MACQUARIE University
SYDNEY·AUSTRALIA

MASTER OF RESEARCH - RESEARCH PROJECT

"EXPLORING NATURAL YEAST DIVERSITY FOR HIGH PROTEIN SECRETION"

JOHN TIMMINS

Department of Molecular Sciences
Macquarie University

01 May 2018

This work has not been submitted for a higher degree to any other university or institution.
Signed: …… John Timmins.

ACKNOWLEDGMENTS

I am indebted to my supervisor, Dr Heinrich Kroukamp, for his wonderful advice, guidance and supervision of my Master of Research Project over the last 10 months and for his patient teaching of new test protocols and instruction in the use of new equipment. I am also grateful to Professor Ian Paulsen for his sage advice in directing me to undertake this subject as my research topic and for his constructive comments during the conduct of the Year 2 progression. In acknowledging the whole Paulsen Group, both in the laboratory and in the office, I thank them sincerely for their generous time in always being willing to answer questions and to show me new or better approaches to problems that arose.

TABLE OF CONTENTS

CHAPTER THREE – RESULTS and DISCUSSION

LIST OF TABLES AND FIGURES

TABLES

FIGURES

ABSTRACT

Ethanol is seen as a sustainable biofuel, reducing reliance on non-sustainable fossil fuels. Cellulose is an abundant component in agricultural waste and is composed of glucose, linked by β-1,4-glycosidic bonds. If this waste biomass could be utilised as a potential source of glucose for fermentation to ethanol, it would lower production costs. However, cellulose is recalcitrant to enzymatic hydrolysis, requiring high enzyme levels for efficient sugar liberation. In this project, the natural variation of divergent *Saccharomyces cerevisiae* isolates was utilised to identify high heterologous cellobiohydrolase secreting strains. In preparation for quantitative trait loci identification, 18 haploid *S. cerevisiae* strains were generated from these isolates. Expression cassettes were constructed to allow the single copy integration and expression of two closely related cellobiohydrolase genes from *Rasamsonia emersonii* (Te) and *Trichoderma reesei* (Tr), and the ability of each of the 18 haploid strains to secrete the two cellobiohydrolases was evaluated. There were distinct secretion phenotypes between different strains secreting the same CBH1, with up to a 5-fold difference between the lowest and highest secretors. These high secretion phenotypes resulted in the identification of different strains efficient at secreting either the Te or Tr CBH1. Strains with superior secretion characteristics were combined with reference strains, and representative high secreting progeny were selected for downstream beneficial allele identification.

CHAPTER ONE

INTRODUCTION

1. ETHANOL AS A RENEWABLE ENERGY SOURCE

Ethanol produced from plant biomass is seen as a sustainable and green biofuel, reducing reliance on polluting and non-sustainable fossil fuels (den Haan et al, 2013; Lynd et al, 2002). Ethanol is the fermented product from glucose; glucose results from β-glucosidase hydrolysis of oligosaccharides; oligosaccharides, such as cellobiohydrolase, are produced from the hydrolysis of complex carbohydrates by endoglucanases and exoglucanases. In this way, a complex carbohydrate biomass, of which cellulose is an example, can be broken down to fermentable sugars to produce a biofuel (den Haan et al, 2013). Cellulose, in being the most plentiful polymer on the planet (Lynd et al, 2002), is composed of the monosaccharide glucose, but linked in chains to produce rigid and semi-rigid fibres; furthermore, it is one of the most abundant components in agricultural waste and, importantly, it does not compete directly with food (Sticklen, 2008). If this glucose-polymer biomass could be efficiently utilised in the biofuel production process, it would lower production costs while also reducing waste (la Grange et al, 2010; Hasunuma and Kondo, 2012; Lynd et al, 2002; van Zyl et al, 2007). Using cellulosic material as feedstock poses several limitations; its recalcitrance to enzymatic degradation is a major challenge to its commercial application (Coughlan, 1990; Lynd et al, 2002; van Zyl et al, 2007). One approach to overcome this recalcitrance is to increase the activity of hydrolytic enzymes on cellulose complexes through matching chemical and physical properties to the type of biomass and, thereby, to maximize potential biological activity (Vinzant et al, 2001). Other approaches might include examining the feasibility to incorporate in a fermenting micro-organism, such as the yeast *Saccharomyces cerevisiae*, multiple heterologous genes possessing the ability to express a number of hydrolytic enzymes for simultaneous secretion in a system called consolidated bioprocessing (CBP). Indeed, until such biomass can be efficiently hydrolysed by overcoming the recalcitrance of this cellulosic material and to have the heterologous hydrolytic enzymes expressed at a viable level, the creation of a CBP industry for the production of ethanol as a biofuel from cellulose will be hard to compete at an economic level with the supplies of ethanol from sources from the traditional non-sustainable petrochemical industries with the current low oil prices (Yamada et al, 2013).

2. CELLULASES AND CONSOLIDATED BIOPROCESSING

Proposals to utilise agricultural biomass as a source for conversion to biofuels are not new and go back to the 1970s, eventually resulting in the concept of CBP (Parisutham et al, 2014). These developments include proposals for various substrates mixed with different micro-organisms, including the yeast *Saccharomyces cerevisiae* which could be transformed with cellulase-expressing genes from fungi. However, the ideal scenario of integrating a range of enzyme-expressing genes into one singular organism to hydrolyse the biomass step-by-step and then to ferment the sugars to ethanol is still to be realised. While recombinant yeast strains hydrolysed cellobiose and a β-glucosidase secreting strain produced ethanol from the product of this hydrolysis, the complete conversion of insoluble cellulosic substrates to ethanol by heterologous yeast strains has yet to be achieved and the ideal expression system yet to be designed (den Haan et al, 2015). As den Haan and colleagues state, the focus in the future should be on constructing a yeast strain transformed with cellulase-expressing fungal genes; such secretion of heterologous enzymes should be at levels which match or approach those levels secreted by the cellulolytic fungus in its native form (den Haan et al, 2015). Adding to the research in the field of cellulases, Liu and colleagues demonstrated that a cellulase, CBH1, selectively hydrolyses the crystalline faces of cellulose which are hydrophobic (Liu et al, 2011); this latter point of hydrophobic 'shielding' of the target faces of cellulose may indeed contribute to its observed slow rate-limiting hydrolysis.

A number of fungi are capable of secreting a variety of cellulases which hydrolyse cellulosic material in nature (den Haan et al, 2013). One such example is the fungus *Trichoderma reesei* which secretes a composite of enzymes consisting of two cellobiohydrolases, CBH1 and CBH2, five different endoglucanases and two β-glucosidases (Herpoël-Gimbert et al, 2008). Of these, CBH1 is the most abundant secreted enzyme at ~60 % of all cellulase proteins secreted by *T. reesei* (Zhang and Lynd, 2004) and this is the enzyme we have focussed on to demonstrate the cellulase-secreting ability and variability of the strains of *Saccharomyces cerevisiae*. While CBH1 is required in the highest abundance to hydrolyse cellulose, it has only been produced in low levels in yeast, thereby providing a test case for secretory comparisons and for industrial purposes. Like other cellobiohydrolases, CBH1 possesses a tunnel-like catalytic domain (CD) and, in some cases, a carbohydrate-binding module (CBM). This cellobiohydrolase CBH1, as one of the three main cellulases secreted from fungal sources, can be transformed into *S. cerevisiae* to create the first step in the construction of a CBP (van Zyl et al, 2007).

The hydrolysis of cellulosic material can be assisted by heat and acid pre-treatment which opens up the crystalline structure, allowing access to the rigid or semi-rigid linear cellulose fibres held together by van der Waals forces (van Rensburg et al, 1998), with cellulases then acting on the exposed crystalline cellulose fibres (Zhang and Lynd, 2004); but large quantities of enzymes added to industrial processes can be a significant cost component (Lynd et al, 2005). In order to reduce this cost component of added enzymes, the concept of incorporating saccharification and fermentation into the one organism through CBP could overcome this problem (Parisutham et al, 2014; den Haan et al, 2013). This concept, applied to the cellulosic material of a biomass, would involve a step-wise process of cellulose hydrolysis by a series of enzymes in a three-step process (den Haan et al, 2013): 1) cellobiohydrolases (CBHs), as exoglucanases, act to release cellobiose – a β-1,4 linked two glucose moiety – from the cellulose chain reducing ends for CBH1s and from the non-reducing ends for CBH2s (Brady et al, 2015; Gusakov et al, 2005); 2) complementing the CBHs by creating more chain-ends are endo-β-1,4-glucanases (EGs), acting at amorphous regions within the cellulose chain (Ilmén et al, 2011); 3) lastly, β-glucosidases (BGLs) break down the cellobiose to glucose (Lynd et al, 2002) (Figure 1).

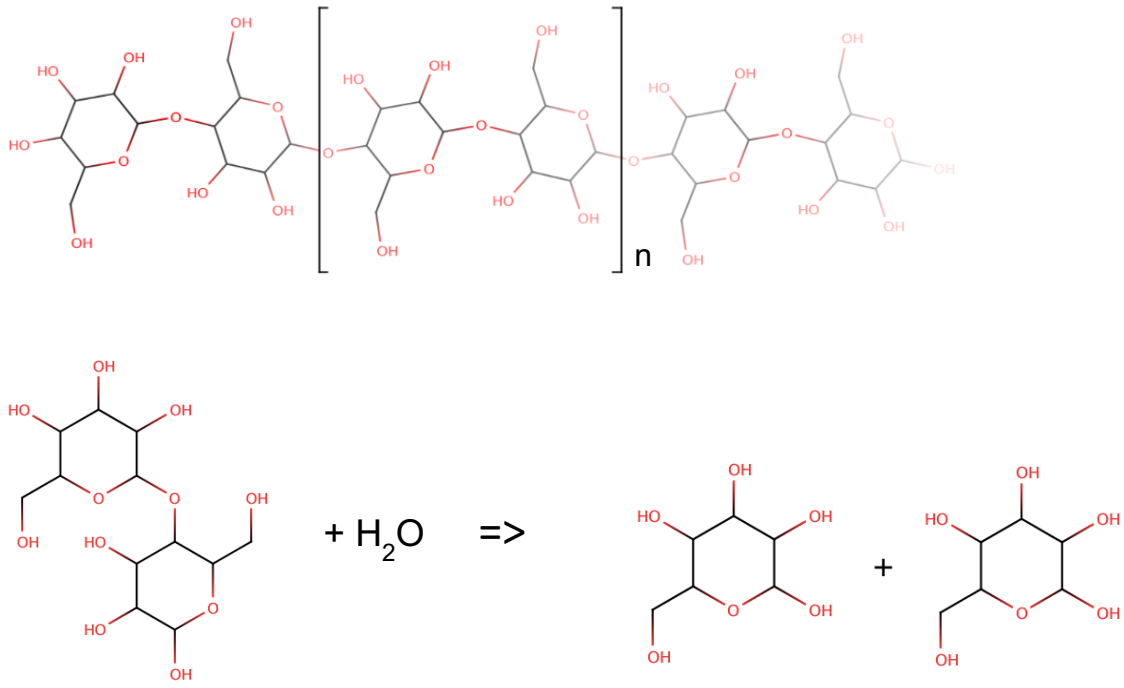


Figure 1. Cellulose chain of cellobiose, then hydrolysed to two glucose molecules by β-glucosidase

The CBH1 encoding genes, from two different cellulolytic fungal sources, can be transformed into the yeast *S. cerevisiae* to test the feasibility of a first step in constructing multiple different enzyme-expressing genes as one type of CBP (den Haan et al, 2013). The two fungal sources of cellobiohydrolases used in this experiment are *Talaromyces emersonii* (Te) (now *Rasamsonia emersonii*) and *Trichoderma reesei* (Tr). While the expressed enzymes are similar in composition at 66% (Ilmén et al, 2011), they exhibit different properties during secretion; contributing factors may include differences in the number of *N*-glycosylation sites and folding secretion signals, resulting in differing tertiary structures (den Haan et al, 2013; Ilmén et al, 2011).

Another factor for consideration in formulating this consolidated bioprocess is that any biomass pre-treatment by heat and acid, as referred to above, can produce toxic compounds which are able to inhibit microbial growth and enzyme activity (Hasunuma and Kondo, 2012). Furthermore, the sources of the cellulosic material can be from diverse plant biomasses and, therefore, require consideration in developing a tailor-made approach to match such pre-treatment with possibly different but still efficient species suitable for the presented biomass (Parisutham et al, 2014).

3. *SACCHAROMYCES CEREVISIAE* AS A 'CELL FACTORY'

The genome of the *Saccharomyces cerevisiae* was sequenced in 1996, defining 5885 protein-encoding genes in its 16 chromosomes (Goffeau et al, 1996); it was the first eukaryote and the largest to be completely sequenced at that time. Since that time, it has been one of the mainstay organisms for recombination technologies and biotechnology systems, with its unicellular structure and its ability to be grown on defined media. With the detailed genetic map of this yeast, manipulation of its genome with mutant alleles, insertion of heterologous genes or gene deletions, molecular biologist are provided with an accessible tool for wide-ranging genetic studies, as well as undertaking a function analysis of its own ~6000 genes.

Saccharomyces cerevisiae has long been used to leaven bread and to produce ethanol in various guises, such as from fermenting malted barley to produce beer and as the unknown ingredient (up until Pasteur) responsible for the conversion of grape juice to wine (Fay and Benavides, 2005). The advantage of *S. cerevisiae* is the high rate of glucose conversion to ethanol (3.3 g/L/h) (den Haan et al, 2013) and its inherent tolerance to high levels of ethanol and other stress conditions, including

low pH and high sugar (Swinnen et al, 2012), which enable it to complete the fermentation of grape juice to dryness, ie all hexose sugars to ethanol or to associated pathway products such as glycerol. But these days, the interest in biological conversion of sugars to ethanol has moved beyond food stuffs and beverages to biofuels in an attempt to replace unsustainable fossil fuels (Sticklen, 2008). Now there is a focus on how *S. cerevisiae* can play its part in also fermenting the cellulosic material of a biomass when it does not secrete the enzymes necessary to break down the polysaccharides to glucose moieties for potential fermentation.

S. cerevisiae is already the preferred organisms as a 'cell factory' for industrial production of ethanol as a biofuel and of many other products, such as bulk and fine chemicals, namely vanillin, isoprenoids or opiates (Kavšček et al, 2015). To maximise the ability of *S. cerevisiae* to act as a 'cell factory', van Dijken and colleagues (van Dijken et al, 2000) set out the following desired properties to be sought in target strains: 1) fast growth in defined mineral media without supplements other than vitamins, with a wide range of carbon and nitrogen sources for growth, yielding a high biomass yield on a carbon source; 2) fast aerobic, respiratory growth in glucose-limited chemostat cultures and growth in defined media under strictly anaerobic conditions; 3) high sporulation efficiency, spore viability and mating efficiency and with high transformation efficiency which is genetically stable and providing a good production of heterologous proteins, both intra- and extra-cellularly. However, use of consolidated bioprocessing for enzyme secretion, utilising *Saccharomyces cerevisiae,* is not a new concept; but there have been issues in its applied outcome with a relatively low titre of secreted heterologous cellulases, especially cellobiolydrolases, together with a highly variable range of such enzyme secretion, from 0.002 to >1% of total cell protein (den Haan et al, 2013) with other findings reporting up to 4% of total cell protein being of one particular type of enzyme, CBH2 (Ilmén et al, 2011).

Although CBH1 secretion, in conjunction with other types of hydrolytic enzymes such as CBH2, EG and BGL, has still not been able to fully hydrolyse crystalline cellulose, significant improvements in the rates of CBH1 secretion have been reported over time (Ilmén et al, 2011). Therefore, it can be seen that a sufficient level of heterologous protein expression in yeast is a limiting factor, when protein-encoding genes are transformed into yeast; this limitation could be significantly improved through strain engineering (Kroukamp et al, 2017a; Lynd et al, 2005).

4. BROAD RANGE OF YEAST STRAINS TESTED FOR PROTEIN SECRETION

A collection of 24 natural yeast isolates of *S. cerevisiae* was obtained from the Agricultural Research Service Culture Collection (NRRL), Peoria, Illinois, USA. This collection exhibits a broad range of samples from diverse industries and sources, gathered from across the globe (Figure 2).

Figure 2. Locations of the diverse world-wide sources for yeast samples.

The environmental origins of these samples include yeast isolated from ale and wine to sugar cane juice and molasses liquor, from spoiled banana to fermented food; and there is also a yeast sample from soil. The experiments also included using yeast samples of laboratory strains – Cen.PK 113-7D, S288c and BY4742 and AWRI-1631. The aim of sampling these strains from such different environments across various countries was to provide genetically and phenotypically diverse strains to compare protein secretion (Gallone et al, 2016; Mukherjee et al, 2014) and, with this collection, this project attempts to check such divergence within the different *S. cerevisiae* strains through sequencing of selected strains. Indeed, the recent *Nature Research Article* examining the genome evolution of 1,011 isolates of *S. cerevisiae* presented insights into such genotype-phenotype relationships, reviewing genetic variants including single nucleotide polymorphisms (SNPs), copy-number variants (CNVs) and non-reference variable open reading frames (ORFs) (Jackson et al, 2018).

5. ALLELE IDENTIFICATION

To map genetically the phenotypic traits of the strains of interest, applications such as quantitative trait loci (QTL) analysis can be employed, as most phenotypic diversity in natural populations are displayed by differences in degree, rather than in kind (Steinmetz et al, 2002). This has been demonstrated for a number industrially relevant phenotypes, like high ethanol tolerance (Pais et al, 2013) and acetic acid tolerance (Meijnen et al. 2016). The progress in QTL analysis since 2002 was outlined by Swinnen et al (2012), stressing the quantitative nature of such phenotypic traits in *S. cerevisiae* strains, versus Mendelian qualitative traits. In 2002, intrinsic deficiencies had been suggested when a single gene per locus approach was employed for closely-linked loci (Steinmetz et al, 2002) but now the application of high density oligonucleotide arrays with detection of large numbers of molecular markers and whole genome sequencing approaches have been achieved with resultant high-resolution QTL mapping of the entire genome, thus overcoming this deficiency. QTL has subsequently been refined for allele determination through high-throughput approaches using three independent techniques of individual analysis of meiotic segregants, bulk segregant analysis and reciprocal hemizygosity scanning (Wilkening et al, 2014).

6. SCOPE OF THE PROJECT

The scope of this research project is to explore the heterologous protein secretion capacities of genetically diverse, naturally isolated *Saccharomyces cerevisiae* strains. The protein to be used as the reporter protein in this case, from two separate species of cellulolytic fungal organisms, is the cellobiohydrolase, CBH1. Variations in the secretion levels of the various strains are to assessed, and the particular allele/s responsible for high secretion levels identified using next generation sequencing (NGS) and quantitative trait loci (QTL) analysis. By transforming known industrial high performing yeasts with these alleles, the recalcitrance of crystalline cellulose to hydrolysis by cellulolytic enzymes could be overcome and the usefulness of biomass in the production of biofuel at an industrial level increased.

7. Prior relevant work conducted at Macquarie University, contiguous with this project immediately before the Master of Research

The experimental work covered by this thesis is part of a bigger project, of which part has been completed prior to this work. The experimental steps and the work package covered in this thesis are illustrated in the project workflow diagram, Figure 3.

Genetically diverse yeast isolates form the the Agricultural Research Service Culture Collection (NRRL) in the USA were confirmed to be *S. cerevisiae* with RFPL analysis of the rDNA gene internal transcribed spacer (ITS) regions, as previously described (Clemente-Jimenez et al, 2004; Suranska H et al, 2016) (data not shown). The analysis of mating types revealed the presence of both *MATa* and *MATα* loci. In order to simplify the SNP segregation analysis and generate mating competent strains, haploid progeny were generated from a selection of diverse *S. cerevisiae* isolates. At least one copy of the *HO* allele was disrupted with a cassette containing a G418-resistance maker in each isolate before sporulation. The absence of HO expression demolishes the strain's ability to switch mating type and subsequent homozygous diploid formation. Colonies originating from a single spore, which had resistance to G418, were evaluated for presence of the disrupted *HO* allele in the presence of a single mating allele with PCR.

Cellobiohydrolase I expressing cassettes were constructed, to allow the integration of a single cassette into the yeast genome for stable constitutive expression. Synthetic *Rasamsonia emersonii* (previously known as *Talaromyces emersonii*), designated Te, and *Trichoderma reesei*, designated Tr, CBHI encoding ORFs were cloned downstream of a constitutive *S. cerevisiae PGK1* promoter. These constructs were subsequently cloned into pJet1.2 cloning vector containing a hygromycin B resistance cassette (under the control of the yeast *GAL1* promoter). The CBH expression cassette and resistance marker were flanked with sequences homologous to the *FLO8* gene. The final constructs are illustrated in Figure 4.

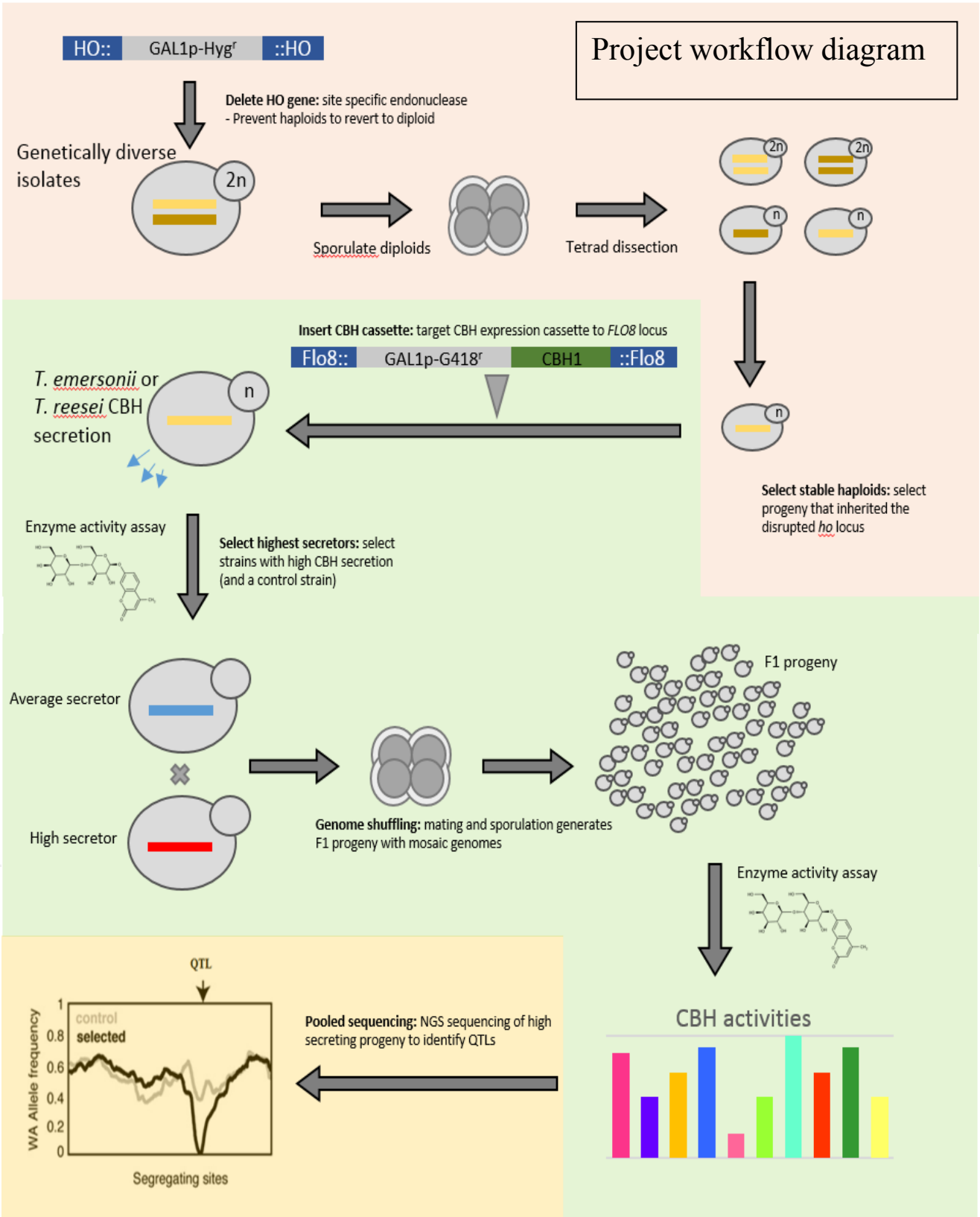
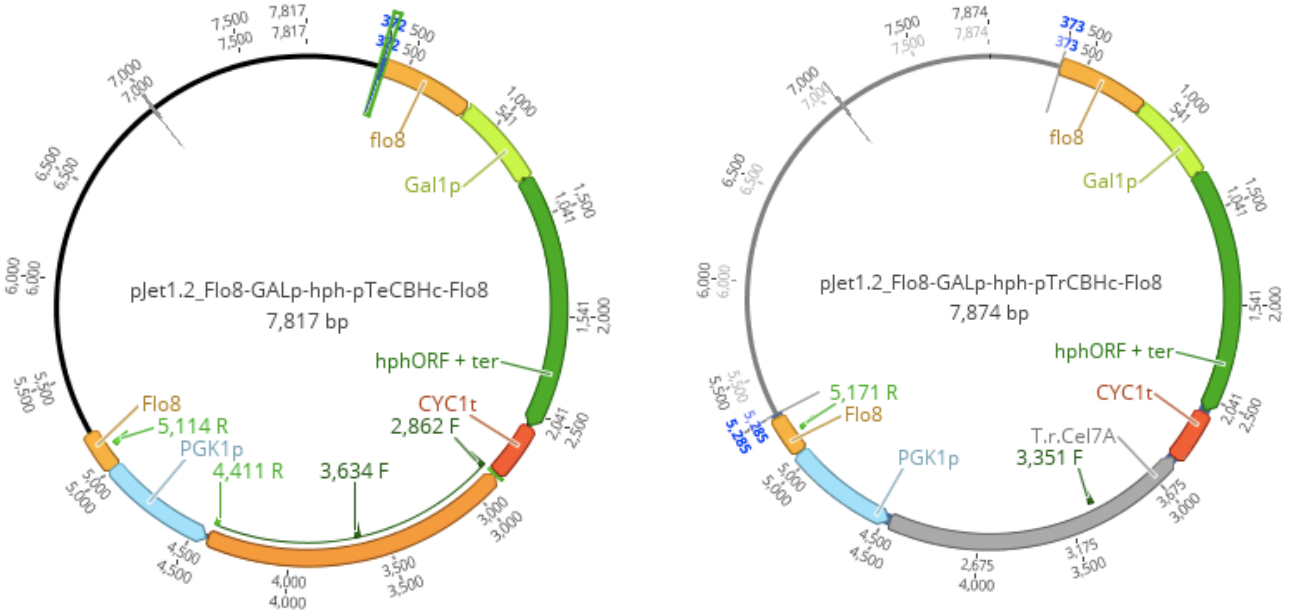


Figure 3. Project work flow. Steps in orange were completed prior to this study. Steps in green represents the work completed and presented in this thesis. Steps in yellow remains to be completed.



A. B.

Figure 4: Te (A) and Tr (B) cassettes containing two constructed sequences, a hygromycin marker under the transcriptional control of an inducible *GAL1* promoter and the *cbh1* gene under the constitutive *PGK1* promoter; the cassettes are flanked by sequences homologous to *flo8* (from Geneious Software).

CHAPTER TWO

MATERIALS and METHODS

1. MEDIA

YPD medium was prepared with 1 % yeast extract, 2 % bacteriological peptone (Oxoid Ltd, UK) and 2 % glucose; YPD agar was prepared by adding 2 % agar to the former. YPGal medium was prepared with 1 % yeast extract, 2 % bacteriological peptone (Oxoid Ltd, UK) and 2 % galactose (Amresco, USA). Minimal Medium consisted of 1.36 % yeast nitrogen base without amino acids, 2 % succinic acid, 1.2 % sodium hydroxide, 1.6 % yeast synthetic drop-out medium supplement, 0.038 % L-leucine and 2 % glucose, with the pH adjusted to 6.0 with sodium hydroxide mini-pearls (Bacto Laboratories Pty Ltd, Australia). Sporulation plates were prepared containing 1 % potassium acetate in 2 % agar. LB medium was prepared with 0.5 % yeast extract, 1 % tryptone (Amyl Media Pty Ltd, Australia) and 1 % sodium chloride. Source of all ingredients was Sigma-Aldrich, Australia, unless noted otherwise.

2. POLYMERASE CHAIN REACTION

Reagents for the construction of plasmid constructs were purchased from NEB (New England BioLabs Inc., USA) and used as specified by the supplier. The primers are set out in Table 1.

Table 1. Primers used in this study.

Type of primer	**Sequence of primer**
MATa Forward	5'-ACT CCA CTT CAA GTA AGA GTT TG-3'
MATα Forward	5'-GCA CGG AAT ATG GGA CTA CTT CG-3'
MAT locus Reverse	5'-AGT CAC ATC AAG ATC GTT TAT GG-3'
Flo8_transform_F	5'-ATG AGT TAT AAA GTG AAT AGT TCG T-3'
Flo8_transform_R	5'-TCA GCC TTC CCA ATT AAT AA-3'
New FloCon_F	5'-GAT ATT GCT ACT AAC TTC CGC AAT CAA GC-3'
New TeCBH_F	5'-CAC CAC AGT CAT CGC CGG AAC AC-3'
New TrCBH_F	5'-TCA GCG GTA CAG TAG TCG TCG TTC-3'
New FloCon_R	5'-ACT GGT TTG CTG AGG ACC CAA AGT TG-3'

3. PLOIDY VERIFICATION

Saccharomyces cerevisiae is a heterothallic yeast, with its mating type depending on the presence of the *MATa* or *MATα* allele at the *MAT* locus on chromosome III (al Safadi et al, 2010). The mating type – either MATa or MATα – was determined by PCR amplification (~540 bp product for MATa and ~450 bp product for MATα), using specific MATa Forward and MATα Forward primers plus a universal reverse locus primer, MAT locus Reverse (see Table 1), uniquely amplifying one or the other type; gel electrophoresis with 1 % agarose was used to resolve the PCR products.

4. YEAST TRANSFORMATIONS

The cassettes, containing *CBH1* genes from one of two fungal sources: *Rasamsonia emersonii* (formerly *Talaromyces emersonii*) – designated Te – and *Trichoderma reesei* – designated Tr – had earlier been constructed in *Escherichia coli* within a pJet1.2 plasmid backbone, as outlined in Section 1.7. These cellobiohydrolase-expressing gene cassettes were transformed into a series of genetically diverse strains of *Saccharomyces cerevisiae* (Table 2 in Section 3.2).

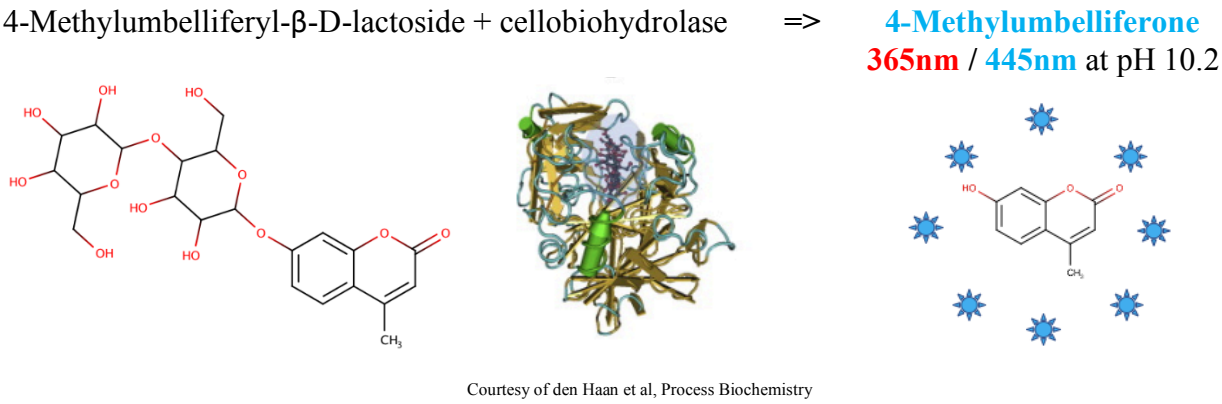
E. coli colonies were routinely inoculated in LB medium and incubated overnight at 37 ºC with rotation. Plasmids were extracted using Invitrogen Purelink Quick Plasmid Miniprep Kit and digested with *Xho*I and *Xba*I restriction enzymes, targeting the plasmid backbone to include the full cassette with flanking *flo8*START and *flo8*END; the samples were then PCR amplified using Eppendorf vapo.protect Mastercycler proS, followed by 0.8% agarose gel electrophoresis. PCR was used to amplify the cassette with Flo8_transform_F and Flo8_transform_R primers to target sites within the *flo8*START and *flo8*END regions (Table 1, Figure 2) providing the cassette band at 4850 bp on 0.8% agarose for excision. DNA recovery was done using Zymo Research's Zymoclean Gel DNA Recovery Kit.

The yeast cells were transformed using the PEG/LiOAc Yeast Transformation method (Green and Sambrook, 2012). To the yeast pellet was added 0.5 mL PEG/LiOAc Solution, 20 µL of carrier ssDNA (salmon sperm DNA, heated to 95 ºC for 10 min immediately before use) and 1 µg of the DNA recovered from the excised gel. This mixture was then vortexed for one minute, incubated overnight at room temperature, heat shocked at 42 ºC for 10 minutes, then centrifuged at 4000 rpm for 1 minute, discarding the supernatant. The pellet was washed with MilliQ Water and 100 µL was plated on YPGal selection plates with hygromycin 400 µg / mL. The plates were incubated for 2-3 days until colonies were visible. The putative transformants were evaluated using PCR,

interrogating both the presence of the cassettes and the successful targeting to the *flo8* locus.

5. Enzymatic assays

To determine secreted cellobiohydrolase 1 (CBH1) activity, confirmed yeast transformants were screened using the MULac Assay, with 4-methylumbelliferyl-β-D-lactoside (MULac) as substrate (Carbosynth Ltd, UK); this indicated differences between strains in cellobiohydrolase activity through their quantitative ability to hydrolyse MULac to release fluorescent 4-methylumbelliferone (4-MU), shown in Figure 5. The hydrolysed 4-MU, under excitation with a red laser at 365 nm, will fluoresce at 445 nm in the blue spectrum, at a pH of 10.2 (Boschker and Cappenberg, 1994):



Courtesy of den Haan et al, Process Biochemistry

Figure 5: Hydrolysis of MULac by CBH to produce the fluorescent 4-MU

MULac Assays were conducted using three solutions: a 50 mM sodium acetate solution buffered at pH 5.0 using glacial acetic acid; a 1.0 M sodium carbonate solution; and a 50 mM 4-methylumbelliferone (4-MU) (Sigma-Aldrich, NSW, Australia) solution in DMSO. The 4-MU solution was further diluted in 1.0 M sodium carbonate solution to give a 0.2 mM 4-MU stock solution for preparation of the standard curve between values of 0.1 µM and 10.0 µM 4-MU (Supplementary Material, Table S1 and Figure S1).

Colonies of the 18 strains for the two separate CBH1 groups were inoculated into 5 mL of YPD medium and incubated at 30 ºC for 24 hours. Five µL aliquots of these precultures were added to 1 mL of YPD in 24 well clear Greiner CELLSTAR plates (Merck, Darmstadt, Germany), together with the empty cassette control strain, Cen.PK 2-1C, and incubated at 30 ºC for 48 hours; this was done in triplicate. Secreted enzyme activities for the strains were then determined using soluble 4-methylumbelliferyl-β-D-lactoside (MULac, Carbosynth Ltd, Berkshire, UK) in 50 mM sodium

acetate (pH 5.0) buffer. Of the culture supernatant 25 µL was added to the 25 µL sodium acetate buffer containing 4.0 mM MULac in 96 well Eppendorf PCR plates (Eppendorf South Pacific Pty Ltd, NSW, Australia) and incubated at 50 °C (Kroukamp et al, 2017a). Culture supernatants were diluted as required to keep emission values within the standard curve range: Te strain supernatants were diluted 1:10, while the Tr strain supernatants were used undiluted.

Reactions were stopped after 15 minutes of incubation for the Te strains and after 5 hours for the Tr strains by adding 50 µL 1 M Na_2CO_3 (pH 10). 80 µL was transferred to black, flat clear bottom microtitreplates (Costar, Corning Inc., NY, USA). Liberation of 4-methylumbelliferone was detected by fluorescence measurement (excitation wavelength = 365 nm, emission wavelength = 445 nm) at pH 10.2 with a PHERAstar spectrofluorometer (BMG Labtech, Ortenberg Germany).

The procedure was repeated for assaying the F1 progeny from the mating of the high secretor strain Te YB-428 with the average secretor strain Te Y-234 to determine enzymatic activity of individual spores.

6. Cell growth evaluation

To compare cell growth with the levels of protein secretion for each strain and type of fungal gene expression, OD_{600} absorbance testing was conducted. Following the MULac Assays of triplicates of the 18 strains of Te and Tr types in 24 well plates, the volume of supernatant removed for the Assay was replaced with a solution of ethylenediaminetetraacetate (EDTA) to produce a final concentration of 12.4 mM, to disperse flocculating yeast cells for more consistent optical density readings. Samples of 10 µL were transferred to 96 well clear Greiner CELLSTAR plates (Merck, Darmstadt, Germany) and diluted with 190 µL of YPD medium. The plates read in the PHERAstar spectrofluorometer. Following sampling of supernatant from the 96 well plates for the MULac Assay, the dry cell weight (DCW) was determined by transferring the cells and remaining supernatant from the wells into previously weighed Eppendorf tubes, centrifuging and removing the supernatant from the cell pellets. The pellets were washed with water and dried in an oven overnight at 65 ºC. The resultant dried cells were weighed and the DCW calculated.

7. WHOLE GENOME SEQUENCING

The selected strains were cultivated in YPD medium for 48 hours before genomic DNA extraction. To isolate gDNA for use in sequencing, the ThermoScientific Yeast DNA Extraction Kit was used. A selection of strains, comprising the top 10 high and average secretors from each of the Te and Tr groups, were identified as being of interest for sequencing. The DNA pellet was resuspended in sterile MilliQ water and left for 5 minutes to fully dissolve. RNaseA was added and incubated for 2 hours. The samples were subjected to gel electrophoresis to display clear bands free of streaking and for minimal RNA presence. The samples were submitted to Macrogen (Macrogen Korea, Seoul, Republic of Korea) for next generation sequencing (NGS) with Illumina Hiseq to generate 100 bp reads, mapped to a reference genome (S288c) using Geneious software (Biomatters Ltd, New Zealand) and SPAdes 3.11.1 version (Centre for Algorithmic Biotechnology, St Petersburg State University, Russia) for de novo assembly.

8. MOSAIC STRAIN GENERATION

Isolates of a high CBH1 secretor (H) and an average CBH1 secretor (A) of the opposite mating type from the Te strains were cross-streaked on YPD plates and incubated at 30 ^{o}C for 72 hours. The process was repeated for high secretors and average secretors of the Tr strains. The resultant crosses were for Te strains: YB-428 (H) x Y-234 (A) and Y-5997 (H) x Y-12683 (A); and for the Tr strains: Cen.PK 113-7D (H) x Y-5997 (A) and YB-369 (H) x YB-1188 (A). The co-cultured cells were streaked for single colonies on fresh YPD plates. From these, diploid colonies were identified with PCR, as previously described.

9. RANDOM SPORE ISOLATION

Confirmed diploid colonies were inoculated into 5 mL YPD medium and grown at 30 ^{o}C for 24 hours in rotating 50 mL falcon tubes. The cells were then spun down, the supernatant poured off and the cells resuspended in the remaining liquid. This was pipetted onto sporulation plates and spread over the entire plate surface and incubated at ambient temperature for one week. Ascus formation was confirmed by microscopy for all four mated types: Te strains: YB-428 x Y-234 and Y-5997 x Y-12683; and for the Tr strains: Cen.PK 113-7D x Y-5997 and YB-369 x YB-1188. For individual spore isolation and identification, scrapings of the four sporulation plates were taken and added to 500 µL of sterile water in 1.5 mL Eppendorf tubes. 10 µL of zymolyase 10,000 U (1,000 U dissolved in 100 µL) (Zymo Research, California, USA) was added to each of these four tubes.

The tubes were then incubated at ambient temperature for two hours. 15 mL of sterile water and 10 mL of acid-washed glass beads (425-600 μm) (Sigma-Aldrich, NSW, Australia) were autoclaved in baffled plastic conical flasks. 20 μL of 2-mercapto-ethanol (Sigma-Aldrich) was added to each flask. The zymolyase-treated asci were added to the conical flasks and rotated at 200 rpm overnight. The liquid was poured off the beads, centrifuged, the spore-containing pellets washed with sterile water, centrifuged and resuspended in 100 μL of water (Treco and Winston, 2008). Serial dilutions of original suspension were carried out on the four samples at 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4} and were plated out on YPD plates, and these were incubated at 30 ^{o}C for 2-3 days.

CHAPTER THREE

RESULTS and DISCUSSION

1. CBH1 EXPRESSION CASSETTE CONSTRUCTION

Expression cassettes of cellobiohydrolase *cbh1* genes from two fungal sources, *Talaromyces emersonii* (Te) [now *Rasamsonsia emersonii*] and *Trichoderma reesei* (Tr), had earlier been constructed and inserted into the cloning vector pJET1.2 using *Escherichia coli*. The two expressed enzymes from these fungal genes in *Saccharomyces cerevisiae* have previously been shown to exhibit different properties, despite being 66% identical on amino acid sequence level (Ilmén et al, 2011). There are various hypotheses on these differences, such as differing tertiary architecture, varying glycosylation sites and levels (den Haan et al, 2013), different induced secretion stress (de Ruijter et al, 2016) and the overall compatibility between source and host organisms (Ilmén et al, 2011); beyond these hypotheses, other challenges to secretion of heterologous proteins, such as configurational changes due to environmental factors (pH, temperature and ionic strength) have been identified (Stephanopoulos, 2007). These findings suggest that certain gene candidates are more suited for expression in different strains of *S. cerevisiae* than others and it was for this reason that we used two enzymes-expressing genes from separate fungal species.

2. EXPLORING NATURAL YEAST DIVERSITY FOR COMPARATIVE SECRETORS

The 24 different *S. cerevisiae* isolates were obtained from the NRRL in the USA. As can be seen from Table 2, the sources from which they were obtained are wide and varied, from soil and fruit to fermented foods and drinks. Colonies of different strains exhibited varying visual phenotypes, such as white to light gold in colour, smooth-edged to indented, and flocculating to non-flocculating (results not shown). This approach of seeking different phenotypes for evaluation of desirable trains aligns with other studies in the field of bioethanol production (Davison et al, 2016; Mukherjee et al, 2014).

A hypothesis has been advanced that diverse environments for *S. cerevisiae* promote diverse genetic development in the species (Brown, 1988). In fact, this species diversity has been demonstrated in a recent article examining the evolution of genomes from 1,011 isolated of *S. cerevisiae* (Jackson et al, 2018). This analogy of diverse environments giving rise to diverse

genetics was also evident from our study, as can be seen from the diverse single nucleotide polymorphisms (SNPs) present in our selected strains (Table 3).

Table 2: Yeast strains used in this study.

Source	Wild Type	
Buttermilk	Y-2222	
Sorghum Brandy	Y-1526	
Sauterne wine, 1896	Y-645	Of the original 24 wild type strains,
Coconut sap	YB-4084	9, highlighted in grey, were eliminated
Grain	Y-383	due to poor plate growth,
Sake	Y-11572	ambiguity of *HO* gene deletion,
Drosophila	Y-2191	difficulties in sporulation, problems in
Bantu beer	Y-7184	haploid formation and spore viability
Rainbow trout	Y-27105	
Ale	Y-11879	Abandoned due to heavy flocculation
Dry white wine	Y-583	
Dry claret	Y-582	
Sugar refinery	Y-11857	
(Unknown)	YB-369	
Spoiled banana	YB-210	
Molasses liquor	Y-234	
Vinegar starter	Y-559	14 genetically diverse WT strains
Citrus juice	YB-1188	used in final assessment
Rum fermentation	YB-428	
Ragi	Y-5997	
Soil	Y-12638	
Loa chao starter	Y-12683	
Cane juice	Y-11878	
Fermented food	Y-12603	
Laboratory strains	Cen.PK 113-7D	Four laboratory strains selected for use
	S288c	
	BY4742	
	AWRI-1631	

3. TRANSFORMATION OF CBH1 EXPRESSION CASSETTE INTO YEAST STRAINS

Yeast strains were transformed with the *CBH1* expression cassettes, containing flanking regions homologous to the *S. cerevisiae FLO8* gene (Figure 6).

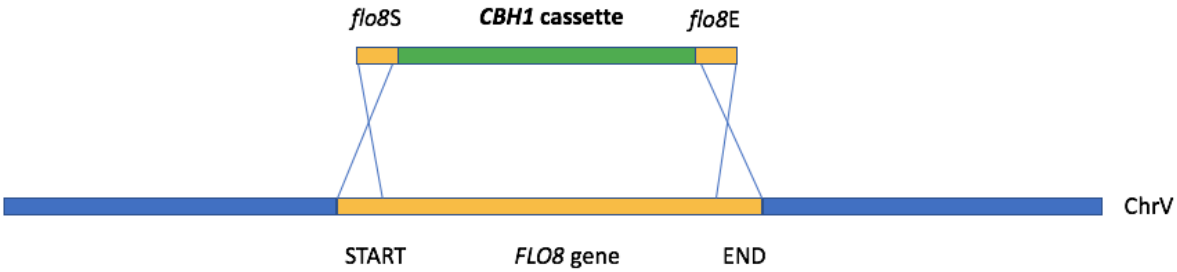


Figure 6: Transformation of *CBH1* expression cassettes

The target *FLO8* gene provides a convenient, non-essential gene to disrupt (Kobayashi et al, 1996), while also reducing the flocculation of the strains to provide better dispersion, to allow accurate sampling (Liu et al, 1996). Furthermore, this *FLO8* target gene has not previously been shown to alter secretion when disrupted (Kroukamp et al, 2017a). The *FLO8* gene is a regulatory gene controlling the expression of the *FLO1* gene in *S. cerevisiae*; together with *FLO5*, *FLO9* and *FLO10*, the *FLO1* gene is considered one of the four dominant genes, but 33 genes in total have been described as being involved in either flocculation or cell aggregation (Teunissen and Steensma, 1995). Hence, while the *FLO8* gene were disrupted, flocculation of cells may not be totally curtailed. This was evident from our results, as some of the *Δflo8* strains still displayed a cell aggregation phenotype. Putative transformed colonies from each of the strains were confirmed to have been successfully transformed with either the Te or Tr *cbh1* cassettes (Supplementary Material, Table S2) by PCR (Figure 7).

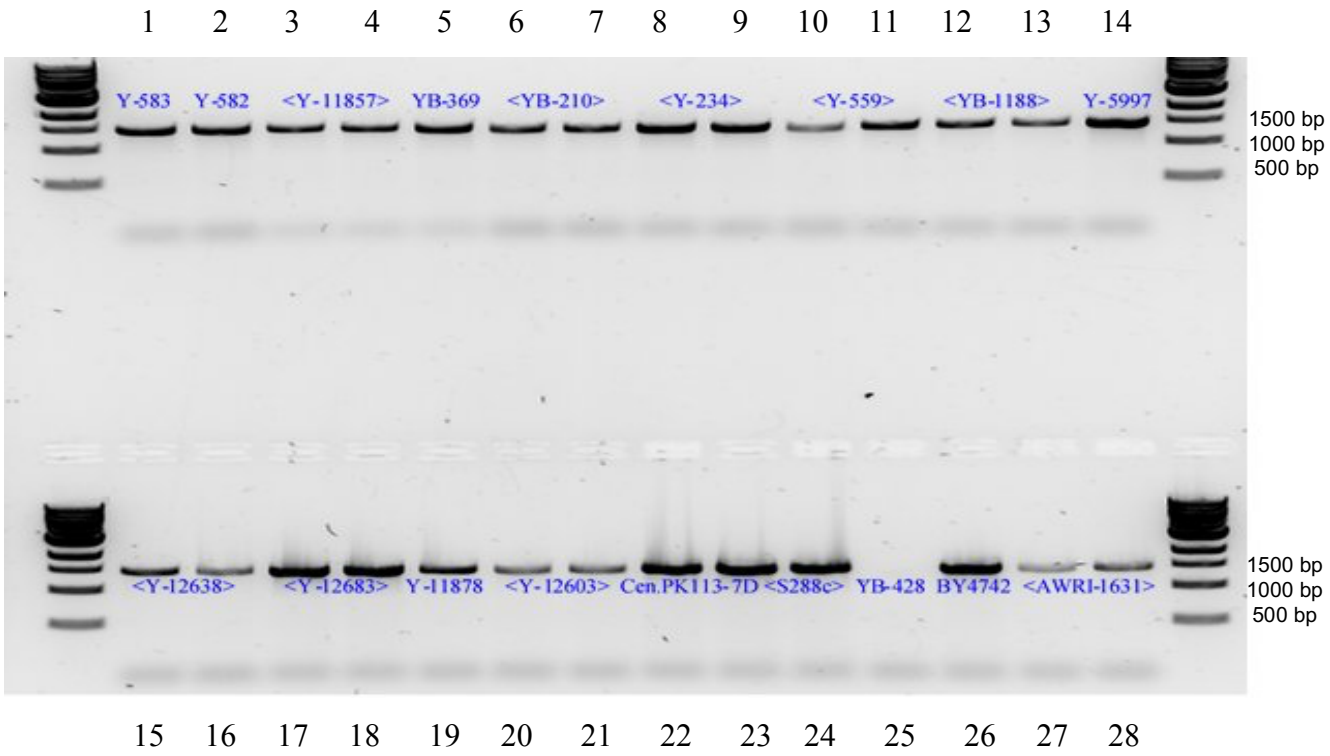


Figure 7: Example of verification of 27 samples from 17 strains of *S. cerevisiae* transformed with the Te *cbh1* cassette at the expected fragment size of 1485 bp. (Notes: the 'no show' at lane 25, Te strain YB-428, was subsequently shown to have the cassette also correctly inserted, as were all the 18 Tr strains.)

4. Secretion capacity of strains

In determining secreted cellobiohydrolase 1 (CBH1) activity, the strains were screened using the MULac Assay, with 4-methylumbelliferyl-β-D-lactoside (MULac) as substrate; this indicated differences between strains in cellobiohydrolase activity – as representative of protein secretion – through their quantitative ability to hydrolyse MULac to release fluorescent 4-methylumbelliferone (4-MU).

The enzymatic activities for cells grown in YPD for the Te group ranged between 130 – 414 katal, and for the Tr group range between 0.108-1.07 katal. Hence, there is a significant difference in secretory ability – across a range of 18 strains of yeast – for both evaluated CBH1s. This superiority of Te strains of *S. cerevisiae* for enzyme activity over other samples of *S. cerevisiae* transformed with different fungal genes expressing cellulases has been demonstrated previously by other researchers (Ilmén et al, 2011). The 18 strains from both sets Te and Tr showed diversity in their

secretion capability; the Te strains had higher enzyme assay results than the Tr strains, with the former strains requiring a ten-fold dilution for the results to fall within the detectable range, whereas the Tr strains were readable undiluted. Within the Te group (Figure 8), the two highest secretors (green columns) were YB-428 (414 katal) and Y-5997 (315 katal), while the two selected average secretors (yellow columns) were in the range of 149-172 katal. In the Tr group (Figures 9), the highest secretors (green columns) were Cen.PK 113-7D (1.07 katal) and YB-369 (0.51 katal), while for the average secretors it was 0.140-0.147 katal (yellow columns). These results suggested that the diverse genetics of the different strains allowed for differential expression of their heterologous proteins and indicated a multi-factorial response to the expression and secretion of the cellobiohydrolases from each of the two fungal sources and from within the 18 strains themselves.

These results answered the two questions posed for these comparative studies, showing differences *between* the Te and the Tr gene expressions with significant differences between the two groups in the reaction time – 15 minutes versus 5 hours, respectively – and in the concentration of the reaction enzyme for detectable fluorescence; as well, differences *within* the 18 strains of *S. cerevisiae* for each of the two fungal gene transformations was demonstrated. These variances in reaction time may indicate fundamental differences in the secretion efficacy of two similar but different proteins, as reviewed in the literature (Romanos et al, 1992). However, different enzymes have different enzyme activities, eg an active enzyme at low secreted levels can have the same assay values as one with low activity at high secreted levels; this will be investigated later in the project.

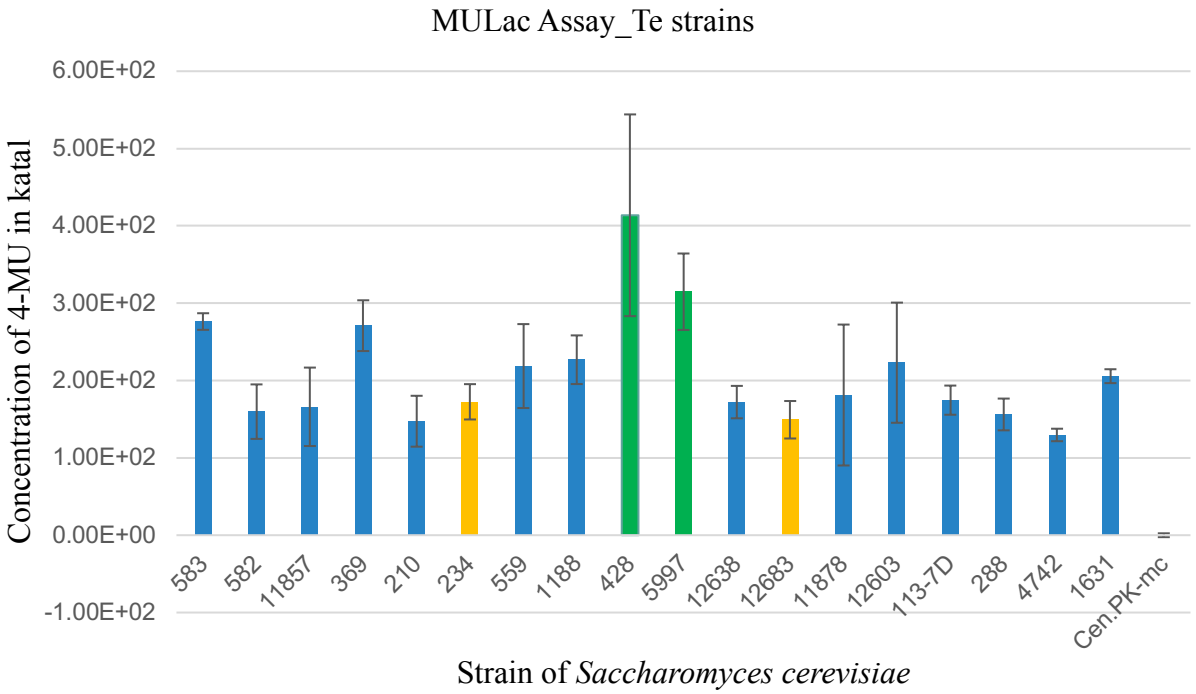


Figure 8: Supernatant CBH1 activity of Te strains incubated in YPD medium after 15 min at 50 ºC. Cen.PK-mc was an empty cassette control strain. The error bars represent the mean averages of triplicate samples. Green columns represent high secretors, yellow represent average secretors.

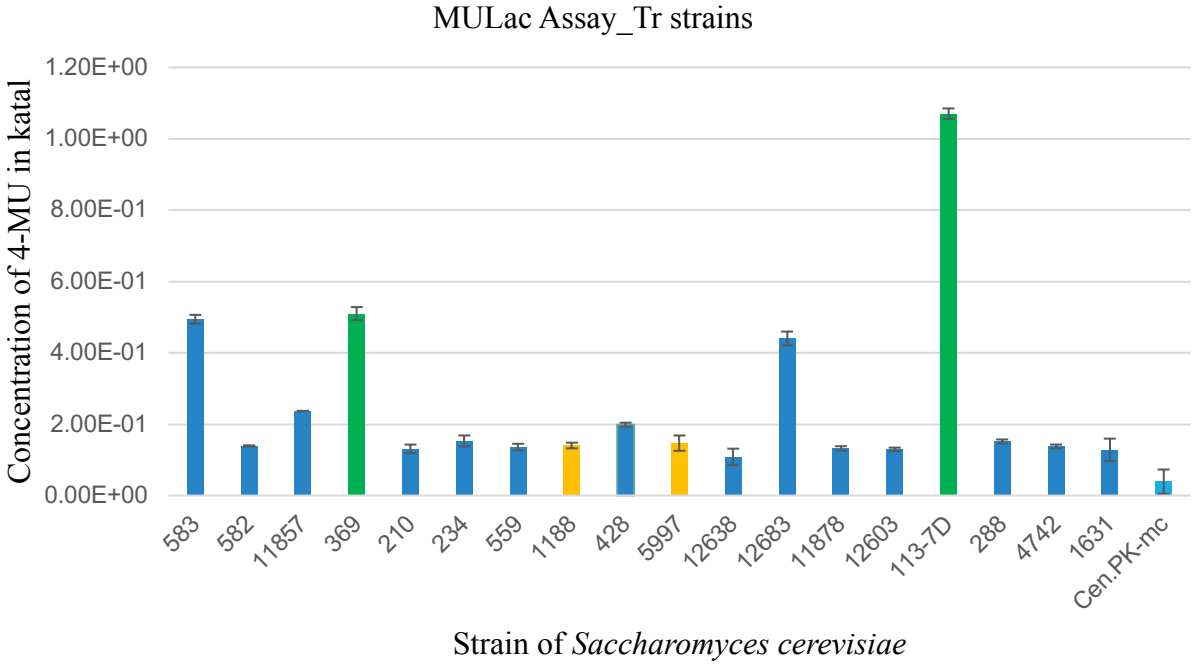


Figure 9: Supernatant CBH1 activity of Tr strains incubated in YPD medium after 5 h at 50 ºC. Cen.PK-mc was an empty cassette control strain. The error bars represent the mean averages of triplicate samples. Green columns represent high secretors, yellow represent average secretors.

Based on these results, as well as on available mating type complementarity, two of the highest secretors and two appropriate average secretors from each of the Te and Tr groups, plus 12 other strains, were later selected for sequencing and mating (Supplementary Material, Figure S2).

The causes of differences *between* the Te and Tr strains of *S. cerevisiae* may include two versus three *N*-glycosylation sites, respectively, (Jeoh et al, 2008) as well as the UPR response (Snapp, 2012), particularly in the Tr strains (Ilmén et al, 2011); another proposal put forward for this apparent inconsistency was the level of compatibility of the gene candidate with the yeast in which they are expressed (Ilmén et al, 2011; Kroukamp et al, 2017a). The same arguments could apply for *within*-group variations as has been mentioned for *between*-group variations, with *N*-glycosylation and *O*-glycosylation variations between samples. Ilmén and colleagues also demonstrated that, with the Tr secreted cellobiohydrolases, only a fraction of the potential secreted protein was enzymatically active due to post-transcriptional events, such as either over-glycosylation or misfolding, providing another reason for its exhibited diminished reactivity (Ilmén et al, 2011). The results from this present study showed a variety of secretory capacities from the 18 strains used in the final assessments with the cellobiohydrolase secretion by the strains evaluated for their innate capacity to express and secrete two different cellobiohydrolases which aligns with the work of Ilmén and colleagues.

The various yeast strains were also incubated in Minimal Medium and the MULac Assay results compared with those where the strains had been incubated in YPD. It is reported that growth rate and final cell densities is higher in complex media, such as YPD, than in minimal media (Narendranath et al, 2001). This was seen in the secretion of the CBH1 with the Te strains in YPD, with the levels being significantly higher – range 130-414 katal – than in Minimal Medium – range 64-140 katal. However, the situation was reversed with the Tr strains in YPD producing 0.108-1.07 katal, compared with 1.88-14.9 katal in Minimal Medium (Figures 10, 11). The YPD and minimal medium results are also combined for side-by-side comparison (Figures 12, 13) where the Te strains show the superiority of their secretion in YPD (blue columns) compared to minimal medium (orange columns) on the same scale; however, the Tr strains show the minimal medium results being superior to the YPD. In the Te strains with better secretion in YPD, the more complex medium containing components such as yeast extract and peptone can provide a degree of protection against stress conditions in the ER. With the opposite effect in the Tr strains, some other factor or factors must be at work to counter this effect. One could surmise that, with a poorer medium slowing growth, this would lower flux through the secretory pathway, the protein has more

time to fold correctly and the ER is less subject to any stress caused from accumulation of unfolded proteins which is now less likely to occur, resulting in more active enzyme being secreted; so the protein can be afforded a smooth transition through the ER, to the Golgi, through the cytosol to the cell membrane for secretion.

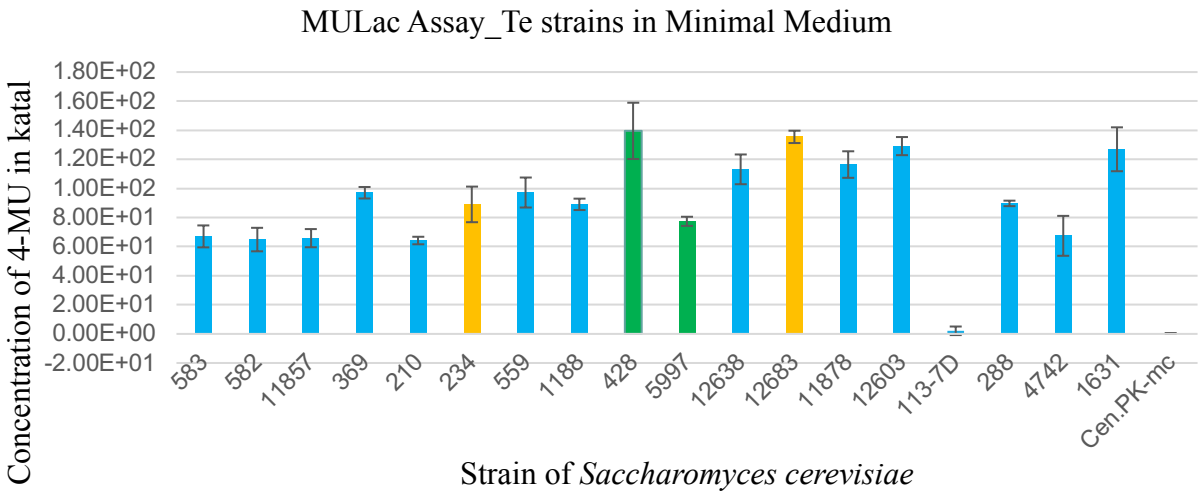


Figure 10: Supernatant CBH1 activity for Te strains incubated in Minimal Medium for 15 minutes at 50 ºC. Cen.PK-mc was an empty cassette control strain. The error bars represent the mean averages of triplicate samples. Green columns represent previously identified high secretors in YPD, yellow represent average secretors in YPD.

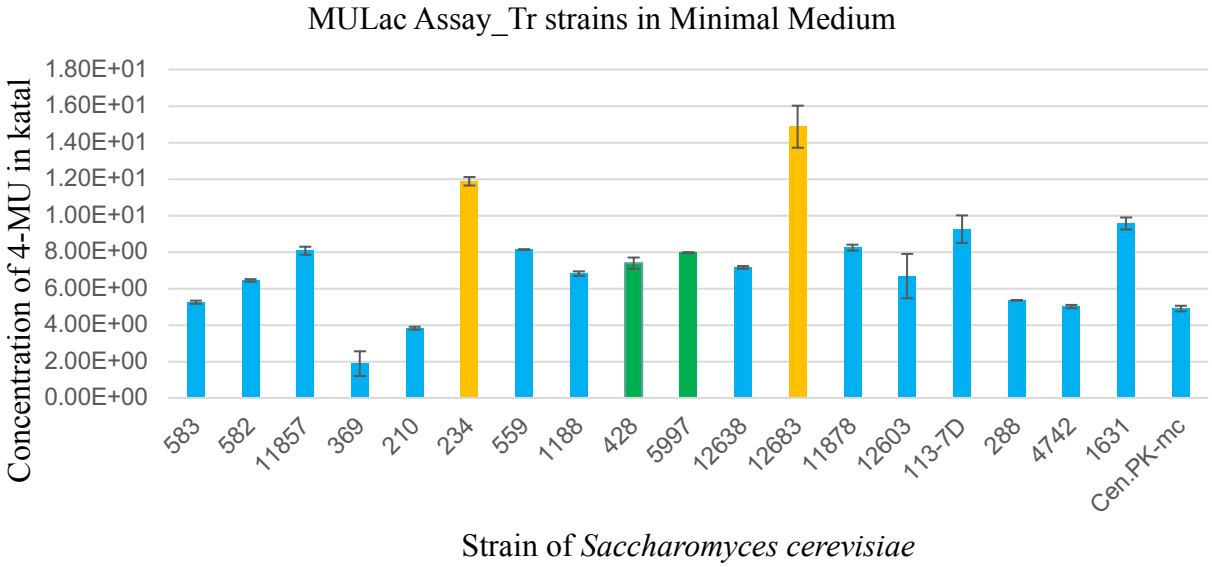


Figure 11: Supernatant CBH1 activity for Tr strains incubated in Minimal Medium for 5 h at 50 ºC. Cen.PK-mc was an empty cassette control strain. The error bars represent the mean averages of triplicate samples. Green columns represent previously identified high secretors in YPD, yellow represent average secretors.

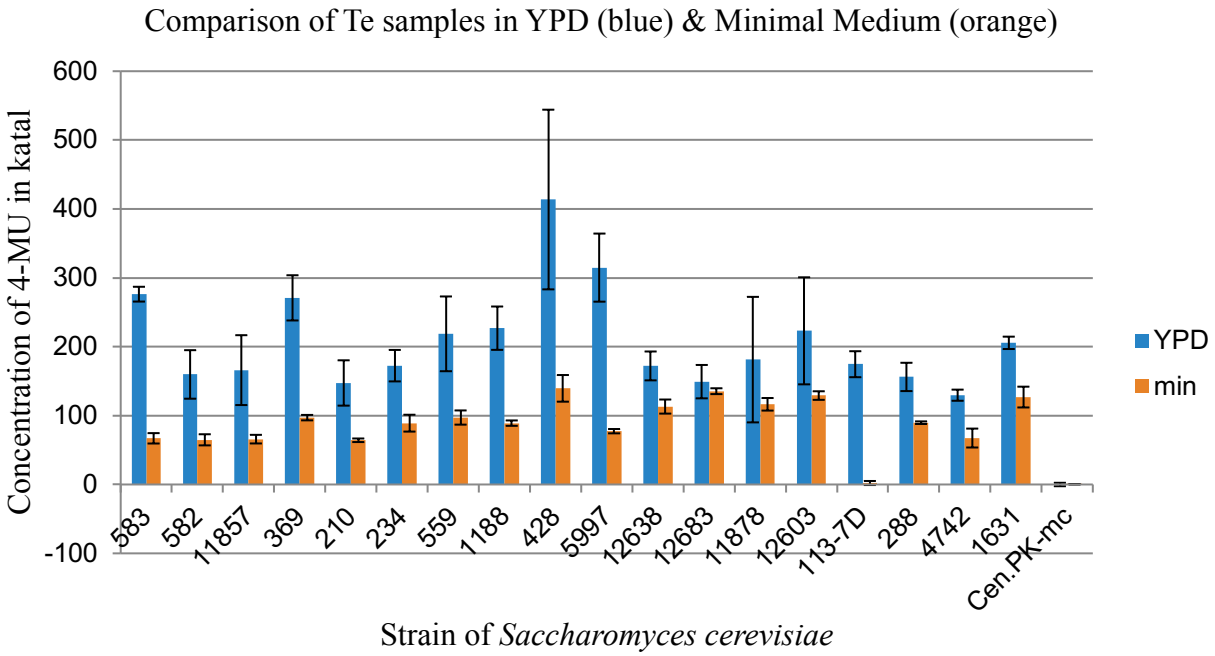


Figure 12: Comparison of Assay results for Te strains incubated in YPD and Minimal Medium for 15 minutes at 50 ºC. Cen.PK-mc was an empty cassette control strain. The error bars represent the mean averages of triplicate samples.

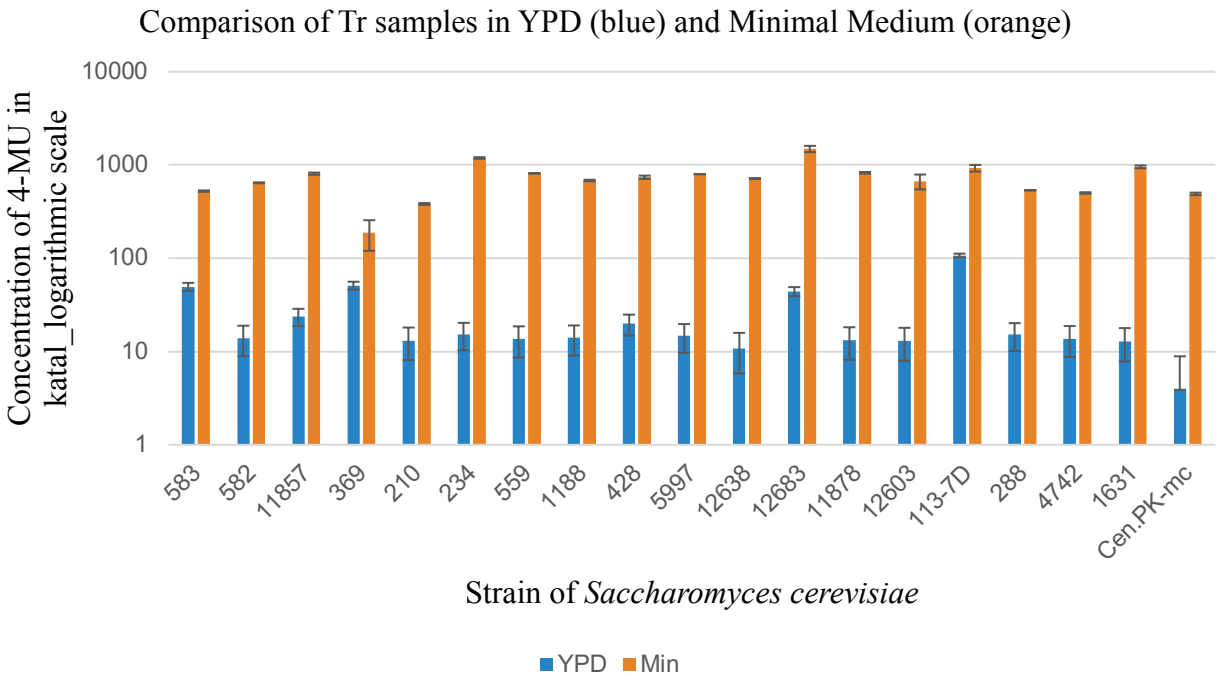


Figure 13: Comparison of Assay results for Tr strains incubated in YPD and Minimal Medium for 5 hours at 50 ºC. Cen.PK-mc was an empty cassette control strain. The error bars represent the mean averages of triplicate samples.

5. Differences between protein secretion and cell growth

In addition to the enzyme activity experiments, the remaining cells, adjusted to the original volume in the plate wells after supernatant sampling, were analysed for optical density (OD_{600}) (Figures 14, 15). The secretion of proteins in yeast has been shown to occur predominately at the bud tip (Puxbaum et al, 2016); from the experimental results here, it would appear that cells which are capable of secreting high titres of protein are sometimes compromised in their ability to grow. This is seen in the contrary results of high secretors from the Tr strains having low optical density, and vice versa; for example, Cen.PK 113-7D is the highest cellobiohydrolase secretor of the Tr stains, yet it has the lowest OD_{600} results of the subject strains; the exception is Y-12683 which exhibits both high secretion and good growth. However, this premise does not always hold up, as can be seen with the Te strains; for example, two of the top secretors are among the top five strains with the highest OD_{600} results. As consistently observed, the strain Y-583 displayed poor growth in any medium for both Te and Tr groups.

The OD_{600} results (Figures 9 and 10) and dry cell weights (DCW) results (Supplementary Material Figure S2) correlated. Due to the OD_{600} and DCW results being comparable and well correlated, the more direct and quicker process of OD_{600} to determine cell growth was used instead of DCW.

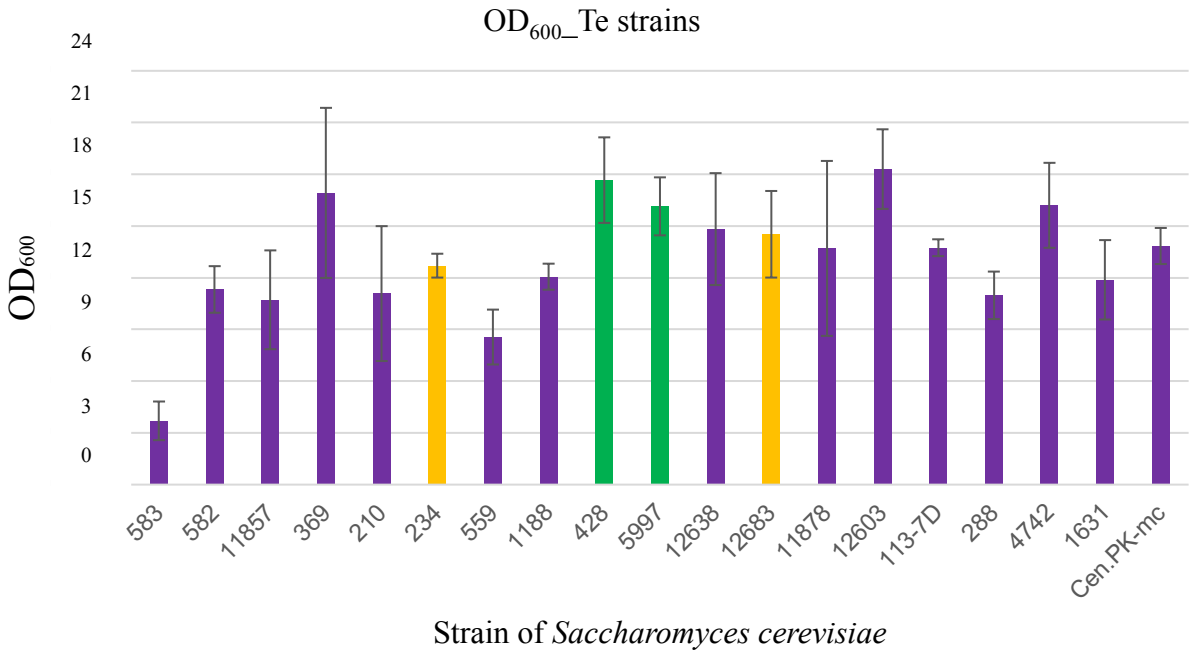


Figure 14: Optical Density of Te strains incubated in YPD medium. Cen.PK-mc was an empty cassette control strain. The error bars represent the mean averages of triplicate samples. Green columns represent high secretors, yellow represent average secretors.

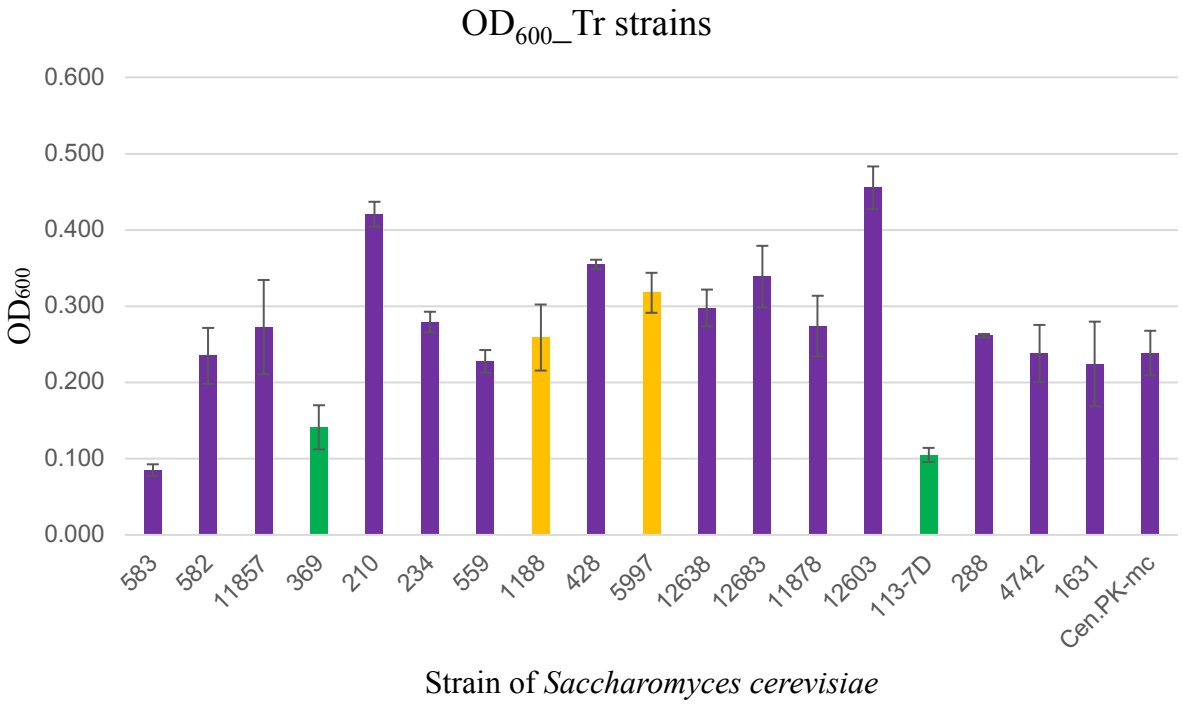


Figure 15: Optical Density of Tr strains incubated in YPD medium. Cen.PK-mc was an empty cassette control strain. The error bars represent the mean averages of triplicate samples. Green columns represent high secretors, yellow represent average secretors.

6. Results of sequencing of original haploids, mated diploids & F1

Many proteins are involved in the function and regulation of the yeast secretion pathway. In order to identify which alleles then confer the ability of high protein secretion, the best secretors were mated with an average secretor and sporulated (Deutschbauer & Davis, 2005) to allow genomic rearrangement. Having confirmed the mating type of each strain through PCR and gel electrophoresis (Figure 16), the MATa haploid of the high secretor was mated with the MATα haploid of the average secretor, or vice versa (Supplementary Material, Table S2). The mated haploid strains produced diploid *Saccharomyces cerevisiae* (Figure 17), with one confirmed diploid colony being selected for sporulation; spore formation was then induced on sporulation plates; first generations (F1) of haploid segregants were obtained through random spore isolation (Kroukamp et al, 2017a), after asci selection and dissection. The 30 haploids with the highest secreting activity of the first generation (F1) progeny – as a statistical minimum – were then isolated, again using the MULac Assay for selection. The pools of 30 best and 30 random secretors, plus parents as reference strains, were subsequently sequenced using next generation sequencing (NGS).

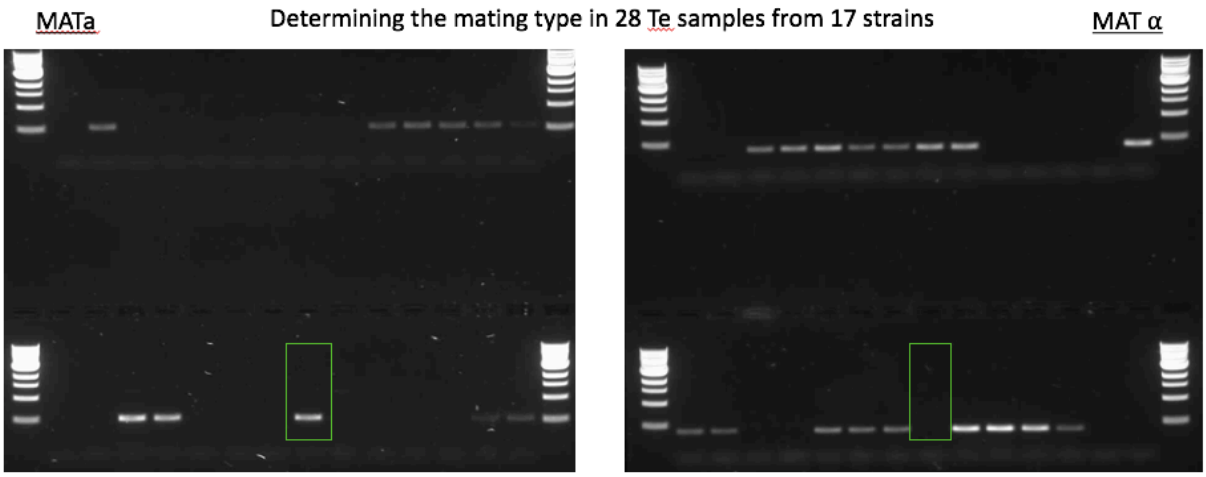


Figure 16: Mating type determination with PCR, showing mutual exclusion (green boxes) of mating types in a sample of yeast strains (exemplified by this Te set of strains: left MATa top L-R Y-582x1, Y-559x2, YB-1188x2; left bottom L-R Y-12683x2, Cen.PK113-7Dx1[boxed], AWRI-1631x2; right MATα top L-R Y-11857x2, YB-369x1, YB-210x2, y-234x2, Y-5997x1, bottom L-R Y-12638x2, Y-11878x1, Y-12603x2, empty box [Cen.PK113-7D], S288cx2, By4742x2).

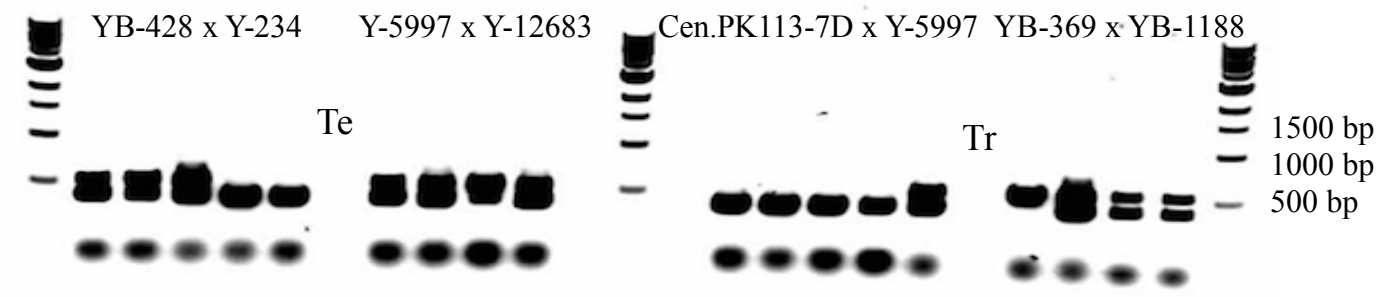


Figure 17: Mating in at least one colony for each group was confirmed with both MATa (~ 540 bp band size) and MATα (~ 450 bp) bands confirming diploid status.

Further examination of the sequencing results revealed large genomic variations between the different strains of *S. cerevisiae* in single nucleotide polymorphisms (SNPs), as indicated by the numbers listed (Table 3). In addition to identification of strain differentiation, these SNPs will also serve as genomic markers for the downstream allele identification. From these differences, the SNPs unique to the high secreting strains will lead to the alleles responsible for the genetic variations of these phenotypes through quantitative trait loci (QTL) analysis. The software used for

the SNPs detection through genome assembly was Geneious. In a recent article (Jackson et al, 2018) examining the genome evolution of 1,011 strains of *S. cerevisiae*, it was stated that the accumulation of SNPs drives the genome evolution in wild isolates but that these SNPs are very low in occurrence; interestingly, the greater changes in phenotypic effect are associated with copy-number changes, rather than SNPs (Jackson et al, 2018). Table 3 also lists the ploidy number in each chromosome of the sequenced strains. Whilst there are some 2n and '1.5'n evident, especially in Chromosomes 1 and 3, there are displays of 1n for the majority of strains. Chromosomes 1, 3 and 6 are considered more unstable than the other 14 chromosomes with more variability in terms of what the yeast can tolerate (Kumaran et al, 2013). The '1.5'n may represent mixtures of 1n and 1n+1 ploidy states in the same sampled population.

As a quality control step, copy number integration of the cassette into the yeast was determined through sequencing; this confirmation was a necessary step for determining actual high secretors as multiple CBH1 cassette integrations could alter the levels of secreted enzyme and skew the interpretations of the results (Ilmén et al, 2011). The results of the whole genome sequencing of 21 selected samples from the 18 Te strains and the 18 Tr strains were obtained from Macrogen and the data examined to determine copy number integration. The average sequence depth of the assembled genomes was ~85-105 times coverage.

The sequencing results of selected original haploids showed that the CBH1 cassette had successfully integrated as one copy into the genomic DNA of all the Te yeast strains without duplication, partial integration or misplaced location. For the Tr strains, one copy of the cassette was also demonstrated, except for Tr Y-12683 and Tr Y-234 strains, in which duplications were found. However, some of the Tr sequences were contaminated with Te DNA which, with the latter's higher enzyme activity, will inflate the assay results. This is shown with some of the high cellobiohydrolase activities observed, especially that of the Tr Cen.PK 113-7D strain (cf Figure 9).

Table 3. Single nucleotide polymorphisms of *S. cerevisiae* wild type strains, relative to laboratory strain S288c, plus ploidy of chromosomes.

Strains of	**TOTAL**	**Ploidy of Chromosomes**															
S. cerevisiae	**no. of SNPs**	**Chr 1**	**Chr 2**	**Chr 3**	**Chr 4**	**Chr 5**	**Chr 6**	**Chr 7**	**Chr 8**	**Chr 9**	**Chr 10**	**Chr 11**	**Chr 12**	**Chr 13**	**Chr 14**	**Chr 15**	**Chr 16**
Y-559	51216	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
Y-11878	62444	2	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1
Y-12683	73253	1.5	1	1.5	1	1.5	1	1	1	1	1	1	1.5	1	1	1	1
Y-234	64056	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Y-583	56864	1.5	1	1.5	1	1	1	1	1	1	1	1	1.5	1	1	1	1
Y-582	52199	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Y-5997	73394	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
YB-369	60454	2	1	2	1	1	2	1	1	1.5	1	1	1	1	1	1	1
YB-1188	64755	1.5	1	1.5	1	1	1.5	1	1	1	1	1	1	1	1	1	1
YB-428	60175	1.5	1	1	1	1	1.5	1	1	1	1	1	1	1	1	1	1

7. MULac Assay of selected haploid, diploid and mated Te strains

The haploid, diploid and mated versions of the selected Te strains Y-234 and YB-428 were assayed for enzymatic activity (Figure 18). The results for the mated version did not reflect the sum of the two contributing haploid results nor even the same as the higher secretor YB-428 (H), but rather very close to the result of the average secretor, Y-234 (A). One conclusion to be drawn from these results is that, in the mated version Y-234 x YB-428, the H gene expressing the cellobiohydrolase in YB-428 is recessive to the dominant A gene in Y-234, resulting in enzyme activity resembling an average secetor. Furthermore, a doubling of the dominant gene in the 2n diploid produced nearly twice the enzyme activity, whereas doubling the recessive gene decreases this activity, albeit not halving it.

These results will help to explain the activities of the F1 progeny in searching for those which exhibit the highest level of enzyme secreting activity during further screening for the highest secretors of cellobiohydrolase.

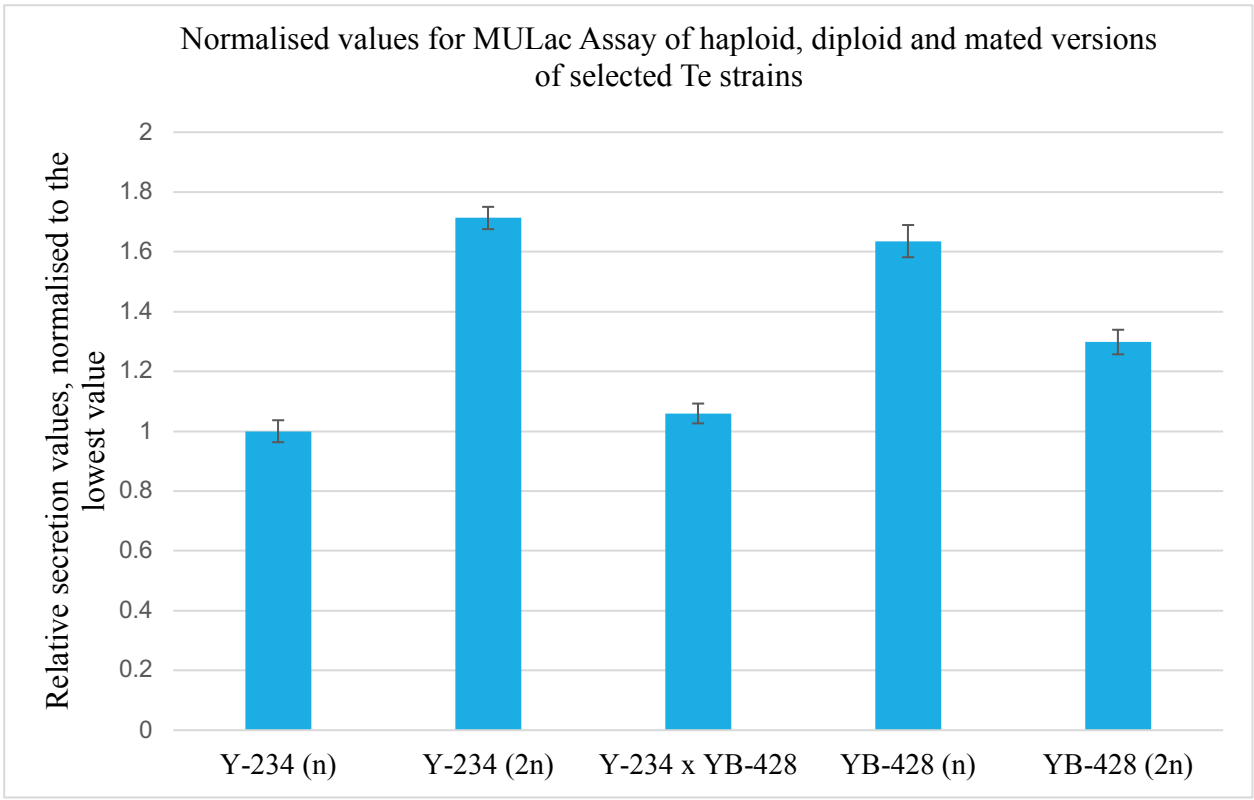


Figure 18. Supernatant CBH1 activity of haploid, diploid and mated versions of selected Te strains.

8. PHENOTYPE EVALUATION OF F1 PROGENY

Around 900 spores were dissected from the asci of the mated strains Y-234 x YB-428, grown on YPD plates and cellobiohydolase secretion and enzyme activity tested by the MULac Assay. The results provide a range of activity from 4.62 katal to 96.0 katal (Figure 19). From these, the genomes of a top selection of the highest secretors can be sequenced (post thesis submission) and examined with QTL analysis for the alleles responsible for high enzyme secretion.

With the secretion patterns of the various yeast strains having been determined, the genomes of the high secretors, average secretors and their progeny – once mated – can be sequenced (the sequencing of the progeny will take place as on-going work, following submission of this thesis). From these data, the genomes can be examined and differences highlighted between high and average secretors through SNP analysis, copy-number variants and variable ORFs. The alleles responsible for high protein secretion can then be identified through quantitative trait loci (QTL) analysis.

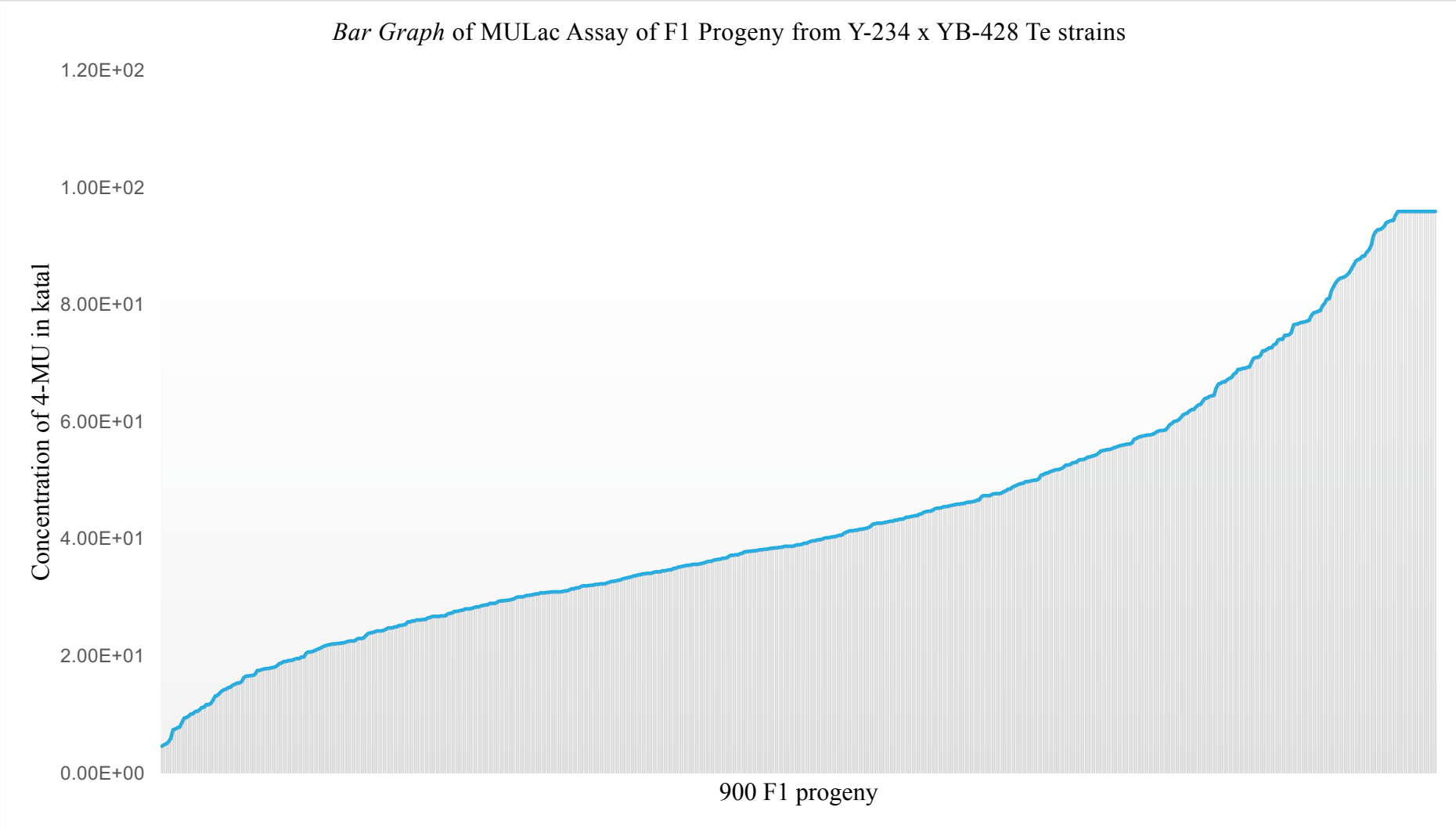


Figure 19. Supernatant CBH1 activity of F1 progeny from Y-234 x YB-428 Te strains.

CHAPTER FOUR

CONCLUSIONS

The worldwide demand for limited fossil fuels is driving new initiatives in green fuel technology. For the efficient utilisation of renewable waste biomass as a feedstock for yeast cell factories, high levels of cellulolytic enzymes are required, with cellobiohydrolases being reported as the rate limiting step.

Yeast strains isolated from different environments have been shown to display phenotypic heterogeneity, usually as a result of genomic alterations allowing for a competitive advantage. Here we evaluated the ability of genetically diverse *S. cerevisiae* strains to produce heterologous proteins. To this end, we constructed and introduced cellobiohydrolase-gene-expressing cassettes into 18 strains and evaluated their cellobiohydrolase secretion capacity. Even without any obvious natural selection to high cellulase secretion, we found a large spread of protein secretion capabilities between the different strains, with differences within a single strain, depending on the heterologous protein being expressed. Analysis of the segregation of the high *Rasamsonia emersonii* CBH secretion phenotype in the F1 progeny revealed that this phenotype is a complex trait, with multiple alleles contributing to this single characteristic.

The dissection of these alleles will allow a greater understanding of the genetic factors that govern high protein secretion in yeast and will provide tools for the improvement of current protein production yeast platforms. This study highlights how the naturally occurring yeast diversity can provide answers for improving the next generation of industrial yeast strains.

REFERENCES:

Boschker HTS and Cappenberg TE, 1994. A sensitive method using 4-methyl-umbelliferyl-β-cellobiose as a substrate to measure (1,4)-β-glucanase activity in sediments. Applied and Environmental Microbiology, 60:3592-3596.

Brady SK, Sreelatha S, Feng Y, Chundawat SP, Lang MJ, 2015. Cellobiohydrolase 1 from *Trichoderma reesei* degrades cellulose in single cellobiose steps. Nature Communications, 6:10149.

Brown JH, 1988. Species diversity. Analytical Biogeography, Chapter 3:57-89.

Coughlan MP, 1990. Cellulose degradation by fungi. Microbial Enzymes and Biotechnology, 1-29.

Davison SA, den Haan R, van Zyl WH, 2016. Heterologous expression of cellulase genes in natural *Saccharomyces cerevisiae* strains. Applied Microbiology and Biotechnology, 100:8241-8254.

Deutschbauer AM and Davis RW, 2005. Quantitative trait loci mapped to single nucleotide resolution in yeast. Nature Genetics, 37:1333-1340.

van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C, Giuseppin MLF, Heijnen JJ, Hoare M, Lange HC, Madden EA, Niederberger P, Nielsen J, Parrou JL, Petit T, Porro D, Reuss M, van Riel N, Rizzi M, Steensma HY, Verrips CT, Vindelov J, Pronk JT, 2000. An inter-laboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. Enzyme and Microbial Technology, 26:706-714.

Fay JC and Benavides JA, 2005. Evidence for domestication and wild populations of *Saccharomyces cerevisiae.* PLoS Genetics, 1:66-71.

Gallone B, Steensels J, Prahl T, Soriaga L, Saels V, Herrera-Malaver B, Merlevede A, Roncoroni M, Voordeckers K, Miraglia L, Teiling C, Steffy B, Taylor M, Schwartz A, Richardson T, White C, Baele G, Maere S, Verstrepen KJ, 2016. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. Cell, 166:1397-1410.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG, 1996. Life with 6000 genes. Science, 274:546-567.

la Grange DC, den Haan R, van Zyl WH, 2010. Engineering cellulolytic ability into bioprocessing organisms. Applied Microbiology and Biotechnology, 87:1195-1208.

Green MR and Sambrook J, 2012. Molecular Cloning: Laboratory Manual, Fourth Edition. Cold Spring Harbor.

Gusakov AV, Sinitsyn AP, Salanovich TN, Bukhtojarov FE, Markov AV, Ustinov BB, van Zeijl C, Punt P, Burlingame R, 2005. Purification, cloning and characterisation of two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense*. Enzyme and Microbial Technology, 36:57-69.

den Haan R, Kroukamp H, van Zyl J-HD, van Zyl WH, 2013. Cellobiohydrolase secretion by yeast: current state and prospects for improvement. Process Biochemistry, 48:1-12.

den Haan R, van Rensburg E, Rose SH, Gorgens JF, van Zyl WH, 2015. Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing. Current Opinion in Biotechnology, 33:32-38.

Hasunuma T and Kondo A, 2012. Development of yeast factories for consolidated bioprocessing of lignocellulose to bioethanol through cell surface engineering. Biotechnology Advances, 30:1207-1218.

Herpoël-Gimbert I, Margeot A, Dolla A, Jan G, Molle D, Lignon S, Mathis H, Sigoillot J-C, Monot F, Asther M, 2008. Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains. Biotechnology for Biofuels, doi:10.1186/1754-6834-1-18.

Ilmén M, den Haan R, Brevnova E, McBride J, Wiswall E, Froehlich A, Koivula A, Voutilainen SP, Siika-aho M, la Grange DC, Thorngren N, Ahlgren S, Mellon M, Deleault K, Rajgarhia V, van Zyl WH, Penttila M, 2011. High level secretion of cellobiohydrolases by *Saccharomyces cerevisiae*. Biotechnology for Biofuels, 4:1-15.

Jackson P, de Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, Cruaud C, Labadie K, Aury J-M, Istace B, Lebrigand K, Barbry P, Engelen S, Lemainque A, Wincker P, Liti G, Schacherer J, 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature Research Article, doi.org/10.1038/s41586-018-0030-5.

Jeoh T, Michener W, Himmel ME, Decker SR, Adney WS, 2008. Implications of cellobiohydrolase glycosylation for use in biomass conversion. Biotechnology for Biofuels, 1:10-21.

Kavšček M, Stražar M, Curk T, Natter K, Petrovič U, 2015. Yeast as a cell factory: current state and perspectives. Microbial Cell Factories, 14:94-103.

Kobayashi O, Suda H, Ohtani T, Sone H, 1996. Molecular cloning and analysis of the dominant flocculation gene *FLO8* from *Saccharomyces cerevisiae.* Molecular and General Genetics, 251:707-715.

Kroukamp H, den Haan R, la Grange DC, Sibanda N, Foulquié-Moreno MR, Thevelein JM, van Zyl WH, 2017a. Strain breeding enhanced heterologous cellobiohydrolase secretion by *Saccharomyces cerevisiae* in a protein specific manner. Biotechnology Journal, DOI: 10.1002/biot.201700346.

Kumaran R, Yang S-Y, Leu J-Y, 2013. Characterization of chromosome stability in diploid, polyploid and hybrid yeast cells. PLoS ONE, 8:1:10.

Liu H, Styles CA, Fink GR, 1996. *Saccharomyces cerevisiae* S288c has a mutation in *FLO8*, a gene required for filamentous growth. Genetics, 144:967-978.

Liu Y-S, Baker JO, Zeng Y, Himmel ME, Haas T, Ding S-Y, 2011. Cellobiohydrolase hydrolyses crystalline cellulose on hydrophobic faces. Journal of Biological Chemistry, 286:11195-11201.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS, 2002. Microbial cellulose utilisation: fundamentals and biotechnology. Microbiology and Molecular Biology Reviews, 66:506-577.

Lynd LR, van Zyl WH, McBride JE, Laser M, 2005. Consolidated bioprocessing of cellulosic biomass: an update. Current Opinion in Biotechnology, 16:577-583.

Meijnen J-P, Randazzo P, Foulquié-Moreno MR, van den Brink J, Vandecruys P, Stojiljkovic M, Dumortier F, Zalar P, Boekhout T, Gunde-Cimerman N, Kokosar J, Stajdohar M, Curk T, Petrovic U, Thevelein JM, 2016. Polygenic analysis and targeted improvement of the complex trait of high acetic acid tolerance in the yeast *Saccharomyces cerevisiae*. Biotechnology for Biofuels, 9,5:1-18.

Mukherjee V, Steensels J, Lievens B, van de Voorde I, Verplaetse A, Aerts G, Willems KA, Thevelein JM, Verstrepen KJ, Ruyters S, 2014. Phenotypic evaluation of natural and industrial *Saccharomyces* yeasts for different traits desirable in industrial bioethanol production. Applied Microbiology ad Biotechnnology, 98:9483-9498.

Narendranath NV, Thomas KC, Ingledew WM, 2001. Effects of acetic acid and lactic acid on the growth of Saccharomyces cerevisiae in a minimal medium. Journal of Industrial Microbiology and Biotechnology, 26:171-177.

Pais TM, Foulquié-Moreno MR, Hubmann G, Duitama J, Swinnen S, Goovaerts A, Yang Y, Dumortier F, Thevelein JM, 2013. Comparative polygenic analysis of maximal ethanol accumulation capacity and tolerance to high ethanol levels of cell proliferation in yeast. PLoS Genetics, 9,6:1-18.

Parisutham V, Kim TH, Lee SK, 2014. Feasibilities of consolidated bioprocessing microbes: from pre-treatment to biofuel production. Bioresource Technology, 161:431-440.

Puxbaum V, Gasser B, Mattanovich D, 2016. The bud tip is the cellular hot spot of protein secretion in yeasts. Applied Microbial and Cell Physiology, 100:8159-8168.

van Rensburg P, van Zyl WH, Pretorius IS, 1998. Engineering yeast for efficient cellulose degradation. Yeast, 14:67-76.

Romanos MA, Scorer CA, Clare JJ, 1992. Foreign gene expression in yeast: a review. Yeast, 8:423-488.

de Ruijter JC, Koskela EV, Frey AD, 2016. Enhancing antibody folding and secretion by tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum. Microbial Cell Factories, 15:87.

al Safadi R, Weiss-Gayet M, Briolay J, Aigle M, 2010. A polyploid population of *Saccharomyces cerevisiae* with separate sexes (dioecy). FEMS Yeast Research, 10:757-768.

Snapp EL, 2012. Unfolded protein responses with or without unfolded proteins? Cells, 1:926-950.

Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW, 2002. Dissecting the architecture of a quantitative trait locus in yeast. Nature, 416:326-330.

Stephanopoulos G, 2007. Challenges in engineering microbes for biofuels production. Science, 315:801-804.

Sticklen MB, 2008. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nature Reviews – Genetics, 9:433-443.

Suranska H, Vranova D, Omelkova J, 2016. Isolation, identification and characterisation of regional indigenous *Saccharomyces cerevisiae* strains. Brazilian Journal of Microbiology, 47:181-190.

Swinnen S, Thevelein JM, Nevoigt E, 2012. Genetic mapping of quantitative phenotypic traits in *Saccharomyces cerevisiae*. FEMS Yeast Research, 12:215-227.

Teunissen AWRH and Steensma HY, 1995. Review: the dominant flocculation genes of *Saccharomyces cerevisiae* constitute a new subtelomeric gene family. Yeast, 11:1001-1013.

Treco DA and Winston F, 2008. Growth and manipulation of yeast. Current Protocols in Molecular Biology, 82:13.2.1-13.2.12

Vinzant TB, Adney WS, Decker SR, Baker JO, Kinter MT, Sherman NE, Fox JW, Himmel ME, 2001. Fingerprinting *Trichoderma reesei* hydrolases in a commercial cellulase preparation. Applied Biochemistry and Biotechnology, 91:99-107.

Wilkening S, Lin G, Fritsch ES, Tekkedil MM, Anders S, Kuehn R, Nguyen M, Aiyar RS, Proctor M, Sakhanenko NA, Galas DJ, Gagneur J, Deutschbauer A, Steinmetz LM, 2014. An evaluation of high-throughput approaches to QTL mapping in *Saccharomyces cerevisiae*. Genetics, 196:853-865.

Yamada R, Hasunuma T, Kondo A, 2013. Endowing non-cellulolytic microorganisms with cellulolytic activity aiming for consolidated bioprocessing. Biotechnology Advances, 31:754-763.

Zhang Y-H and Lynd LR, 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: non-complexed cellulase systems. Biotechnology and Bioengineering, 88:797-824.

van Zyl WH, Lynd LR, den Haan R, McBride JE, 2007. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. Advances in Biochemical Engineering Biotechnology, 108:205-235.

SUPPLEMENTARY MATERIAL

Table S1: A series of concentration levels of 4-MU for the construction of a Standard Curve.

µM 4-MU	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
µL 4-MU	0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
µL Na_2CO_3	1000	999.5	999	998.5	998	997.5	997	996.5	996	995.5
Well volume	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL

µM 4-MU	**1.0**	**2.0**	**3.0**	**4.0**	**5.0**	**6.0**	**7.0**	**8.0**	**9.0**	**10.0**
µL 4-MU	5	10	15	20	25	30	35	40	45	50
µL Na_2CO_3	995	990	985	980	975	970	965	960	955	950
Well volume	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL

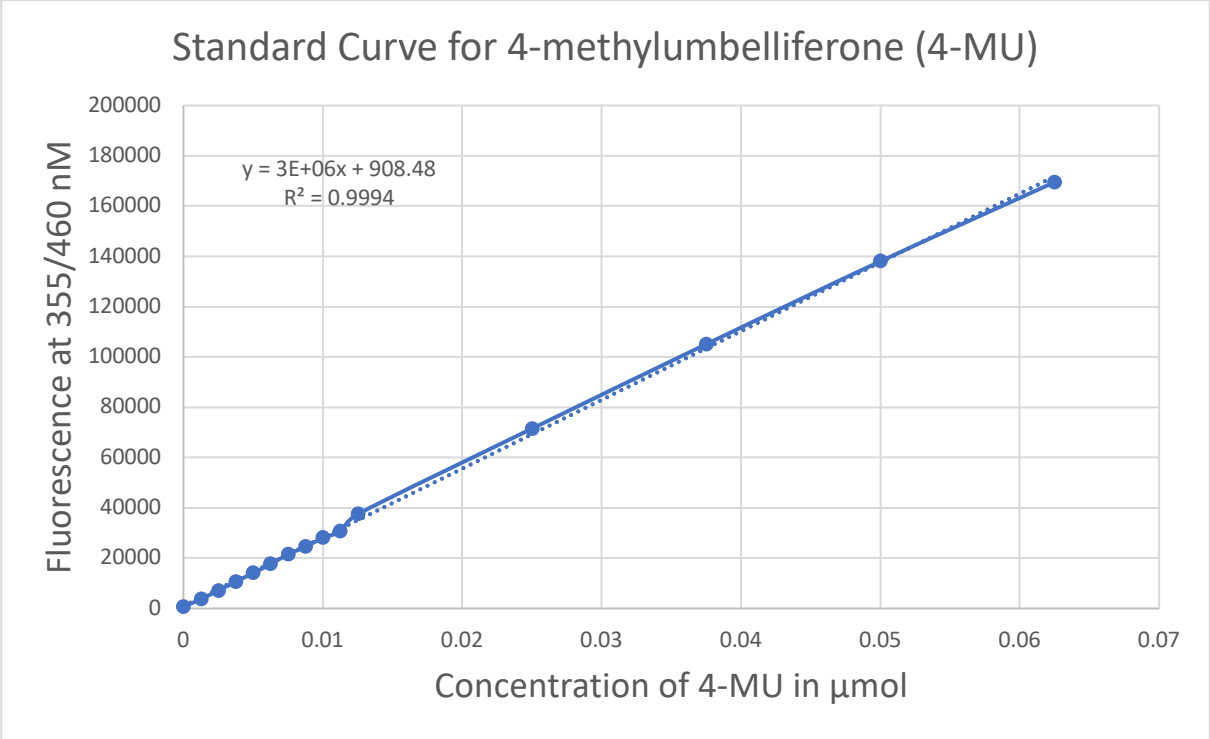


Figure S1: Standard Curve of concentration of 4-methylumbelliferone against its fluorescence.

Table S2. List of *S. cerevisiae* strains with their confirmation of transformation and mating type and matched pairs of selected strains.

Strain	Transformation		Te - Mating type		Tr - Mating type	
WT, lab, wine	Te.CBH	Tr.CBH	a	α	a	α
Y 583	Y	Y		α		α
Y 582	Y	Y	a		a	
Y 11857	Y	Y		α		α
YB 369	Y	Y		α		α
YB 210	Y	Y		α		α
Y 234	Y	Y		α		α
Y 559	Y	Y	a		a	
YB 1188	Y	Y	a		a	
YB 428	Y	Y	a		a	
Y 5997	Y	Y		α		α
Y 12638	Y	Y		α		α
Y 12683	Y	Y	a		a	
Y 11878	Y	Y		α		α
Y 12603	Y	Y		α		α
CenPK 113-7D	Y	Y	a		a	
S288c	Y	Y		α		α
By 4742	Y	Y		α		α
AWRI 1631	Y	Y	a		a	

Table S3. List of *S. cerevisiae* strains with their data in preparation for sequencing.

	Sequencing data and designations					
Strain	tube	no.	Number	Nanodrop	260/280	260/230
WT, lab, wine	Te	Tr	in order	reading	RNA	Protein
Y 583	6	16	1	4097	2.11	2.31
Y 582	7	17	2	1365	2.18	1.82
Y 11857			3	2272	2.17	2.28
YB 369	9	12	4	2815	2.15	1.99
YB 210			5	2481	2.12	2.06
Y 234	3	20	6	1641	2.09	2.01
Y 559	10	19	7	4037	2.14	2.28
YB 1188	5	14	8	2764	2.15	2.22
YB 428	1	18	9	1720	2.10	2.21
Y 5997	2	13	10	2957	2.10	2.19
Y 12638			11	1336	2.14	2.13
Y 12683	4	15	12	789	2.09	1.4
Y 11878	21		13	649	2.04	1.53
Y 12603			14	1341	2.12	1.73
Cen.PK 113-7D	8	11	15	2793	2.08	2.09
S288c			16	4772	2.13	2.28
By 4742			17	1292	2.11	2.05
AWRI 1631			18	1748	2.10	2.19
			19	1895	2.13	1.83
			20	1999	2.13	2.13
			21	2247	2.17	2.21

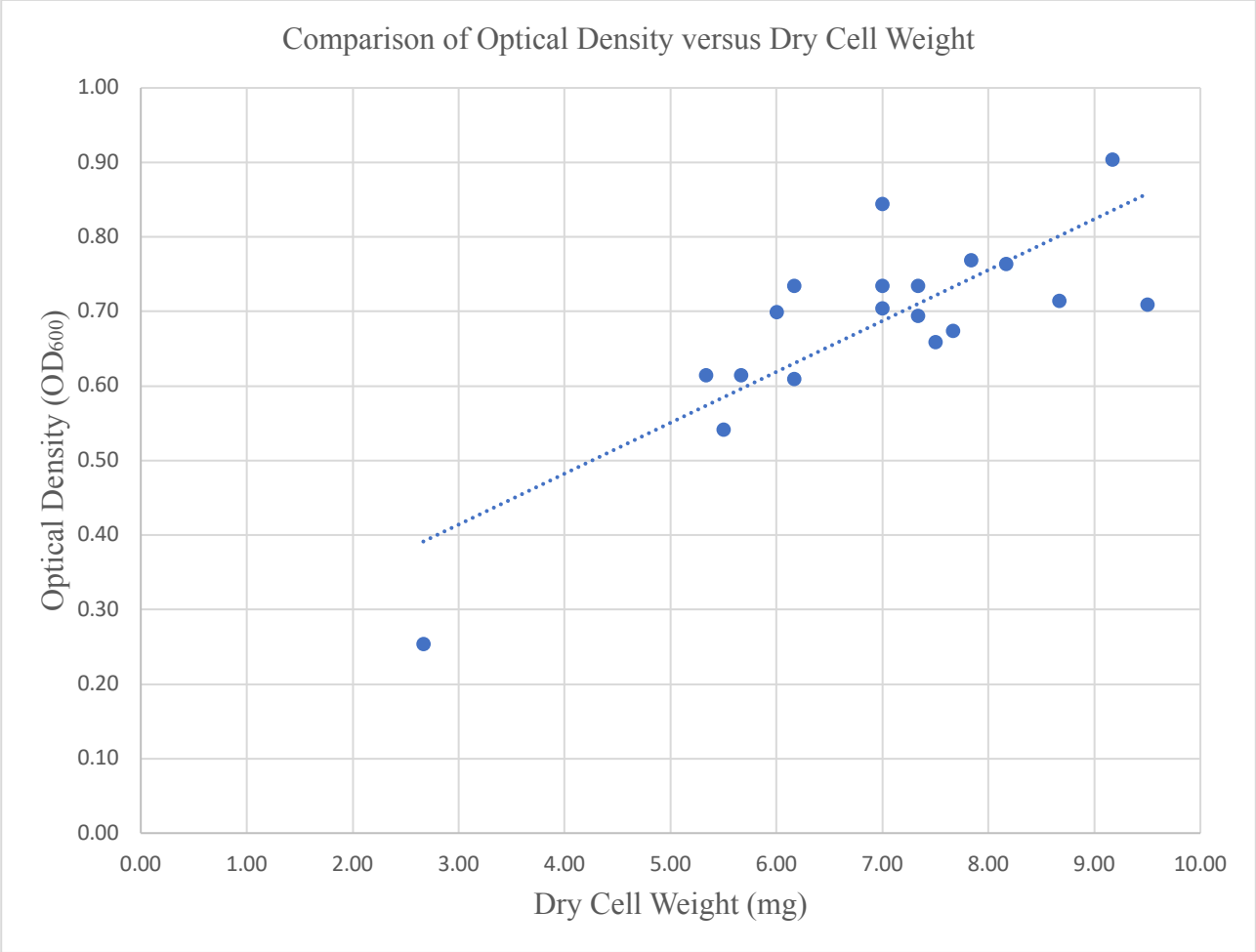


Figure S2. Linear regression of optical density versus dry cell weight of Te-expressing strains.