineering, 2009. **11**(6): p. 328-334.
96. Ding, M.-Z., et al., *Proteomic Research Reveals the Stress Response and Detoxification of Yeast to Combined Inhibitors.* PLoS ONE, 2012. **7**(8): p. e43474.
97. Riego, L., et al., *GDH1 expression is regulated by GLN3, GCN4, and HAP4 under respiratory growth.* Biochemical and Biophysical Research Communications, 2002. **293**(1): p. 79-85.
98. DeLuna, A., et al., *NADP-Glutamate Dehydrogenase Isoenzymes of Saccharomyces cerevisiae.* Journal of Biological Chemistry, 2001. **276**(47): p. 43775-43783.
99. Tang, Y., A. Sieg, and P.J. Trotter, *13C-metabolic enrichment of glutamate in glutamate dehydrogenase mutants of Saccharomyces cerevisiae.* Microbiological Research, 2011. **166**(7): p. 521-530.
100. Bro, C., B. Regenberg, and J. Nielsen, *Genome-wide transcriptional response of a Saccharomyces cerevisiae strain with an altered redox metabolism.* Biotechnology and Bioengineering, 2004. **85**(3): p. 269-276.
101. Han, E.-K., et al., *Characterization of AGT1 encoding a general α-glucoside transporter from Saccharomyces.* Molecular Microbiology, 1995. **17**(6): p. 1093-1107.
102. Hounsa, C.-G., et al., *Role of trehalose in survival of Saccharomyces cerevisiae under osmotic stress.* Microbiology, 1998. **144**(3): p. 671-680.
103. Cheng, Q. and C.A. Michels, *MAL11 and MAL61 encode the inducible high-affinity maltose transporter of Saccharomyces cerevisiae.* Journal of Bacteriology, 1991. **173**(5): p. 1817-1820.
104. Stambuk, B.U., et al., *Expression of high-affinity trehalose–H+ symport in Saccharomyces cerevisiae.* Biochimica et Biophysica Acta (BBA) - General Subjects, 1998. **1379**(1): p. 118-128.
105. Plourde-Owobi, L., et al., *AGT1, Encoding an α-Glucoside Transporter Involved in Uptake and Intracellular Accumulation of Trehalose inSaccharomyces cerevisiae.* Journal of Bacteriology, 1999. **181**(12): p. 3830-3832.
106. Salema-Oom, M., et al., *Maltotriose Utilization by Industrial Saccharomyces Strains: Characterization of a New Member of the α-Glucoside Transporter Family.* Applied and Environmental Microbiology, 2005. **71**(9): p. 5044-5049.
107. Alves, S.L., et al., *Molecular Analysis of Maltotriose Active Transport and Fermentation by Saccharomyces cerevisiae Reveals a Determinant Role for the AGT1 Permease.* Applied and Environmental Microbiology, 2008. **74**(5): p. 1494-1501.
108. Smit, A., et al., *The Thr505 and Ser557 residues of the AGT1-encoded α-glucoside transporter are critical for maltotriose transport in Saccharomyces cerevisiae.* Journal of Applied Microbiology, 2008. **104**(4): p. 1103-1111.
109. Meister, A. and M.E. Anderson, *Glutathione.* Annual review of biochemistry, 1983. **52**(1): p. 711-760.

110. Izawa, S., Y. Inoue, and A. Kimura, *Oxidative stress response in yeast: effect of glutathione on adaptation to hydrogen peroxide stress in Saccharomyces cerevisiae.* FEBS Letters, 1995. **368**(1): p. 73-76.
111. Grant, C.M., F.H. MacIver, and I.W. Dawes, *Glutathione is an essential metabolite required for resistance to oxidative stress in the yeast Saccharomyces cerevisiae.* Current genetics, 1996. **29**(6): p. 511-515.
112. Liu, H., R. Lightfoot, and J.L. Stevens, *Activation of Heat Shock Factor by Alkylating Agents Is Triggered by Glutathione Depletion and Oxidation of Protein Thiols.* Journal of Biological Chemistry, 1996. **271**(9): p. 4805-4812.
113. Nikko, E. and H.R.B. Pelham, *Arrestin-Mediated Endocytosis of Yeast Plasma Membrane Transporters.* Traffic, 2009. **10**(12): p. 1856-1867.
114. Harada, Y., Y. Tamura, and T. Endo, *Identification of yeast Art5 as a multicopy suppressor for the mitochondrial translocator maintenance protein Tam41.* Biochemical and Biophysical Research Communications, 2010. **392**(2): p. 228-233.
115. Lin, C.H., et al., *Arrestin-Related Ubiquitin-Ligase Adaptors Regulate Endocytosis and Protein Turnover at the Cell Surface.* Cell, 2008. **135**(4): p. 714-725.
116. Winters, M.J., et al., *A Membrane Binding Domain in the Ste5 Scaffold Synergizes with Gβγ Binding to Control Localization and Signaling in Pheromone Response.* Molecular Cell, 2005. **20**(1): p. 21-32.
117. Sette, C., et al., *Mutational Analysis Suggests That Activation of the Yeast Pheromone Response Mitogen-activated Protein Kinase Pathway Involves Conformational Changes in the Ste5 Scaffold Protein.* Molecular Biology of the Cell, 2000. **11**(11): p. 4033-4049.
118. Marcus, S., et al., *Complexes between STE5 and components of the pheromone-responsive mitogen-activeated protein kinase module.* Proc. Natl. Acad. Sci. USA, 1994. **91**: p. 7762-7766.
119. Printen, J.A. and G.F. Sprague, *Protein-protein interactions in the yeast pheromone response pathway: Ste5p interacts with all members of the MAP kinase cascade.* Genetics, 1994. **138**(3): p. 609-19.
120. Chol, K.-Y., et al., *Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in S. cerevisiae.* Cell, 1994. **78**(3): p. 499-512.
121. Kranz, J.E., B. Satterberg, and E.A. Elion, *The MAP kinase Fus3 associates with and phosphorylates the upstream signaling component Ste5.* Genes & Development, 1994. **8**(3): p. 313-327.
122. Zalatan, J.G., et al., *Conformational Control of the Ste5 Scaffold Protein Insulates Against MAP Kinase Misactivation.* Science, 2012. **337**(6099): p. 1218-1222.
123. Malleshaiah, M.K., et al., *The scaffold-protein Ste5 directly controls a switch-like mating decision in yeast.* Nature, 2010. **465**: p. 101-105.
124. Duncan, K., R.M. Edwards, and J.R. Coggins, *The Saccharomyces cerevisiae ARO1 gene An example of the co-ordinate regulation of five enzymes on a single biosynthetic pathway.* FEBS Letters, 1988. **241**(1–2): p. 83-88.
125. Duncan, K., R. Edwards, and J. Coggins, *The pentafunctional arom enzyme of Saccharomyces cerevisiae is a mosaic of monofunctional domains.* Biochemical Journal, 1987. **246**(2): p. 375.
126. Bauer, B.E., et al., *Weak organic acid stress inhibits aromatic amino acid uptake by yeast, causing a strong influence of amino acid auxotrophies on the phenotypes of membrane transporter mutants.* European Journal of Biochemistry, 2003. **270**(15): p. 3189-3195.

127. Yoshikawa, K., et al., *Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in Saccharomyces cerevisiae.* FEMS Yeast Research, 2008. **9**(1): p. 32-44.
128. Matsumoto, R., et al., *The stress response against denatured proteins in the deletion of cytosolic chaperones SSA1/2 is different from heat-shock response in Saccharomyces cerevisiae.* BMC genomics, 2005. **6**(1): p. 141.
129. Verghese, J., et al., *Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (Saccharomyces cerevisiae) as a Model System.* Microbiology and Molecular Biology Reviews, 2012. **76**(2): p. 115-158.
130. Hahn, J.S., et al., *Genome-wide analysis of the biology of stress responses through heat shock transcription factor.* Molecular and Cellular Biology, 2004. **24**(12): p. 5249-5256.
131. Eastmond, D.L. and H.C.M. Nelson, *Genome-wide analysis reveals new roles for the activation domains of the Saccharomyces cerevisiae heat shock transcription factor (Hsf1) during the transient heat shock response.* Journal of Biological Chemistry, 2006. **281**(43): p. 32909-32921.
132. Kim, S., et al., *Folding in vivo of a newly translated yeast cytosolic enzyme is mediated by the SSA class of cytosolic yeast Hsp70 proteins.* Proceedings of the National Academy of Sciences, 1998. **95**(22): p. 12860-12865.
133. Park, S.H., et al., *The cytoplasmic Hsp70 chaperone machinery subjects misfolded and endoplasmic reticulum import-incompetent proteins to degradation via the ubiquitin–proteasome system.* Molecular Biology of the Cell, 2007. **18**(1): p. 153-165.
134. Heck, J.W., S.K. Cheung, and R.Y. Hampton, *Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1.* Proceedings of the National Academy of Sciences, 2010. **107**(3): p. 1106-1111.
135. Becker, J., et al., *Functional interaction of cytosolic hsp70 and a DnaJ-related protein, Ydj1p, in protein translocation in vivo.* Molecular and Cellular Biology, 1996. **16**(8): p. 4378-4386.
136. Juretschke, J., et al., *The Hsp70 chaperone Ssa1 is essential for catabolite induced degradation of the gluconeogenic enzyme fructose-1, 6-bisphosphatase.* Biochemical and Biophysical Research Communications, 2010. **397**(3): p. 447-452.
137. Sharma, D. and D.C. Masison, *Single methyl group determines prion propagation and protein degradation activities of yeast heat shock protein (Hsp)-70 chaperones Ssa1p and Ssa2p.* Proceedings of the National Academy of Sciences, 2011. **108**(33): p. 13665-13670.
138. Horton, L.E., et al., *The yeast hsp70 homologue Ssa is required for translation and interacts with Sis1 and Pab1 on translating ribosomes.* Journal of Biological Chemistry, 2001. **276**(17): p. 14426-14433.
139. Corbett, K.D., et al., *The Monopolin Complex Crosslinks Kinetochore Components to Regulate Chromosome-Microtubule Attachments.* Cell, 2010. **142**(4): p. 556-567.
140. Johzuka, K. and T. Horiuchi, *The cis Element and Factors Required for Condensin Recruitment to Chromosomes.* Molecular Cell, 2009. **34**(1): p. 26-35.
141. Waples, W.G., et al., *Putting the Brake on FEAR: Tof2 Promotes the Biphasic Release of Cdc14 Phosphatase during Mitotic Exit.* Molecular Biology of the Cell, 2009. **20**(1): p. 245-255.
142. Harsay, E. and R. Schekman, *Avl9p, a Member of a Novel Protein Superfamily, Functions in the Late Secretory Pathway.* Molecular Biology of the Cell, 2007. **18**(4): p. 1203-1219.
143. Ekena, K., et al. *The VPS1 protein is a dynamin-like GTPase required for sorting proteins to the yeast vacuole*. in *Ciba Found. Symp*. 2008.

144. Yu, X. and M. Cai, *The yeast dynamin-related GTPase Vps1p functions in the organization of the actin cytoskeleton via interaction with Sla1p.* Journal of Cell Science, 2004. **117**(17): p. 3839-3853.
145. Smaczynska-de Rooij, I.I., et al., *A role for the dynamin-like protein Vps1 during endocytosis in yeast.* Journal of Cell Science, 2010. **123**(20): p. 3496-3506.
146. Nothwehr, S.F., N.J. Bryant, and T.H. Stevens, *The newly identified yeast GRD genes are required for retention of late-Golgi membrane proteins.* Molecular and Cellular Biology, 1996. **16**(6): p. 2700-2707.
147. Vizeacoumar, F.J., et al., *YHR150w and YDR479c encode peroxisomal integral membrane proteins involved in the regulation of peroxisome number, size, and distribution in Saccharomyces cerevisiae.* The Journal of Cell Biology, 2003. **161**(2): p. 321-332.
148. Rad, M.R., et al., *Saccharomyces cerevisiae Apl2p, a homologue of the mammalian clathrin AP beta subunit, plays a role in clathrin-dependent Golgi functions.* Journal of Cell Science, 1995. **108**(4): p. 1605-1615.
149. Hua, Z., P. Fatheddin, and T.R. Graham, *An Essential Subfamily of Drs2p-related P-Type ATPases Is Required for Protein Trafficking between Golgi Complex and Endosomal/Vacuolar System.* Molecular Biology of the Cell, 2002. **13**(9): p. 3162-3177.
150. Jochum, A., et al., *Yeast Ysl2p, homologous to Sec7 domain guanine nucleotide exchange factors, functions in endocytosis and maintenance of vacuole integrity and interacts with the Arf-Like small GTPase Arl1p.* Molecular and Cellular Biology, 2002. **22**(13): p. 4914-4928.
151. Barbosa, S., et al., *Oligomeric Dop1p is Part of the Endosomal Neo1p-Ysl2p-Arl1p Membrane Remodeling Complex.* Traffic, 2010. **11**(8): p. 1092-1106.
152. Gillingham, A.K., et al., *Mon2, a Relative of Large Arf Exchange Factors, Recruits Dop1 to the Golgi Apparatus.* Journal of Biological Chemistry, 2006. **281**(4): p. 2273-2280.
153. Pascon, R.C. and B.L. Miller, *Morphogenesis in Aspergillus nidulans requires Dopey (DopA), a member of a novel family of leucine zipper-like proteins conserved from yeast to humans.* Molecular Microbiology, 2000. **36**(6): p. 1250-1264.

Appendices

Appendix A: Making Yeast Electrocompetent cells and electroporating

Adapted from: Nucleic Acids Research, 2009, Vol. 37, No. 2 e16 by Dr. Andy Ekins

1. Inoculate 25ml of YPD broth in a 125ml flask with 0.5ml of an overnight culture and incubate at 30°C with shaking (200rpm).

2. Grow the cells to an OD_{600} ~0.75-1.0 (~5h)

3. Pellet the cells at 2000XG for 1 min at room temperature in a centrifuge.

4. Wash the cells with 25ml of cold autoclaved nanopure water and pellet as above.

5. Resuspend the cells with 1-2ml cold autoclaved 1M sorbitol and pellet at 2000XG for 1 min in a centrifuge.

6. Repeat the washes 3 more times.

7. Finally resuspend the pellet with 250µl of 1M Sorbitol*.

8. Use 50µl of cells for each transformation in a 2mm gap electroporation cuvette (blue caps) and add DNA (salt free, i.e. dialyzed). Keep the cuvettes on ice.

9. Electroporate the cells using the following conditions 25µF, 200OHMS, 1.5KV.

10. Immediately add 1ml of YPD and incubate at 30°C for 1h.

11. Pellet the cells via centrifugation at 2000XG for 1 min. If you are plating cells onto minimal media (i.e. YNB) wash the cells with 1M sorbitol a few times to wash away the nutrients.

*If you wish, aliquots of competent yeast can be frozen at -80℃, just include 15% (w/v) glycerol or 10% DMSO in the final wash solution.

Appendix B: Making electrocompetent *E.coli* cells and electroporation

Protocol supplied by Elena Fossati

1. Inoculate 5ml LB with DH5α *E.coli* and incubate overnight at 37°C and 200 rpm
2. Use the 5ml overnight culture to inoculate 500ml of LB.
3. Incubate at 37°C and 200 rpm until the cells have reached an OD600 of ~0.7-0.8.
4. Chill the cells on ice as 50ml aliquots for 10 min.
5. Centrifuge in a cold (4°C) rotor at 2500XG for 10 min.
6. Remover supernatant and resuspend the pellet in 10ml ice-cold autoclaved ddH_2O.
7. Let sit 5 minutes on ice.
8. Centrifuge as above.
9. Remove supernatant and resuspend as 2 aliquots of 10ml ice-cold autoclaved ddH_2O (ie you should only have 2 tubes of 10 ml).
10. Let sit for 5 min.
11. Centrifuge as above.
12. Repeat step 9, 10, 11
13. Remove supernatant and pool samples with 5ml pre-chilled 10% DMSO.
14. Let site for 5 min on ice.
15. Centrifuge as above.
16. Repeat steps 13-15 two more times
17. Remove supernatant and resuspend in 1ml 10% DMSO.
18. Aliquot as 50µl aliquots and freeze at -80°C

*To electroporate, use 50µl in a 2mm gap electroporation cuvette. Electroporate at 25µF, 200OHMS and 2.5 KV.

Appendix C: Methylene Blue staining

1. Making methylene Blue stain:
    a. 0.1M glycine buffer pH 10.6
    b. 1/10 Methylene Blue
2. Pipette 0.1ml of culture into 0.9ml of 1x PBS
3. Centrifuge at 3200XG for 2 minutes
4. Discard supernatant and resuspend pellet in 0.5ml Methylene blue stain.
5. Incubate 15min at room temperature.
6. Count cells on a hemocytometer.
    a. Blue cells are dead
    b. Clear cells are alive.

Appendix D: Transformation of circular plasmids in yeast (rapid protocol)

Modified from Jansen et al. Gene 2005 344:43-51 by Elena Fossati

1. Streak out yeast cells from -80°C stock on YPD agar plates and incubate at 30°C for 48 hours.
2. Scrape yeast cells from plate (1 to 3 day old plate) and suspend in 150 µl of reaction mix:
    a. 15 µl LiAc 1M (0.1M final concentration)
    b. 75 µl PEG4000 80% (40% final)
    c. DNA suspended in 60 µl of sterile ddH_2O
3. Incubate the solution for at least 2 hours to overnight at 30°C
4. Plate on selective media and incubate at 30°C for 48 hours.