A systems biology approach to annotate novel open reading frames in the *Saccharomyces cerevisiae* Enoferm M2 wine yeast strain

by

Andrew Taplin Ferguson

B.Sc., University of Victoria, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS OF THE DEGREE OF
MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Food Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2017

ABSTRACT

Saccharomyces cerevisiae has been extensively studied both for its commercial applications in alcoholic beverages, bread and biofuel, and as a genetic model for higher eukaryotes; a function, however, has still not been described for over 10% of *S. cerevisiae*'s genes, and furthermore novel open reading frames (ORFs) have recently been identified in many yeast strains isolated from wine, beer and bioethanol industries. The present study used a systems biology approach, incorporating metabolomics, transcriptomics and proteomics, to systematically analyze novel ORFs and poorly-annotated genes during wine fermentation. Genes were either deleted or constitutively expressed in a commercial wine yeast, Enoferm M2, and the mutant strains were used to ferment Chardonnay grape must. Primary metabolites were monitored throughout the fermentation by high-performance liquid chromatography (HPLC), volatile metabolites were analyzed in the final wine by gas chromatography mass spectrometry (GCMS), and the transcriptome and proteome of the fermenting yeast were analyzed by ribonucleic acid (RNA) microarray and isobaric tagging for relative and absolute quantification (iTRAQ), respectively. The expression of fourteen novel ORFs from the Enoferm M2 genome was confirmed for the first time, and data was generated for the effect of mutation of these novel ORFs on wine fermentation. In addition to the novel ORFs, four poorly-annotated genes (YLL054C, YKL222C, *PHD1* and *HMS1*) were found to positively regulate the allantoin metabolic pathway, and had additional effects on the cell wall and transmembrane transport. The *GEP5* gene was found to negatively regulate the sulphur metabolic pathway, which could have significant effects on both sulphite tolerance and the production of off-flavours such as hydrogen sulphide during wine production. The systems biology approach used in conjunction with systematic gene mutation and industrially-relevant conditions was able to improve the annotation of novel and poorly-annotated genes in the *S. cerevisiae* genome.

LAY SUMMARY

Wine yeast has been extensively studied both for its commercial applications and as a genetic model for larger organisms; a function, however, has still not been described for over 10% of yeast's genes, and furthermore novel genes have been identified in yeast isolated from human industries. The present study used a multi-facetted approach to systematically analyze poorly-described genes during a wine fermentation. Data was generated for the first time for novel genes only found in a commercial wine yeast. In addition, a cluster of genes were found to regulate an aspect of nutrient metabolism, and another gene was found to affect sulphur metabolism, which could have significant effects on off-flavours produced during wine production. The multi-facetted analyses in conjunction with industrially-relevant conditions improved the description of novel yeast genes.

PREFACE

This thesis is an original and unpublished work conducted by the author, Andrew Ferguson, under the supervision of Dr. Hennie J.J. van Vuuren and the guidance of his supervisory committee members: Dr. Vivien Measday, Dr. Simone Castellarin and Dr. Christopher J. R. Loewen.

The main concept for this research project and the overarching research design was conceived by Dr. van Vuuren. The research was funded by a Natural Sciences and Engineering Research Council Collaborative Research and Development (NSERC CRD) grant to Dr. van Vuuren in partnership with Mark Anthony Group Inc. The majority of the work presented in this thesis was conducted in the Wine Research Centre at the University of British Columbia.

The novel Open Reading Frames (ORFs) investigated in this study were selected from previously identified novel ORFs in the thesis work of Mayumi Iwashita, a graduate of the van Vuuren laboratory. Additional ORFs investigated in this study were selected based on a previous study published by the van Vuuren laboratory (Marks *et al.*, 2008).

The construction of 17 non-novel null mutants was conducted by Dr. Christopher J. Walkey, and the construction of two constitutive expression mutants was conducted by Mayumi Iwashita. All other mutants described in this study were constructed by A. Ferguson.

Metabolic analysis of the fermented wine by high-performance liquid chromatography (HPLC) and gas chromatography mass spectrometry (GCMS) was conducted under the technical guidance of the Wine Research Centre Mass Spectrometry Manager, Lufiani L. Madilao, at the University of British Columbia. Proteomic analysis was conducted by the Technical Operations Manager, Derek Smith, at the University of Victoria Genome BC Proteomics Centre. Microarray analysis of 16 of the non-novel null mutants was conducted by Dr. Zongli Luo at the University of British Columbia. All other experiments were conducted at the University of British Columbia by Andrew Ferguson.

Dr. van Vuuren and Dr. Measday contributed to the editing of the thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
2D-PAGE	2-Dimensional PolyAcrylamide Gel Electrophoresis
ANOVA	Analysis Of Variance
ATP	Adenosine Triphosphate
BCA	Bicinchoninic Acid
BLAST	Basic Local Alignment Search Tool
bp	Base pair
cDNA	Complementary Deoxyribonucleic Acid
CDS	Coding DNA Sequence
CE	Constitutive Expression
CNV	Copy Number Variation
cRNA	Complementary Ribonucleic Acid
DE	Differential Expression/Differentially Expressed
DNA	Deoxyribonucleic Acid
dsDNA	Double-stranded Deoxyribonucleic Acid
FC	Fold Change
FDR	False Discovery Rate
FSR	Fermentation Stress Response
GCMS	Gas Chromatography Mass Spectrometry
GMO	Genetically Modified Organism
HCD	Higher-energy Collisional Dissociation
HOG	High-Osmolarity Glycerol
HPLC	High Performance Liquid Chromatography
HSD	Honest Significant Difference
indel	Insertion or Deletion
iTRAQ	isobaric Tagging for Relative and Absolute Quantification
ITS	Internally Transcribed Spacer
IVT	*In Vitro* Transcription
LC-MS	Liquid Chromatography – Mass Spectrometry
MMTS	S-Methyl methanethiosulphonate
mRNA	messenger Ribonucleic Acid
MS	Mass Spectrometry
NCBI	National Center for Biotechnology Information
NGS	Next-Generation Sequencing
NSERC CRD	National Sciences and Engineering Research Council Collaborative Research and Development
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
Pfam	Protein family
qRT-PCR	Quantitative Real-Time Polymerase Chain Reaction
QTL	Quantitative Trait Loci
rDNA	Ribosomal Deoxyribonucleic Acid

RNA	Ribonucleic Acid
RNA-seq	Ribonucleic Acid sequencing
SAGE	Serial Analysis of Gene Expression
SGD	Saccharomyces Genome Database
SNP	Single Nucleotide Polymorphism
SPME	Solid-Phase MicroExtraction
TAE	Tris, Acetic Acid and Ethylenediaminetetraacetic acid
TCEP	Tris(2-carboxylethyl)phosphine
TEAB	Triethylammonium bicarbonate
tRNA	Transfer Ribonucleic Acid
WGD	Whole Genome Duplication
YMM	Yeast Minimal Media
YNB	Yeast Nitrogen Base
YPD	Yeast extract, Peptone and Dextrose

1 INTRODUCTION

1.1 *Saccharomyces cerevisiae*, a model organism

1.1.1 History of laboratory studies of *S. cerevisiae*

Saccharomyces cerevisiae is an ascomycetous yeast long associated with human activity. The species name when translated from its Latin roots roughly means "sugar-mold of beer". Historically, *S. cerevisiae* has been associated with the production of bread and alcoholic beverages, such as wine and beer. More recently, *S. cerevisiae* has been used in the production of bioethanol and other industrially significant metabolites (McIlwain *et al.*, 2016). Food, beverage and biofuel industries utilize the ability of *S. cerevisiae* to rapidly ferment sugars into ethanol and carbon dioxide, even in the presence of oxygen. The close association of this yeast with human activity has led to a wealth of research, beginning in the late 1800s in the Carlsberg Laboratory in Denmark when pure cultures of yeast were first used for the production of beer (Verstrepen, Chambers and Pretorius, 2006). Scientific research with *S. cerevisiae* accelerated in the 1930s and 1940s through the work of Øjvind Winge and Carl Lindegren. Their approaches differed, with Winge working primarily with homothallic isolates, while Lindegren used heterothallic strains for breeding studies. Lindegren is credited with the popularization of the direct progenitor of one of the most utilized lab strains of *S. cerevisiae*, S288C, from strains originally isolated by Emil Mrak in the late 1930s from rotting figs in Merced, California (Mortimer and Johnston, 1986). The type strain S288C was developed from Lindegren's isolates by Mortimer with the intension of producing a yeast that could grow easily and predictably in the laboratory. While the primary lab strain of *S. cerevisiae* was isolated from figs, Mortimer (2000) has speculated that this strain may have in fact been transferred to the fig by wasps from a nearby vineyard.

1.1.2 First eukaryote to have its genome fully sequenced

While physically small and unicellular, *S. cerevisiae* is nonetheless a eukaryotic organism. As such, many of the cellular processes are the same as larger, multicellular eukaryotes, such as plants and humans. The relatively small genome size (~13 megabases) and large number of chromosomes (16)

make *S. cerevisiae* an ideal model for the study of eukaryotic genetics. The genomic features, combined with a rapid generation time, compact size, simple nutritional requirements and easy gene editing techniques, have helped to make *S. cerevisiae* such a ubiquitously used model organism. *Saccharomyces cerevisiae* became the first eukaryotic organism to have its genome fully sequenced in 1996, through a collaborative effort by scientists across the globe (Goffeau *et al.*, 1996). The original report described 12,068 kilobases of deoxyribonucleic acid (DNA), coding for 5,885 open reading frames (ORFs), as well as 140 ribosomal ribonucleic acids (RNAs), 40 small nuclear RNAs and 275 transfer RNAs. One of the striking features of the yeast genome is how compact it is; roughly 70% of the genomic material encodes a putative protein, while in the human genome there is about 15 times as much non-coding DNA (Goffeau *et al.*, 1996).

While the original report identified 5,885 ORFs, after discarding a further 390 ORFs that were deemed unlikely to yield proteins, the actual number of protein-coding ORFs in *S. cerevisiae* has been slowly defined since. Not all ORFs – sequences containing a start codon and stop codon – encode a translated protein. ORFs have three levels of classification based on the likelihood of being actively transcribed: "dubious" ORFs are non-conserved and have no experimental evidence of being transcribed, "uncharacterized" ORFs are conserved among multiple species, and "verified" ORFs have experimental messenger RNA (mRNA) or protein evidence (Fisk *et al.*, 2006). ORFs shorter than 150 codons or situated within a longer ORF are generally thought to be non-coding, although modern techniques have identified functional proteins smaller than 100 codons in size (Yagoub *et al.*, 2015). Using mathematical modelling, Kowalczuk *et al.* (1999) estimated the number of protein-coding ORFs in *S. cerevisiae* to be closer to 4800 genes. This number is much closer to the number of genes predicted by early transcriptomic studies. The seminal transcriptome produced by Velculescu *et al.* (1997) identified 4665 genes based on 60,633 detected transcripts. The strong base of yeast genome sequencing has since been updated with a complete resequencing of the S288C genome in 2010, and updated versions of the genome continue to be produced (Engel *et al.*, 2014). Evolving technologies have allowed for improved resolution in genomic, transcriptomic and proteomic studies, and these developments will be discussed in

the following sections. See Engel *et al.* (2014) for a complete review of the history of the *S. cerevisiae* type strain S288C and the reference genome.

1.2 Genetic diversity

1.2.1 Strains of *S. cerevisiae*

1.2.1.1 Methods of defining strains.

Intraspecific genotype variation can have a strong influence on the phenotype of a strain, and is important to consider for experimental design. Moreover, a domesticated species such as *S. cerevisiae* can display variations in phenotype and genotype due to geographic or reproductive differentiation. *S. cerevisiae* relies predominantly on an asexual lifestyle, reproducing clonally with infrequent (one in 1000 divisions) sexual reproduction and only periodic (one in ~50,000 divisions) outcrossing events (Bisson, 2012). Consequently, yeast scientists tend to discuss strains of *S. cerevisiae* that are populations of isogenic cells. It is necessary to have robust methods to differentiate strains, and multiple lines of evidence have been used to do so over the years.

Early studies of yeast used morphology and growth requirements to differentiate species. These techniques remain the basis of well-accepted yeast taxonomies (Barnett, Payne and Yarrow, 1990). Nuclear ribosomal internally transcribed spacer (ITS) lengths can be amplified with yeast-specific primers and used to differentiate a number of yeast species and identify hybrids, although this technique does not possess the resolution to go beyond the species level (Bradbury *et al.*, 2006; Schoch *et al.*, 2012). Alternative techniques, sequencing the highly variable D1/D2 domain of the 26S ribosomal DNA (rDNA) or sequencing the entire 18S rDNA, are also able to differentiate genera and species, but no further. These techniques helped redefine older taxonomies by identifying yeasts that had phenotypic differences, but were genetically equivalent in these regions (Kurtzman and Robnett, 1998).

Phenotypic traits continue to be used regularly to differentiate strains, and are the most important characteristic for industry applications. A wide range of phenotypic variation exists within *S. cerevisiae*, and this variation continues to be characterized (Kvitek, Will and Gasch, 2008; Skelly *et al.*, 2013). As

researchers assess natural isolates for their potential abilities, phenotypic traits such as growth rate and stress tolerance continue to be important metrics for the differentiation of novel strains (Hyma *et al.*, 2011; Treu, Campanaro, *et al.*, 2014). Most recently, next-generation sequencing (NGS) techniques have allowed for the high-throughput sequencing of full yeast genomes. NGS-based studies have allowed for the high-resolution differentiation of strains, rapidly advancing our understanding of the relationships between strains, and between their genotypes and phenotypes.

1.2.1.2 Next-generation sequencing.

Next generation sequencing (NGS) is the overarching term for a suite of technologies that have allowed for the rapid and high-resolution sequencing of large amounts of DNA or RNA by a combination of genome fragmentation, tagging, amplification and automated detection. In an example, a general Illumina sequencing experimental flow begins with the extraction of genomic DNA from the target strain. The DNA is then randomly fragmented, and the fragments ligated to specific adapter sequences. These adapter sequences correspond to a lawn of primers on a flow cell. A polymerase enzyme and the denatured, tagged DNA fragments are added to the flow cell and nucleotides are incorporated to produce double-stranded DNA (dsDNA), amplifying the potential signal. The dsDNA is then denatured back to single stranded DNA fragments. Primers, DNA polymerase and fluorescently-labelled nucleotides are then added to the flow cell, and one base is added to each strand on DNA. A laser detects the identity of the first base, then another reaction cycle proceeds, and the second base is added and recorded, and so on. In this manner, the sequence of millions of 100-250 bp fragments can be determined rapidly and accurately (Illumina, 2010). The result is a massive amount of data that must be assembled and interpreted with bioinformatic techniques (DePristo *et al.*, 2011; McIlwain *et al.*, 2016).

Since the advent of NGS, there has been a rapid progression in sequencing technology, with an increase in the speed and coverage and reduces the associated cost. Particularly for yeasts, which possess a relatively small genome, it is no longer a prohibitive cost to undertake large amounts of sequencing. Early efforts used whole-genome shotgun sequencing, a method which cuts the genomic DNA into small

fragments that are sequenced and then assembled back together. This method relies on a robust reference genome for scaffolding as telomeric and repeated regions are poorly assembled. Early efforts were only able to sequence two or three yeasts by this method (Kellis *et al.*, 2003). A genome reduction approach by Cromie *et al.* (2013), using restriction site-associated sequencing allowed 1% of the genome of 262 diverse strains of *S. cerevisiae* to be sequenced and compared. More recent studies have focused on the sequencing of strains with particular industrial importance (Argueso *et al.*, 2009), improving the quality of coverage by integrating multiple sequencing technologies (McIlwain *et al.*, 2016), or generating genome sequences for large numbers of divergent strains (Borneman *et al.*, 2016; Gallone *et al.*, 2016). Recently a single study sequenced 157 strains of *S. cerevisiae* in their natural ploidy from industrial environments, including 102 beer isolates (Gallone *et al.*, 2016). This was followed by a similar study that sequenced 212 strains of *S. cerevisiae*, with a greater focus on wine-related strains (Borneman *et al.*, 2016). There is now a wealth of publicly available information available for comparative genomic studies on a massive scale. A complete list of *S. cerevisiae* genome sequences available from the National Center for Biotechnology Information (NCBI, Bethesda MD, USA, www.ncbi.nlm.nih.gov) as of November 30th, 2017, can be found in Appendix A.1.

1.2.1.3 Genetic diversity of S. cerevisiae.

The extensive sequencing of the *S. cerevisiae* genome has given us a better understanding of its genetic diversity. Strains are typically diploid, although many laboratory strains are haploid to facilitate manipulation, and many industrial strains are aneuploid (Bradbury *et al.*, 2006; Bisson, 2012). Wine strains are often homothallic diploids, meaning that after sporulation, haploid cells can switch their mating type and mate with cells derived from the same meiotic division (Mortimer, 2000; Bisson, 2012). A consequence of homothallism is that a large proportion of wine yeast is homozygous as a result of inbreeding (Bradbury *et al.*, 2006). Brewing strains, on the other hand, have a greater propensity to be aneuploid or polyploid and highly heterozygous (Borneman *et al.*, 2011). While many studies sequence haploid derivatives (Borneman *et al.*, 2008; Argueso *et al.*, 2009; Liti *et al.*, 2009), or remove heterozygous regions from the final data from simplicity (Novo *et al.*, 2009), recent studies have aimed to

sequence strains in their natural ploidy in order to retain information about heterozygosity (Borneman *et al.*, 2011, 2016; Dunn *et al.*, 2012; Cromie *et al.*, 2013). Sequenced strains of *S. cerevisiae* show a relatively low level of heterozygosity, but this could reflect a bias from the greater likelihood of sequencing haploid derivatives (often used in laboratory studies) and homothallic strains (common in wine isolates). Cromie *et al.* (2013) found a large number of heterozygous strains associated with fruit sources, which could be the result of breeding between geographically diverse commercial strains and natural occurring strains, and this hypothesis is reinforced by a greater-than-expected level of heterozygosity found in human-associated yeast populations compared to wild isolates. The most comprehensive survey of heterozygosity in *S. cerevisiae* found a much higher degree of heterozygosity in ale and baking yeasts, which could be attributed to their frequency of higher ploidy numbers. Homozygosity was also found to be artificially induced in some commercial strains, as a result of techniques used in the development of the strain (Borneman *et al.*, 2016). The reproductive life history appears to have a strong influence on the zygosity of yeast strain.

Apart from intra-strain diversity, a large amount of inter-strain diversity exists in the forms of single nucleotide polymorphisms (SNPs), insertions or deletions (indels), retrotransposon (Ty) elements, or strain specific loci. SNPs and indels may create allelic differences in genes, or even play a role in regulation when present in non-coding regions. An early study of wine, lab and pathogenic yeast found SNPs to occur as frequently as every 150 bp on average (Borneman *et al.*, 2008), but the inclusion of more strains (12 strains vs. 3 strains), including brewing isolates, decreased the median distance between SNPs to 37 bp (Borneman *et al.*, 2011). These numbers appear to be dependent on the method of detection, data processing, and strains used, as a study using 63 strains and tiling microarrays found a reduced frequency of SNPs at one per 357 bp (Schacherer *et al.*, 2009). What can be said conclusively, however, is that genetic polymorphisms are not uniformly distributed, and are less likely to occur in coding regions and more likely to occur in subtelomeric regions (Schacherer *et al.*, 2009; Borneman *et al.*, 2011). In addition to small polymorphisms, many strains contain large (10 to over 100 kb) regions of DNA that cannot be placed in the reference genome (Borneman *et al.*, 2016). These regions have been

shown to encode functional proteins, homologous ORFs, novel ORFs and dubious ORFs. Many of these gene clusters are common to a number of related strains, such as the biotin-prototrophy genes found in sake strains and the "RTM1-cluster" found in brewing and hybrid strains (Borneman *et al.*, 2016). These non-conserved genomic loci tend to be located in subtelomeric regions, which are hotbeds of genetic plasticity in yeast (Borneman *et al.*, 2008, 2011, 2016).

1.2.2 Origin of diversity in *S. cerevisiae*

1.2.2.1 Natural diversity.

While *S. cerevisiae* is often discussed in the context of its industrial applications, it is important to reflect on its geographic distribution and the natural diversity that existed prior to conscious human intervention. A whole genome duplication (WGD) event occurred in the ancestry of *Saccharomyces* (Wolfe and Shields, 1997); post-WGD divergence of ohnologues gave rise to the system of fermentative and respiratory metabolism present in modern day *Saccharomyces* (Thompson and Cubillos, 2017). It has become generally accepted after extensive genomic comparisons that *S. cerevisiae* strains cluster into five clean lineages: European, North American, Sake, Malaysian and West African (Liti *et al.*, 2009; Warringer *et al.*, 2011; Bisson, 2012) (Figure 1.1). Apart from these "clean" lineages, a small number of strains, often associated with industry, show evidence of mosaic genomes, with varying contributions from different lineages (Liti *et al.*, 2009). When creating phylogenetic trees of *S. cerevisiae*, *S. paradoxus*, the closest related species, is often used as an outgroup. While *S. cerevisiae* is far less genetically diverse than *S. paradoxus*, *S. cerevisiae* demonstrates far greater phenotypic variation and niche diversification (Warringer *et al.*, 2011). Beyond geographic associations, *S. cerevisiae* lineages also cluster according to technological use, a topic which will be discussed thoroughly in the next section.

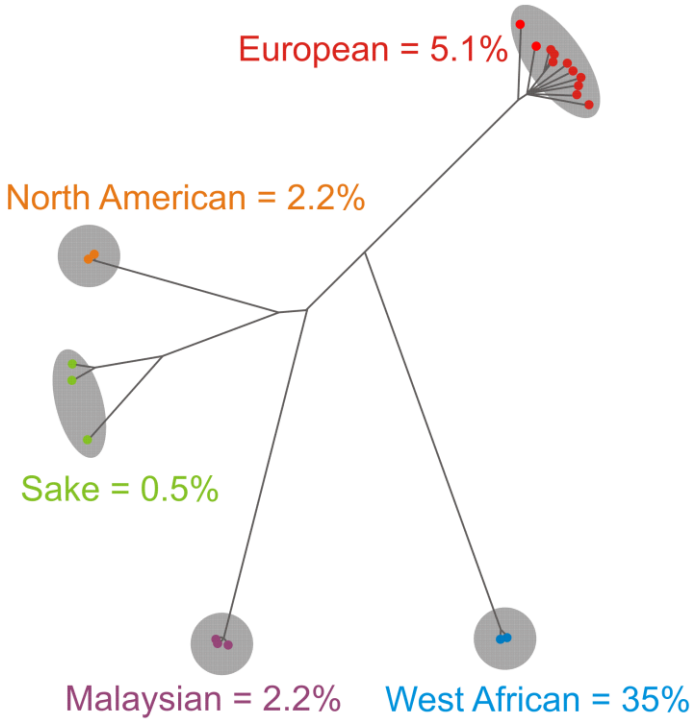


Figure 1.1 Phylogenetic separation of *S. cerevisiae* strains. Strains cluster cleanly by geographic origin. Percentages indicate frequency of population specific traits, a measure of distance from all other populations. Image is reproduced from Warringer et al. (2011) under a Creative Commons license.

So-called "natural" or "wild" isolates of *S. cerevisiae* are found in soil, on fruit, associated with oak, as human pathogens, and isolated from a myriad of natural fermentations (Liti *et al.*, 2009; Borneman and Pretorius, 2015; Gonçalves *et al.*, 2016). In one peculiar instance, *S. cerevisiae* has been isolated from bertam palm (*Eugeissona tristis*) nectar, where the alcohol content was as high as 3.8 %; a rare case where an alcoholic beverage is produced without human intervention (Wiens *et al.*, 2008). Natural isolates of *S. cerevisiae* vary widely in their phenotypes, which may have contributed to the wide variety of described species in the *Saccharomyces* genus prior to DNA-based phylogenies (Barnett, Payne and Yarrow, 1990; Kvitek, Will and Gasch, 2008; Warringer *et al.*, 2011). For instance, the ability to ferment galactose was historically used to differentiate *Saccharomyces* species, but recent work sequencing the GAL genes in various species has indicated that Gal^+ and Gal^- phenotypes are more widespread than previously thought, and not an appropriate measure for species identification (Warringer *et al.*, 2011; Dulermo *et al.*, 2016). One difficult aspect of studying natural variation in yeasts is the inherent bias towards the isolation of strains from human associated activities, as these are thought to be

most likely to possess commercially relevant traits. The continued isolation of yeasts from diverse environments and geographies will expand our knowledge of *S. cerevisiae* ecology and potential reservoirs of genetic diversity.

1.2.2.2 Domestication and selection of S. cerevisiae.

In addition to geographic lineages, *S. cerevisiae* strains also cluster by technological application (Legras *et al.*, 2007; Hyma *et al.*, 2011; Bisson, 2012; Borneman *et al.*, 2016; Gallone *et al.*, 2016; Gonçalves *et al.*, 2016). This is particularly true for strains associated with the production of wine, beer, bread and sake, all of which have been associated with human activity for millennia (Legras *et al.*, 2007; Gallone *et al.*, 2016; Gonçalves *et al.*, 2016). Many have reported this phylogenetic separation as a direct consequence of domestication (Legras *et al.*, 2007; Hyma *et al.*, 2011), although it has also been hypothesized that so-called "domestication phenotypes" may simply be the result of population bottlenecks and genetic drift (Warringer *et al.*, 2011). Some traits are more easily attributed to domestication than others; *CUP1* amplifications in *S. cerevisiae* are found wine and sake strains and are associated directly with the application of copper as a fungicide during the production of grapes and rice (Warringer *et al.*, 2011). This relationship shows a direct cause and effect between a human action and a selection for a particular yeast trait. Other phenotypes, however, are less clear in their causal relationship. For instance, "wild" yeast strains isolated from oak were found to produce wines with less desirable aromas when compared to wine associated yeast strains (Hyma *et al.*, 2011). This does not necessarily imply that humans have selected for desirable aromas from wine yeast; in fact, it is equally likely that we have come to associate certain yeast characteristics as desirable because we associate them with a particular wine region.

Life history has played an important role in the domestication and selection of yeast strains. Beer and bread yeast are used in a dramatically different manner than wine yeast, and this has consequences for certain traits. Beer and bread are largely a perennial endeavour, and yeast is transferred from batch to batch by using a small portion of the prior batch as an inoculum for the next batch – this is termed "back-slopping". By back-slopping, yeast are maintained almost constantly in a nutrient-rich medium with

moderate stresses from ethanol and changes in osmotic pressure. Sexual reproduction declines significantly; 44.4% of the primary beer clade of yeast strains are obligate asexual (Gallone *et al.*, 2016). Wine strains, on the other hand, spend much of the year in low nutrient conditions and then are annually exposed to the high stresses of wine production. To facilitate this cycle, wine strains demonstrate far greater long-term survival rates as well as increased sexual reproduction (Gallone *et al.*, 2016). A study of yeast populations in Canadian vineyards and wineries found surprisingly low persistence of strains from year to year, and identified introduced commercial yeast as a major driver of yeast population in the modern winemaking environment (Martiniuk *et al.*, 2016). For industrial beverage production, yeast cells are typically added in the "active dry yeast" form; these yeast have been cultured to high cell number, then carefully dried to preserve cell viability. The drying process can result in oxidative stress to the cells, as well as the physical stress of desiccation (França, Panek and Eleutherio, 2007). Not all yeast strains are suited to the production of active dry yeast, and this added selection step has altered the gene pool of commercial yeast (Reed and Chen, 1978).

Commercialized strains of *S. cerevisiae*, while numerous, represent a very small portion of the overall genetic diversity found in nature. Many "different" commercially available strains are highly inbred or even genetically equivalent (Borneman *et al.*, 2016). This may be due to repeated isolation of the same strain from the vineyard or winery, or more likely due to the licensing of the same strain to multiple commercial entities. As the desirable qualities for a wine yeast are common across most wine applications (rapid fermentation, desiccation tolerance, sulphite resistance, *etc.*), it makes sense that the selected isolates share many genetic similarities (Verstrepen, Chambers and Pretorius, 2006).

1.2.3 S288C as a reference strain

1.2.3.1 Origin of S288C.

The *S. cerevisiae* reference strain, S288C, was purpose-bred by Robert Mortimer for use in laboratory experiments. S288C is a haploid *gal2* mutant with a reasonably well documented genealogy. The primary contributor to this strain is thought to be a yeast designated as EM93, isolated from rotting figs

near Merced, California by Emil Mrak in the 1930s. The genotype of this isolate was determined to be *MAT*a/*MAT*α *SUC2/SUC2 GAL2/gal2 MAL/MAL mel/mel CUP1/cup1 FLO1/flo1*, and it is thought that the *gal2* and *flo1* of S288C are contributed from this parent, leading to a non-galactose fermenting, non-flocculent phenotype (Mortimer and Johnston, 1986). The heterothallic nature of the parent EM93 isolate made it an ideal candidate for genetic crosses, and this isolate was used extensively by Carl Lindegren is his early genetic work (Engel *et al.*, 2014). A collegial atmosphere among yeast geneticists during the 50s and 60s led to the sharing of yeast strains among researchers. Through Cornelius Tobias, Mortimer obtained a derivative of EM93 that became one of the direct parents of S288C. EM93 is estimated to contribute 88% of the genetic material of S288C, with the rest being contributed by other laboratory strains popular at the time (see Mortimer and Johnston (1986) for a full review). Mortimer deliberately bred S288C to be non-flocculent, and to have minimal nutritional requirements, making it ideal for biochemical studies in the laboratory. The crossing experiments that led to the isolation of S288C also resulted in the creation of the first genetic maps, starting with Lindegren in 1949 (Figure 1.2). This work was built upon with the first comprehensive genetic map by Mortimer and Schild (1980), who argued for the existence of 17 chromosomes, although by 1992 only 16 chromosomes had been determined (Mortimer, Contopoulou and King, 1992). S288C was distributed widely, and became a natural candidate for the yeast genome project of the early 90s. The original reference genome published in 1996 is based solely on the strain S288C, and its direct progeny, AB972 and FY1679 (Goffeau *et al.*, 1996).

1.2.3.2 S288C is not a representative genome.

While S288C may have been a strong candidate for the first complete yeast genome due to its ubiquity in laboratory use, recent lines of evidence indicate that this strain is not an ideal representation of the *S. cerevisiae* species. Whole genome analysis of many *S. cerevisiae* strains and isolates cluster S288C most closely with wine yeast, which makes sense given the fruit-based origin of the S288C parental isolate (Borneman *et al.*, 2008). Mortimer (2000) has hypothesized that the original isolate was actually transported to the fig from a nearby vineyard by wasps, making EM93 a “wine-related” yeast (Mortimer, 2000). The selection and crossing process by which S288C was isolated, however, has resulted in a

number of unique features for its genome. Wide-scale genomic comparison has revealed that S288C represents a conserved, core genome with little in the way of unique genetic material, although it does

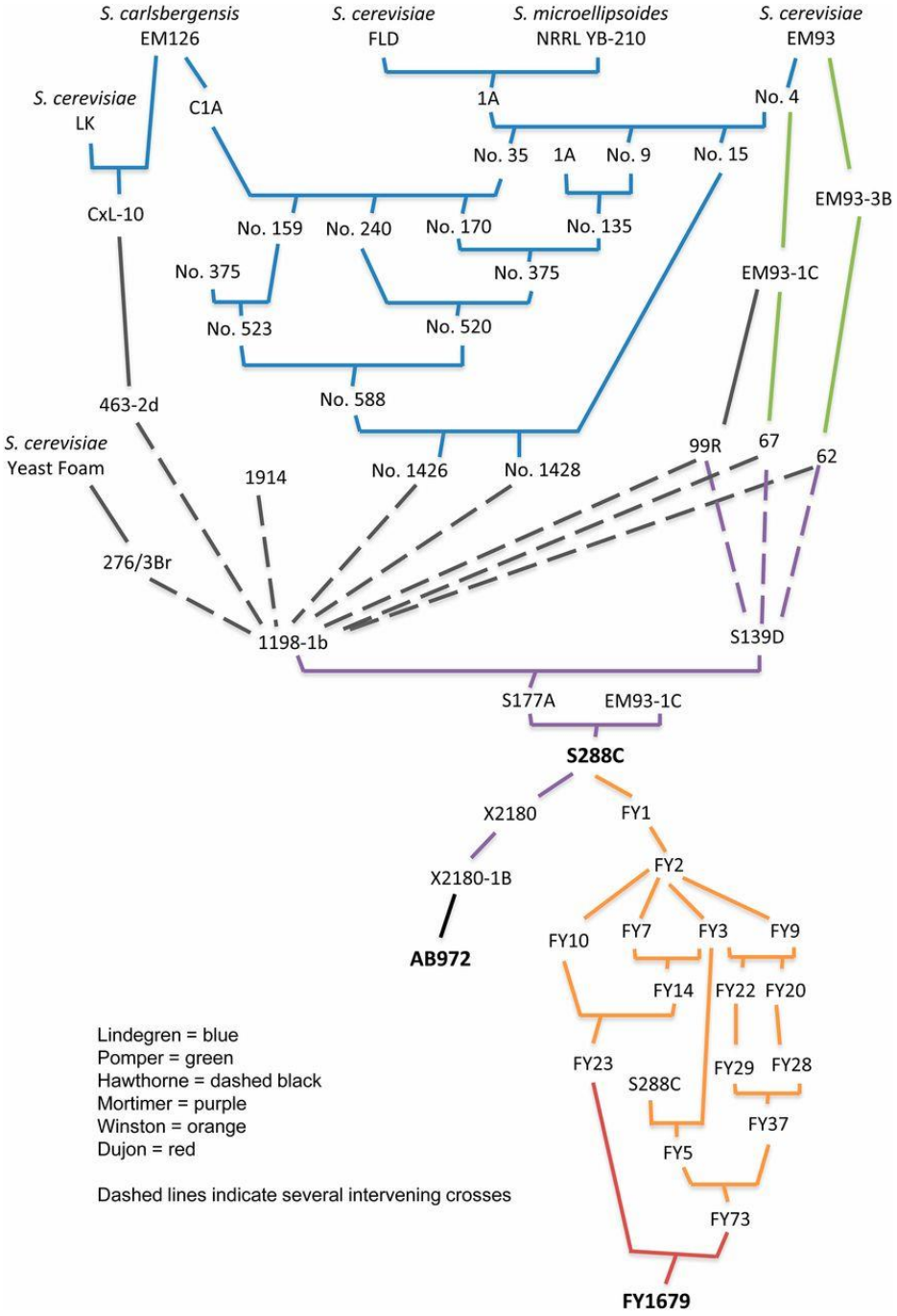


Figure 1.2 The lineage of the S. cerevisiae reference strain S288C. The complex genealogy and contributing scientists that led to the breeding of the S288C strain, and some of its derivatives. Image is reproduced from Engel et al. (2014) under a Creative Commons license.

contain an unusually high number of Ty elements (Borneman *et al.*, 2008). On the other hand, over 200 kb of DNA has been found from natural or industrial yeast isolates that is not present in the S288C genome (Borneman and Pretorius, 2015). Mortimer (2000) has noted that "natural wine yeast have tremendous diversity and about two-thirds [of natural yeast strains] has one or more heterozygosities". Among other commonly used laboratory strains such as W303, up to 0.36% nucleotide divergence was found from the S288C reference strain (Schacherer *et al.*, 2007). Efforts are being made to address these inadequacies of the S288C reference strain, including the incorporation of the current genomic knowledge of as many isolates as possible into a pan-genome that reflects the range of possible genetic elements in *S. cerevisiae* and not just a conserved core (Engel *et al.*, 2014). These efforts should aid in the annotation of the non-reference genetic material, and the understanding of the role of genomic variability among strains and isolates.

1.3 Annotation of the genome

1.3.1 A significant portion of the yeast genome remains unannotated

Despite being over two decades since the publication of the complete *S. cerevisiae* genome, 791 ORFs of the 6691 identified ORFs on the SGD remain uncharacterized. A small portion of these (55) are "dubious" ORFs, but this still indicates that there is a sizeable portion of the *S. cerevisiae* genome that remains uncharacterized. A great number of strategies have been employed over the years to annotate the genome, including large-scale genome-wide screens as well as more targeted approaches. The yeast deletion collection (discussed in greater detail in the following section) was successful in identifying growth phenotypes for many homozygous and heterozygous deletion mutants, but is limited to the S288C genetic background (Giaever *et al.*, 2002). More recent expansion of the number of growth conditions tested has greatly increased the number of annotated phenotypes from the yeast deletion collection (Hillenmeyer *et al.*, 2008). Large-scale protein tagging has been employed to tie sub-cellular localization of the encoded protein with the biological function of each ORF, with some success (Huh *et al.*, 2003). Researchers have also leveraged tandem affinity purification and mass spectrometry to conduct large-

scale screens of protein-protein interactions, and have identified over 7000 protein-protein interactions and over 500 protein complexes (Gavin *et al.*, 2006; Krogan *et al.*, 2006). A great number of ORFs are also assigned "putative functions" based on sequence similarity to or co-localization with well annotated ORFs. Current annotation strategies will be discussed, along with some of the shortcomings that have led to an incomplete annotation of the genome.

1.3.2 The yeast deletion collection

One of the many resources developed to aid the annotation of the *S. cerevisiae* genome is the yeast deletion collection: a strain library in which most of the annotated ORFs of the *S. cerevisiae* genome have been systematically deleted. The library construction was accomplished by replacing the ORF of interest with a *kanMX* antibiotic resistance gene flanked by two 20-nucleotide "barcodes" that uniquely identify each strain. This construction allows the entire pool of deletion strains to be cultured simultaneously, and enriched or depleted strains can be identified by comparing the unique barcodes from the starting culture to the final culture (Giaever *et al.*, 2002). Homozygous haploid and diploid deletion strains were created for non-essential genes, while heterozygous deletion strains were created for both non-essential and essential genes. The yeast deletion collection encompasses approximately 96% of the reference genome, with some ORFs not able to be deleted due to high similarity with other sequences in the genome making it difficult to create appropriate primers (Giaever *et al.*, 2002). A second collection has also been made with double gene deletions, using the *natMX* antibiotic resistance marker to eliminate a second gene; this extensive collection of over 23 million double mutants was recently used to map the "interactome" of *S. cerevisiae*, identifying positive and negative interactions of double gene deletion strains (Costanzo *et al.*, 2016).

Using rich growth medium at the standard growth temperature of 30 °C, only 15% of homozygous diploid deletion mutants result in a slow growth phenotype compared to wild type (Giaever *et al.*, 2002). Many of these deletions only yield a phenotype under specific conditions, and a comprehensive study that systematically exposed the deletion collection to 1144 chemical compounds identified a growth

phenotype for 97% of the yeast deletion collection (Hillenmeyer *et al.*, 2008). Another study applied continuous culture conditions in bioreactors to the yeast deletion collection to specifically look for genes involved in the High-Osmolarity Glycerol (HOG) pathway (Gonzalez *et al.*, 2016). The continuous culture approach has been used to mimic the early conditions of a wine fermentation and identify genes that are important early in the wine fermentation process (Novo *et al.*, 2013). The yeast deletion collection has also been used in a simulated wine fermentation, and fermentation phenotypes were identified for 139 previously uncharacterized genes (Piggott *et al.*, 2011). A complementary approach to the yeast deletion collection has also been taken; high-copy plasmids with unique "barcodes" have been used to individually overexpress each ORF of the reference genome (Ho *et al.*, 2009). This plasmid library has the advantage of being able to be transformed into multiple different *S. cerevisiae* strain backgrounds to understand interactions between the overexpression of certain genes and the genetic background. The high-copy plasmid library was used to identify key genes involved in ethanol tolerance (Anderson *et al.*, 2012).

1.3.3 Transcriptomic approaches to annotate genes

Whole genome transcription studies, termed "transcriptomics", analyze the average total transcribed RNA content of a cell in a given condition at a given point in time. Serial analysis of gene expression (SAGE) was the first technology to effectively allow the identification of transcript abundance for thousands of transcripts simultaneously (Velculescu *et al.*, 1995). SAGE technology uses biotinylated cDNA reverse synthesized from isolated mRNA. The cDNA is cleaved by a restriction enzyme and bound to a streptavidin bead, leaving the restriction site exposed, at which point the cDNA pool is divided in two. Both pools are tagged at the restriction site with one of two linkers, and a tagging enzyme is used to cleave the linker with a small fragment of cDNA from the beads. The pools are combined and ligated, which creates short fragments with known linker sequences on either end which can be amplified by PCR. These fragments are then amplified, cloned into a plasmid vector, and manually sequenced. The resulting sequences can be matched back to genomic data, and the clone frequency can be related to

transcript abundance (Velculescu *et al.*, 1995). SAGE was used to produce the first transcriptome of *S. cerevisiae* shortly after the original genome sequence was released. The first transcriptome produced using SAGE identified 4665 transcribed genes when cells were growing logarithmically in rich media, and was able to provide a picture of relative gene expression level (Velculescu *et al.*, 1997). SAGE technology was also employed to analyze the dynamics of the yeast transcriptome during wine fermentation (Varela *et al.*, 2005). With the release of the S288C reference genome, a microarray technology was pioneered using 260,000 specific 25 base pair probes covering all predicted ORFs from the genome and was first used to characterize the diauxic shift from fermentation to respiration (DeRisi, Iyer and Brown, 1997; Wodicka *et al.*, 1997). The basic premise of DNA microarray chips for the analysis of gene expression is the cross-linking of unique probes for every known ORF for a given species to a small chip, which is then bound with fluorescently labelled cDNA samples. The chip is scanned and the fluorescence of a given probe corresponds to the expression of that transcript (Lashkari *et al.*, 1997). Microarray technology has been used extensively to map the transcriptome of *S. cerevisiae* and many studies have focused specifically on wine conditions or wine-related stresses (Alexandre *et al.*, 2001; Brem, 2002; Marks *et al.*, 2008; Rossouw and Bauer, 2009; Rossouw *et al.*, 2009; Bessonov *et al.*, 2013; Barros de Souza *et al.*, 2016). This technology has the advantage of being relatively comparable across experiments when the same microarray chip technology is used. Later manifestations of microarray technologies feature improved probesets in terms of coverage and specificity, as well as improved means of normalization among probes and between chips.

Microarray technology is useful when the same species and strain is being used, as the current gold standard GeneChip Yeast Genome 2.0 chip probeset is based on the 5,845 genes published in the current S288C reference genome available from SGD (Affymetrix, Santa Clara, California). The probeset may not reflect unique ORFs in a particular strain, and completely new chips would have to be designed for new species. Because of these reasons, and also because of the falling cost of sequencing technologies, the field of transcriptomics has begun to move away from microarray based technologies towards sequencing-based technologies. The advent of high-throughput RNA sequencing (RNA-seq)

technologies has been the next step in transcriptomics. RNA-seq can be used without a reference genome and hence eliminates the need to design hybridizing probes, is highly specific and does not saturate easily. Multiple samples can easily be multiplexed into a single run, allowing a massive amount of data to be generated efficiently (Wang, Gerstein and Snyder, 2009). RNA-seq technology has successfully been applied to study the transcriptional response of *S. cerevisiae* fermenting white wine at low temperature (Deed, Deed and Gardner, 2015). RNA-seq technology is more important, however, when used to analyze the transcriptional response of multiple non-reference strains that may contain non-reference transcripts (Treu, Campanaro, *et al.*, 2014; Treu, Toniolo, *et al.*, 2014). RNA-seq technology is also easily applied to species without a robust reference genome (Lertwattanasakul *et al.*, 2015) and experiments where multiple species are cultured in the same experiment (Tronchoni *et al.*, 2017).

Although high-throughput gene expression analysis can help to predict gene function, the differential expression (DE) of a transcript does not necessarily imply that the deletion of the gene will produce the expected phenotype. For example, when four different stress conditions were tested, less than 7% of genes with upregulated transcripts showed any growth phenotype when deleted (Giaever *et al.*, 2002). The lack of phenotype could be due to subtle changes not detectable by the growth assay, or due to genetic redundancy because another gene is compensating for the gene deletion (Gu *et al.*, 2003). While many papers have analyzed the transcriptional response to particular growth conditions and yeast stresses, only a handful have gone on to characterize the function of poorly annotated genes with follow-up experiments. Genes identified as being components of the "fermentation stress response" were found to have roles in the regulation of glycogen levels and the regulation of acetate production during fermentation (Marks *et al.*, 2008; Walkey *et al.*, 2011, 2012).

1.3.4 Proteomic approaches to annotate genes

The field of proteomics analyzes the average total protein content of a cell at a given point of time in a given condition. Individually, protein quantities are often measured by Western blotting, but other techniques exist that attempt to analyze all proteins at once. Early proteomic work used 2-dimensional

polyacrylamide gel electrophoresis (2D-PAGE) to separate protein extracts, followed by the identification of proteins by some combination of excision of spots and direct protein sequencing, the excision of spots and analysis by protein digestion and mass spectrometry, and matching of gel spots to reference protein gels (Trabalzini *et al.*, 2003). These early methods were labour intensive and had difficulty in unambiguously identifying proteins due to overlapping spots. Furthermore, the selective identification of proteins yields datasets that only reflect a fraction of the overall yeast proteome and favours the more abundant proteins (Gygi *et al.*, 1999; de Groot *et al.*, 2007; Rossignol *et al.*, 2009).

Isobaric tagging for relative and absolute quantification (iTRAQ) is a method that was developed in late 2004 and has been used extensively in proteomics experiments since (Ross *et al.*, 2004). The workflow begins with the extraction of total protein from a sample, followed by trypsin digestion, the tagging of fragments with 4-plex or 8-plex isobaric tags "such that all derivatized peptides are isobaric and chromatographically indistinguishable, but yield signature or reporter ions following [collision-induced dissociation] that can be used to identify and quantify individual members of the multiplex set" (Ross *et al.*, 2004), the pooling of samples into 4- or 8-plex runs, a peptide fractionation step by HPLC and finally online MS/MS detection of the peptides (Noirel *et al.*, 2011) (Figure 1.3). This method provides accurate and precise relative quantification of proteins across a number of samples, although there is an inherent trend to minimize the magnitude of differences between samples (Evans *et al.*, 2012). While iTRAQ is the method used most often in multi-sample proteomics experiments, label-free methods are currently gaining popularity (Evans *et al.*, 2012).

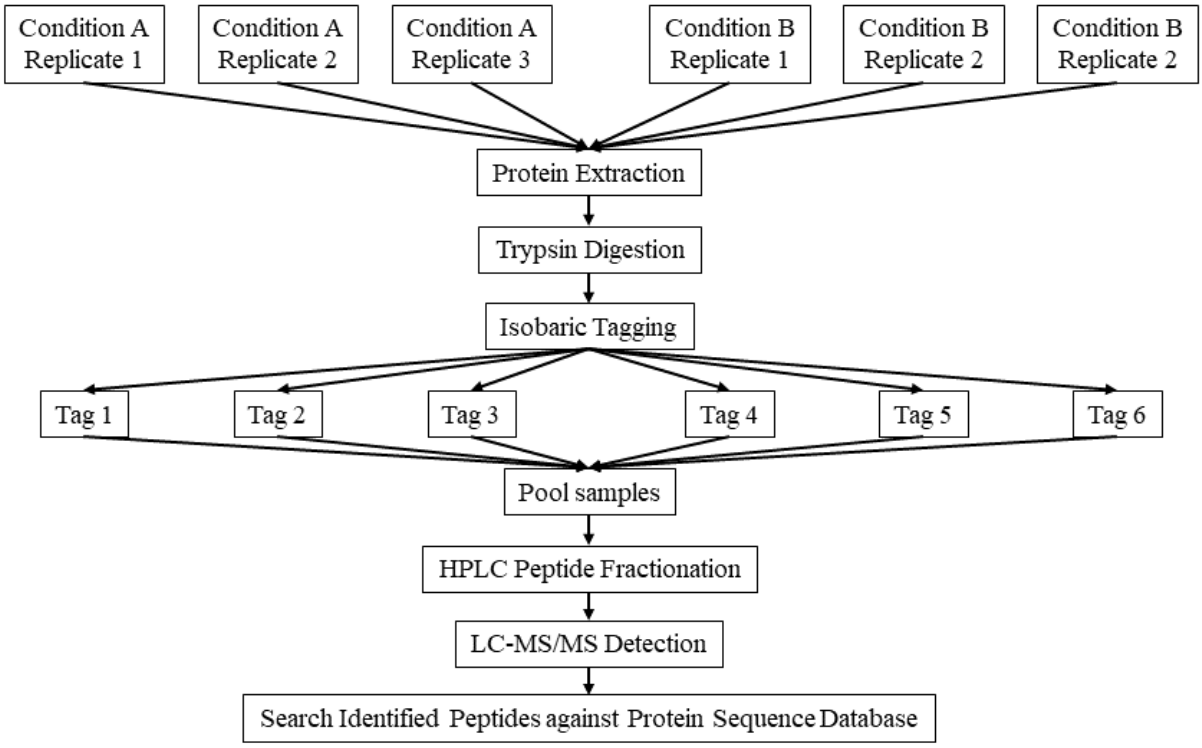


Figure 1.3 General workflow for iTRAQ analysis. A standard iTRAQ experiment for relative proteomic quantification can multiplex up to eight different samples in the same experimental run using unique tags.

Various methods have been used to characterize the yeast proteome during fermentation, or during fermentation-relevant stresses. The 2D-PAGE method was used to characterize the proteomic response of a wine yeast to fermentation stresses, and found effects of stress-induced autoproteolysis on glycolytic enzymes (Trabalzini *et al.*, 2003). The proteome of commercial wine yeast has been monitored by 2D-PAGE and by iTRAQ methods during a simulated wine fermentation using synthetic must (Rossignol *et al.*, 2009; Rossouw *et al.*, 2010). The iTRAQ method has also been successfully applied to *Acetobacter* to characterize the proteomic response to elevated acetic acid levels during rice vinegar fermentation (Xia *et al.*, 2016).

1.3.5 Agreement between the proteome and the transcriptome

With a number of transcriptomic and proteomic studies of yeast during the wine fermentation process, it is possible to investigate the correlation of the transcriptome and proteome. One would expect a high degree of correlation as mRNA transcripts are directly responsible for the translation of cellular

proteins, but a number of factors come into play that may affect the assessed transcriptome and proteome. Transcript stability is variable and can affect the duration of the transcript to be available for translation. Translation efficiency may be affected by the availability of amino acids, tRNAs, ribosomal subunits and sequestration of mRNAs in p-bodies. Post-translational modification of proteins, such as ubiquitination, can also have a substantial impact on the duration of the protein in the cell.

While the methods and precise results vary, multiple studies have concluded that mRNA abundance (the transcriptome) is a poor predictor of the proteome (Gygi *et al.*, 1999; de Groot *et al.*, 2007; Rossignol *et al.*, 2009; Rossouw *et al.*, 2010). These results point towards a major contribution from post-transcriptional mechanisms in shaping the yeast proteome and, inevitably, the cellular response to fermentation stress. Recent advances have pioneered techniques that allow the monitoring of mRNA translation by sequencing ribosome-protected mRNA fragments (Ingolia *et al.*, 2009; Brar *et al.*, 2012). Ribosome-associated noncoding RNAs have also been shown to bind to ribosomes in a stress-dependent manner and alter protein synthesis (Bąkowska-Żywicka, Kasprzyk and Twardowski, 2016). Future studies with a correlational analysis of the transcriptome and proteome will continue to provide insights into possible mechanisms and points of cellular regulation during stress responses.

1.3.6 Genotypic approaches to gene annotation

The underlying purpose of much of the study of yeast genomics is to unravel the genetic underpinning of phenotypic traits. Sequencing technologies have improved in speed and throughput, and have drastically reduced in cost. Consequently, there is now a large number of *S. cerevisiae* strains with genome sequences available. The difficulty is now reliably relating the genetic features with a predictable gene function. The S288C reference strain of *S. cerevisiae* has already been extensively studied, therefore recent studies have focused on the comparative genomics of industrial isolates or commercial strains of yeast.

Unfortunately for the advancement of annotation of genes not in the S288C genome, many comparative genomics studies rely on the alignment of newly sequenced yeast to the S288C reference

genome. For instance, a comprehensive genetic, transcriptomic and proteomic study of cold-tolerant and cold-sensitive yeast during low temperature wine fermentation attributed changes in low-temperature tolerance to non-synonymous SNPs in the *MET28* and *MET4* genes that encode two sulphur metabolism related transcription factors. The methods used, however, completely eliminated the possibility of identifying novel genes, as the transcriptomic and proteomic analysis relied on the S288C reference genome, and the genomic comparison only analyzed SNPs, indels and copy number variations (CNVs) for the regions aligned to the S288C reference genome (García-Ríos, López-Malo and Guillamón, 2014). A more recent study by the same group, however, used quantitative trait loci (QTL) mapping to investigate low-temperature wine fermentation and identified the highly plastic subtelomeric regions as important contributors to the tolerance phenotype (García-Ríos *et al.*, 2017). García-Ríos *et al.* (2017) further identified the need for long-read sequencing technologies to be applied to more yeast genomes in order to resolve the subtelomeric regions, which often cannot be correctly aligned with high-throughput short read sequencing, as these regions are common sites of gene duplication and evolution (Brown, Murray and Verstrepen, 2010).

A simple exploratory study with *Kluyveromyces marxianus* sought to relate the genome to the transcriptome in four different growth conditions. While the study did not go into great depth on specific gene function, this approach represents a perfect starting point for identifying candidate genes required for particular stress tolerance mechanisms (Lertwattanasakul *et al.*, 2015). For example, this study identified several transcripts with high similarity to known xylose transporters from *Scheffersomyces stipitis*, but which showed no induction specific to growth on xylose media. Conversely, the study also identified putative transporters (based on similarity) that were induced under xylose conditions in *K. marxianus*, illustrating the complementation of gene homology and transcriptomic evidence in annotating genes (Lertwattanasakul *et al.*, 2015).

1.4 Choice of yeast strain for study

1.4.1 Industrial relevance

When selecting a strain, or strains, of *S. cerevisiae* for research it is important to consider how the results will be applied. Industrial isolates have been shown to contain genetic material not present in laboratory strains, which is an important consideration when designing an experiment (Borneman *et al.*, 2011). Researchers are beginning to see the value of using multiple strains, or industrially specific strains, particularly when investigating traits such as stress response and fermentation, which are directly related to industry (Ivorra, Pérez-ortín and del Olmo, 1999; Dunn, Levine and Sherlock, 2005). The importance of strain background is highlighted by the EBY.VW4000 strain of *S. cerevisiae*, which was constructed to test the function of hexose transporters as all of the native transporters have been deleted from this strain (Wieczorke *et al.*, 1999). Whole genome sequencing has revealed that this strain has had widespread genome rearrangements due to the multiple *LoxP* scars created during the deletion strategy, resulting in the loss of some genetic information and the translocation of other elements (Solis-Escalante *et al.*, 2015). The EBY.VW400 genetic background, derived in part from S288C and in part from the CEN.PK2 lab strain, is vastly different from industrial yeast isolates, many of which contain *HXT* gene sequence variation and CNV (Wieczorke *et al.*, 1999; Dunn, Levine and Sherlock, 2005). Similarly, the reference strain S288C is a poor representation of industrial isolates and not suitable for all types of study (Engel *et al.*, 2014). Therefore, multiple lines of evidence point towards the need to select yeast strains with the appropriate industrial background when conducting research on industrially-relevant characteristics.

1.4.2 Ease of use for laboratory methods

While industrial isolates and commercial yeast strains are abundant, many of them are not well suited to standard laboratory methods. Many strains are aneuploid, heterozygous or have poor sporulation efficiency. Natural isolates generally do not possess nutritional auxotrophies that can be used as markers, so antibiotics must be used for selection, although some auxotrophic industrial isolates have been

obtained for breeding programs (Fernández-González, Úbeda and Briones, 2015). Fortunately, many industrial isolates have been screened for use in laboratory settings. Wine yeasts have been identified that are homozygous diploids with good sporulation rates, making them amenable to sequencing and genetic manipulation (Bradbury *et al.*, 2006; Fernández-González, Úbeda and Briones, 2015). In some cases, protocols such as the standard lithium acetate transformation have been slightly modified to improve the results when not using standard strains (Walkey *et al.*, 2011). It is important to consider possible limitations and compatibility with standard laboratory methods when selecting industrial isolates for further study.

1.4.3 Enoferm M2

Enoferm M2 is a commercial wine yeast, suitable for the production of red and white wine, originally isolated from South Africa and commercialized by Lallemand Inc. (for more background information, see Materials and Methods). Whole genome sequencing has revealed that this strain is likely isogenic with AWRI796 and closely related to several other commercial strains (Borneman, Pretorius and Chambers, 2013; Borneman *et al.*, 2016). Enoferm M2 was identified as an ideal candidate for laboratory study of wine yeast from a screen of 45 different commercial yeast (Bradbury *et al.*, 2006). This strain possesses a number of qualities that make it amenable to laboratory work, such as good sporulation and homothallism, and consequently it has been used extensively in a number of published reports (Deed, Van Vuuren and Gardner, 2011; Walkey *et al.*, 2011, 2012; Anderson *et al.*, 2012; Bessonov *et al.*, 2013; Huang, Roncoroni and Gardner, 2014; Deed, Deed and Gardner, 2015). Specifically, Enoferm M2 has been used in transcriptomic studies to investigate the effects of low temperature fermentation (Deed, Van Vuuren and Gardner, 2011) and general wine fermentation (Bessonov *et al.*, 2013). Enoferm M2 has also been used as the genetic background to express a high copy plasmid library and identify genes that aid in ethanol resistance (Anderson *et al.*, 2012). Finally, Enoferm M2 has also been used as the genetic background for the investigation of genes related to the regulation of nutrient uptake and secondary metabolite production during wine fermentation (Walkey *et al.*, 2011, 2012; Huang, Roncoroni and

Gardner, 2014; Deed, Deed and Gardner, 2015). Overall, Enoferm M2 is a commercial, industrially-relevant strain with established protocols and history of use in genetic and genomic studies.

1.5 Choice of conditions for gene annotation

Once the choice of an appropriate yeast strain for pursuing gene annotation has been decided, the next step is to identify the culture conditions and sampling strategy, both of which can have profound effects on the results.

1.5.1 Synthetic vs. natural media

The vast majority of yeast laboratory studies use defined growth media, such that each component is known and can be easily replicated. Defined laboratory growth media are a model, or simplification of the complex growth media that yeasts grow on in natural or industrial environments. Synthetic wine media are necessary for the study of particular aspects of wine fermentation, such as nitrogen assimilation, where the goal is to determine the effect of media on the yeast (Varela, Pizarro and Agosin, 2004). While time and effort has gone into analyzing complex media, such as grape juice, to be able to replicate its components with reasonable accuracy, it is impractical to include every component that would be found in the original complex mixture. MS300 is a commonly used synthetic wine medium, which contains 20% glucose, trace minerals and 300 mg/L of assimilable nitrogen from various sources (Bely, Sablayrolles and Barre, 1990; Riou *et al.*, 1997; Salmon and Barre, 1998; Rossignol *et al.*, 2003; Beltran *et al.*, 2008). This media traditionally uses only glucose as a carbon source, when natural grape must is typically a 50-50 mixture of glucose and fructose; since fructose is metabolized slower than glucose and affects fermentation dynamics, this simple problem has been rectified in some more recent studies (Varela, Pizarro and Agosin, 2004; Varela *et al.*, 2005; Novo *et al.*, 2009). Furthermore, the MS300 synthetic media is often modified for each specific study (lipid composition, nitrogen sources, *etc.*), which reduces our ability to compare results between studies.

While synthetic wine media is touted by many researchers as being practically equivalent to natural grape must, some studies have found discrepancies in gene expression levels between synthetic media and

natural grape must (Riou *et al.*, 1997), while others have found comparable results (Puig and Pérez-Ortín, 2000; Rossouw and Bauer, 2009). The researcher is then faced with a trade-off between using synthetic media and being able to easily replicate an experiment without relying on a particular grape must, or using natural grape must and accurately depicting a commercial wine fermentation. Natural grape must has been used successfully in transcriptomic studies of wine yeast (Marks *et al.*, 2008; Rossouw and Bauer, 2009), while to date no proteomic studies of wine yeast have used natural grape must despite no technical limitations restricting the use of natural grape must for proteomic studies.

1.5.2 Time point analysis

Sampling strategy can have enormous implications for the results of transcriptomic and proteomic studies. While the transcriptome of yeast contains nearly 6000 different transcripts, and the theoretical proteome is slightly smaller, in practice not all of these transcripts and proteins will be detected at a single sampling point in a single growth condition. Sampling at multiple time points of a fermentation has shown us that the *S. cerevisiae* transcriptional and proteomic responses are highly dynamic (Marks *et al.*, 2008; Rossouw *et al.*, 2010). Patterns of differential gene expression are strain dependent, and consequently a single time point "snapshot" may not accurately capture important differences between strains or mutants (Rossouw *et al.*, 2009). Time points selected for analysis may correspond to the growth phases of the yeast (lag phase, early exponential growth, late exponential growth, stationary phase) (Varela *et al.*, 2005; Rossouw *et al.*, 2010), or to changes in the growth medium caused by the fermentation process (depletion of nutrients, increase in ethanol) (Marks *et al.*, 2008).

1.5.3 Continuous culture vs. batch culture

Continuous culture experiments use a bioreactor to continuously maintain culture conditions at a steady state. A steady culture state is maintained by monitoring multiple aspects of the culture, such as biomass, metabolites and pH, and using this information to dilute the culture and feed it with fresh stocks. Continuous cultures are often used in competition experiments, using the barcoded yeast deletion collection, as this method allows multiple generations of yeast to be grown under the same conditions

(Novo *et al.*, 2013). Continuous cultures have also been applied to transcriptomic and proteomic work in yeast (García-Ríos, López-Malo and Guillamón, 2014). In this case, the goal was to compare the transcriptome of different yeast strains growing in the exact same conditions, and by achieving a steady state, a single time point can be sampled for all cultures.

A typical wine fermentation is a good example of a batch culture: the growth medium is inoculated with yeast and the fermentation is allowed to proceed without intervention. The growth medium is vastly different at the endpoint of the culture compared to the starting point, and the yeast goes through several different growth phases. During batch culture the proteome and transcriptome is highly dynamic, and some difficulties may arise when comparing strains or mutants with different growth rates (Regenberg *et al.*, 2006). With proper sampling strategies, however, batch cultures have been used successfully in transcriptomic analyses of wine and simulated wine fermentations (Marks *et al.*, 2008; Rossouw *et al.*, 2009, 2010). Comparative studies have shown that yeast cells behave quite differently whether they are grown in continuous or batch cultures (Ding *et al.*, 2009). It is important, therefore, to consider the biological question at hand when designing an experiment and select the appropriate culture conditions. When attempting to replicate the stresses of an industrial setting such as wine fermentation, a batch culture is more accurate in depicting industrial conditions.

1.6 Opportunities for strain development

An expanded knowledge of yeast genomes and their corresponding phenotypes opens up many avenues for the development of new strains that are better suited to industrial tasks. There are many ways to exploit this new knowledge, each with their advantages and drawbacks.

1.6.1 New strains from wild isolates

Current genome sequencing projects have attempted to provide a catalogue of *S. cerevisiae* diversity on a genomic level (Borneman *et al.*, 2016; Gallone *et al.*, 2016). Due to convenience and industrial relevance, most isolates studied in these large-scale genome sequencing endeavours are of wine or beer origin. Industrial fermentation strains of *S. cerevisiae* appear to be highly inbred and only represent a

small portion of the overall species diversity as indicated by wild isolates (Liti *et al.*, 2009; Steensels *et al.*, 2014; Borneman *et al.*, 2016). Wild strains – hereby defined as strains isolated from non-human related sources – of *S. cerevisiae* can be isolated from a variety of fruits, tree surfaces (mainly oak), soil samples and the guts of insects (Hyma *et al.*, 2011). Researchers have used mixed cultures of vineyard isolates to produce high quality wine (Terrell, Cliff and Van Vuuren, 2015). Researchers have also used isolates from oak and soil to ferment wine, which can be readily distinguished from wine made with classic wine strains, although the sensory qualities were not ideal (Hyma *et al.*, 2011). Efforts have also been made to identify strains that have appropriate qualities, such as good sporulation and spore viability, to make the strains amenable to breeding programs (Fernández-González, Úbeda and Briones, 2015). Wild yeast isolates are unlikely to be ideally suited to industrial environments, but they represent a pool of genetic diversity that may be able to aid in the identification of particular phenotypes and provide expanded genetic stock for breeding programs.

1.6.2 Directed evolution

Another approach to develop improved yeast strains for industrial applications has been the use of directed evolution. With this approach, strains are continuously grown for multiple generations under selective conditions, with the expectation that spontaneous natural mutations will occur and be selected for if they confer an advantage in the given condition. In some cases, a mutagenizing agent may be employed to increase the natural rate of mutation, although this has the trade-off of also introducing deleterious mutations into the population (Steensels *et al.*, 2014). To compensate for potential deleterious mutations, the "evolved" strain is often backcrossed to the parent several times to maintain the desired phenotype but eliminate undesirable changes in the genome. This technique is useful when the desired outcome is improved stress tolerance, but is more difficult to apply when the desired phenotype is not easily selectable (for a particular secondary metabolite, for example). Directed evolution has been applied with great success using both chemostat (continuous) and batch cultures, and applying a variety of different stressors, with evolved strains performing up to 1000 times better than the parent strain in terms

of stress resistance (Çakar *et al.*, 2005). Directed evolution provides a means to change the genome and introduce potentially novel coding sequences, without the yeast being considered a genetically modified organism (GMO).

1.6.3 Hybridization

Hybridization of yeast strains, or even of yeast species, presents an opportunity to combine several desirable phenotypes into a single strain of yeast. While the simplest form of hybridization involves a mating of two spores of *S. cerevisiae* of opposite mating types (direct mating), a number of techniques exist to allow the hybridization of strains and species that may not hybridize naturally or easily. For strains with low sporulation and reduced capacity for sexual mating, techniques such as protoplast fusion and cytoduction can be used to overcome reproductive barriers (Pretorius, 2000; Steensels *et al.*, 2014). Homothallic strains present a challenge for hybridization as well, but techniques have been developed to facilitate the mating of homothallic strains by the selection for inherent traits that would only be present in progeny of both mating strains (Ramírez, Peréz and Regodón, 1998). Alternatively, selectable plasmids can be inserted into the parental strains and selected for when mating, and the plasmids can be removed later from the progeny, although this method is more likely to run into regulatory issues regarding genetic modification in some jurisdictions (Da Silva *et al.*, 2015).

Wine yeast have been deliberately improved for decades using classic breeding programs and direct mating (Romano *et al.*, 1985). Multiple studies have identified strains of *S. cerevisiae* that possess technological properties (rapid fermentation, good sporulation, high ethanol tolerance, diploid, homozygous) that make them suitable breeding stock for commercial yeast strains (Bradbury *et al.*, 2006; Fernández-González, Úbeda and Briones, 2015). With a stock of strains with clearly identified desirable technical properties as a base, novel phenotypes can be bred into commercial strains by mating; this has been done successfully to introduce novel yeast phenotypes that benefit the wine fermentation process (Da Silva *et al.*, 2015). When a single trait is to be introduced from one strain to another, a backcross

approach to the parental strain can be taken to minimize the genetic contribution of the donor strain to just that needed for the desired phenotype (Marullo *et al.*, 2009).

1.6.4 Genome editing

The most straightforward and perhaps most controversial technique for the development of novel yeast strains is direct genome editing. Genetic modification of yeast generally comes in two forms: the expression of genes on a circular plasmid, or the integration/deletion a of portion of yeast DNA. Both methods have advantages and disadvantages. Plasmids do not disrupt the natural DNA sequence of a cell, whereas direct integration may have some effect depending on the choice of integration site. Plasmids may be present in one to many copies, depending on the plasmid type and selection method, whereas direct integration is generally only one (heterozygous) or two (homozygous) copies. Once integrated, direct integration events are generally stable, whereas plasmids typically require continuous selective pressure by antibiotics or auxotrophies to avoid being lost (Steensels *et al.*, 2014). Gene deletion cannot be accomplished with plasmids and requires direct editing of the genome. Typically, a selectable marker is inserted in place of the gene of interest using homologous overlapping DNA segments. Techniques exist to then remove the marker gene, although this often leaves a small "scar" in the genome, which may have undesirable downstream effects (David and Siewers, 2015; Solis-Escalante *et al.*, 2015).

Genome editing has been successfully employed to exploit the current knowledge of the yeast genome and develop several yeast strains with enhanced technological properties. Specifically, a strain was constructed that constitutively expresses both *DUR1,2* and *DUR3* genes in commercial wine and sake backgrounds to effectively reduce the levels of ethyl carbamate, a known mutagen, in the final product (Dahabieh, Husnik and van Vuuren, 2009; Dahabieh, Husnik and Van Vuuren, 2010). Another yeast strain was constructed and commercialized that is capable of carrying out malolactic fermentation, a process that traditionally requires the inoculation of the bacterium *Oenococcus oeni* (Husnik *et al.*, 2007). While this strain has been approved for commercial use and produces wines with excellent sensory qualities, genetically modified yeast strains have failed to gain traction in the wine and beer industries due

to poor customer acceptance and variable regulation of GMOs (Steensels *et al.*, 2014). To avoid regulatory issues, a self-cloning technique has been improved such that no non-*Saccharomyces* DNA is required for the genetic modification. This technique has been applied to improve the freeze tolerance of baking yeast for frozen dough applications (Nakagawa *et al.*, 2017). While still controversial for food production, genetic modification of yeast strains for the production of biofuel and novel chemical compounds continues to occur with great success and these synthetic biology endeavours benefit greatly from the improved annotation of industrial yeast genomes (Steensels *et al.*, 2014).

1.7 Research Proposal

1.7.1 Rationale for study

While there has been a significant increase in genome sequencing of yeast strains from a variety of industries and natural sources, a portion of genes remains unannotated. This applies particularly to regions of the yeast genome that contain ORFs, but have only been identified in industrial isolates and not in the *S. cerevisiae* reference genome. Novel, non-reference ORFs have been identified in the commercial wine yeast Enoferm M2, some of which are unique to the strain and some of which are shared with other commercial wine yeasts.

Systematic approaches, such as the yeast deletion collection and high copy plasmid overexpression, have been used successfully to improve the annotation of many yeast genes by identifying associated phenotypes in various conditions. These approaches, however, only apply to the conserved core of genes found in the *S. cerevisiae* reference strain, S288C. Genetic background can have a strong modifying role on gene interactions, and the S288C is not considered to be a good representation of the consensus genome of commercial wine yeasts. Comparative genomic studies have investigated the genetic differences that are associated with the performance of various industrial yeast strains, such as truncations, SNPs and CNVs, but this method does little to resolve the function of ORFs that are present in only a handful of strains. Transcriptomic and proteomic studies have been used in conjunction with stress conditions to identify candidate genes/proteins that are differentially regulated in a given condition.

Poorly annotated genes have often been identified as differentially regulated under stress conditions in transcriptomic studies, but are rarely followed up with further analysis. Furthermore, these studies are typically conducted under synthetic conditions, rather than in stress conditions that replicate industrial conditions. Combined proteomic and transcriptomic studies have identified discrepancies in results that indicate a significant role for post-transcriptional and post-translational mechanisms in the regulation of the cell during stress response.

The present study was designed to systematically investigate potential functions of novel ORFs found in the commercial wine strain Enoferm M2, during the fermentation of Chardonnay grape must, using a variety of phenotypic measures as well as transcriptomic and proteomic analysis. The study is intended to address knowledge gaps regarding the role of these specific novel ORFs. The transcriptomic and proteomic data sets generated will also allow the investigation of potential points of post-transcriptional and post-translational regulation during a commercial wine fermentation.

1.7.2 Proposed research

The proposed research will create null mutants and constitutive expression (CE) mutants for 15 novel, non-reference ORFs identified in the Enoferm M2 genome by prior analysis of the genome. As a positive control the yeast gene *HXK2*, a gene with known roles in glycolysis, will also be deleted and constitutively expressed. These stable mutants, along with a wild type control, will be inoculated into sterile Chardonnay grape must to ferment as small-scale wine fermentations. Samples of both the wine and the yeast cells will be taken aseptically at multiple time points. Wine samples will be analyzed for primary and secondary yeast metabolites by HPLC and GCMS, and cells will be used to isolate RNA and proteins for transcriptomic and proteomic analysis, respectively (Figure 1.4).

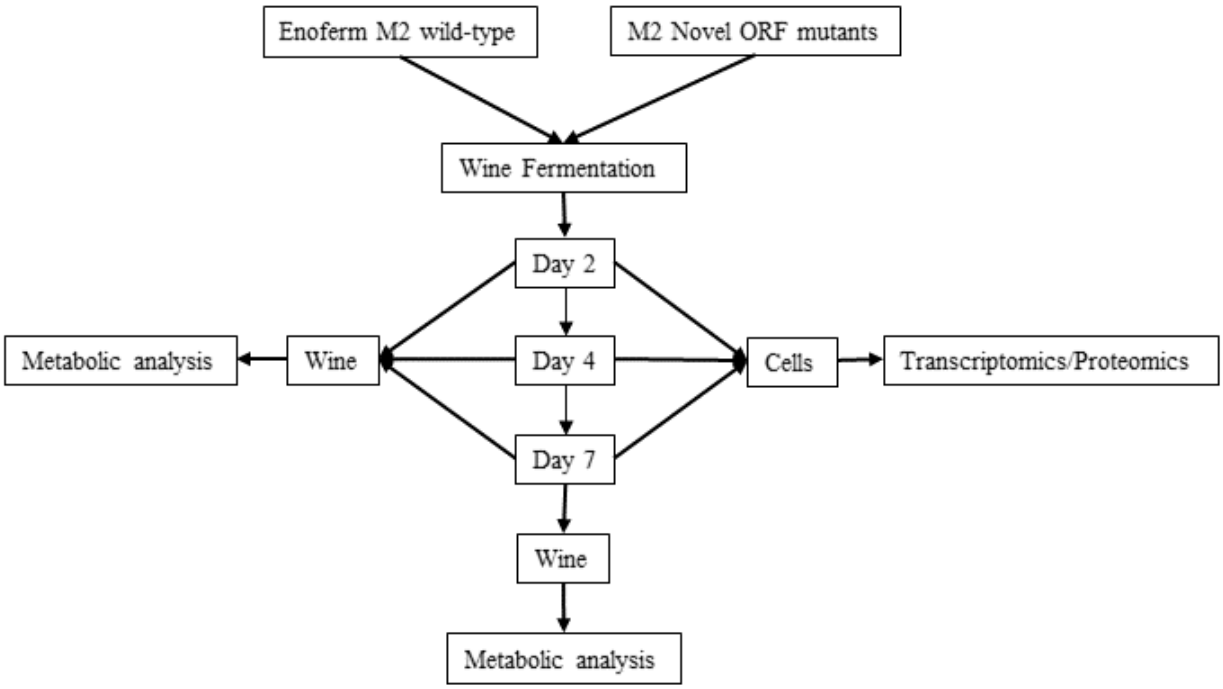


Figure 1.4 Proposed research workflow. The proposed workflow for the systematic annotation of novel Enoferm M2 ORFs by wine fermentation and transcriptomic, proteomic and metabolomic analysis.

Mutant strains will be compared against wild type for all phenotypic measures tested, including glucose and fructose consumption, ethanol production, organic acid production and the production of various secondary metabolites. Quantitative real-time PCR will be performed on RNA isolated from wild type cells with primers specific to each novel ORF to confirm transcript expression during the wine fermentation. RNA isolated from three different time points (Figure 1.4) will be analyzed by DNA microarray technology and protein isolated from the corresponding samples will be analyzed by isobaric tagging for relative and absolute quantification (iTRAQ). Mutants will be compared against wild type to determine differentially expressed (DE) transcripts and differentially expressed proteins. These will be analyzed for ontology enrichments that may indicate specific assays that could be applied to further annotate these novel ORFs.

The three-point time course response to wine fermentation stress in wild type Enoferm M2 will be analyzed in terms of the transcriptome and proteome at each time point. Differential expression between time points will be analyzed to characterize the strain specific response to wine fermentation stress.

These results will be compared against published data for other wine strains in similar conditions to determine if there are any responses unique to Enoferm M2. Lastly, the transcriptome and proteome will be correlated to determine which proteins may be under post-transcriptional or post-translational regulation during stress response.

1.7.3 Research hypotheses

The following statements shall govern the focus of the proposed research.

Hypothesis 1: Novel ORFs present in industrial wine yeast have a function in fermentation and/or stress tolerance. The functions of novel ORFs will be revealed by testing under enological conditions.

Hypothesis 2: The deletion or constitutive expression of novel ORFs will have effects on the transcriptome and/or proteome of the fermenting mutant cells.

Hypothesis 3: Time course study of the Enoferm M2 transcriptome and proteome will identify strain specific responses to fermentation stress.

1.7.4 Research objectives

The following objectives will be achieved by the proposed research.

Objective 1: Determine a fermentation phenotype for the deletion or constitutive expression of 15 novel ORFs from the genome of the wine strain Enoferm M2.

Objective 2: Assess the differences in the transcriptome and proteome of deletion/constitutive expression mutants.

Objective 3: Determine upregulated and downregulated transcripts and proteins for each sampling time. Assess agreement between transcriptomic and proteomic data. Compare results with published data from other strains.

2 MATERIALS AND METHODS

2.1 Strain of Saccharomyces cerevisiae

The mutant strains used in this study are based on the commercially available wine yeast Enoferm M2 (a gift from Dr. Richard C. Gardner, from the University of Auckland, NZ). The strain was originally isolated from a vineyard in Stellenbosch, South Africa, and selected by the University of Stellenbosch from the University Massey culture collection as a candidate for commercialization (Lallemand, 2015). Enoferm M2 is described as a moderate fermenter, with high nitrogen requirements for robust fermentation and an alcohol tolerance around 15% v/v. This strain produces relatively high amounts of volatile acidity, but low amounts of glycerol and sulphur compounds. It is thought to enhance citrus notes in white wines, and berry notes in reds, but produces low levels of esters and is generally regarded as a "neutral" yeast. Enoferm M2 is genomically equivalent to AWRI796 (AB Mauri, Sydney, Australia), another commercially available wine strain, although the commercial descriptions highlight slightly different technical capabilities (Borneman *et al.*, 2016). The genome sequence of Enoferm M2 has been deposited at NCBI (BioSample: SAMN03417849, Sample name: UOA_M2), and has also been re-sequenced and assembled to the EC1118 wine strain scaffold by the van Vuuren laboratory (unpublished).

2.2 Standard culture and storage of yeast

For routine culture, yeast was grown in YPD broth (1% yeast extract, 2% peptone, 2% dextrose, Sigma Aldrich, St. Louis, MO) unless otherwise specified. For growth on solid media, YPD was supplemented with 2% agar (Sigma Aldrich, St. Louis, MO). Unless otherwise noted, liquid cultures were grown at 30°C aerobically with shaking. Strains were stored as 1 ml aliquots at -80°C in 15% glycerol solution.

2.3 Construction of mutant stains

2.3.1 PCR amplification of cassettes for null mutants

For the creation of null mutants in the Enoferm M2 background, the *kanMX* cassette (conferring resistance to the antibiotic G418) was amplified by polymerase chain reaction (PCR) from the pUG6

plasmid (full sequence in Appendix A.3) using primers (see Appendix A.4 for list of primers used) designed with 60 bp overlapping ends corresponding to the regions immediately upstream of the start codon and immediately downstream of the stop codon for each ORF of interest. Primer DNA (Integrated DNA Technologies, Coralville, IA, USA) was dissolved to a 100 µM concentration in sterile, deionized water and stored at -20 °C, and diluted ten-fold prior to use. PCR was performed with the iProof polymerase (Bio-Rad, Hercules, CA) according to the manufacturer's recommendations. Briefly, 10 µL of iProof buffer, 1 µL of 10 mM dNTPs, 2.5 µL of each primer, 1 µL (~ 2 ng) of plasmid DNA, 0.5 µL of iProof polymerase and 32.5 µL of water were combined on ice. The reaction mix was thermocycled as follows: 98 °C for 30 seconds to denature, then 98 °C for 10 seconds, 54 °C for 20 seconds and 72 °C for 60 seconds, repeated 30 times, and finally 72 °C for five minutes. PCR products were confirmed by agarose gel electrophoresis.

2.3.2 PCR amplification of cassettes for overexpression mutants

Constitutive expression mutants were constructed for five novel ORFs (M13, M18, M21, M23, M28) in the Enoferm M2 background. To do so the *PGK1*-promoter-*kanMX* cassette was amplified from the pCW1 plasmid (full sequence in Appendix A.3) using primers (see Appendix A.4 for list of primers used) designed with 60 bp overlapping ends corresponding to the region including and immediately downstream of the start codon and the region immediately upstream of the start codon. *PGK1* has been previously found to maintain constitutive strong expression during wine fermentation (Marks *et al.*, 2008). This method effectively inserts the *PGK1* constitutive expression promoter immediately upstream of the start codon to drive the expression of the ORF of interest, as well as inserting a *kanMX* selectable marker upstream of the promoter. Primer DNA (Integrated DNA Technologies, Coralville, IA) was dissolved to a 100 µM concentration in sterile, deionized water and stored at -20 °C, and diluted ten-fold prior to use. PCR amplification of the cassettes was accomplished using Q5 Hot Start High-fidelity 2X Master Mix (New England BioLabs, Whitby, ON) according to the manufacturer's recommendations. Briefly, 20 µL of 2X Q5 Master Mix, 1 µL of each primer, 0.5 µL of plasmid DNA template and 17.5 µL

of water were combined on ice. The reaction mix was thermocycled as follows: 98 °C for 30 seconds to denature, then 98 °C for 10 seconds, 52 °C for 20 seconds and 72 °C for 60 seconds, repeated 30 times, and finally 72 °C for two minutes. PCR products were confirmed by agarose gel electrophoresis.

2.3.3 Transformation of yeast

Wild type Enoferm M2 was transformed with PCR amplified cassettes using a modified lithium acetate transformation protocol (Gietz and Schiestl, 2007). The M2 yeast was inoculated into YPD media from glycerol stocks and grown overnight. Cultures were diluted to 0.2 OD600 and grown for a further 3-4 generations (3-4 hours). Roughly five OD600 cells were used for each transformation. Cultured cells were collected by centrifugation and washed once with sterile water. Water was removed, and cells were suspended in 50 µL of 0.1 M lithium acetate and incubated at 30°C for 10 minutes. Lithium acetate was removed by centrifugation, and cells were suspended in the reaction mix (240 µL 50% polyethylene glycol, 36 µL 1 M lithium acetate, 10 µL 10 µg/mL denatured cDNA, 40 µL water, and 25 µL of PCR product). Tubes were vigorously vortexed to suspend the pellet, then incubated for 30 minutes at 30°C and 20 minutes at 42°C. The reaction mix was then removed by centrifugation, the cells were suspended in 1 mL of YPD and outgrowth was performed for 2 hours at 30°C with shaking. Cells were diluted 100- or 1000- fold for plating on YPD agar containing 200 µg/mL of G418. Plates were incubated at 30°C until colonies were clearly visible.

Colonies were tested by colony PCR for the presence of the *kanMX* cassette in the correct location. Primers were designed to amplify a ~400 bp region spanning the first 200 bp of the *kanMX* cassette and 200 bp upstream of the insertion site. Positive colonies were re-streaked to G418 selective agar. Single colonies were then grown overnight in YPD. Overnight culture (200 µL) was centrifuged and ~190 µl of supernatant removed. The remaining supernatant was used to suspend the pellet and transfer it to a 2% potassium acetate plate supplemented with 2% agar. Plates were incubated at 19°C for 5 days to allow for the sporulation of yeast cells. Sporulation was confirmed by visualization under a microscope. A light swab of cells was treated with 60 U/ml Zymolase for 10 minutes at 30°C. Cells were

diluted in 1 mL of water and a 10 µL aliquot was placed on the edge of a YPD plate and the plate was tilted to create a line (Figure 2.1). Four to five tetrads were dissected for each transformation, and plates were incubated at until visible colonies formed. Plates were then replica plated with velvets to G418 selection plates.

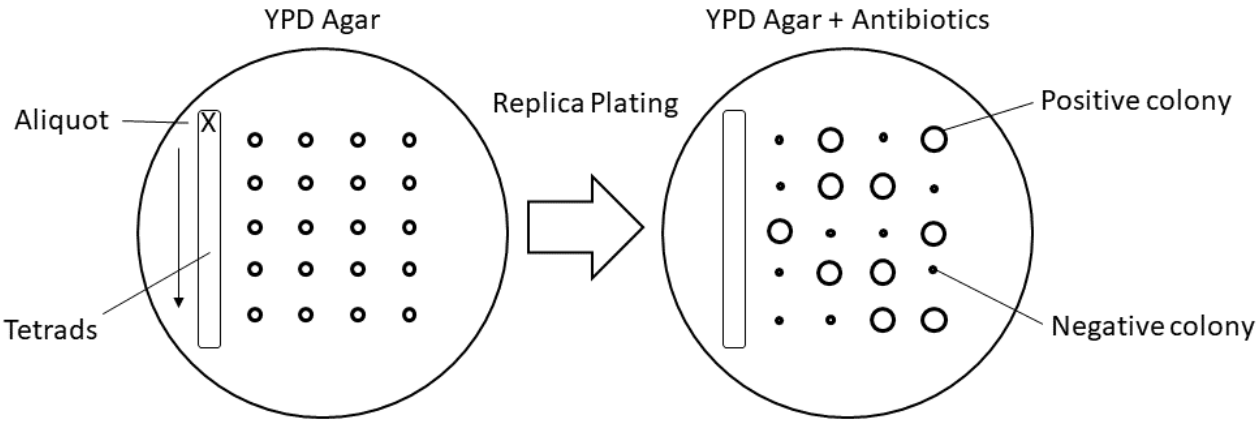


Figure 2.1 Tetrad dissection of transformed yeast. After Zymolase treatment, an aliquot was placed on YPD agar plates and tilted to form a line, from which tetrads were dissected. Haploid spores undergo mating-type switching and self-mating, and after replicate plating to a selection plate only colonies homozygous for the mutation are able to grow.

All four colonies from a tetrad were verified by colony PCR for both the presence of the *kanMX* cassette and the absence of the ORF of interest. Successful transformants had robust growth on the selection plate and yielded no band for the novel ORF. Wild type Enoferm M2 was used as a control, to confirm that primers produced no band for the marker and a positive band for the novel ORF in wild type (Figure 2.2). Confirmed colonies (*kanMX*/*PGK1*prom positive, ORF negative, and growth on G418) were inoculated into YPD, grown overnight and frozen at -80°C as glycerol stocks. All mutants used in this study were homozygous diploids. No growth phenotype was observed for any null mutants when grown under standard conditions (YPD broth/agar at 30 °C).

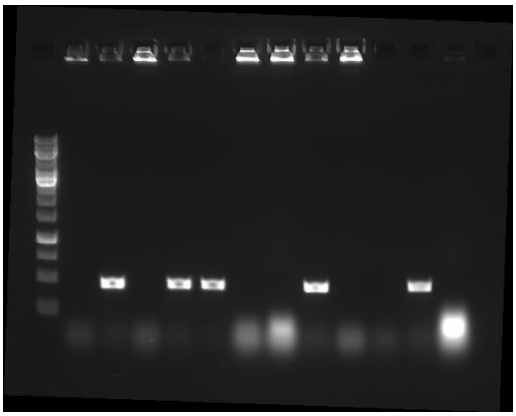
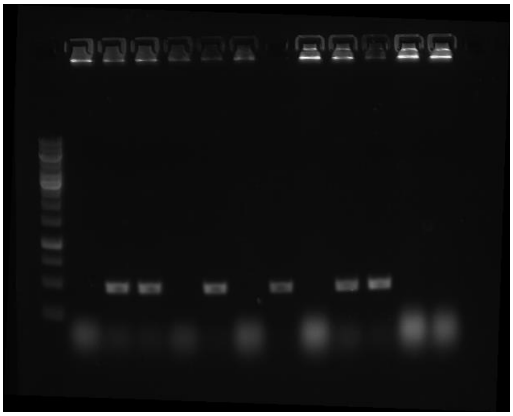
a)  b) 

Figure 2.2 Confirmation of null mutants. Example DNA electrophoresis gels of the confirmation of null mutants for M15 (a) and M6 (b). Primers were designed to amplify from ~200 bp within the original ORF and from ~200 bp upstream of the start codon. Blanks represent colonies that were successfully deleted, and all blanks correspond to colonies that grew well on G418 supplemented plates and had the correct insertion of the *kanMX* marker confirmed by PCR (not shown here). Positive bands are from colonies that were unable to grow on selection media. Three tetrads (12 colonies) were checked for each mutant.

2.4 Poorly-annotated ORF null mutants

In addition to the novel ORF null mutants and constitutive expression mutants, a number of other strains were also used for a parallel wine fermentation. This included null mutants for ten putative transcription factors (*SEF1*/YBL066C, *NRG2*/YBR066C, *TBS1*/YBR150C, YFL052W, *PHD1*/YKL043W, YKL222C, YLL054C, YLR278C, YNR063W, *HMS1*/YOR032C), as well as seven other poorly annotated genes selected based on previous transcriptomic studies in wine yeast (*PAR32*/YDL173W, YBR056W, *GEP5*/YLR091W, *PDR18*/YNR070W, YPL225W, YDR089W, *HXK2*/YGL253W) (Marks *et al.*, 2008) (Table 2.1). These null mutants were constructed previously in the laboratory by Dr. Christopher Walkey and Mayumi Iwashita using the same protocol as this study, and the strains were stored in 15% glycerol stocks at -80 °C until use.

Table 2.1 Non-novel null mutants. Null mutants for other ORFs of interest were identified from previous transcriptomic studies of the wine fermentation and descriptions are taken from the Saccharomyces Genome Database.

ORF Name	Gene Name	Description
YGL253W	*HXK2*	Hexokinase enzyme
YLR091W	*GEP5*	Protein of unknown function
YNR070W	*PDR18*	Putative ATP-binding cassette transporter
YDL173W	*PAR32*	Protein of unknown function
YBR056W	-	Putative glycoside hydrolase
YPL225W	-	Protein of unknown function

ORF Name	Gene Name	Description
YDR089W	*VTC5*	Vacuolar transporter chaperone
YBL066C	*SEF1*	Putative transcription factor
YBR066C	*NRG2*	Transcriptional repressor, involved in mediation of glucose repression
YBR150C	*TBS1*	Protein of unknown function
YFL052W	*ZNF1*	Zinc cluster transcription factor, involved in regulation of respiratory growth
YKL043W	*PHD1*	Transcriptional activator that enhances pseudohyphal growth
YKL222C	-	Protein of unknown function
YLL054C	-	Putative protein of unknown function
YLR278C	-	Zinc cluster protein
YNR063W	-	Putative protein of unknown function
YOR032C	*HMS1*	Similar to myc-family transcription factors

2.5 Wine fermentations

Wine fermentation trials were conducted in 80 ml of Chardonnay grape juice (23.4 Brix, pH 3.48, TA 0.57 g/L, SO2 50 ppm, sourced as Scarlett Ranch Chardonnay from Lanza Vineyard in Suisun Valley California, M & M Wine Grape Co, CT). Juice was aliquoted from 23 L pails to one and two L bottles and maintained frozen at -20 °C prior to use. Upon use, juice was pooled and pre-filtered using glass fibre prefilters (Merck Millipore, Cork, Ireland), then sterile-filtered through 0.22 micron filters before being aliquoted to individual bottles.

Strains were inoculated into 3 ml of YPD from -80 °C glycerol stocks and grown for 15-17 hours. Cultures were then diluted to 0.2 OD600 in 8 ml of fresh YPD and grown for approximately three generations to mid-log phase (1-1.5 OD600). Eight OD600 cells (~ 6 mL) were collected and centrifuged to remove YPD, then inoculated into the grape juice to yield a final OD600 of 0.1. Bottles were stoppered with rubber bungs fitted with S-shaped airlocks filled with 2 ml of sterile water to allow the escape of gas while maintaining sterility. Fermentations were maintained still at 19 °C in a controlled chamber. Fermentation progress was monitored by weight loss to measure the production of CO_2, or by HPLC to analyze the rate of sugar consumption and alcohol production. All sampling was conducted aseptically using a sterile needle to pierce the rubber bung. An uninoculated grape juice control was used to monitor for losses due to evaporation and potential contamination. Three separate fermentation trials

were conducted, each with their own wild type control: an initial fermentation using the null mutants monitored by weight loss, a second fermentation of null mutants monitored by HPLC during which cell samples were taken, and a third fermentation using constitutive expression mutants monitored by HPLC during which cell samples were taken.

2.6 High performance liquid chromatography (HPLC) analyses

2.6.1 Run protocol

Samples of wine collected during the fermentation were sterile-filtered (0.22 micron) into 1.5 ml amber auto sampler vials. HPLC analyses were conducted based on a previously validated method (Walkey *et al.*, 2012; Terrell, Cliff and Van Vuuren, 2015). Four µL of sample was injected into an Agilent 1100 series HPLC equipped with an Agilent 1260 Infinity refractive index detector. An isocratic run of 0.00425 M sulphuric acid was used in conjunction with a nucleogel Ion 300OA valco type column (300 x 7.8 mm inner diameter) (Macherey-Nagel, Düren, Germany) at a mobile phase flow rate of 0.75 mL/min for a ~28 minute total run time. Samples were injected at ambient temperature (21-25 °C) and the column was maintained at 71 °C. Data was collected using LC/MS Chemstation (Rev.A.09.03).

2.6.2 HPLC data analyses

Data was exported to as a .xls file and analyzed in R Studio (R Development Core Team, 2017). All samples were assembled into a data.frame and two-way analysis of variance (ANOVA) was performed modelling a given metabolite against genotype and day of fermentation. When a significant difference was found, Tukey's honest significant difference (HSD) test was applied to determine if a single genotype was significantly different from wild type.

2.7 Gas chromatography mass spectrometry (GCMS) analyses

2.7.1 Sample preparation and run protocol

At the end of fermentation, the final wine was preserved with the addition of 0.78 mM potassium metabisulphate (Sigma Aldrich), allowed to settle and carefully transferred from the yeast lees to a smaller bottle to minimize headspace. Any remaining headspace was flushed with nitrogen gas and

samples were stored at 4 °C until immediately before analysis. For analysis, 5 mL of sample was transferred to a 20 mL vial with 1.5 g of sodium chloride and 100 µL of 0.565 mg/L 3-octanol (internal standard).

GCMS analyses were conducted with an Agilent 7890A GC coupled to an Agilent 5975C inert XL MSD with triple-axis detector, and equipped with a GC sampler 80, to analyze the headspace of the wine samples. Injection was by solid-phase microextraction (SPME) with pre-incubation at 30 °C for 300 s while agitated at 500 rpm (3 seconds on, 3 seconds off). Samples were extracted for 300 s and desorbed for 150 s. The fibre was conditioned between each injection. The column was a DB-WAX fused silica open tubular (29.84 m x 250 µm x 0.25 µm) (J&W Scientific, Palo Alto, CA, USA). The oven program was as follows: 40 °C for 4 min, then 3 °C/min to 110 °C for 0 min, then 3 °C/min to 150 °C for 0 min, then 25 °C/min to 230 °C for 10 min (total run time = 53.867 min). Injection mode was split 3:1, and the carrier gas was ultra-high-purity helium at a constant total flow of 0.92665 mL/min. The MSD acquisition was set to scan mode with low and high mass cut-odds of 33 and 450, respectively. Data were analyzed with MSD Chemstation E.02.02.1431 (Agilent) software.

2.7.2 GCMS data analyses

Data were exported to as a .xls file and analyzed in R Studio (R Development Core Team, 2017). All samples were assembled into a data.frame and one-way ANOVA was performed modelling a given metabolite against genotype. When a significant difference was found, Tukey's honest significant difference (HSD) test was applied to determine if a single genotype was significantly different from wild type.

2.8 Microarray analyses

2.8.1 Sample collection and RNA Extraction

Samples of wine were taken aseptically by hypodermic needle through the rubber bung during fermentation. One millilitre of wine was removed and centrifuged, and the supernatant removed. The pellet was suspended in 400 µl of RNAlater ICE (Ambion, Austin, Texas) and frozen at -80 °C until

processing. All surfaces and equipment were cleaned with RNAse AWAY (Ambion) prior to processing samples. RNA was extracted from cell pellets using the RiboPure RNA Purification Kit, yeast (Ambion), with a reduced volume protocol.

Briefly, 200 µl of cell culture suspension was centrifuged and the RNAlater removed. The pellet was suspended in 240 µl of lysis buffer and 24 µl of 10% sodium dodecyl sulphate (SDS), and added to a screw-top tube containing ~400 µl of ice-cold zirconia beads and 240 µl of phenol:chloroform:isoamylacetate. Tubes were vortexed for 10 minutes, and centrifuged for 5 minutes at 16,000 rpm. The aqueous phase (150 µl) was then transferred to a fresh tube, to which 538 µl of binding buffer and 354 µl of pure ethanol was added. This mixture was applied to a filter cartridge in two rounds, with a 30 second centrifuge and flow-through discarded. The cartridge was washed once with 700 µl of Wash Solution 1 and washed twice with 500 µl of Wash Solution 2/3. Cartridges were centrifuged dry for one minute to remove any excess fluid, then transferred to a fresh collection tube. RNA samples were eluted with two 35 µl aliquots of 95°C Elution buffer, with a one-minute centrifuge for each aliquot. Samples were then treated with 7 µl of 10x DNAse buffer and 2 µl of DNAse I, and incubated at 37 °C for 30 minutes. Eight µl of inactivation reagent was added to stop the reaction, and tubes were incubated at room temperature for 5 minutes. Tubes were centrifuged for 2.5 minutes and the clear supernatant (isolated RNA) transferred to a screw-top cryogenic tube for storage at -80 °C.

The quality of RNA was assessed by loading a 2 µl sample into a 1% agarose TAE gel and running at 80-100 volts until bands were clearly separated. Gels were checked for the relative intensity of the 28S and 18S bands, as well as a general lack of smearing from degraded RNA. Concentration of RNA was determined by a ND-1000 spectrophotometer (Nanodrop, Wilmington, Delaware).

2.8.2 Hybridization to microarray chips

RNA samples were 3' IVT labelled and purified according to the Affymetrix protocol, using the '3 IVT Express Kit scaled for reduced volumes. To do so, RNA samples were diluted to 25 ng/µL in nuclease-free water, and PolyA RNA controls were diluted according to the manufacturer's instructions.

First strand complementary DNA (cDNA) synthesis was achieved by mixing 0.5 µL of PolyA control RNA, 2 µL of first strand buffer, 0.5 µL of first strand enzyme and 2 µL of diluted sample RNA on ice. The reaction mix was incubated at 42 °C for two hours, then 4 °C for at least two minutes. Second strand cDNA synthesis was achieved by removing 2.5 µL from the first reaction, then mixing in 3.25 µL of nuclease-free water, 1.25 µL of second strand buffer, and 0.5 µL of second strand enzyme on ice. This reaction mix was incubated at 16 °C for one hour, then 65 °C for ten minutes, and finally 4 °C for at least two minutes. Biotin labelling by *in vitro* transcription (IVT) was achieved by mixing 1 µL of IVT biotin label, 5 µL of IVT buffer, and 1.5 µL of IVT enzyme at room temperature and adding this to the second strand synthesized DNA mix. This reaction mixture was incubated at 40 °C for 16 hours, then 4 °C until use purification.

Biotin-labelled complimentary RNA (cRNA) samples were purified with magnetic beads and ethanol washes. To do so, 25 µL of purification beads was added to the 15 µL cRNA sample and the mixture was transferred to a 96-well plate and incubated at room temperature for ten minutes. The plate was transferred to a magnetic stand and the beads were allowed to pellet. Supernatant was removed, and the beads were washed three times with 80% ethanol. Beads were air dried, and the plate was removed from the magnets. To each sample, 30 µL of nuclease free water pre-heated to 65 °C was added and mixed. After one minute, the plate was returned to magnets to pellet the beads. The supernatant was collected for fragmentation, or stored at -80 °C. Quality and concentration was assessed by ND-1000 spectrophotometer.

The cRNA samples were hybridized to Yeast GeneChip 2.0 microarray chips according to the Affymetrix protocol, using 6 µg of starting cRNA (Affymetrix, Santa Clara, CA, USA). To begin, samples were fragmented by combining 6 µg of cRNA, water to 32 µL and 8 µL of fragmentation buffer, and incubating the mix at 94 °C for 35 minutes, then 4 °C until use. Pre- and post-fragmentation samples were visualized by agar gel electrophoresis to confirm complete fragmentation. Fragmented cRNA was stored at -80 °C until required for hybridization to microarray chips. For hybridization of cRNA to microarray chips, a master mix was prepared with 1.7 µL of B2, 5 µL of 20x, 50 µL of 2x and 10 µL of

DMSO, per sample. The master mix (66.7 µL) was combined with 33.3 µL of each sample, and the reaction mixture was incubated at 99 °C for five minutes, then 45 °C until use. The microarray chips, at room temperature, were loaded with 80 µL of pre-hybridization buffer and incubated for 15 minutes in a rotisserie at 60 rpm. The pre-hybridization buffer was removed from chips and 80 µL of the sample cocktail was loaded, and the chips were incubated at 40 °C for 16 hours in the rotisserie at 60 rpm. Sample cocktail was removed, and chips were loaded with 120 µL (excess) wash A before proceeding to the fluidics station for staining.

2.8.3 Analyses of microarray data

Data files from the microarray reader were normalized (Affymetrix GeneChip Command Console). These files were then analyzed with Transcriptome Analysis Console 3.0. Since three time points were tested, the mutants were tested against wild type for each time point and differentially expressed (linear fold-change > 2, ANOVA p-value < 0.05) transcripts were identified. In addition, wild type samples for each time point were tested against each other to provide a picture of day-to-day changes in transcript abundances.

2.9 Proteomic analyses

2.9.1 Sample collection and protein extraction

Wine trials were swirled to homogenize, and 5 ml was harvested aseptically by hypodermic needle through the rubber bung. Samples were processed for protein extraction based on a modification of a protocol by Hebert *et al*. (2014). Briefly, the wine was centrifuged at 3600 rpm for 10 minutes at 4 °C. The supernatant was decanted, and the pellet suspended in 1 ml of chilled water and transferred to a microcentrifuge tube. The pellet was washed twice with chilled water, and then suspended in lysis buffer (50 mM Tris, 75 nM sodium chloride, 100 mM sodium butyrate, 8 M urea, cOmplete inhibitor tablet (Roche) and PhoStop inhibitor tablet (Roche)) and transferred to a FastPrep Matrix C tube. Cells were beaten four times at 4.0 m/s for 30 seconds with 3-5 minute rests in between. Tubes were centrifuged at 16,000 rpm for 2 minutes and the supernatant was transferred to a cryogenic tube and frozen at -80 °C.

2.9.2 iTRAQ

The bicinchoninic acid (BCA) assay was used to determine the protein concentration of each sample (Sigma, ON, Canada). The volume containing 100 µg of protein was acetone precipitated in 10 volumes of ice-cold acetone overnight at -20 °C. The precipitated samples were centrifuged and the acetone was removed. The protein pellets were re-solubilized in 30 µL, 0.5M triethylammonium bicarbonate (TEAB) and 3 µL, 2% SDS. Samples were rehydrated for four hours at 4 °C. Two µL, 50 mM tris(2-carboxyethyl)phosphine (TCEP) was added and samples were incubated for one hour at 60 °C and then allowed to cool to room temperature; 1 µL, 200 mM S-methyl methanethiosulphonate (MMTS) was then added and samples were incubated at room temperature for 10 minutes. Eighty micrograms of trypsin (Promega, Sequencing Grade Modified Trypsin) was re-suspended in 250 µL, 100 mM TEAB and 120 µL (10 µg) was added to each sample. Samples were incubated overnight at 37 °C. The digested samples were dried by vacuum centrifugation (Savant Instruments, Holbrook, NY) and then 30 µL, 0.5 M TEAB was added to each dried sample followed by the addition of 50 µL isopropanol. The iTRAQ label (AB Sciex, ON, Canada) was added to each sample and then incubated at room temperature for two hours. The labeled peptides were pooled and vacuum centrifuged until the final volume was approximately 100 µL.

An Agilent 1290 HPLC (Agilent, CA, USA) was equipped with an XBridge C18 BEH300 (Waters, MA, USA) 250 mm X 4.6 mm, 5 µm, 300 A HPLC column. Buffer A was 10 mM Ammonium hydroxide (pH 10). Buffer B was 80% Acetonitrile, 10 mM Ammonium hydroxide (pH 10). The flow rate was set to 0.75 ml/min. Samples were brought up to 1.8 mL with buffer A and injected onto the column. The column was allowed to equilibrate for 5 min in buffer A before a gradient was applied; 5-45% B in 75 minutes. Fractions were collected every minute for 96 minutes. The HPLC fractions were then reduced in volume by lyophilization and concatenated into 24 fractions by combining every 24th fraction (*e.g.* fractions 1, 25, 49, and 73 were combined).

After C18 StageTip concentration each sample was rehydrated with 20 µl (2% Acetonitrile, 0.5%

Formic acid). A 5 µl aliquot of the peptide solution was separated by on-line reversed phase liquid chromatography using a Thermo Scientific EASY-nLC 1000 system with a reversed-phase pre-column Magic C18-AQ (100 µm I.D., 2 cm length, 5 µm, 100 Å), and an in-house prepared reversed phase nano-analytical column Magic C-18AQ (75 µm I.D., 15 cm length, 5 µm, 100 Å, Michrom BioResources Inc, Auburn, CA), at a flow rate of 300 nL/min. The chromatography system was coupled on-line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a Nanospray Flex NG source (Thermo Fisher Scientific). Solvents were A: 2% acetonitrile, 0.1% formic acid; B: 90% acetonitrile, 0.1% formic acid. After a 300 bar (~ 12µL) pre-column equilibration and 300 bar (~ 4µL) nanocolumn equilibration, samples were separated using a 120-minute gradient (0 min: 5% B; 100 min: 40% B; 5 min: 80% B; 2 min: 100% B; 13 min: 100% B). The Orbitrap Fusion Tribrid instrument parameters were as follows: Nano-electrospray ion source with spray voltage 2.5 kV, capillary temperature 275 °C. Survey MS1 scan m/z range 380-2000 profile mode, resolution 60,000 FWHM at 200 m/z one microscan with maximum inject time 50 ms. The Lock mass Siloxane 445.120024 was used for internal calibration. Data-dependent acquisition Orbitrap survey spectra were scheduled at least every 3 seconds, with the software determining "Top-speed" number of MS/MS acquisitions during this period. The automatic gain control (AGC) target values for FTMS and MSn were 400,000 and 50,000 respectively. The most intense ions m/z range 380-2000, charge state 2-7 exceeding 50,000 counts were selected for higher-energy collisional dissociation (HCD) Orbitrap trap MSMS fragmentation with detection in centroid mode. Dynamic exclusion settings were: repeat count: 1; exclusion duration: 30 seconds with a 2ppm mass window. The ddMS2 OT HCD scan used a quadrupole isolation window of 1.6 Da; Orbitrap resolution 15,000 FWHM @ 200m/z, 105 m/z first mass range, centroid detection, 1 microscan, 32 ms maximum injection time, parallelizable option enabled and stepped HCD collision energy 35 $\pm$ 5% (Senko *et al.*, 2013).

Raw files were created by XCalibur 3.0.63 (Thermo Scientific) software and analyzed with Proteome Discoverer 1.4.0.228 software suite (Thermo Scientific). Spectrum Selection was used to generate peak lists of the HCD spectra (parameters: activation type: HCD; s/n cut-off: 1.5; total intensity threshold: 0;

minimum peak count: 1; precursor mass: 350-5000 Da). The peak lists were submitted to an in-house database search using Mascot 2.4 (Matrix Science), and were searched against the following database; Yeast (6,147 sequences; 3,049,991 residues). Search parameters were: precursor tolerance 10 ppm; MS/MS tolerance 15 mmu (for FT MS/MS HCD data); enzyme Trypsin; 2 missed cleavage; instrument type FT-ICR ESI; fixed modification: Methylthio (C), iTRAQ8plex (K), and iTRAQ8plex (N-term); variable modifications: Oxidation (M), Deamidated (NQ), iTRAQ8plex (Y). The HCD Percolator settings were: Max delta Cn 0.05; Target FDR (false discovery rate) strict 0.01, Target FDR relaxed 0.05 with validation based on q-Value.

2.9.3 Analyses of proteomic data

The results from iTRAQ proteomic analysis were processed using Scaffold 4 software (Version 4.8.3, Proteome Software, Inc., Portland, OR, USA). Individual BioSamples were merged into a single experiment file. Tagged proteins were identified with a 99.0% protein threshold, and a minimum of two peptides with a 95% peptide threshold. For quantification, an intensity-based normalization and analysis was selected. The experimental design was between-subjects, with a common reference for all BioSamples (one biological replicate of wild type). Each sample was assigned to its own category as a biological replicate, as they were all taken from individual fermentations. The minimum dynamic range was set such that only reporter ion peaks that were above 5% of the highest peak in a spectrum were included in the analysis. The normalization was conducted based on the median, blocking level at unique peptides, reference type set to individual spectrum, and spectrum quality filter requiring reference values. A Mann-Whitney test with Benjamini-Hochberg correction ($p > 0.05$) was conducted between each mutant and the wild type control to test for significant differences in relative protein abundances, with a 1.5-fold change cut off imposed on the results.

2.10 YeastMine Lists for Gene Ontology and Pathway Enrichment

To analyze lists of differentially expressed transcripts and proteins, the corresponding gene identifiers were used to create lists in the YeastMine tool available through the Saccharomyces Genome

Database (Stanford University, Stanford, CA, USA). The Gene Ontology Enrichment and Pathway Enrichment widgets were applied to the lists, both using the Holm-Bonferroni test correction and a maximum *p*-value of 0.05. Genes without ontology or pathway annotations were excluded from these lists.

2.11 Quantitative real-time PCR (qRT-PCR)

Quantitative real-time polymerase chain reaction (qRT-PCR) was used to quantify the transcription of novel ORFs in wild type Enoferm M2 during Chardonnay fermentation. To do so, RNA was isolated as described for microarray analysis (Section 2.7). RNA was used to synthesize cDNA using Superscript VILO Master Mix (Invitrogen, Carlsbad, CA, USA) as per an adaption of the manufacturer's protocol. Briefly, for each sample 4 µL of Superscript VILO Master Mix was combined with 100 ng of RNA and made up to 20 µL with nuclease-free water. The reaction mix was incubated at 25 °C for ten minutes, 42 °C for one hour, then 85 °C for five minutes. cDNA checked by Nanodrop spectrophotometer for quantity and quality (concentration range from 1700-1925 ng/µL), and was stored at -20 °C until needed.

Primers (Integrated DNA Technologies, Coralville, IA, USA) were designed specific to each novel ORF, generally 20-22 bp in length (annealing temperature ~54 °C), amplifying a segment ~ 100 bp in length (available in Appendix A.5). The constitutively expressed gene *TAF10* was used as a reference gene, as its constant expression in wine fermentation has been previously validated (Walkey *et al.*, 2012). Primers were checked by standard PCR and visualized by agarose gel electrophoresis to ensure they amplified a single amplicon of the correct size. The reaction mix was 10 µL of PowerUp SYBR Green 2X Master Mix (Applied Biosystems, Forster City, CA, USA), 0.2 µL of each primer (10 µM), 7.6 µL of nuclease-free water, and 2 µL of sample cDNA. Reactions were conducted in 96-well plates in a thermocycler (Applied Biosystems 7500 Real Time PCR System) with 40 cycles of 15 seconds at 95 ºC denaturation and 60 seconds at 60 ºC anneal/extension. Data were normalized by the expression of *TAF10*, using cDNA from day two of fermentation as the baseline expression. All sample/primer combinations were run in duplicate.

2.12 Spot assays for phenotypic traits

Spot assays were used to screen for potential phenotypes of the novel ORF deletion mutants. The base approach for all assays began with the inoculation of mutant and wild type cultures from glycerol stocks into YPD. These cultures were grown approximately 16 hours at 30 °C with shaking, then one OD600 of cells was harvested, pelleted and suspended in 1 mL of water. Ten times serial dilutions were performed four times, resulting in OD600 values of 1, 0.1, 0.01, 0.001 and 0.0001 for each sample. From these dilutions, 4-6 µL was spotted onto the appropriate agar plate, which was then sealed with parafilm and incubated at 30 °C for 48 hours, or 19 °C for 5-7 days until colonies were easily visible. Plates were then photographed and visually evaluated for differences in growth between mutants and wild type. Initial trials were conducted with no replicates; if a difference was seen, the assay was conducted a second time in triplicate to confirm results.

The agar plates used for phenotyping were either carbon source based, or inhibitor based. Carbon source plates contained 0.67% w/v Difco yeast nitrogen base (YNB) (Becton, Dickinson and Company, Sparks, MD, USA), 2% w/v agar and one of the following: 2% w/v glycerol, 2% w/v ethanol, 2% w/v dextrose, or 2% w/v galactose. Inhibitor plates contained 2% w/v YPD, 2% w/v agar and one of the following: 11% v/v ethanol, 15% v/v ethanol, 0.1 g/L copper (II) sulphate, 0.04% w/v potassium metabisulphite, or 1% w/v potassium metabisulphite. Additional tolerance tests were conducted for freeze/thaw resistance, with outgrowth on standard YPD agar plates. Freeze/thaw resistance was tested by harvesting 1 OD600 of cells from overnight cultures and suspending them in yeast minimal media (YMM, 0.67% w/v Difco YNB, 2% w/v glucose). Tubes were placed at -80 °C for one hour, then thawed at 30 °C for 20 minutes. Cell cultures were then serially diluted and spotted to plates as described previously.

2.13 Bioinformatic analyses of novel ORFs

2.13.1 BLAST homology

The nucleotide sequence of each novel ORF was checked for homologues in the *S. cerevisiae* S288C reference genome with the Basic Local Alignment Search Tool (BLAST) (Boratyn *et al.*, 2013). The BLAST search was conducted with the default BLASTx 2.6.1 parameters (expect threshold = 10, word size = 11, match/mismatch scores = 2, -3, gap costs = existence: 5 extension: 2, low complexity regions filtered) but with maximum target sequences limited to 10. The BLASTx tool queries the translated nucleotide sequence against a protein sequence database. The nucleotide sequence of each novel ORF was also queried against all non-redundant GenBank coding DNA sequence (CDS) translations by BLASTx (expect threshold = 10, word size = 6, match/mismatch scores = 2, -3, gap costs = existence: 11 extension: 1, low complexity regions filtered).

2.13.2 Pfam domain homology

The translated protein sequence of each novel ORF was checked for conserved protein domains by searching against the protein family (Pfam) database using HMMER web version 2.13.4 (Finn *et al.*, 2015). The search was run with the default settings (-E 1 --domE 1 --incE 0.01 --incdomE 0.03 --mx BLOSUM62 --pextend 0.4 --popen 0.02 --seqdb uniprotrefprot), which searches the UniProt reference database with an e-value cut-off of one.

3 RESULTS AND DISCUSSION

3.1 The Enoferm M2 genome contains novel ORFs

3.1.1 Enoferm M2 novel ORFs are not found in the S288C reference genome

Prior work had assembled the Enoferm M2 genome (van Vuuren laboratory, unpublished), scaffolded to the genome of EC1118, another commonly used commercial wine strain. Several contigs were identified that were unable to be accurately mapped to the genome, likely due to being located in subtelomeric regions. Subtelomeric regions contain repeated elements that make them hot spots of recombination and difficult to sequence (Strope *et al.*, 2015). The Enoferm M2 contigs contained a number of putative ORFs, some of which have been investigated in previous thesis work (Iwashita, 2016). The presented thesis work involves 15 ORFs identified from Enoferm M2 contigs not mapped to the EC1118 genome. To determine their novelty compared to the reference genome, the 15 ORFs were aligned using the BLAST tools to determine best matches in the S288C reference genome (Table 3.1).

Fourteen out of the 15 novel ORFs had at least one BLAST hit within the given parameters. Only M25 had no significant hit with an E-value cut-off of 10. The maximum percentage identity was 94.6, and the maximum percentage positives (includes synonymous SNPs) was 97.2, both for M23. The M23 ORF is a putative hexose transporter; hexose transporters are a large family in *S. cerevisiae* encoded by many closely related genes. The fact that no ORF possesses more than 95% identity with the reference genome leads us to the conclusion that the 15 ORFs, while potentially paralogues, are novel and not present in the reference genome.

Table 3.1 Novel ORF BLAST results. Sequences of fifteen novel ORFs from the Enoferm M2 genome were searched against the reference database of S. cerevisiae protein encoding sequences using BLAST. All lengths are reported in amino acids.

Novel ORF	S288C Best Protein Match	Percent Identity	Percent Positives	Novel ORF Length	S288C Length	Alignment Length	Mismatches	Gap Opens	E-value	Bit Score
M6	*DIP5*	36.87	56.51	447	608	453	274	6	9.5E-81	259
M13	*DUR1,2*	28.06	44.90	641	1835	196	125	6	2.0E-09	59
M14	*EXO5*	32.50	47.50	190	585	40	27	0	4.3E+00	26
M15	*AFI1*	35.56	55.56	266	893	45	24	1	5.3E+00	27

Novel ORF	S288C Best Protein Match	Percent Identity	Percent Positives	Novel ORF Length	S288C Length	Alignment Length	Mismatches	Gap Opens	E-value	Bit Score
M16	*AAD16*	60.22	77.96	228	153	186	60	2	4.0E-77	234
M17	*ADH7*	80.22	91.64	359	361	359	71	0	0.0E+00	552
M18	*RDS1*	56.30	75.15	827	832	833	357	4	0.0E+00	934
M19	*AAD16*	66.29	85.14	216	153	175	57	2	4.3E-84	251
M20	*AAD14*	84.53	93.33	375	376	375	58	0	0.0E+00	676
M21	*THI73*	32.56	50.23	447	523	430	267	11	2.4E-54	187
M22	*TDP1*	31.25	62.50	304	544	32	22	0	7.3E-01	30
M23	*HXT17*	94.59	97.19	481	564	462	25	0	0.0E+00	897
M23	*HXT13*	94.59	97.19	481	564	462	25	0	0.0E+00	896
M25	None found.	-	-	175	-	-	-	-	-	-
M26	*ATO3*	47.48	62.23	275	275	278	129	6	3.1E-48	159
M28	*CSS1*	88.46	88.46	299	995	26	3	0	2.0E-03	38

3.1.2 Enoform M2 ORFs share similarity with other wine strains and non-*Saccharomyces* yeast

In addition to searching against the *S. cerevisiae* to determine the novelty of the Enoform M2 ORFs, the nucleotide sequences were also queried against the GenBank non-redundant CDS database to determine the best similarity. The GenBank database contains sequences from yeast species, as well as thousands of other eukaryotic and prokaryotic species, and this homology search identified hits from the yeast strain AWRI796 as the top match for many of the novel ORFs. The AWRI796 strain was also isolated from the Stellenbosch region of South Africa, and may share recent lineage with Enoform M2. In all, AWRI796 was the best match for nine novel ORFs, as well as being the second-best match for two other novel ORFs. For all the matches, AWRI796 showed 100% identity match although this did not always cover the full alignment length, with the exception of two ORFs, M6 and M16, which had slightly lower percent identities of 99.55 and 93.55, respectively. Three novel ORFs (M13, M21 and M22) show 100% positive amino acid matches with the strain R008, which was isolated from the Piave Appellation of Origin in Italy (Treu, Campanaro, *et al.*, 2014). The M14 and M15 ORFs are highly similar to *S. cerevisiae MEL5*/*MEL6* genes, but these are verified genes classified as "not in the systematic sequence of S288C".

In addition to matches within *S. cerevisiae*, many of the novel ORFs have strong matches with other fungal and bacterial genes (Table 3.2). The M6 ORF's top hit with 100% positives and a full alignment length was a tyrosine permease from *Saccharomyces pastorianus*. The M16 ORF's second hit with 84.2% positives, but only a partial alignment length was an oxidoreductase from *Lachancea thermotolerans*. The M17 ORF's top hit with 93.9% positives was an ADH7-like protein from *Saccharomyces kudriavzevii*. The M19 ORF had a strong hit, with 85.7% positives and moderate alignment length, which was a putative protein from *Lachancea dasiensis*. The M21 ORF had a strong hit, with 80.0% positives and good alignment length, which was a putative allantoate permease from the methylotrophic yeast *Ogataea parapolymorpha*. The M22 ORF had a strong hit, with 80.9% positives and good alignment length, which was a class IV aminotransferase from *Methylobacterium platani*. This collection of significant hits in different species and genera may indicate possible xenologs that have introgressed into Enoferm M2, or orthologues that have been lost in the reference genome.

Table 3.2 Novel ORF BLAST results against CDS database. DNA sequences of fifteen novel ORFs were queried against the non-redundant CDS database using BLAST algorithm with default settings. Shaded entries represent strong homology to the wine strain AWRI796. All lengths are reported in amino acids.

Query ID	Subject ID	Percent Identity	Percent Positives	Alignment Length	Mis-matches	Gap Opens	E-value	Bit Score	Notes
M6	BAF57236.1	99.776	100	447	1	0	0	857	tyrosine permease [*Saccharomyces pastorianus*]
M6	EDV10956.1	99.553	99.78	447	2	0	0	851	AWRI796
M13	EGA72492.1	100	100	641	0	0	0	1209	AWRI796 (amidase)
M13	EWG83447.1	99.834	100	602	1	0	0	1129	R008
M13	CAA85738.1	98.73	99.09	551	7	0	0	1016	ORF near MEL [*Saccharomyces cerevisiae*]
M14	CAA85740.1	99.474	100	190	1	0	1.57E-132	388	Alpha-galactosidase [*Saccharomyces cerevisiae*] *MEL5*
M14	CAA85739.1	99.474	100	190	1	0	1.59E-132	388	*MEL6*
M15	CAA85740.1	100	100	266	0	0	0	563	Alpha-galactosidase [*Saccharomyces cerevisiae*] *MEL5*
M16	EGA72495.1	93.548	93.55	186	0	1	2.76E-119	352	AWRI796, predicted oxidoreductase

Query ID	Subject ID	Percent Identity	Percent Positives	Alignment Length	Mis-matches	Gap Opens	E-value	Bit Score	Notes
M16	CAR21362.1	69.565	84.24	184	44	1	1.59E-88	273	*Lachancea thermotolerans*, predicted oxidoreductase
M17	EJT44343.1	84.401	93.87	359	56	0	0	575	*ADH7*-like protein [*Saccharomyces kudriavzevii*]
M18	EGA72498.1	100	100	647	0	0	0	1299	AWRI796
M18	AJQ43916.1	56.543	75.51	833	355	4	0	959	*RDS1* [*Saccharomyces cerevisiae*]
M19	EGA72500.1	100	100	152	0	0	8.27E-107	318	AWRI796
M19	SCU81064.1	74.857	85.71	175	44	0	3.01E-90	277	*Lachancea dasiensis*
M20	EWG83449.1	100	100	375	0	0	0	780	*AAD4* AWRI796
M20	ONH78529.1	87.2	94.13	375	48	0	0	687	*AAD14 Saccharomyces cerevisiae*
M21	EWG83450.1	100	100	432	0	0	0	802	R008
M21	EGA72502.1	100	100	320	0	0	0	629	AWRI796
M21	ESW99896.1	65.412	80	425	139	2	2.79E-176	511	putative allantoate permease [*Ogataea parapolymorpha*]
M22	EWG83451.1	100	100	304	0	0	0	600	R008
M22	KMO11958.1| KMO13983.1	68.792	80.87	298	93	0	6.32E-140	406	aminotransferase class IV [*Methylobacterium platani*]
M23	EGA72506.1	100	100	463	0	0	0	944	*HXT13* AWRI796
M25	EGA72507.1	100	100	116	0	0	4.70E-66	206	AWRI796
M26	EGA72508.1	100	100	262	0	0	8.68E-158	448	AWRI796
M26	EHM99598.1	84.388	88.61	237	36	1	5.18E-109	323	*ATO3* VIN7
M28	EGA74324.1	100	100	60	0	0	1.01E-31	127	AWRI796

3.1.3 Pfam homologies complement BLAST alignment results

Complementary to the BLAST searches, the novel ORFs were queried against the Pfam database using the HMMER tool, allowing the identification of conserved protein domains that may give clues to the function(s) of the novel ORF. Significant hits were only found for 13 of the 15 novel ORFs queried, with M25 and M28 yielding no conserved protein domains (Table 3.3). For the most part, the Pfam results validated the BLAST results, which makes sense as the Pfam database is a key tool in the

automated annotation of new genomes. The Pfam results, however, gave a better indication of the potential function of the ORF, while the BLAST results gave more information about the relationship of the ORF to ORFs found in other organisms.

Table 3.3 Pfam homology of novel ORFs. Novel ORF amino acid sequences were searched against the Pfam database using the HMMER tool. Only top hits from distinct family ids are reported.

ORF	Family id	Family Accession	Bit Score	Independent E-value	Conditional E-value	Description
M6	AA_permease	PF00324.20	266.59	3.30E-79	4.00E-83	Amino acid permease
M13	Amidase	PF01425.20	119.43	1.80E-34	1.10E-38	Amidase
M14	Melibiase_2	PF16499.4	21.58	9.60E-05	5.80E-09	Alpha galactosidase A
M15	Melibiase_2	PF16499.4	152.11	1.60E-44	1.90E-48	Alpha galactosidase A
M16	Aldo_ket_red	PF00248.20	101.66	3.80E-29	2.30E-33	Aldo/keto reductase family
M17	ADH_N	PF08240.11	91.9	2.00E-26	6.00E-30	Alcohol dehydrogenase GroES -like domain
M17	ADH_zinc_N	PF00107.25	63.32	1.80E-17	5.50E-21	Zinc-binding dehydrogenase
M17	AlaDh_PNT_C	PF01262.20	25.62	5.90E-06	1.80E-09	Alanine dehydrogenase/PNT
M18	Zn_clus	PF00172.17	31.23	1.60E-07	9.50E-12	Fungal Zn(2)-Cys(6) binuclear cluster domain
M19	Aldo_ket_red	PF00248.20	116.72	9.80E-34	5.90E-38	Aldo/keto reductase family
M20	Aldo_ket_red	PF00248.20	237.97	1.10E-70	6.70E-75	Aldo/keto reductase family
M21	MFS_1	PF07690.15	121.46	3.50E-35	2.10E-39	Major Facilitator Superfamily
M22	Aminotran_4	PF01063.18	131	5.30E-38	3.20E-42	Amino-transferase class IV
M23	Sugar_tr	PF00083.23	501.18	2.50E-150	3.00E-154	Sugar (and other) transporter
M23	MFS_1	PF07690.15	88.27	4.40E-25	5.20E-29	Major Facilitator Superfamily
M26	Grp1_Fun34_YaaH	PF01184.18	230.69	1.20E-68	7.30E-73	GPR1/FUN34/yaaH family

Some ORFs had multiple hits to protein domains with the same categorization, as many proteins are multi-domain. For instance, the top hits for M6 were both amino acid permease domains. The top hit for M13 was an amidase domain. The M14 and M15 ORFs both had alpha galactosidase A or melibiase domains, complementary to their BLAST results. The M14 and M15 ORFs are thought to encode a single functional protein, but they are separated by a stop codon in many strains of *S. cerevisiae*, including Enoferm M2. The M16, M19 and M20 ORFs all possessed the same protein domain that belongs to the aldo/keto reductase family. The M17 protein coding sequence had multiple significant hits for related dehydrogenase families, including an alcohol dehydrogenase GroES-like domain, a zinc-binding dehydrogenase domain and an alanine dehydrogenase domain. The M18 ORF putatively encodes a protein that matched a fungal Zn(2)-Cys(6) binuclear cluster domain, which binds DNA and could be

indicative of a transcription factor. The M21 ORF codes for a domain that belongs to the Major Facilitator Superfamily, a family which includes an allantoate permease as suggested by BLAST homology. The M22 ORF had an aminotransferase class IV domain. The M23 ORF also contains a Major Facilitator Superfamily domain, but also more specifically had a sugar transporter domain, consistent with this ORF's strong homology to known hexose transporters. Finally, the M26 ORF possessed a GPR1/FUN34/yaaH domain, consistent with this ORF's BLAST homology to ATO3, a putative yaaH family ammonium permease from the yeast strain Vin7. No results were found for either M25, which also had no BLAST results, or M28, which had the shortest BLAST alignment.

3.1.4 Novel ORFs are transcribed during wine fermentations

Isolated RNA from wild type Enoferm M2 was analyzed by qRT-PCR for the expression of the novel ORFs during exponential growth in YPD, as well as at four time points during a Chardonnay wine fermentation. Primer pairs were designed to be specific to each novel ORF, and all primer pairs amplified a single amplicon when tested by colony PCR of Enoferm M2. Novel ORF M25 not amplifiable using the designed primers. The lack of amplicon, combined with the lack of any sequence homology for this novel ORF, indicates that M25 is likely a dubious ORF.

When grown in YPD, expression of the novel ORFs was generally poor, and many primer sets failed to amplify any transcripts at all. When grown in Chardonnay grape juice, however, the novel ORF transcripts could be amplified at all time points tested. To determine the relative expression pattern of the transcripts, day two of fermentation was taken as a reference value of one, with expression levels expressed as relative values compared to this reference point. It should be noted that the qRT-PCR method employed in this study does not yield an absolute transcript quantity, only relative expression between samples. The standard deviation of samples also generally increased as the fermentation progressed, which could be a sign of RNA degradation later in cell life.

The novel ORFs showed different patterns of expression during the fermentation. Eleven out of the fourteen novel ORFs demonstrated an overall increase in expression towards the end of fermentation,

although the magnitude and timing of the increase varied (Figure 3.1). Of note, five ORFs (M13, M14, M15, M18 and M28) increased in expression more than five-fold from day two – day seven of fermentation, with M14, M15 and M28 increasing in expression more than 15-fold. Six ORFs (M6, M16, M17, M19, M20 and M23) had moderate increases in expression, ranging from 2.4- to 4.7-fold increases. Three ORFs (M21, M22 and M26) were expressed relatively constantly throughout the fermentation, indicating constitutive expression, although the quantity of transcripts per cell is not known.

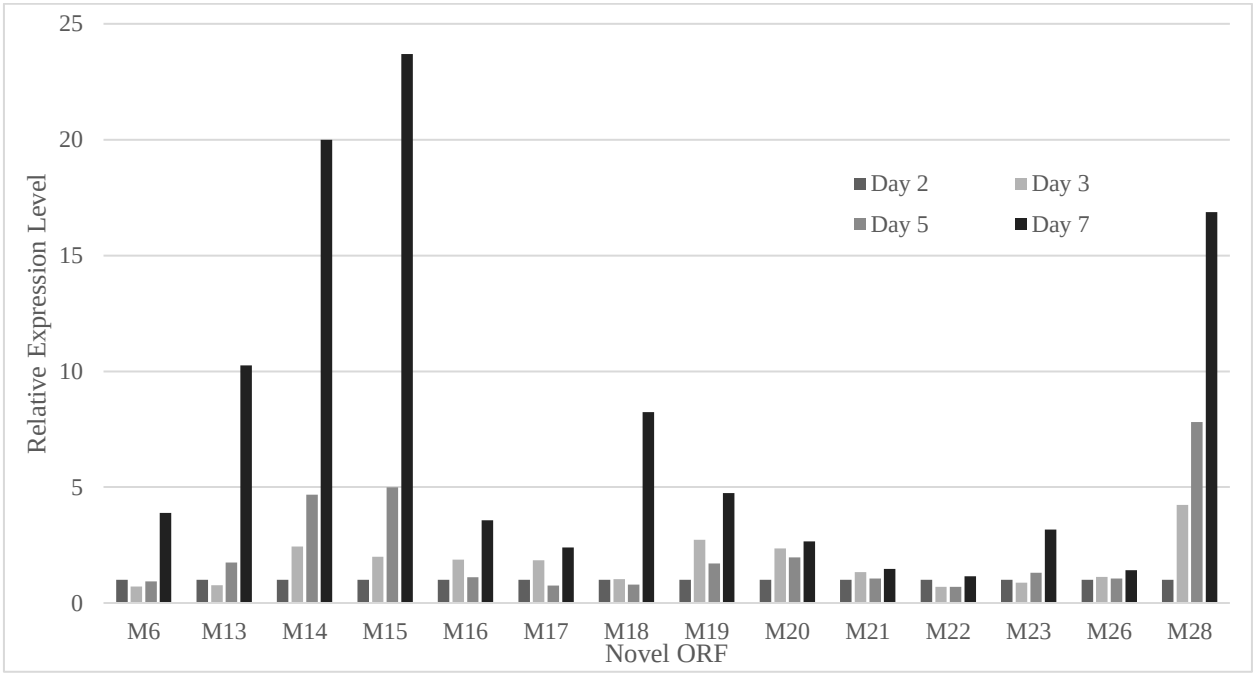


Figure 3.1 Transcript expression of novel ORFs. Wild type M2 mRNA was isolated and tested by qRT-PCR to determine relative transcript levels of novel ORFs at four time points during Chardonnay fermentation. Day two was taken as the reference value of one, and other time points are expressed as relative expression levels.

By designing primers specific to the novel ORFs, we were able to use qRT-PCR to determine the mRNA transcript based expression level of the novel ORFs at multiple time points of a wine fermentation. In this manner, we confirmed that fourteen of the fifteen novel ORFs studied were transcribed during wine fermentation. By using multiple time points it was possible to assess when these novel ORFs were most highly expressed during the fermentation. Furthermore, we had difficulty amplifying the mRNA transcripts of the ORFs from wild type Enoferm M2 grown under optimal conditions (aerobic growth in YPD broth at 30 ºC), which is a potential indicator that the expression of the novel ORFs investigated is specific to wine fermentation or other stress conditions. Overall, the

evidence gathered from the qRT-PCR portion of this study can be used to justify the classification of these ORFs as "uncharacterized" at a minimum, and to justify the inclusion of these ORFs in the *S. cerevisiae* pan-genome.

3.2 The mutation of ORFs has varied impacts on wine fermentations

3.2.1 Deletion of novel ORFs does not affect wine fermentation rate

Chardonnay wine fermentations were conducted using the homozygous diploid null mutants for novel ORFs and the M2 wild type as a control. Progress was measured by weight loss, a measure of carbon dioxide produced during the fermentation process (Figure 3.2), and ANOVA was used to determine whether there was any significant difference between the null mutants and wild type. No significant difference was found for genotype ($p = 0.91$), indicating that the homozygous deletion of any novel ORF did not have a significant impact on fermentation performance. In addition, as the weight loss trials were conducted in three batches to facilitate handling, each with five novel ORF mutants and wild type, batch effects were also tested. No significant difference was found for batch ($p = 0.73$), indicating that results from fermentations inoculated at different times but conducted under the same conditions are directly comparable.

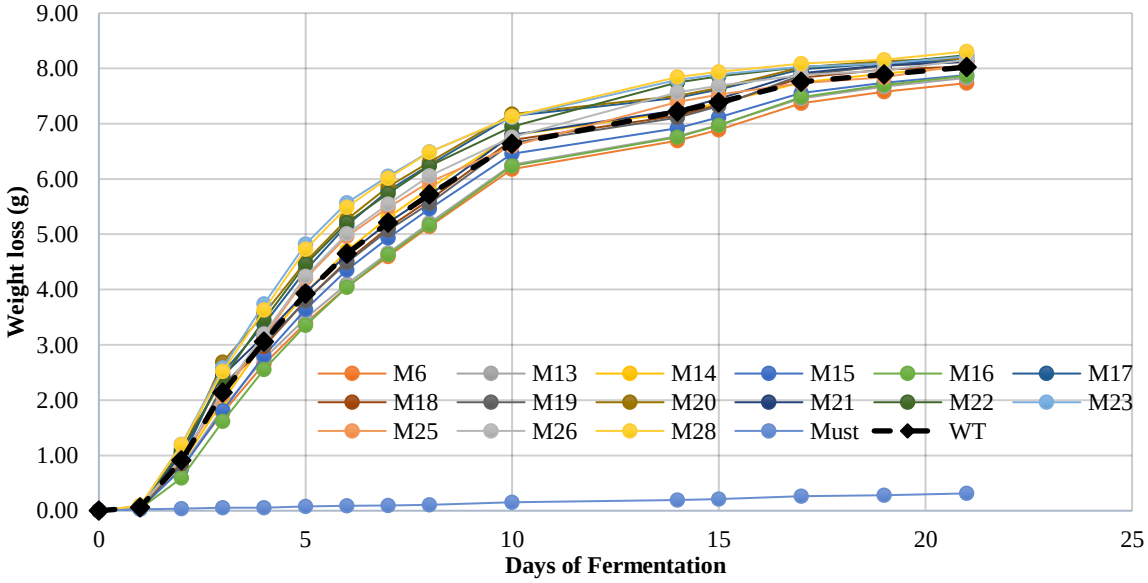


Figure 3.2 Wine fermentation progress. The fermentation of Chardonnay wine by Enoferm M2 and the novel ORF mutants was monitored by weight loss as a measure of CO_2 production, an indicator of fermentation activity. Points are the mean of three biological replicates.

3.2.2 Primary metabolite profiles result in significant differences for some novel ORF mutants

3.2.2.1 Novel ORF null mutants do not result in significant effects on primary metabolite profiles.

Primary metabolites such as glucose, fructose, ethanol, glycerol and acetic acid were monitored throughout a second, independent Chardonnay fermentation (the first fermentation measured weight loss) by analyzing wine samples by HPLC (Figure 3.3), and ANOVA was conducted to determine significant differences between wild type and null mutant genotypes. Two biological replicates of each genotype were used. Genotype was found to cause a significant difference for all five metabolites tested (ethanol, $p = 2.87e^{-9}$; glucose, $p = 4.04e^{-9}$; fructose, $p = 1.04e^{9}$; glycerol, $p = 7.27e^{-6}$; acetic acid, $p = <2.00e^{-16}$), with no significant genotype-day interaction effect found. Tukey's HSD was applied, and this determined that the only mutant with significant differences from wild type for ethanol, glucose, fructose and glycerol was M13. No significant difference was found for acetic acid for any mutant compared to wild type.

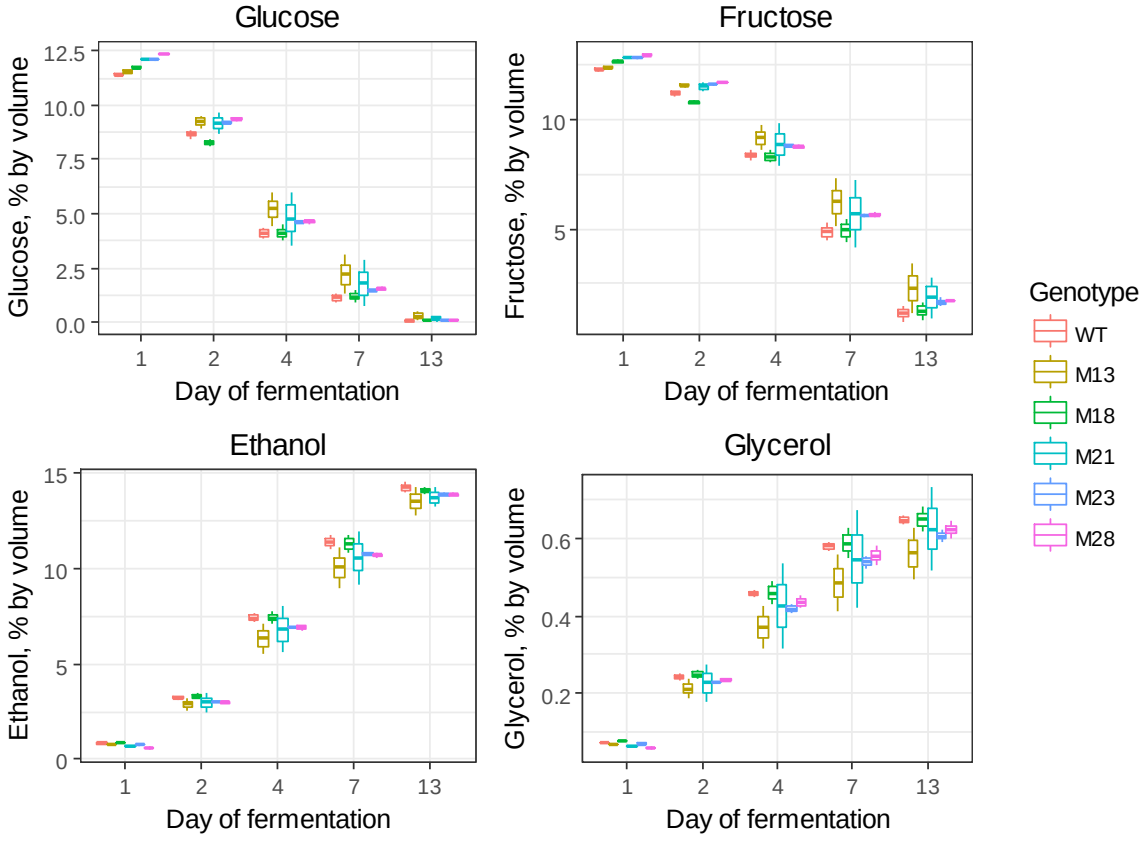


Figure 3.3 HPLC monitoring of wines produced by null mutants. Boxplot (box from 25th to 75th quartiles, horizontal line = median, vertical line = range) summary of key fermentation metabolites profiled by HPLC from samples taken during Chardonnay wine fermentation from Enoferm M2 wild type and select novel ORF null mutants.

A boxplot of 95% confidence intervals for the data, however, identified several outliers (not shown in figure). The most extreme outlier is one of the M13 replicates, which appears to have exhibited a slow growth phenotype not observed during the independently conducted weight loss monitoring.

3.2.2.2 Novel ORF constitutive expression mutants yield significant effects for multiple primary metabolites.

The same primary metabolites were monitored throughout a third, independent Chardonnay fermentation (Figure 3.4), with ANOVA and Tukey's HSD used as described above to determine possible significant differences between novel ORF constitutive expression mutants and wild type. Genotype was found to cause a significant difference for all five metabolites tested (ethanol, $p = 1.53e^{-9}$; glucose, $p = 1.74e^{-5}$; fructose, $p = 5.54e^{-9}$; glycerol, $p = 1.11e^{-6}$; acetic acid, $p = 1.15e^{-12}$), with significant genotype-day interaction effects found for glucose ($p = 2.60e^{-6}$), fructose ($p = 9.08e^{-4}$) and acetic acid ($p = 0.027$). Only the M21 constitutive expression mutant was significantly slower ($p = 0.00250$) than wild type with regards to ethanol production. With regards to glucose consumption, only the M18 constitutive expression mutant was significantly faster ($p = 0.0150$) from wild type. The fructose consumption of both M18 and M23 was significantly faster ($p = 3.37e^{-5}$ and $1.59e^{-3}$, respectively) than wild type. The glycerol production was not significantly different from wild type for any individual mutant; in this case the overall significance of the ANOVA could be from large sample variation. The acetic acid production of the M13, M21 and M28 constitutive expression mutants were all significantly lower (M13, $p = 3.85e^{-5}$; M21, $p = 1.44e^{-3}$; M28, $p = 0.0100$) than wild type.

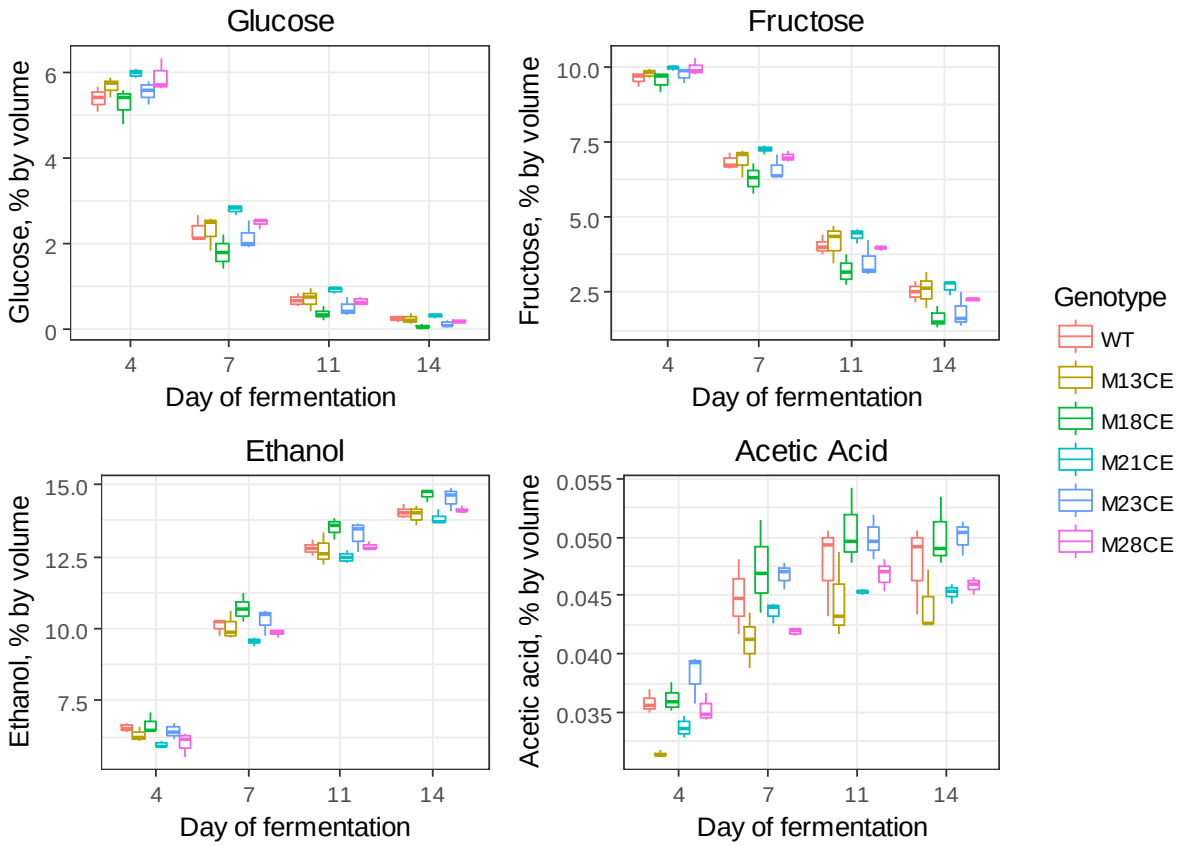


Figure 3.4 HPLC monitoring of wines produced by constitutive expression mutants. Boxplot (box from 25th to 75th quartiles, horizontal line = median, vertical line = range) summary of key fermentation metabolites profiled by HPLC from samples taken during Chardonnay wine fermentation from Enoferm M2 wild type and novel ORF constitutive expression mutants.

3.2.2.3 Transcription factor and other ORF null mutants strongly affect metabolite profiles.

A fourth, independent Chardonnay fermentation was conducted using ten putative transcription factor null mutants and seven other poorly annotated ORFs of enological interest, with samples regularly analyzed by HPLC (Figure 3.5). The ANOVA was highly significant ($p < 1e^{-13}$) for the metabolites analyzed: ethanol, glucose, fructose, glycerol and acetic acid. Glucose was the only metabolite that had a significant interaction effect with the day of sampling ($p = 0.000321$). The strong significance of the ANOVA appeared to be driven by four genotypes, null mutants for YFL052W, *PHD1*, *GEP5* and *HXK2*. Based on Tukey's HSD, these four genotypes are significantly different from wild type for three (YFL052W), four (*PHD1* and *HXK2*) or all five (*GEP5*) metabolites analyzed (Table 3.4). The YFL052W and *PHD1* null mutants had improved fermentation performance over wild type, both consuming glucose faster and producing ethanol faster. In addition, the *PHD1* null mutant produced

more glycerol than wild type. The *HXK2* and *GEP5* null mutants both had slow growth phenotypes, taking longer to consume glucose and produce ethanol than wild type. The *GEP5* null mutant also produced considerably more acetic acid than wild type.

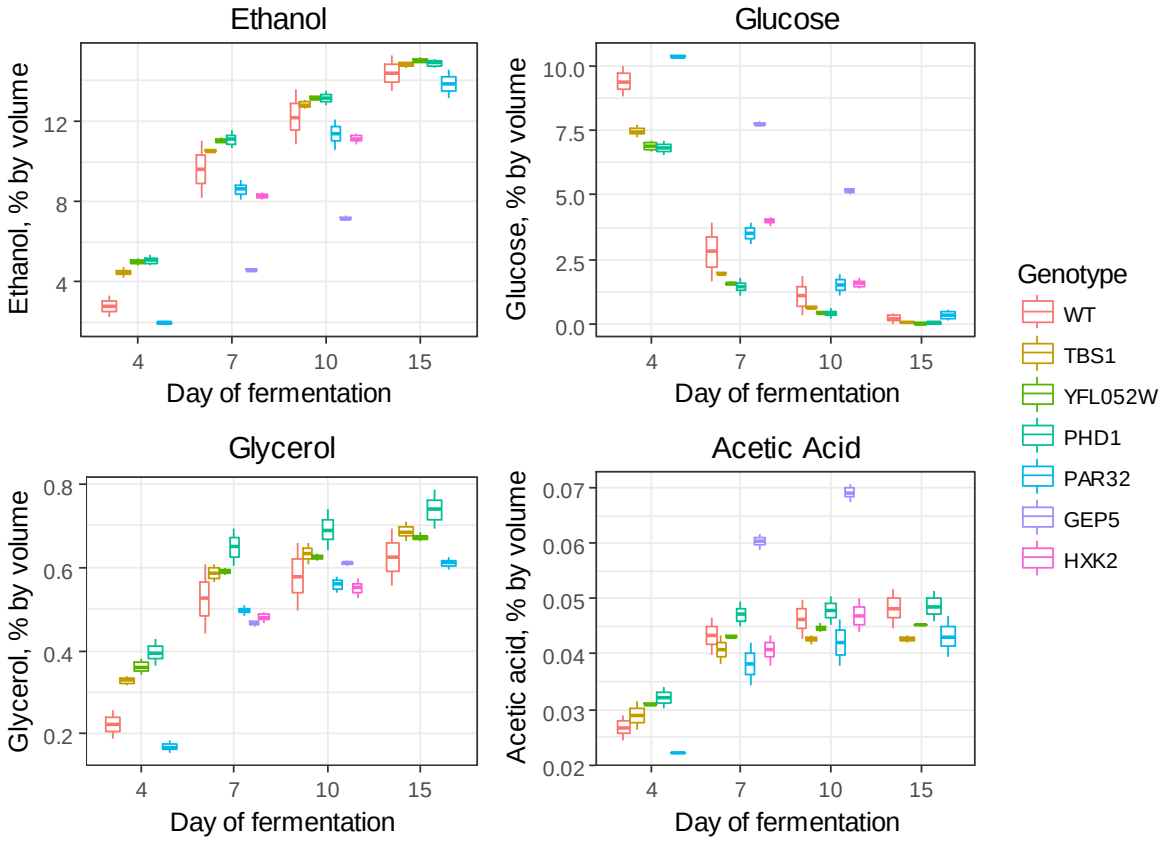


Figure 3.5 HPLC monitoring of wines produced by non-novel null mutants. Boxplot summary (box from 25th to 75th quartiles, horizontal line = median, vertical line = range) of key fermentation metabolites profiled by HPLC from samples taken during Chardonnay wine fermentation from Enoferm M2 wild type and non-novel null mutants. Only mutants with a significantly different result are shown.

Table 3.4 Significant differences in metabolites produced by non-novel ORF null mutants. HPLC analyzed fermentation metabolites from non-novel ORF null mutants were compared by ANOVA, and then Tukey's HSD *post hoc* test. Tukey's HSD *p*-values are reported for significantly different mutants (*p*-value < 0.05).

	Tukey's HSD *p*-value				
Null Mutant	**Ethanol**	**Glucose**	**Fructose**	**Glycerol**	**Acetic Acid**
TBS1	-	0.00218	-	-	-
YFL052W	0.00104	$2.18e^{-6}$	-	0.0241	-
PHD1	0.00130	$8.63e^{-7}$	0.0398	$1.25e^{-6}$	-
YLL054C	-	0.0453	-	-	-
YLR278C	-	0.0454	-	-	-
PAR32	-	-	-	-	0.0381
GEP5	$1.91e^{-6}$	$6.15e^{-5}$	$3.83e^{-6}$	$1.29e^{-6}$	$3.99e^{-10}$
HXK2	$3.96e^{-6}$	$3.99e^{-10}$	-	$9.40e^{-5}$	$3.92e^{-9}$

3.2.3 Volatile metabolites produced were highly variable

Upon completion of the fermentation, wine samples were analyzed by SPME-GCMS to detect volatile compounds present in the headspace of the wine, which would be perceived as aromatic or flavouring compounds by a human. Without running a standard curve for each compound, absolute quantification cannot be reported, but by using a spiked internal standard the relative quantity of individual compounds can be compared across samples. Twenty compounds were accurately detected from the wine produced by the novel ORF null mutants, although ethanol and carbon dioxide were excluded from the analysis, leaving 18 compounds to be analyzed for differences. Twenty-four compounds were accurately detected from the wines produced by the constitutive expression mutants, leaving 22 compounds for comparison after excluding ethanol and carbon dioxide. While significant differences were found for some compounds, many compounds had large standard deviations both between genotypes and between biological replicates. Given the small sample size (n = 3 per genotype), the large standard deviation made it harder to assess significant differences.

3.2.3.1 Novel ORF null mutants produced significantly different volatiles than wild type.

Eleven of the 18 compounds produced by the novel ORF null mutants and detected by GCMS had significant ($p < 0.05$) ANOVA results, with p-values ranging from 0.0218 down to $7.26e^{-6}$ (Table 3.5). When Tukey's HSD was applied to these results, no mutants were significantly different than wild type with respect to three compounds (isobutyl acetate, ethyl hexanoate and hexanol) despite yielding a significant ANOVA p-value. Of the remaining eight compounds, four compounds (ethyl acetate, ethyl butyrate, isoamyl acetate, isobutyl acetate) were significantly different for the M23 null mutant, two compounds (ethyl octanoate, ethyl decanoate) were significantly different for the M15 null mutant, two compounds (2,3-butanediol, *p*-ethylacetophenone) were significantly different for the M28 null mutant and one compound (ethyl dodecanoate) was significantly different for the M19 null mutant (Table 3.5).

Table 3.5 GCMS analysis of wines produced by novel ORF null mutants. Eighteen volatile compounds were compared between wild type and novel ORF null mutants by ANOVA and Tukey's HSD. Three biological replicates were used for each genotype. Only *p*-values less than 0.05 are reported as significant.

Compound	ANOVA *p*-value	Significant mutants vs. wild type	Tukey's HSD *p*-value
Ethyl acetate	0.00102	M23	0.00622
Isobutyl acetate	0.00595	M23	0.0420
Ethyl butyrate	0.000319	M23	0.0194
Propanol	0.713	N/A	N/A
Isobutyl alcohol	0.0766	N/A	N/A
Isoamyl acetate	0.00686	M23	0.0247
Isopentyl alcohol	0.424	N/A	N/A
Ethyl hexanoate	0.0218	None	N/A
Hexyl acetate	0.214	N/A	N/A
Hexanol	0.0166	None	N/A
Ethyl octanoate	0.0825	M15	0.00981
2,3-Butanediol	0.0288	M28	0.0241
Ethyl decanoate	0.0131	M15	0.00412
Ethyl 9-decanoate	$1.26e^{-5}$	None	N/A
Acrylic acid, 2-phenylethyl ester	0.141	N/A	N/A
Ethyl dodecanoate	0.0102	M19	0.00321
Phenylethyl alcohol	0.344	N/A	N/A
P-ethylacetophenone	$7.26e^{-6}$	M28	0.000133

3.2.3.2 Novel ORF M23 affects both sugar uptake and ester formation.

When the headspace of the wine samples was analyzed for differences in volatile secondary metabolites only null mutants, and not constitutive expression mutants, caused any significant change from wild type. One null mutant, M23, stood out as it had significantly altered volatile abundances compared to wild type for four different compounds. The M23 null mutant demonstrated significantly higher levels of four esters: ethyl acetate, ethyl butyrate, isobutyl acetate and isoamyl acetate. The M23 constitutive expression mutant also had altered metabolic profiles, but no difference in volatile composition, and particularly no concurrent decrease in any of the four esters that were increased in the null mutant. The

M23 constitutive expression mutant had significantly greater fructose consumption than wild type; the null mutant had slightly less fructose consumption than wild type, but not significantly so.

The sequence of M23 is highly similar to known polyol and hexose transporters *HXT17* and *HXT13*, which could account for a slight increase in fructose consumption when M23 is constitutively expressed. While the hexose transporters *HXT6* and *HXT7* have been implicated in the fermentation stress response (FSR) (Marks *et al.*, 2008), little is known of the biological role of *HXT13* and *HXT17*. The genome of the popular wine strain EC1118 also contains an *HXT13*-like ORF within a region containing other ORFs not found in the S288C reference sequence (Novo *et al.*, 2009). A recent study identified the most likely role for *HXT17* and *HXT13* to be specifically in facilitative transport of mannitol (Jordan *et al.*, 2016). This study also noted that *HXT17* and *HXT13* are located in similar chromosomal regions as YNR073C and *DSF1*, which are known mannitol dehydrogenases, and this co-localization phenomenon is associated with paired duplication and functionalization of associated genes. The M23 ORF, similarly, co-localizes on a contig with M17, which shares 80% identities with *ADH7* and may be an alcohol dehydrogenase. When deleted, however, there was no significant decrease in sugar uptake by the M23 null mutant, and so it is unclear by what mechanism this deletion could be affecting the production of volatile esters. The HXT transporters are known to be somewhat promiscuous with regards to substrate (Wieczorke *et al.*, 1999; Biswas *et al.*, 2013; Jordan *et al.*, 2016), and it could be that co-localization of the transporter with a particular alcohol dehydrogenase results in the improved conversion of the transported substrate. This could create minor effects in glycolytic fluxes, resulting in altered production of volatile esters.

3.2.3.3 Constitutive expression of novel ORFs yields no significant changes in volatiles.

None of the five constitutive expression mutants produced significantly different relative quantities of volatiles for any of the 23 different compounds measured by GCMS. The ANOVA was significant ($p < 0.05$) for two compounds, hexanol and hexanoic acid, but this appears to be because of general variation between the mutants, as Tukey's HSD found no significant difference between any one mutant and wild type. While not statistically significant, both hexanol and hexanoic acid appear to increase in the M21 constitutive expression mutant (Figure 3.6).

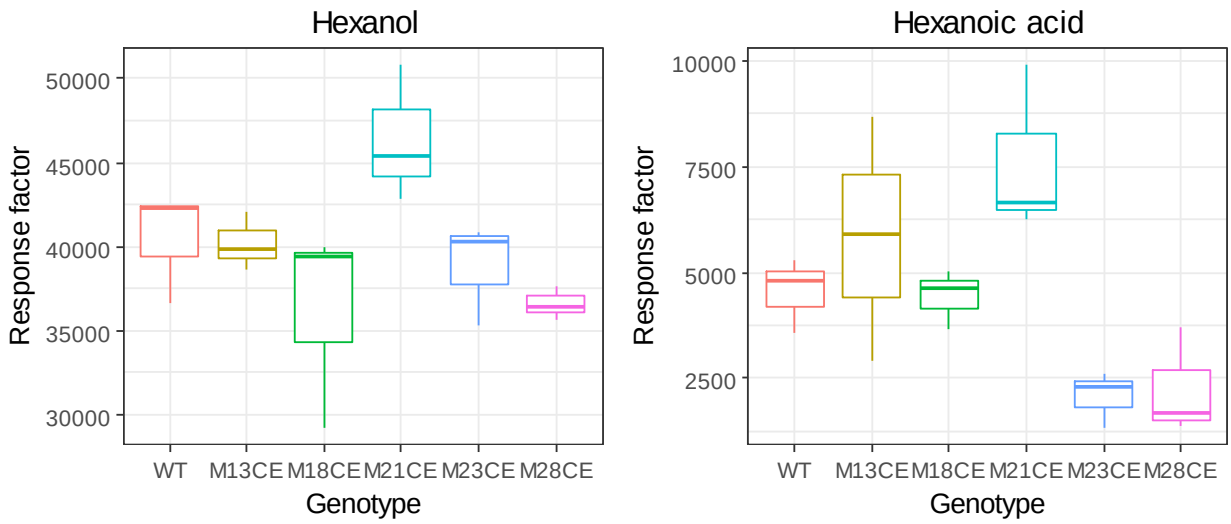


Figure 3.6 GCMS analysis of wines produced by novel ORF constitutive expression mutants. Boxplot summaries (box from 25th to 75th quartiles, horizontal line = median, vertical line = range) of hexanol and hexanoic acid quantities as analyzed by GCMS in final Chardonnay wine samples fermented by novel ORF constitutive expression mutants. Each genotype was analyzed in biological triplicate.

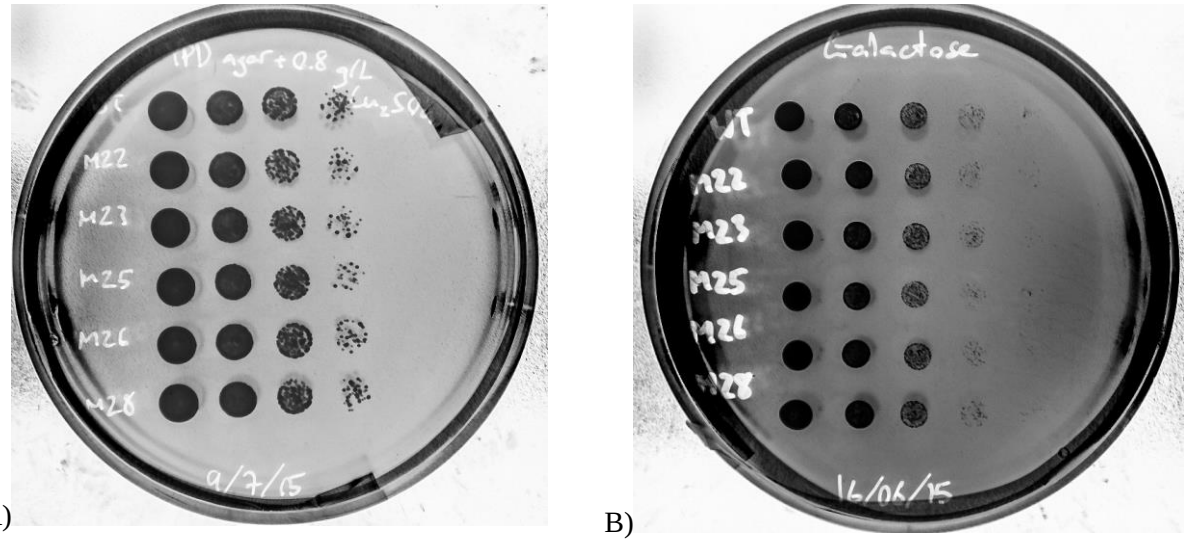
3.2.3.4 Volatile secondary metabolites are strongly affected by storage conditions.

Wines produced by transcription factor null mutants, as well as null mutants for other poorly-annotated ORFs of interest were not analyzed systematically by GCMS upon completion of fermentation. Even with sulphites added to the final wine samples and storage at 4 ºC, storage time was found to affect the consistency of results, and the slow fermentation phenotype of some of the transcription factor mutants made direct comparisons with wild type harder to accomplish.

3.2.4 Additional phenotypic assays yield no observable phenotype

Given the lack of major phenotype for the novel ORF null mutants when used to perform a standard wine fermentation, spot assays were conducted using various carbon sources (glycerol, ethanol, dextrose, galactose), inhibitors (ethanol, copper (II) sulphate, potassium metabisulphite) and stress tests (freeze/thaw) to determine if the mutant strains exhibited a visible phenotype when grown at either 19 ºC or 30 ºC. Initial tests were conducted as a single screening run, with a follow-up in triplicate to confirm if a growth phenotype was observed initially. No consistent growth phenotype was observed for any of the novel ORF null mutants tested (Figure 3.7). Given the large amount of labour required and the difficulty

in objectively quantifying the results from this method, the null ORF constitutive expression mutants and transcription factor null mutants were not subjected to the same series of phenotypic assays.



A) B)

Figure 3.7 Spot assays for identification of phenotypes. Examples of spot assays for sensitivity to 0.8 g/L copper (II) sulphate (A) and growth on 2% galactose media (2% galactose, 0.67% Difco YNB, 2% agar). Example plates contain, from top to bottom: Enoferm M2 wild type control, and null mutants for M22, M23, M25, M26 and M28. From left to right the spots are diluted 1x, 10x, 100x, 1000x and 10000x from an initial concentration of 1 OD600 cells.

3.2.5 The mutation of novel ORFs resulted in few observable phenotypes

In this study, the expression of the novel ORFs was confirmed, and the null mutation and constitutive expression of these genes was investigated. While the novel ORFs studied are only present in industrial wine strains of *S. cerevisiae*, the deletion or constitutive expression of these novel ORFs had only minimal effects on the metabolite profile of a Chardonnay wine fermentation. No significant effect was seen on the fermentation rate for any mutant, as measured by weight loss as a proxy for CO_2 release. While ANOVA found genotype to cause a significant difference for ethanol, glucose, fructose, glycerol and acetic acid, this was attributed to a single outlier for one biological replicate of the M13 null mutant which was inconsistent with other fermentations conducted with the same null mutant. Because of this, it was concluded that the deletion of the fifteen novel ORFs caused no difference in basic metabolism during wine fermentation.

While constitutive expression was not employed for all novel ORFs, these mutants caused some small but significant differences in the metabolite profile of the wine produced. No mutant caused a

difference for more than two metabolites, which is consistent with a specific alteration rather than a general slow fermentation phenotype. The constitutive expression of M18 increased the consumption rate of both glucose and fructose. Consistent with what might be hypothesized, the constitutive expression of the putative hexose transporter M23 produced a slight increase in the rate of fructose consumption. These results are consistent with some of the novel ORFs having a beneficial effect on fermentation performance when constitutively expressed in the Enoferm M2 background.

Despite some promising results, the majority of the null mutants tested through fermentation yielded no discernable phenotype. The lack of strong phenotype for many of the null mutants is not entirely surprising, as previous work has noted that only 15% of viable homozygous deletions in *S. cerevisiae* result in a growth phenotype in rich media (Giaever *et al.*, 2002). A more thorough attempt to identify a phenotype for all genes of the *S. cerevisiae* deletion collection (approximately 6000 heterozygous and 5000 homozygous gene deletions) used 1144 chemical assays and was able to identify a measurable growth phenotype for 97% of the gene deletions (Hillenmeyer *et al.*, 2008). This means that despite the immense number of conditions tested, there were still roughly 180 genes that demonstrated no measurable phenotype. In a recent, massive effort to further elucidate gene networks in the cell, a collection of double deletion mutants was created to identify positive and negative genetic interactions. This work identified ~900,000 genetic interactions between 5416 genes of the *S. cerevisiae* genome (Costanzo *et al.*, 2016). Furthermore, the creation of double deletion mutants can help identify a phenotype for genes which are duplicates and likely to be compensated for by another gene; it has been estimated that a quarter of gene deletions without phenotype are compensated for by a duplicate gene (Gu *et al.*, 2003). These three studies were all conducted with a lab strain base; a comparative phenomic study of 22 *S. cerevisiae* strains found that more than 50% of the transcript, protein, metabolite and morphological values (over 14,000 total values for each strain) varied significantly between strains grown to mid-log phase in rich media (Skelly *et al.*, 2013). Given such a large, documented variation among strains, it is highly likely that the phenotypes associated with particular gene deletions would be varied depending on the genetic background (*i.e.* strain) in which the gene was deleted.

Gene deletion or overexpression, combined with alternative growth conditions, has been used to successfully identify growth phenotypes in non-annotated genes, and provide further annotation for previously annotated genes. Anderson *et al.* (2012) used a molecular barcoded yeast open reading frame high copy plasmid (overexpression) library to identify ethanol tolerance genes in both a lab strain and a wine strain of *S. cerevisiae*. Using two strains of yeast allowed the identification of two genes that were able to confer resistance to ethanol in both genetic backgrounds. A study by Walkey *et al.* (2012) deleted the 62 non-annotated FSR genes from a wine strain of *S. cerevisiae*, and used the resulting strains to conduct a trial wine fermentation; an experimental method that was used in the present thesis work. The null mutant of one FSR gene, YML081W, produced less acetic acid than wild type, and was furthermore found to increase the expression of two acetaldehyde dehydrogenases when YML081W was constitutively expressed (Walkey *et al.*, 2012). The results from these two studies validate the importance of yeast strain genetic background and relevant growing conditions when attempting to annotate a phenotype to ORFs.

3.3 Transcriptomic changes during fermentation are both time and genotype dependent

The yeast transcriptome was measured at day two, day four and day seven of the Chardonnay wine fermentation for Enoferm M2 wild type and a selection of null and constitutive expression mutants. From day two, when glucose concentrations were around 8.7 % w/v and ethanol was ca. 3.2 % v/v, the fermentation progressed to day seven, when glucose was down below 2 % w/v and ethanol was over 10 % v/v (Table 3.6). During this period, the fructose also declined from 11.2 % w/v at day two to 4.9 % v/v at day seven, but is still present in concentrations that could be contributing to osmotic stress. Two general types of analysis were conducted: A time point analysis of wild type to investigate the overall changes in the transcriptome from day two of fermentation through to day seven of fermentation, and an analysis of the differences between wild type and mutants at each time point.

Table 3.6 Primary metabolite profile of wild type Enoferm M2 during fermentation. Wild type wine samples were analyzed by HPLC on the three days of fermentation sampled for transcriptomic and proteomic analyses (days two, four and seven).

Day	Ethanol (% v/v)	Glucose (% v/v)	Fructose (% v/v)	Total Sugars (glucose + fructose % v/v)
Two	3.24	8.67	11.16	19.83
Four	7.42	4.10	8.35	12.45
Seven	11.38	1.12	4.86	5.98

3.3.1 Enoferm M2 wild type undergoes large transcriptional changes during fermentation

The largest magnitude of changes in the yeast transcriptome are seen when comparing wild type samples at the three sampling points during the fermentation: Day two, day four and day seven. As can be seen from the metabolite profiling (Figure 3.3) and weight loss curves (Figure 3.2), these time points are taken during the exponential phase of the fermentation process (which does not necessarily correspond to the exponential growth phase of the yeast). By day two, a fermentation stress response should have been elicited (Marks *et al.*, 2008), and by day seven we should expect to see a further response as nutrients are depleted and ethanol increases. Transcript abundances were analyzed on each day and categorized into groups based on their expression pattern over the three time points (Figure 3.8). Each category was then analyzed for gene ontology and pathway enrichments to assess the transcriptomic changes happening in the yeast during fermentation. Results are summarized here, and the entire list of ontology enrichments can be found in Appendix F, Table A.4.

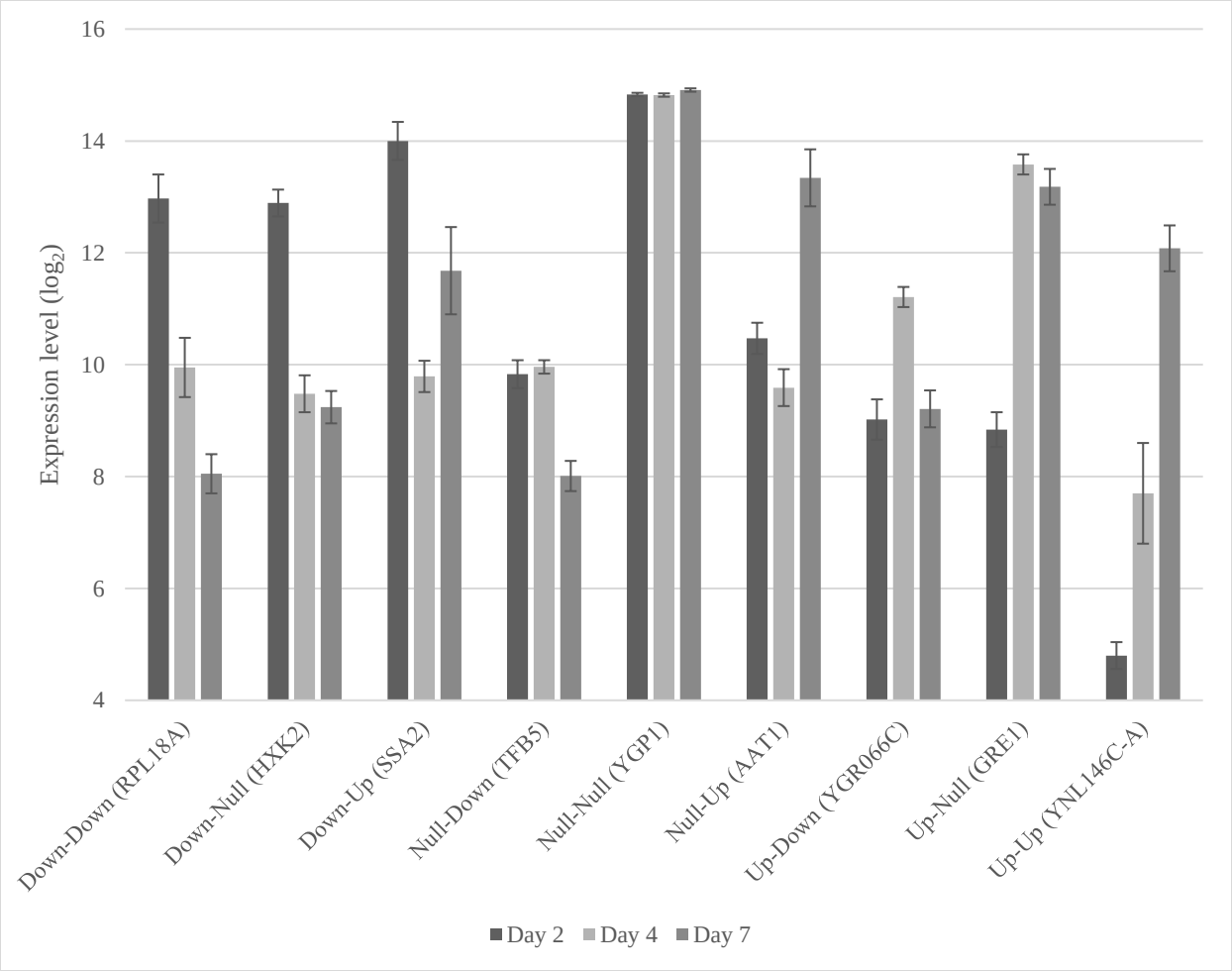


Figure 3.8 Expression patterns of transcripts over three time points. Wild type gene expression was clustered by differential expression patterns over the three days of fermentation, and each possible pattern is represented here by expression data from one gene from each cluster. Expression values are the mean of three biological replicates and bars represent one standard deviation.

The first thing that is observed is that from day two – day seven of fermentation, the majority of genes (65.4%) remain unchanged in their expression (Figure 3.9). This "null-null" category shows significant enrichment in genes related to cellular organization and localization, primary metabolic processes, response to stimuli and intracellular transport. The vast majority of these genes (2806, 73.3%) are associated with membrane-bound organelles. In terms of molecular function, the "null-null" category shows significant enrichment in protein and ion binding activities, as well as catalytic and transferase activity. While these genes remain unchanged in their expression, they may be constitutively expressed at low or high levels, but likely are important in general cell homeostasis.

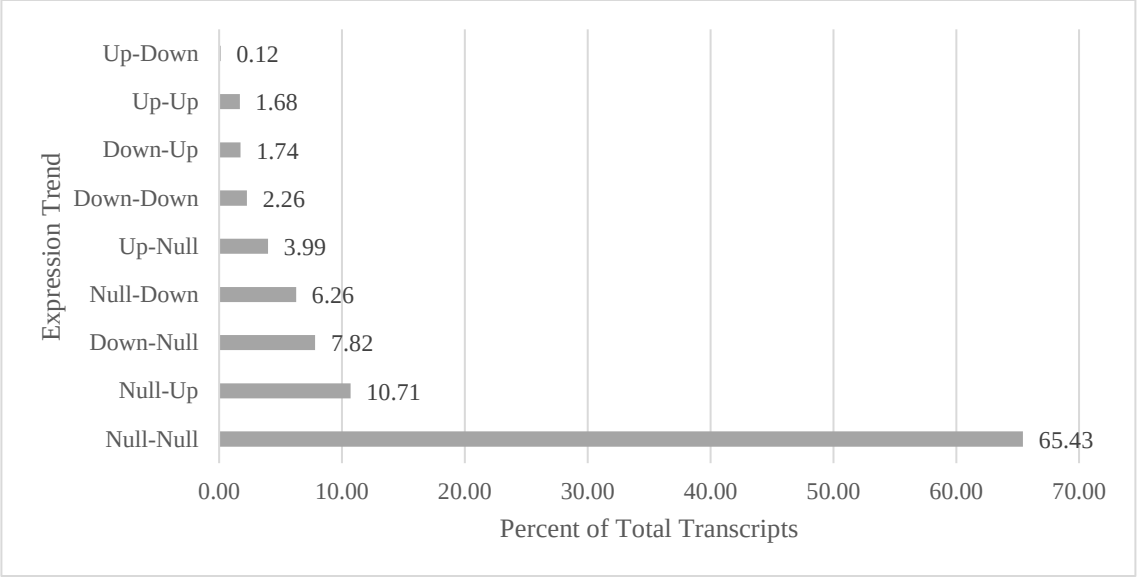


Figure 3.9 Expression trends from day two - day seven of wine fermentation. Transcripts were binned into one of nine expression categories based on their differential expression in wild type Enoferm M2 between days two, four and seven of Chardonnay wine fermentation. Quantities are shown as a percentage of the 5,846 total transcripts measured.

On the other hand, the two of the smallest categories were "up-down" and "down-up" containing only 0.12% and 1.74% of all the genes, respectively. Neither category was significantly enriched for any gene ontology category, although manual curation taking the magnitude of fold-changes of transcripts into account revealed some interesting findings. Notably, five genes (*SSA2*, *HMG1*, *FTR1*, *DIP5* and YER053C-A) known to be induced by DNA replication stress were among the genes with the largest fluctuations from day two – day seven; all five of these genes decreased expression from day two – day four and increased expression from day four – day seven. It is possible that DNA replication stress was high at day two due to the high sugar concentration, decreased at day four as sugars were consumed, and then increased again at day seven due to the increased ethanol content of the media. Only seven genes followed the "up-down" regulation pattern, four of which encode putative proteins. Of the remaining three characterized genes, two (*APN2* and *TKL2*) had relatively minor (< 3-fold) fold-changes between any two days. The final gene, *PMA2*, had a 2.88 fold-change from day two – day four and a -7.05 fold-change from day four – day seven. The *PMA2* gene encodes a proton pump responsible for the maintenance of cytoplasmic pH and membrane potential. While some of the genes in the "up-down" and

"down-up" categories may represent genes that are induced both by the stresses of early and late fermentation, but not during mid-fermentation, some of these genes may also be false-positives caused by variable gene expression and transcript stability.

The other two smallest categories were "up-up" and "down-down", containing 1.68% and 2.26% of all genes, respectively. The "up-up" and "down-down" categories did not necessarily represent genes with the largest fold changes between day two and day seven, but rather genes with a consistent step-wise expression change and an intermediate expression level at day four. The "down-down" category was significantly enriched for genes related to the organo-nitrogen compound metabolic process, such as the synthesis of proteins. The down regulation of protein synthesis was consistent with an enrichment for ribosomal genes to also be down-regulated. The "down-down" category was also significantly enriched for the molecular function category of "aminoacyl-tRNA ligase activity". Overall, the three separate gene ontology enrichments all pointed towards a steady decline in protein synthesis from day two – day seven of fermentation. Despite containing 98 genes, the "up-up" category showed no significant gene ontology enrichment. Of note, however, were six putative proteins of unknown function (YNL067W-B, YMR317W, YNL146C-A, YNL277W-A, YKL107W, YMR175W-A) that were induced 12 to 28-fold from day two – day seven. In addition, there were many upregulated genes related to the meiotic cell cycle (*RDH54*, *DMC1*, *SPO11*, *TNG2*, *SPO22*, *HOP2*, *MND1*, *SLZ1*, *GAC1*, *MEK1*), transcription (*NRG2*, *XBP1*, *RTR2*, *UPC2*, *ADR1*, *CAT8*, *LEG1*, *USV1*, *CSR2*, *RRT15*) and oxidative stress (*FRM2*, *XBP1*, *SRX1*, *ECM5*, *MEK1*).

The final four categories all represented a moderate proportion of the overall number of yeast genes: 7.82 % were "down-null", 6.26% were "null-down", 10.71 % were "null-up", and 3.99 % were "up-null". These categories represented genes that either had a sharp change in expression early in the fermentation which is maintained to the end, or they maintained expression until the middle of fermentation and then changed sharply towards the end. Overall, these genes made up the largest number of dynamic changes during the fermentation, as well as the largest magnitude of changes (up to 28.3 fold-change for *BAR1*).

To begin with the "up-null" category, containing a modest 233 genes, a significant enrichment was seen for the broad gene ontology term of "single-organism process". Within the overarching "single-organism process" enrichment, the most common "child" terms were oxidation-reduction process (20 genes), sporulation (12 genes) and cell cycle (9 genes). A significant enrichment was also seen for genes localizing to the peroxisome, consistent with genes involved in oxidation-reduction processes. Of note were two genes, *SPG1* and *GRE1*, that had greater than 20-fold induction from day two – day four of fermentation. The *SPG1* gene is required for high temperature survival during stationary phase, and *GRE1* is a hydrophilin required for the desiccation-rehydration process, as well as tolerance to numerous other stressors. Interestingly, *GRE1*'s paralogue, *SIP18*, falls into the same regulation category but was expressed at much higher levels beginning at day two and only experienced a moderate increase for the remainder of the fermentation. In contrast, *GRE1* was dramatically induced between day two and day four, highlighting differences in regulation between these two otherwise very similar proteins.

The "null-down" category also had relatively little in the way of gene ontology enrichment. There was a significant enrichment for the cytoplasmic cellular location ontology. In addition, the molecular function of "catalytic activity" was also significantly enriched and accounted for 45.9 % of the genes in this category. The most abundant catalytic activities were hydrolase (48 genes) and transferase (43 genes) activities. These enrichments would be consistent with a reduction in cytoplasmic enzymatic activity towards the end of fermentation. While the decrease in transcript abundance for these genes was significant between day four and day seven, the greatest fold-change was a modest -4.54 fold.

The "null-up" category was the largest category of genes with dynamic changes in transcript abundance from day two – day seven of the wine fermentation. The "null-up" category was up-regulated late in the wine fermentation when nutrients had been depleted and alcohol had reached 10% v/v. This cluster had a significant gene ontology enrichment for genes involved in reproduction, the cell cycle and regulation of biological processes, and many of these genes were localized to the cell periphery, chromosomes or the cytoskeleton. The molecular function gene ontologies were enriched for double-stranded DNA binding and chromatin binding. The "null-up" category included the most upregulated

gene of the data set, *BAR1*, which encodes for an aspartyl protease which cleaves the alpha mating factor and assists with cell mating. All of the "null-up" responses were consistent with cells preparing for sexual reproduction in the later stages of fermentation.

Aside from the large "null-null" category, the "down-null" category produced the greatest number of gene ontology enrichments. The "down-null" cluster of genes demonstrated significant gene ontology enrichment for genes involved in nitrogen compound metabolic processes, translation and ribosome biogenesis. The cellular localization ontologies were significantly enriched for the cytosolic ribosome and the larger intracellular ribonucleoprotein complex. In keeping with the other ontology enrichments, the molecular function ontologies were significantly enriched for structural constituents of the ribosome. Together, these enrichments indicated that there was a strong downregulation of ribosomal machinery and cytoplasmic protein synthesis from day two – day four of the fermentation, and that this downregulation was maintained until the end of the fermentation. The downregulation of protein synthesis likely corresponded with the depletion of nitrogen sources in the grape juice, although this theory was not tested explicitly.

Overall the time course analysis of Enoferm M2 transcription during wine fermentation painted a clear picture of cellular changes that occurred as the fermentation progressed from day two – day seven. The vast majority of genes remained at a steady state from day two until day seven of the fermentation; these "maintenance" genes were involved in cellular organization and primary metabolic processes, keeping the cell alive and functioning. From day two – day four, an induction of peroxisomal genes and genes involved in stress tolerance was observed, along with a dramatic downregulation of genes involved in protein synthesis. From day four – day seven there was an induction of genes involved in reproduction and corresponding downregulation of cytoplasmic enzymes. Throughout the fermentation there was a consistent upregulation of genes involved in DNA replication stress, and downregulation of ribosomal activity and protein synthesis.

3.3.1.1 Osmotic stress response genes largely maintain constitutive expression.

Given the dynamic changes to the media contents between day two and day seven of fermentation, the fate of genes tagged with the "cellular response to osmotic stress" gene ontology (GO: 0071470) was investigated. Ten genes fall into this category, and six of the osmotic stress genes (*CDC14, SLT2, CKA2, HOG1, HSP12, MNN4*) were constitutively expressed during the period of fermentation analyzed. While *HOG1* is thought to be a major player in osmoregulation (Albertyn *et al.*, 1994; Gonzalez *et al.*, 2016), this gene's expression remained constant as the total sugars in the cell decreased substantially. This could be because of additional osmotic stress caused by the increasing ethanol content binding free water later in fermentation, or that levels of *HOG1* remain relatively constant, but its activation state changes during fermentation to allow an osmotic response (Pelet *et al.*, 2011). The *MOT3* and *ROX1* genes, however, which both act as transcriptional repressors in response to hypoxic and hyperosmotic stress (Martínez-Montañés *et al.*, 2013), were both downregulated > 5-fold between day two and day four. The aldose reductase *GRE3*, which has been implicated in multiple stress responses and is regulated by the high-osmolarity glycerol (HOG) pathway (Garay-Arroyo and Covarrubias, 1999) was the only osmotic stress-related gene that was actually downregulated at day seven of the fermentation. The final osmotic-stress related gene, *HSP30*, was upregulated at day four, and remained strongly expressed at day seven, which corroborates this gene's status as a stress-responsive protein (Piper *et al.*, 1997).

3.3.2 Novel ORF null mutant effects are time point dependent

Microarray analysis of transcript abundance was conducted for nine of the 15 novel ORF null mutants (M13, M14, M15, M18, M20, M21, M22, M23 and M28) sampled on day two, day four and day seven of a Chardonnay wine fermentation. Differential expression was measured for each time point based on a wild type Enoferm M2 control fermentation, using an ANOVA p-value < 0.05 and a linear fold-change > 2 as the cut-off for significance.

It was clear from a brief overview of the data generated that differential expression of null mutants with regards to wild type was highly dependent upon the time point sampled. Furthermore, very few

genes were differentially expressed for the same mutant at multiple time points. Finally, at day two of fermentation many of the same genes were differentially expressed in multiple mutants. Upon further investigation, at day two there appeared to be 53 genes that were differentially expressed between wild type and the combination of all mutant strains. This phenomenon did not occur at day four or day seven. To take this anomaly into account, the experimental design was modified for the day two microarrays, and mutant samples were compared to the pool of all other mutants as well as wild type. Since none of the mutants produced a detectable phenotype, this method improved the statistical power to detect significant differences caused by a single mutation by increasing the number of samples used to make a comparison.

3.3.2.1 Only five novel ORFs produced transcriptomic changes at day two of fermentation.

When the RNA transcript abundance of nine novel ORF null mutants was compared against Enoferm M2 wild type on day two of wine fermentation, all nine mutants had significant differences in gene expression. When, however, the mutants were compared against the panel of all other mutants as well as wild type, only five mutants (M13, M15, M21, M23 and M28) produced significant results. The 53 genes that were differentially expressed in wild type compared to the mutants may have been a result of the *kanMX* marker gene present in the null mutants, although this marker has not been known to affect microarray studies previously. The *kanMX* marker used for the creation of the novel ORF null mutants was confirmed to have been inserted specifically into the ORF of interest, none of which overlapped with any other known gene, and therefore off target effects are unexpected. The most highly differentially expressed gene between wild type and the mutants was *BTN2*, which was expressed 7.74-fold lower in wild type. The *BTN2* gene is one of the few genes in wild type that had a strong downregulation from day two – day four, followed by an even stronger (13.5-fold) upregulation from day four – day seven. In addition, this gene had higher than average standard deviation among samples. This gene is a v-SNARE binding protein implicated in protein retrieval from the late endosome to the Golgi, the modulation of arginine uptake and the mediation of pH homeostasis of the vacuole.

Five novel ORF null mutants had significant differential expression of genes compared to all other mutants and wild type on day two of wine fermentation (Table 3.7). The M13 null mutant resulted in four differentially expressed genes: YFR057W, *GAC1*, *THI11* and *SPO22*. The maximum linear fold-change was a 2.37-fold decrease in expression for YFR057W in the M13 null mutant. The *THI11* gene was of interest, as this protein is involved in the synthesis of a thiamine precursor, which is important for maintaining thiamine levels during wine fermentation and belongs to a family of genes encoded in subtelomeric regions. The M15 null mutant resulted in a single differentially expressed gene, *TOS6* (2.54-fold change increase in the mutant), which encodes a glycosylphosphatidylinositol-dependent cell wall protein able to aid lactic acid resistance when decreased. The M21 null mutant resulted in eight differentially expressed genes: *HXT5*, *ARO10*, *ARO9*, *BAP3*, *NEJ1*, *ECM13*, *BIO5* and YGR035C. The gene *BAP3* had the greatest fold-change, with a 2.98-fold decrease in the mutant; *BAP3* gene encodes an amino acid permease. The *ARO10* and *ARO9* genes were upregulated in the mutant and encode enzymes involved in amino acid synthesis. The *HXT5* gene encodes a moderate affinity glucose transporter, which is highly relevant to the fermentation process, and this gene was slightly upregulated in the M21 null mutant. The *ECM13* and YGR035C genes encode proteins of unknown function. The M23 null mutant resulted in ten differentially expressed genes: *UPC2*, *YPS3*, *MIG1*, *MIG2*, *SET6*, *STP4*, YPR015C, YPL272C, YEL076C-A, and YKR075C. The most differentially expressed gene, *UPC2*, was downregulated 3.2-fold in the null mutant. The *UPC2* gene encodes a transcription factor that regulates sterol biosynthetic genes. The *MIG1* and *MIG2* genes encode transcriptional repressors, which work in coordination with each other in glucose repression, and were both downregulated in the M23 null mutant. The *STP4* gene encodes a putative transcription factor, and YPR015C, YPL272C, YEL076C-A, and YKR075C are all proteins or putative proteins of unknown function. The M28 null mutant resulted in 14 differentially expressed genes: *RRT15*, *GAP1*, *MEP2*, *SSA4*, *PRM5*, *HPA2*, *YHB1*, *PAU15*, *PAU10*, *PAU5*, *PAU2*, YDR222W, YGR204C-A, and YJL047C-A. The Rrt15p is a putative protein of unknown function, but it has been identified as having an effect on rDNA transcription. *GAP1*, which encodes a general amino acid permease, showed the most significant fold change, a 3.56-fold increase in the null

mutant. *MEP2* encodes an ammonium permease under nitrogen catabolite repression, which along with *GAP1* is involved in invasive growth. Another notable trend in the M28 null mutant was the upregulation of four PAU genes (*PAU2*, *PAU5*, *PAU10*, *PAU15*), which are members of the seripauperin multigene family encoded in subtelomeric regions. This large family of highly similar genes has been associated with wine fermentations (Rossignol *et al.*, 2003; Marks *et al.*, 2008; Deed, Deed and Gardner, 2015; Tronchoni *et al.*, 2017), and they are structural constituents of the cell wall, but their function is as yet unknown. Proteins encoded by YDR222W, YGR204C-A and YJL047C-A are all putative proteins of unknown function.

Table 3.7 Day two novel ORF null mutant DE genes. Novel ORF null mutants were compared against wild type Enoferm M2 and all other samples. Significant differentially expressed genes were determined as ANOVA *p*-value <0.05 and linear fold-change > 2.

Mutant	Transcript ID	Mutant Mean Signal (log_2)	All Other Samples Mean Signal (log_2)	Linear Fold Change	ANOVA *p*-value	Gene Symbol
M13	YFR057W	4.14	5.39	-2.37	0.0025	YFR057W
M13	YOR178C	9.84	10.91	-2.1	0.0215	*GAC1*
M13	YDL244W	13.42	14.55	-2.18	0.0438	*THI11*
M13	YIL073C	5.33	6.44	-2.16	0.0457	*SPO22*
M15	YNL300W	10.34	9	2.54	0.0238	*TOS6*
M21	YHR096C	9.24	8.14	2.13	0.0019	*HXT5*
M21	YDR380W	11.3	9.87	2.68	0.0043	*ARO10*
M21	YDR046C	9.9	11.48	-2.98	0.0048	*BAP3*
M21	YLR265C	5.89	4.57	2.5	0.0085	*NEJ1*
M21	YHR137W	11.04	9.7	2.52	0.0161	*ARO9*
M21	YBL043W	11.35	12.84	-2.8	0.0333	*ECM13*
M21	YNR056C	9.59	10.9	-2.48	0.0405	*BIO5*
M21	YGR035C	10.87	12.3	-2.7	0.0492	YGR035C
M23	YLR121C	7.26	8.42	-2.23	0.0073	*YPS3*
M23	YGL035C	9.18	10.26	-2.11	0.0114	*MIG1*
M23	YPL165C	6.97	8.06	-2.14	0.0252	*SET6*
M23	YPR015C	7.36	8.59	-2.33	0.0286	YPR015C
M23	YDR213W	9.43	11.11	-3.2	0.0287	*UPC2*
M23	YGL209W	8.92	9.93	-2.01	0.0292	*MIG2*
M23	YPL272C	11.08	12.1	-2.02	0.0323	YPL272C
M23	YEL076C-A	7.85	8.91	-2.1	0.0343	YEL076C-A
M23	YDL048C	10.41	11.47	-2.08	0.0399	*STP4*
M23	YKR075C	9.32	10.51	-2.28	0.0447	YKR075C
M28	YLR162W-A	7.14	5.75	2.62	0.0000	*RRT15*

Mutant	Transcript ID	Mutant Mean Signal (log_2)	All Other Samples Mean Signal (log_2)	Linear Fold Change	ANOVA *p*-value	Gene Symbol
M28	YDR222W	11.4	10.11	2.43	0.0018	YDR222W
M28	YGR204C-A	9.86	8.73	2.19	0.0085	YGR204C-A
M28	YIL117C	10.17	8.99	2.26	0.0116	*PRM5*
M28	YPR193C	11.55	9.94	3.04	0.0143	*HPA2*
M28	YKR039W	12.77	10.94	3.56	0.0163	*GAP1*
M28	YJL047C-A	8.6	9.8	-2.29	0.0179	YJL047C-A
M28	YCR104W	11.75	10.6	2.23	0.0180	*PAU15*
M28	YBL108C-A	12.9	11.88	2.03	0.0281	*PAU10*
M28	YER103W	12.91	11.29	3.09	0.0333	*SSA4*
M28	YNL142W	12.87	11.5	2.58	0.0355	*MEP2*
M28	YEL049W	13.6	12.45	2.22	0.0370	*PAU2*
M28	YGR234W	11.46	10.45	2.01	0.0401	*YHB1*
M28	YFL020C	14.08	13	2.12	0.0486	*PAU5*

3.3.2.2 Eight novel ORFs produced transcriptomic changes at day four of fermentation.

All novel ORF null mutants, with the exception of M22, resulted in significant differences in transcription when compared to wild type yeast at day four of Chardonnay wine fermentation (Table 3.8). Wild type did not have any significant differences in transcription when compared to the pool of all null mutants combined (except for the *kanMX* marker gene, as expected), although there were seven genes that were differentially expressed in multiple null mutants. YLR154W-E was differentially expressed in four null mutants (M13, M15, M20, and M28), and is a dubious ORF situated anti-sense to ribosomal DNA. Another putative protein with no known function from the same region, YLR154C-H, was differentially expressed in three null mutants (M13, M20, M28). In both cases, YLR154W-E and YLR154C-H transcripts were upregulated in the M28 null mutant, but downregulated in all others. The *BAP3* gene, encoding an amino acid permease, was significantly upregulated in three null mutants (M13, M20, and M23). The remaining four genes, *CYC1*, *DIP5*, *HMG1* and *PUT1*, were all differentially expressed in only two null mutants and in all four cases the gene was either downregulated or upregulated in both null mutants.

Table 3.8 Day four novel ORF null mutant DE genes. Novel ORF null mutants were compared against wild type Enoferm M2. Significant differentially expressed genes were determined as ANOVA *p*-value <0.05 and linear fold-change > 2.

Mutant	Transcript ID	Mutant Mean Signal (log_2)	Wild Type Mean Signal (log_2)	Linear Fold Change	ANOVA *p*-value	Gene Symbol
M13	YGR055W	11.04	9.91	2.19	0.0002	*MUP1*
M13	YLR154W-E	10.63	12.11	-2.78	0.0072	YLR154W-E
M13	YDR046C	12.64	10.93	3.27	0.0108	*BAP3*
M13	YJR010W	11.63	10.52	2.16	0.0204	*MET3*
M13	YLR154C-H	7.68	8.73	-2.06	0.0250	YLR154C-H
M13	YMR107W	9.39	10.76	-2.58	0.0273	*SPG4*
M14	YJR048W	7.76	9.21	-2.73	0.0002	*CYC1*
M14	YML075C	8.71	9.87	-2.24	0.0006	*HMG1*
M15	YLR154W-E	10.96	12.11	-2.22	0.0048	YLR154W-E
M15	YGL256W	11.96	10.61	2.56	0.0420	*ADH4*
M18	YPL265W	10.22	8.65	2.96	0.0001	*DIP5*
M18	YMR317W	8.35	7.29	2.08	0.0001	YMR317W
M18	YCL026C-A	10.99	9.72	2.42	0.0009	*FRM2*
M18	YER188C-A	6.18	5.11	2.11	0.0020	YER188C-A
M18	YDR522C	6.69	5.26	2.69	0.0041	*SPS2*
M18	YLR265C	6.13	4.71	2.67	0.0059	*NEJ1*
M18	YNL146C-A	9.1	7.59	2.86	0.0071	YNL146C-A
M18	YLR042C	8.83	7.61	2.33	0.0080	YLR042C
M18	YPL121C	4.91	3.18	3.31	0.0098	*MEI5*
M18	YBL008W-A	5.57	4.52	2.07	0.0107	YBL008W-A
M18	YIL046W-A	6.12	4.88	2.36	0.0207	YIL046W-A
M18	YGL041C-B	8.67	7.59	2.12	0.0323	YGL041C-B
M18	YBR219C	5.34	4.31	2.04	0.0354	YBR219C
M18	YMR101C	9.56	8.55	2.01	0.0491	*SRT1*
M20	YLR154W-E	10.99	12.11	-2.17	0.0054	YLR154W-E
M20	YLR154C-H	7.65	8.73	-2.1	0.0099	YLR154C-H
M20	YPL265W	9.7	8.65	2.07	0.0108	*DIP5*
M20	YDR046C	12.26	10.93	2.51	0.0134	*BAP3*
M21	YBR208C	10.63	11.75	-2.18	0.0364	*DUR1,2*
M23	YLR142W	11.9	13.11	-2.31	0.0013	*PUT1*
M23	YDR046C	12.38	10.93	2.73	0.0134	*BAP3*
M28	YML075C	8.43	9.87	-2.72	0.0001	*HMG1*
M28	YJR048W	8.16	9.21	-2.06	0.0003	*CYC1*
M28	YLR162W-A	9.37	7.71	3.15	0.0003	*RRT15*
M28	YLR154C-H	10.63	8.73	3.76	0.0018	YLR154C-H
M28	YML054C	8.27	9.35	-2.12	0.0028	*CYB2*
M28	YBR301W	11.61	9.81	3.48	0.0047	*PAU24*
M28	YLR142W	12.02	13.11	-2.13	0.0077	*PUT1*
M28	YLR154W-E	13.41	12.11	2.47	0.0080	YLR154W-E
M28	YGR177C	8.4	7.3	2.13	0.0094	*ATF2*
M28	YOR010C	12.97	11.82	2.22	0.0098	*TIR2*
M28	YAR020C	12.83	10.83	4.01	0.0107	*PAU7*
M28	YEL049W	12.8	11.72	2.11	0.0157	*PAU2*

Mutant	Transcript ID	Mutant Mean Signal (log_2)	Wild Type Mean Signal (log_2)	Linear Fold Change	ANOVA *p*-value	Gene Symbol
M28	YFL020C	13.79	12.28	2.86	0.0393	*PAU5*
M28	YKR053C	10.84	9.65	2.28	0.0418	*YSR3*
M28	YFL020C	14.12	12.59	2.87	0.0441	*PAU5*

The M13 null mutant resulted in six differentially expressed genes when compared to wild type at day four of wine fermentation. Two of these genes were the dubious ORFs YLR154W-E and putative protein encoding YLR154C-H, that were both differentially expressed in multiple mutants. The *BAP3* gene, which was also differentially expressed in multiple mutants, had the greatest fold-change in the M13 null mutant with a 3.27-fold increase compared to wild type. Of the remaining differentially expressed genes, *MUP1* and *MET3* encode a high affinity methionine permease and an adenosine triphosphate (ATP) sulphurylase involved in methionine metabolism, respectively. The final differentially expressed gene, *SPG4*, is required for high temperature growth at stationary phase.

The M14 null mutant resulted in two differentially expressed genes when compared to wild type at day four of wine fermentation. Both of these genes, *CYC1* and *HMG1*, were also differentially expressed in the same manner in the M28 null mutant. The *HMG1* gene product catalyzes a rate-limiting step in sterol synthesis, while *CYC1* is an important component of cellular respiration. Both genes had relatively small fold-changes (< 2.75-fold), but ANOVA *p*-values less than 0.001.

The M15 null mutant resulted in two differentially expressed genes when compared to wild type at day four of wine fermentation. One of the genes was YLR154W-E, which was differentially expressed in multiple null mutants. The other significantly differentially expressed gene was *ADH4*, a zinc-dependent alcohol dehydrogenase.

The M18 null mutant resulted in 14 differentially expressed genes when compared to wild type at day four of wine fermentation. Of these, eight were putative proteins of unknown function (YLR154W-E, YMR317W, YER188C-A, YNL146C-A, YBL008W-A, YIL046W-A, YGL041C-B, and YBR219C) and one was a cell wall protein of unknown function (YLR042C). The *DIP5* gene was also differentially expressed in the M18 null mutant, as well as the M20 null mutant. The largest fold-change (3.31-fold upregulation) was seen for *MEI5*, which is involved in meiotic recombination. The remaining four

differentially expressed genes, *FRM2*, *SPS2*, *NEJ1* and *SRT1*, are involved in stress response, sporulation, non-homologous end joining and dolichol synthesis, respectively.

The M20 null mutant resulted in four differentially expressed genes when compared to wild type at day four of wine fermentation. All four of these genes had minor fold changes, with the largest being *BAP3* with a 2.51-fold increase in the null mutant. All four differentially expressed genes, YLR154W-E, YLR154C-H, *DIP5*, and *BAP3* were differentially expressed in at least one other null mutant as well.

The M21 null mutant resulted in one differentially expressed gene when compared to wild type at day four of wine fermentation. The *DUR1,2* gene was downregulated 2.18-fold in the null mutant compared to wild type expression. The *DUR1,2* gene encodes an urea amidolyase that degrades urea to carbon dioxide and ammonia, and this gene is sensitive to nitrogen catabolite repression and DNA replication stress.

The M23 null mutant resulted in two differentially expressed genes when compared to wild type at day four of wine fermentation. Both of these genes, *PUT1* and *BAP3*, were differentially expressed in at least one other null mutant. Both are related to nitrogen usage, as *BAP3* encodes an amino acid permease while *PUT1* encodes a proline oxidase.

The M28 null mutant resulted in fourteen differentially expressed genes when compared to wild type at day four of wine fermentation. Five of these genes, YLR154W-E, YLR154C-H, *CYC1*, *HMG1* and *PUT1*, were also differentially expressed in at least one other null mutant and their function has been briefly described previously. In addition, four PAU genes (*PAU2*, *PAU5*, *PAU7* and *PAU24*) were all upregulated in the M28 null mutant. The *PAU7* gene had the largest fold-change of any gene for any null mutant at day four, with a 4.01-fold decrease in transcript abundance compared to wild type. As mentioned previously, these genes are members of the seripauperin multigene family encoded in subtelomeric regions, they are structural components of the cell wall, yet their function is unknown. *PAU2* and *PAU5* were also upregulated at day two. The *RRT15* ORF that encodes a putative protein of unknown function was also upregulated at both day two and day four for the M28 null mutant. For the remaining differentially expressed genes, *TIR2* encodes a putative cell wall mannoprotein known to be

induced by anaerobiosis, *CYB2* encodes cytochrome b2 which is required for lactate utilization, *ATF2* encodes an alcohol acetyltransferase important for volatile ester formation in fermented beverages, and *YSR3* encodes a dihydrosphingosine 1-phosphate phosphatase involved in sphingolipid metabolism.

3.3.2.3 Seven novel ORFs produced transcriptomic changes at day seven of fermentation.

All novel ORF null mutants, with the exception of M14 and M15, resulted in significant differences in transcription when compared to wild type yeast at day seven of Chardonnay wine fermentation (Table 3.9). Wild type did not have any significant differences in transcription when compared to the pool of all null mutants combined (except for the increased expression of the *kanMX* marker gene, as expected), although there were six genes that were differentially expressed in multiple null mutants.

Table 3.9 Day seven novel ORF null mutant DE genes. Novel ORF null mutants were compared against wild type Enoferm M2. Significant differentially expressed genes were determined as ANOVA *p*-value <0.05 and linear fold-change > 2.

Mutant	Transcript ID	Mutant Mean Signal (log_2)	Wild Type Mean Signal (log_2)	Linear Fold Change	ANOVA *p*-value	Gene Symbol
M13	YMR305C	11.45	10.41	2.06	0.0088	*SCW10*
M13	YKR016W	10.21	9.19	2.03	0.0104	*FCJ1*
M13	YPL211W	8.07	6.97	2.14	0.0296	*NIP7*
M13	YOL059W	8.55	9.63	-2.11	0.0493	*GPD2*
M18	YIR015W	9.46	10.56	-2.15	0.0049	*RPR2*
M18	YBL101W-A	12.29	13.73	-2.72	0.0063	YBL100W-A
M18	YOR008C-A	6.2	7.56	-2.57	0.0092	YOR008C-A
M18	YKL095W	8.86	10	-2.2	0.0102	*YJU2*
M18	YCL010C	9.45	10.65	-2.3	0.0122	*SGF29*
M18	YEL076C-A	7.35	8.97	-3.07	0.0186	YEL076C-A
M18	YHR105W	8.02	9.18	-2.23	0.0244	*YPT35*
M18	YML037C	7.94	8.96	-2.03	0.0252	YML037C
M18	YBL112C	8.64	10.76	-4.33	0.0261	YBL112C
M18	YNL245C	9.56	10.74	-2.26	0.0263	*CWC25*
M18	YPL017C	10.86	12.01	-2.22	0.0289	*IRC15*
M18	YOR302W	11.59	12.85	-2.4	0.0294	YOR302W
M18	YHR199C-A	8.51	9.52	-2.02	0.0337	*NBL1*
M18	YMR084W	8.58	10.07	-2.82	0.0386	YMR084W
M18	YEL076C	4.61	6.18	-2.97	0.0468	YEL076C
M20	YPL265W	12.76	11.42	2.54	0.0202	*DIP5*
M21	YLR042C	9.44	10.46	-2.02	0.0110	YLR042C
M21	YBL008W-A	6.57	7.76	-2.28	0.0234	YBL008W-A
M21	YBR184W	7.64	8.93	-2.44	0.0409	YBR184W

Mutant	Transcript ID	Mutant Mean Signal (log_2)	Wild Type Mean Signal (log_2)	Linear Fold Change	ANOVA *p*-value	Gene Symbol
M21	YDR146C	7.21	8.85	-3.12	0.0438	*SWI5*
M21	YML123C	7.2	8.33	-2.2	0.0479	*PHO84*
M22	YML123C	7.02	8.33	-2.48	0.0218	*PHO84*
M23	YNR044W	9.5	8.45	2.07	0.0036	*AGA1*
M23	YMR070W	10.43	9.38	2.07	0.0290	*MOT3*
M23	YLL012W	13.11	11.47	3.11	0.0327	*YEH1*
M23	YBR238C	8.81	9.81	-2.01	0.0327	YBR238C
M23	YNL111C	11.39	9.11	4.86	0.0401	*CYB5*
M23	YGR049W	12.96	11.8	2.25	0.0408	*SCM4*
M28	YDR076W	9.45	10.72	-2.41	0.0045	*RAD55*
M28	5.8SrRNA	9.84	12.17	-5.03	0.0063	5.8S rRNA
M28	YBL008W-A	6.43	7.76	-2.52	0.0110	YBL008W-A
M28	YNL111C	11.83	9.11	6.62	0.0235	*CYB5*
M28	YML123C	7.1	8.33	-2.35	0.0247	*PHO84*
M28	YMR070W	10.47	9.38	2.13	0.0256	*MOT3*
M28	YGR049W	13.14	11.8	2.53	0.0274	*SCM4*
M28	YNR019W	12.2	11.14	2.07	0.0343	*ARE2*
M28	YMR220W	11.56	10.53	2.04	0.0358	*ERG8*
M28	YLL012W	13.03	11.47	2.96	0.0361	*YEH1*
M28	YLR154C	9.12	8.06	2.09	0.0374	*RNH203*
M28	YPL121C	5.31	6.83	-2.87	0.0411	*MEI5*
M28	YPR061C	11.86	10.83	2.03	0.0415	*JID1*
M28	YBR056W-A	10.81	9.79	2.02	0.0431	YBR056W-A
M28	YGL001C	12.51	11.28	2.34	0.0442	*ERG26*

The M13 null mutant resulted in four differentially expressed genes when compared to wild type at day seven of wine fermentation. All four differentially expressed genes had relatively minor fold changes, with a maximum of 2.14-fold upregulation for *NIP7*, a nucleolar protein required for the 60S ribosome subunit. The three other differentially expressed genes were *SCW10*, a cell wall protein that may play a role in mating, *FCJ1*, a component of the mitochondrial inner membrane, and *GPD2*, a glycerol 3-phosphate dehydrogenase.

The M18 null mutant resulted in fifteen differentially expressed genes when compared to wild type at day seven of wine fermentation. Four of these genes (YOR008C-A, YML037C, YMR076C, YER022W) encode putative proteins with no known function. The two most differentially expressed genes, YEL076C-A and YBL112C (3.07-fold and 4.33-fold downregulation, respectively), are also

putative proteins of unknown function and are situated within the telomeric regions of chromosome II and chromosome V, respectively. YEL076C-A is unusual in that it possesses an intron. In addition, YEL076C-A overlaps the coding sequence for YEL076C, a putative helicase, which was also downregulated 2.97-fold. *IRC15* and *NBL1* are both involved in chromosome segregation. The remaining differentially expressed genes had relatively minor fold-changes (< 3-fold change) and no common trend in terms of function or ontology. Differentially expressed genes found in the M18 null mutant on day seven were not found on day two or day four.

The M20 null mutant resulted in one differentially expressed gene when compared to wild type at day seven of wine fermentation; *DIP5* was upregulated 2.54-fold in the null mutant compared to wild type. This dicarboxylic amino acid permease encoding gene was also upregulated in the M20 null mutant at day four of fermentation.

The M21 null mutant resulted in five differentially expressed genes when compared to wild type at day seven of wine fermentation. Two of these genes (YBL008W-A and YBR184W) encode putative proteins of unknown function, and one gene (YLR042C) encodes a cell wall protein of unknown function. YBL008W-A was also downregulated in the M28 null mutant. The *PHO84* gene was downregulated in the same manner in the M21 null mutant as it was in the M22 null mutant. The largest fold-change was observed for *SWI5*, downregulated 3.12-fold, which is a transcription factor that activates the expression of genes in G1 phase and the genes controlling mating type switching.

The M22 null mutant resulted in one differentially expressed gene (*PHO84*) when compared to wild type at day seven of wine fermentation; this gene was downregulated 2.48-fold in the null mutant compared to wild type. The *PHO84* gene, encoding a high-affinity inorganic phosphate transporter, was also downregulated in both the M21 and M28 null mutants.

The M23 null mutant resulted in six differentially expressed genes when compared to wild type at day seven of wine fermentation. Four of these genes (*CYB5*, *YEH1*, *SCM4* and *MOT3*) were also differentially expressed in the M28 null mutant; *CYB5* had the strongest differential expression in this mutant, 4.86-fold higher than expression in wild type. This gene encodes cytochrome b5, which acts as

an electron donor in sterol biosynthesis. The *YEH1* gene encodes a steryl ester hydrolase, *MOT3* encodes a transcriptional repressor/activator responsive to hypoxia, and *SCM4* encodes a mitochondrial protein of unknown function. The remaining two differentially expressed genes, which were both unique to the M23 null mutant, were *AGA1*, an anchorage subunit of a-agglutinin of a-cells, and YBR238C, a mitochondrial membrane protein.

The M28 null mutant resulted in fourteen differentially expressed genes when compared to wild type at day seven of wine fermentation. These genes were significantly enriched in the sterol metabolic pathway ontology ($p = 9.95e^{-6}$). The sterol metabolic pathway genes included *CYB5*, which had the largest fold-change with 6.62-fold upregulation in the M28 null-mutant. Other differentially expressed sterol metabolic pathway genes included *YEH1* and *MOT3*, which were also differentially expressed in the M23 null mutant and described above, as well as *ERG8*, *ERG26* and *ARE2*, which are all enzymes that catalyze various steps of sterol synthesis. The remaining eight differentially expressed genes included YBL008W-A, *PHO84* and *SCM4*, which were all differentially expressed in other null mutants, as well as *MEI5* and *RAD55*, both involved in meiotic recombination, *JID1*, a putative heat-shock protein co-chaperone, *RNH203*, a ribonuclease subunit, and YBR056W-A, a protein of unknown function. An additional feature of the M28 null mutant transcriptome at day seven was the 5.03-fold decrease in 5.8S rRNA.

While there was only a significant ontology enrichment for differentially expressed genes for the M28 null mutant at day seven of Chardonnay fermentation, a few other features stood out from this time point. Proteins of unknown function continued to feature prominently among the lists of differentially expressed genes for the different null mutants. Themes that emerged when looking at the differentially expressed genes included genes involved in recombination, cell wall restructuring and sterol biosynthesis, all of which were relevant to the high stress environment found towards the end of wine fermentation. As with day four, there were again a handful of genes that were differentially expressed, generally with the same regulation pattern, in multiple null mutants. A possible rationale for this phenomenon will be explored in the section 3.3.3.

3.3.2.4 Transcriptomic changes are rarely consistent between time points.

By sampling novel ORF null mutant yeast cells at multiple time points during a wine fermentation, it was possible to determine whether the differential expression of genes compared to wild type would be constitutive or time dependent. From the results, it was clear that the differential expression of genes compared to wild type was highly dependent upon the time of sampling during the fermentation. Two null mutants, M14 and M22, only demonstrated differential expression at a single time point, while five null mutants, M13, M15, M21, M23 and M28, demonstrated differential expression at all three time points. Only two null mutants (M20 and M28) resulted in the same gene being differentially expressed at more than one time point sampled during the wine fermentation. The limited consistency of transcriptomic changes could reflect the changing conditions throughout the wine fermentation, which resulted in transcriptomic changes only being observed at particular time points.

For both M20 and M28, however, the null mutants only had consistently differentially expressed genes for day two and day four (*RRT15* for the M28 null mutant) or day four and day seven (*DIP5* for the M20 null mutant) (Figure 3.10). The Rrt15p is a putative protein of unknown function which has been associated in high-throughput studies with decreased levels of rDNA transcription (Hontz *et al.*, 2009). Curiously, levels of 5.8S rRNA were also significantly decreased 5.03-fold at day seven compared to wild type in the M28 null mutant. The *DIP5* gene encodes a dicarboxylic amino acid permease, which mediates the high-affinity and high-capacity transport of L-glutamate and L-aspartate (Regenberg *et al.*, 1998); *DIP5* was also the best BLAST search hit for the M6 novel ORF. The M20 novel ORF was most similar to the putative aryl-alcohol dehydrogenase *AAD14*, which is thought to be involved in cellular aldehyde metabolism and not amino acid metabolism. It would be worthwhile to test levels of L-glutamate and L-aspartate in the wine during fermentation with the M20 null mutant to see if there is any significant difference compared to wild type.

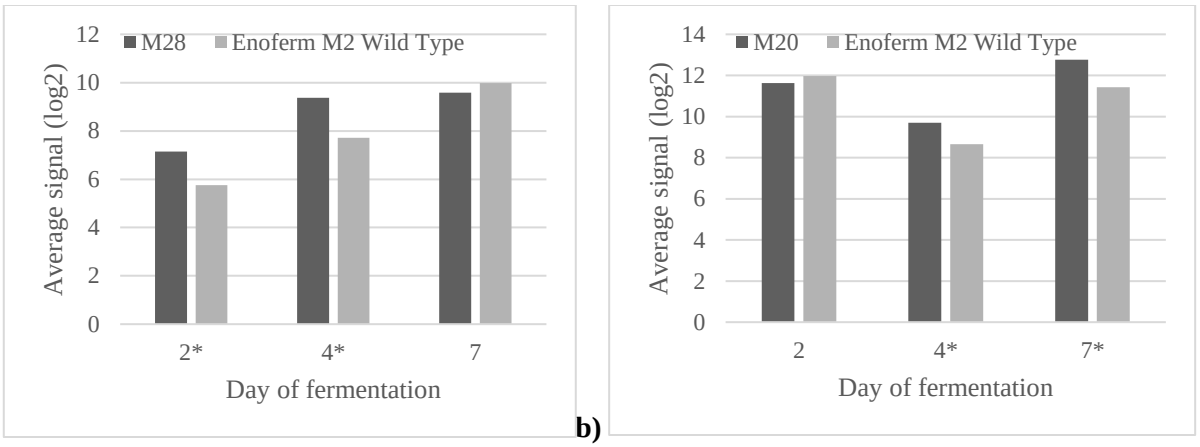


Figure 3.10 DE genes caused by novel ORF null mutants. The average $\log_2$ signal for *RRT15* expression in the M28 null mutant and wild type (a) and for *DIP5* expression in the M20 null mutant and wild type (b), as measured by microarray. Days with significant differences are denoted by *.

3.3.3 The *kanMX* marker may cause transcriptional effects in null mutants

The transcriptomic results for the novel ORF null mutants found many shared DE genes between null mutants at all three days of fermentation. The effect was so noticeable at day two of fermentation that wild type was not able to be used as the single control, and instead the collection of all other mutants and wild type was used to test single mutants for significantly DE genes. A total of 53 genes were DE between all mutants and wild type at day two, whereas at day four there were seven genes that were DE in more than one null mutant, and at day seven there were six genes that were DE in more than one null mutant. The *kanMX* marker has been used extensively in the production of null mutant collections, but has not been previously reported to cause transcriptional changes when used to knock out yeast genes (Giaever *et al.*, 2002; Anderson *et al.*, 2012; Costanzo *et al.*, 2016). The transcriptional changes observed in the present study could be condition dependent, with the use of a strong constitutive promoter (*PGK1* promoter) to drive the expression of the deletion marker causing a minor effect on overall transcription when the cell is exposed to stressful conditions. The *kanMX* marker has been previously reported to significantly affect the copy number of plasmids per cell when used as the selection marker in expression plasmids (Karim, Curran and Alper, 2013), although in the present study the *kanMX* marker was stably integrated into specific ORFs with no overlap of other CDSs.

Batch effects could also be responsible for the differences between wild type and the null mutants, although the normalization of the microarray results should minimize batch effects, unless some transcripts are particularly susceptible to degradation. If this was the case, however, one would expect DE genes to appear to all be regulated in the same manner with respect to wild type, and in fact the clusters of differentially expressed genes are both upregulated and downregulated. Complete randomization of the samples would aid in eliminating batch effects as a potential confounding variable.

One solution could have been to integrate the *kanMX* cassette into a non-coding region of the Enoferm M2 genome for use as a control, although this would have made it more difficult to isolate null mutants from wild type with antibiotics. Alternatively, as the pUG6 plasmid includes *loxP* sites flanking the *kanMX* cassette, it would have been possible to remove the selection marker from the null mutants using the Cre/*loxP* system, effectively leaving the null mutant strains with no foreign DNA being expressed (Gueldener *et al.*, 2002). Removing the antibiotic marker from null mutants would also make the mutant strains more difficult to easily separate from wild type.

3.3.4 Deletion of ORFs encoding putative transcription factors resulted in transcriptomic changes

Microarray analysis of RNA transcript abundance was conducted for null mutants of ten poorly annotated transcription factors (YBL066C/*SEF1*, YBR066C/*NRG2*, YBR150C/*TBS1*, YFL052W, YKL043W/*PHD1*, YKL222C, YLL054C, YLR278C, YNR063W, and YOR032C/*HMS1*) sampled on day two of Chardonnay wine fermentation. These genes are all either confirmed transcription factors with poor annotation, or putative transcription factors based on sequence homology. The null mutants were used to ferment Chardonnay grape juice into wine, and samples were collected at day two of fermentation for microarray analysis of the transcriptome.

The YBL066C/*SEF1* null mutant resulted in two differentially expressed genes: *THI72*, a thiamine transporter, and *YFR1*, a helicase encoded in the subtelomeric region. Both genes were downregulated compared to wild type, with a maximum fold-change of 2.25-fold for *YFR1*.

The YBR066C/*NRG2* null mutant resulted in 21 differentially expressed genes. There was a significant gene ontology enrichment for cell periphery cellular localization; 11 genes fell into this category, including numerous transporters and membrane proteins. Notably, the glucose transporters *HXT2* and *HXT5* were both upregulated, as well as the proline permease *PUT4*. Cell wall structural proteins such as *HPF1*, *PAU2* and *PAU7* were also upregulated compared to wild type. The largest fold-change of 5.29-fold upregulation was observed for *HPF1*, a haze-protective mannoprotein. The differentially expressed genes also included seven proteins of unknown function encoded by *TMA10*, *DIA1*, *RTN2*, *PHM7*, YKR075C, YNL195C, and YNL194C.

The YBR150C/*TBS1* null mutant resulted in three differentially expressed genes, all with minimal fold-changes. The maximum fold-change was a 2.07-fold upregulation for *HXT2*, a high affinity glucose transporter. The *PHM7* gene, encoding a protein of unknown function, and *OPT2*, encoding oligopeptide transporter, were also differentially expressed in the *TBS1* null mutant.

The YFL052W null mutant resulted in 11 differentially expressed genes. There were no significant gene ontology enrichments for the list of differentially expressed genes. There was also no evident theme to the differentially expressed genes, and two genes encoded proteins of unknown function (*PHM7* and YAR068W). The maximum fold-change was for YAR068W, encoding a putative protein, at 2.63-fold upregulation compared to wild type, followed by *PCL1*, a cyclin involved in the cell cycle, with 2.36-fold downregulation compared to wild type.

3.3.4.1 YLR278C is involved in thiamine and methionine metabolism.

The YLR278C null mutant resulted in 66 differentially expressed genes. There was significant ontology enrichment for the sulphur compound biosynthetic process (eight genes: *ARO9*, *ARO10*, *MET28*, *MHT1*, *BIO5*, *THI80*, *THI21*, *THI22*). Most of these genes are directly related to thiamine or methionine metabolism, and all were downregulated in the YLR278C null mutant. There was also significant differential expression for 15 genes encoding proteins of unknown function. The *PCL1* and *CLB6* genes, both encoding cyclins, were the most downregulated genes at 4.86-fold and 5.18-fold, respectively.

Three of the four most upregulated genes are of unknown function, but the fourth is *SSA4*, a stress-induced heat shock protein, which was upregulated 3.37-fold compared to wild type.

Table 3.10 Differentially expressed sulphur compound biosynthesis genes in the YLR278C null mutant. Eight genes with the "sulphur compound biosynthesis process" gene ontology were differentially expressed in the YLR278C null mutant at day two of Chardonnay wine fermentation. Linear fold change is expressed with wild type as a base line, and descriptions are taken directly from the SGD.

Gene Name	Identifier	Linear Fold Change	Description
MHT1	YLL062C	-2.67	S-methylmethionine-homocysteine methyltransferase, functions along with Sam4p in the conversion of S-adenosylmethionine (AdoMet) to methionine to control the methionine/AdoMet ratio
ARO9	YHR137W	-2.54	Aromatic aminotransferase II, catalyzes the first step of tryptophan, phenylalanine, and tyrosine catabolism
ARO10	YDR380W	-2.42	Phenylpyruvate decarboxylase, catalyzes decarboxylation of phenylpyruvate to phenylacetaldehyde, which is the first specific step in the Ehrlich pathway
THI22	YPR121W	-2.41	Protein with similarity to hydroxymethylpyrimidine phosphate kinases; member of a gene family with THI20 and THI21; not required for thiamine biosynthesis
THI21	YPL258C	-2.41	Hydroxymethylpyrimidine phosphate kinase, involved in the last steps in thiamine biosynthesis; member of a gene family with THI20 and THI22; Thi20p also has this activity
MET28	YIR017C	-2.05	Basic leucine zipper (bZIP) transcriptional activator in the Cbf1p-Met4p-Met28p complex, participates in the regulation of sulphur metabolism
BIO5	YNR056C	-2.02	Putative transmembrane protein involved in the biotin biosynthesis pathway; responsible for uptake of 7-keto 8-aminopelargonic acid; BIO5 is in a cluster of 3 genes (BIO3, BIO4, and BIO5) that mediate biotin synthesis
THI80	YOR143C	-2.01	Thiamine pyrophosphokinase, phosphorylates thiamine to produce the coenzyme thiamine pyrophosphate (thiamine diphosphate)

3.3.4.2 YNR063W is involved in DNA replication.

The YNR063W null mutant resulted in 150 differentially expressed genes. There were significant gene ontology enrichments for DNA replication (19 genes: *POL1*, *POL3*, *POL12*, *POL30*, *POL31*, *MCD1*, *DUN1*, *CDC9*, *CDC45*, *RNR1*, *CLB6*, *MCM10*, *RAD27*, *RAD53*, *CSM3*, *CTF18*, *TOF1*, *MSH2*, *DPB2*) and response to stress (43 genes), as well as cellular localization ontology enrichments for the replication fork (12 genes: *POL1*, *POL3*, *POL12*, *POL30*, *POL31*, *CDC9*, *CDC45*, *MCM10*, *CSM3*, *CTF18*, *TOF1*, *DPB2*) and the chromosome (27 genes). All genes involved in DNA replication were downregulated in the null mutant compared to wild type. Once again, *PCL1* and *CLB6* genes, both encoding cyclins, were the most downregulated genes at 5.86-fold and 5.02-fold changes respectively. Strongly upregulated genes included many proteins of unknown function, but also *SSA4* (4.06-fold), *HXT2* (3.64-fold), and

several of the PAU genes. Overall, the evidence points towards YNR063W having a role in the DNA replication stage of the cell cycle.

Table 3.11 DNA replication genes DE in the YNR063W null mutant. Nineteen genes with the "DNA replication" gene ontology were differentially expressed in the YNR063W null mutant at day two of Chardonnay wine fermentation. Linear fold change is expressed with wild type as a base line, and descriptions are taken directly from the SGD.

Gene Name	Identifier	Linear Fold Change	Description
CLB6	YGR109C	-5.03	B-type cyclin involved in DNA replication during S phase; activates Cdc28p to promote initiation of DNA synthesis; functions in formation of mitotic spindles along with Clb3p and Clb4p; most abundant during late G1
MCD1	YDL003W	-3.60	Essential subunit of the cohesin complex required for sister chromatid cohesion in mitosis and meiosis; apoptosis induces cleavage and translocation of a C-terminal fragment to mitochondria; expression peaks in S phase
CDC45	YLR103C	-3.36	DNA replication initiation factor; recruited to MCM pre-RC complexes at replication origins; promotes release of MCM from Mcm10p, recruits elongation machinery; mutants in human homolog may cause velocardiofacial and DiGeorge syndromes
CSM3	YMR048W	-2.72	Protein required for accurate chromosome segregation during meiosis
RNR1	YER070W	-2.56	One of two large regulatory subunits of ribonucleotide-diphosphate reductase; the RNR complex catalyzes rate-limiting step in dNTP synthesis, regulated by DNA replication and DNA damage checkpoint pathways via localization of small subunits
MSH2	YOL090W	-2.55	Protein that forms heterodimers with Msh3p and Msh6p that bind to DNA mismatches to initiate the mismatch repair process; contains a Walker ATP-binding motif required for repair activity; Msh2p-Msh6p binds to and hydrolyzes ATP
TOF1	YNL273W	-2.47	Subunit of a replication-pausing checkpoint complex (Tof1p-Mrc1p-Csm3p) that acts at the stalled replication fork to promote sister chromatid cohesion after DNA damage, facilitating gap repair of damaged DNA; interacts with the MCM helicase
MCM10	YIL150C	-2.41	Essential chromatin-associated protein involved in the initiation of DNA replication; required for the association of the MCM2-7 complex with replication origins
POL30	YBR088C	-2.37	Proliferating cell nuclear antigen (PCNA), functions as the sliding clamp for DNA polymerase delta; may function as a docking site for other proteins required for mitotic and meiotic chromosomal DNA replication and for DNA repair
CTF18	YMR078C	-2.37	Subunit of a complex with Ctf8p that shares some subunits with Replication Factor C and is required for sister chromatid cohesion; may have overlapping functions with Rad24p in the DNA damage replication checkpoint
POL12	YBL035C	-2.37	B subunit of DNA polymerase alpha-primase complex, required for initiation of DNA replication during mitotic and premeiotic DNA synthesis; also functions in telomere capping and length regulation
DUN1	YDL101C	-2.29	Cell-cycle checkpoint serine-threonine kinase required for DNA damage-induced transcription of certain target genes, phosphorylation of Rad55p and Sml1p, and transient G2/M arrest after DNA damage; also regulates postreplicative DNA repair

Gene Name	Identifier	Linear Fold Change	Description
RAD27	YKL113C	-2.28	5' to 3' exonuclease, 5' flap endonuclease, required for Okazaki fragment processing and maturation as well as for long-patch base-excision repair; member of the S. pombe RAD2/FEN1 family
DPB2	YPR175W	-2.19	Second largest subunit of DNA polymerase II (DNA polymerase epsilon), required for normal yeast chromosomal replication; expression peaks at the G1/S phase boundary; potential Cdc28p substrate
POL1	YNL102W	-2.10	Catalytic subunit of the DNA polymerase I alpha-primase complex, required for the initiation of DNA replication during mitotic DNA synthesis and premeiotic DNA synthesis
RAD53	YPL153C	-2.08	Protein kinase, required for cell-cycle arrest in response to DNA damage; activated by trans autophosphorylation when interacting with hyperphosphorylated Rad9p; also interacts with ARS1 and plays a role in initiation of DNA replication
POL3	YDL102W	-2.04	Catalytic subunit of DNA polymerase delta; required for chromosomal DNA replication during mitosis and meiosis, intragenic recombination, repair of double strand DNA breaks, and DNA replication during nucleotide excision repair (NER)
CDC9	YDL164C	-2.04	DNA ligase found in the nucleus and mitochondria, an essential enzyme that joins Okazaki fragments during DNA replication; also acts in nucleotide excision repair, base excision repair, and recombination
POL31	YJR006W	-2.01	DNA polymerase III (delta) subunit, essential for cell viability; involved in DNA replication and DNA repair

3.3.4.3 Proteins encoded by four ORFs may regulate the allantoin metabolic process.

Deletion mutants of four ORFs (YLL054C, YKL222C, *PHD1* and *HMS1*) caused differential expression of genes involved in the allantoin metabolic pathway at day two of wine fermentation. The genes *DAL2*, *DAL3* and *DAL7* were downregulated in all four null mutants, while *DUR1,2* was also downregulated in the YKL222C, *PHD1* and *HMS1* null mutants. These DAL genes encode enzymes that work sequentially to metabolize allantoate to malate, while *DUR1,2* encodes an enzyme that degrades urea to ammonia and carbon dioxide as an alternate fate (Figure 3.11). In addition, *DUR3*, a related urea transporter, was also downregulated in the four null mutants, and *DUR3* is known to be induced by allophanate, which is an intermediate downstream of *DAL3* in the pathway. The specific transcriptional responses for the deletion of each of YLR054C, YKL222C, *PHD1* and *HMS1* are discussed below.

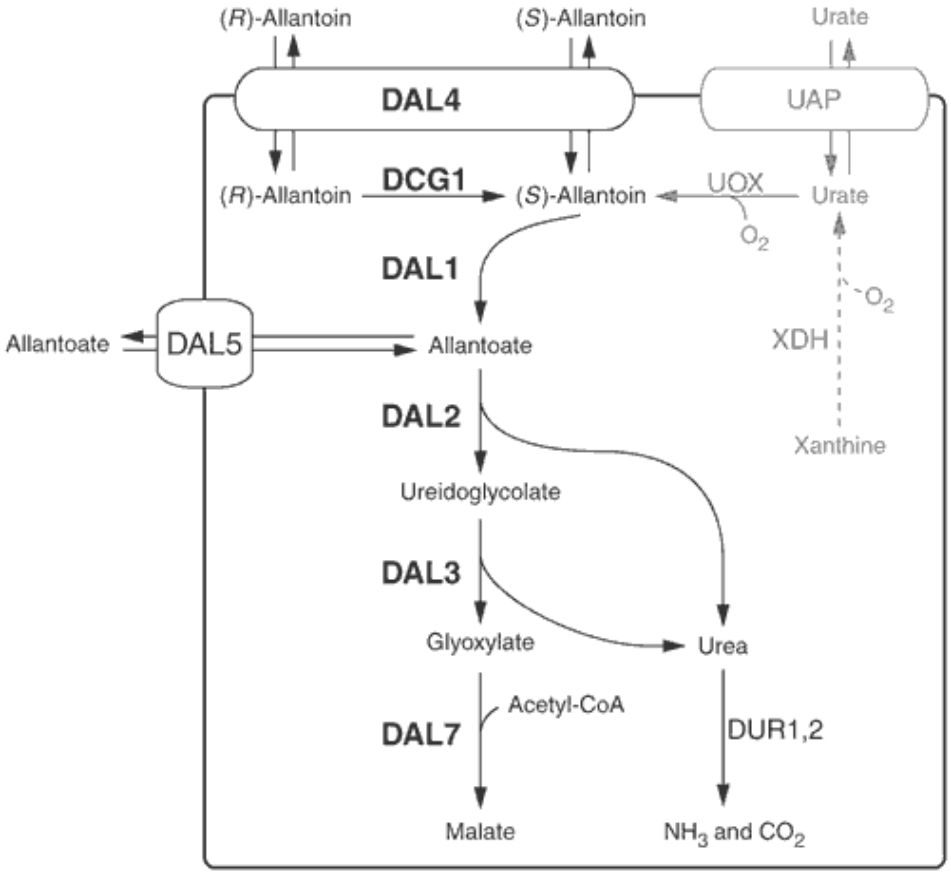


Figure 3.11 The allantoin degradation pathway. The pathway of allantoin metabolism from entry into the cell until formation of either malate or ammonia and carbon dioxide via urea. The lower portion of the pathway, from allantoate onwards, was differentially expressed in the *PHD1*, *HMS1*, YLL054C and YKL222C null mutants. Reproduced with permission from Wong and Wolfe (2005).

The YLL054C null mutant resulted in 66 differentially expressed genes. There were significant gene ontology enrichments for the cellular amide catabolic process (four genes: *DAL2*, *DAL3*, *DAL7*, *DUR3*) as well as molecular function ontology enrichment for structural component of cell wall (six genes: *PAU2*, *PAU3*, *PAU5*, *PAU7*, *PAU9*, *PAU24*). The highest upregulation was seen for PAU genes, with a maximum 4.47-fold upregulation compared to wild type for *PAU15*. *PCL1* and *CLB6*, both cyclins, were the most downregulated genes at 3.79-fold and 3.99-fold changes, respectively.

The YKL222C null mutant resulted in 69 differentially expressed genes. There was a significant gene ontology enrichment for the allantoin metabolic process (four genes: *DAL2*, *DAL3*, *DAL7*, *DUR1,2*), plus the larger organonitrogen compound catabolic process (nine genes: *DAL2*, *DAL3*, *DAL7*, *DUR1,2*, *ARO10*, *ARO9*, *GCV1*, *SWI4*, *HMX1*), as well as a localization ontology enrichment for fungal cell wall

(nine genes: *PAU2*, *PAU3*, *PAU5*, *PAU7*, *PAU9*, *PAU23*, *PAU24*, *TIR3*, *AGA1*), and a molecular function ontology enrichment for structural component of fungal cell wall (eight genes: *PAU2*, *PAU3*, *PAU5*, *PAU7*, *PAU9*, *PAU23*, *PAU24*, *TIR3*). Furthermore, there was a pathway enrichment for the allantoin degradation pathway, as *DAL2*, *DAL3* and *DUR1,2* were all significantly downregulated compared to wild type. The highest downregulation was of *DAL3*, with 4.30-fold lower expression than in wild type. In contrast, the seven PAU genes were all upregulated compared to wild type, including largest fold-change at 4.58-fold upregulation observed for *PAU3*.

The *PHD1* null mutant resulted in 45 differentially expressed genes compared to wild type. There was a significant gene ontology enrichment for the allantoin metabolic process (four genes: *DAL2*, *DAL3*, *DAL7*, *DUR1,2*), plus the larger organonitrogen compound catabolic process (eight genes: *DAL2*, *DAL3*, *DAL7*, *DUR1,2*, *DUR3*, *ARO10*, *ARO9*, *ALD3*). There was also a cellular localization enrichment for integral components of the plasma membrane (13 genes: *UGA4*, *HXT7*, *HXT2*, *MSB2*, *DUR3*, *GAP1*, *YLL053C*, *PHO84*, *MEP2*, *BIO5*, *PNS1*, *PMA2*, *OPT2*), as well as a corresponding molecular function ontology enrichment for substrate-specific transmembrane transport (12 genes: *UGA4*, *HXT7*, *HXT2*, *DUR3*, *GAP1*, *YLL053C*, *PHO84*, *MEP2*, *BIO5*, *PNS1*, *PMA2*, *OPT2*). Transport proteins were both upregulated and downregulated, while the allantoin metabolic genes were specifically downregulated compared to wild type.

The *HMS1* null mutant resulted in 121 differentially expressed genes compared to wild type. There was a significant gene ontology enrichment for the allantoin metabolic process (four genes: *DAL2*, *DAL3*, *DAL7*, *DUR1,2*), as well as a localization ontology enrichment for integral component of the plasma membrane (eleven genes: *UGA4*, *HXT7*, *HXT4*, *HXT2*, *MSB2*, *DUR3*, *PHO84*, *BIO5*, *PNS1*, *PMA2*, *OPT2*), and a molecular function ontology enrichment for pentose transmembrane transporter activity (three genes: *HXT7*, *HXT4*, *HXT2*). Furthermore, there was a pathway enrichment for the allantoin degradation pathway, as *DAL2*, *DAL3* and *DUR1,2* were all significantly downregulated compared to wild type. The hexose transporters *HXT2*, *HXT4* and *HXT7* were all upregulated 2.4 to 4.2-fold compared to wild type. The three most highly upregulated genes (YOL084W, YNL194C, YNL195C) are all

proteins of unknown function, with up regulation ranging from 5.8 to 7.9-fold increase compared to wild type. Once again, *PCL1* and *CLB6*, both cyclins, were the most downregulated genes at 4.3-fold and 3.9-fold changes respectively.

While all four null mutants caused downregulation of the allantoin metabolism genes, they also each resulted in unique transcriptomic changes. Furthermore, the magnitude of the transcriptional change for each of the DAL and DUR genes was different for each of the null mutants (Table 3.12). The YLL054C null mutant resulted in the smallest changes in allantoin metabolic gene expression, but this mutant also caused significant upregulation of PAU genes. The YKL222C null mutant resulted in larger changes in allantoin metabolic gene expression, and also caused significant upregulation of the same PAU genes as YLL054C. The biological function of PAU genes is still not well understood, but they are induced under anaerobiosis, stress conditions and wine fermentation (Rachidi *et al.*, 2000; Luo and van Vuuren, 2009). As the PAU genes are localized to the cell wall (Huh *et al.*, 2003), they could play a role in nutrient sensing of nitrogen sources, which could explain the upregulation of PAU genes being coupled to downregulation of the allantoin metabolic pathway. The YLL054C and YKL222C gene products could mediate this interaction.

Table 3.12 Expression of DAL and DUR genes in transcription factor mutants. The linear fold change of the expression of allantoin metabolic genes determined by microarray as compared to the wild type control at day two of wine fermentation was determined for four transcription factor null mutants: YLL054C, YKL222C, *PHD1* and *HMS1*.

	Linear Change of Gene Expression				
Null Mutant	*DAL2*	*DAL3*	*DAL7*	*DUR1,2*	*DUR3*
YLL054C	-2.54	-3.49	-2.49	NA	-2.16
YKL222C	-3.02	-4.30	-3.93	-2.92	-3.11
PHD1	-3.04	-4.17	-4.25	-2.43	-2.30
HMS1	-3.44	-3.58	-2.97	-2.07	-2.74

The *HMS1* and *PHD1* null mutants both resulted in significant changes in integral components of the plasma membrane, in addition to the effects on the allantoin metabolic pathway. Notably, three high-affinity glucose transporters (*HXT2* and *HXT6/HXT7*) were upregulated in both *HMS1* and *PHD1* null mutants, and a third high-affinity glucose transporter (*HXT4*) was also upregulated in the *HMS1* null

mutant. Several other transporters: *OPT2*, *UGA4*, *BIO5* and *PHO84*, encoding oligopeptide, γ-aminobutyrate, biotin and phosphate transporters, respectively, were all downregulated in the *HMS1* and *PHD1* null mutants. The results indicate that in addition to their role in regulating the allantoin metabolic pathway, *HMS1* and *PHD1* may also have effects on the import and export of various metabolites during wine fermentation.

3.3.5 Constitutive expression and null mutants produce small overlaps in differentially expressed genes.

Microarray analysis of RNA transcript abundance was conducted for constitutive expression mutants, in addition to null mutants, of two poorly annotated transcription factors (YKL043W/*PHD1* and YOR032C/*HMS1*) sampled on day two of Chardonnay wine fermentation. Fermentation conditions and sampling for both the null mutant and constitutive expression mutants were performed in the same manner, and microarray analysis of the transcriptome for all samples was compared against an Enoferm M2 wild type control.

The *PHD1* constitutive expression mutant had only a 1.6-fold increase in expression over M2 wild type at day two of fermentation, although the signal for wild type was 13.81 and the constitutive expression mutant was 14.49 on a $\log_2$ scale. The relatively small increase in expression of *PHD1* in the constitutive expression mutant may simply reflect a strong natural expression at this time point which was not greatly improved by the use of the *PGK1* promoter. Regardless of the minor increase in expression of *PHD1*, the constitutive expression mutant resulted in 99 differentially expressed transcripts compared to wild type. These differentially expressed genes had a significant ontology enrichment for cell wall organization (21 genes: *PAU7*, *PAU9*, *PAU3*, *HLR1*, *PAU2*, *PAU5*, *WSC4*, *SSP1*, *PFS1*, *TIR3*, *PAU14*, *PGU1*, *PAU17*, *PAU23*, *NCW2*, *MID2*, *YPK2*, *HPF1*, *TIR4*, *TIR2*, *SRL1*). There was also a corresponding cellular localization ontology enrichment for the cell periphery (34 genes: *PAU7*, *PAU9*, YAR029W, *KCC4*, *GEX1*, *PAU3*, *GIC2*, *PAU2*, *FCY22*, *FTR1*, *PAU5*, *BUD9*, *YHK8*, *TIR3*, *PDR11*, *PAU14*, *FLO11*, *FAT3*, *PAU17*, *PAU23*, YLR042C, *NCW2*, *MID2*, *HXT2*, *FRE4*, *SKM1*, *HPF1*, *TIR4*,

TIR2, *AUS1*, *THI72*, *SRL1*, *AMF1*, *OPT2*), and molecular function ontology enrichment for structural constituent of the cell wall (12 genes: *PAU2*, *PAU3*, *PAU5*, *PAU7*, *PAU9*, *PAU14*, *PAU17*, *PAU23*, *TIR2*, *TIR3*, *NCW2*). All categories were strongly bolstered by the presence of many differentially expressed PAU genes, encoding seripauperin-family proteins of unknown function that have been implicated in the fermentation stress response. The magnitude of differential expression caused by the constitutive expression of *PHD1* was relatively larger than those caused by the deletion of *PHD1*. The most heavily downregulated gene was *HPF1* (encoding mannoprotein haze-protective factor), which was expressed 19.4-fold less in the mutant than in wild type. The most heavily upregulated genes were *YSR3*, encoding a dihydrosphingosine 1-phosphate phosphatase, and *PAU15*, which were upregulated 10.83-fold and 10.39-fold compared to wild type, respectively.

There were nine genes that were differentially expressed in both the *PHD1* null mutant and the *PHD1* constitutive expression mutant: *PAU7*, *HSP26*, *ARO10*, *HLR1*, *CLB6*, *YOX1*, *HXT2*, *BAG7*, and *OPT2* (Table 3.13). The curious result was that in the case of every gene except for HSP26, the genes were upregulated or downregulated in both the null mutant and the constitutive expression mutant. *HSP26* was the only gene that was downregulated (9.03-fold change) in the constitutive expression mutant and upregulated (2.22-fold change) in the null mutant; *HSP26* encodes a small heat shock protein with chaperone activity and mRNA binding activity. It is unclear why the remaining genes were regulated in the same manner for both constitutive and null mutants, but it could be due to feedback mechanisms caused by the perturbation of this gene, rather than a direct effect.

Table 3.13 Genes differentially expressed in *PHD1* null and constitutive expression mutants. Microarray analysis of *PHD1* null and constitutive expression mutants fermenting Chardonnay wine and sampled at day two were compared to wild type fermentations, and significant differentially expressed genes from both mutant datasets are reported.

Transcript ID	Log2 Fold-Change vs. Wild Type		Gene Symbol
	Null	Constitutive	
YAR020C	2.01	3.74	*PAU7*
YBR072W	2.22	-9.03	*HSP26*
YDR380W	-2.05	-3.62	*ARO10*
YDR528W	-2.35	-2.57	*HLR1*
YGR109C	-2.40	-4.41	*CLB6*

Transcript ID	Log2 Fold-Change vs. Wild Type		Gene Symbol
	Null	Constitutive	
YML027W	-2.17	-2.33	*YOX1*
YMR011W	2.13	2.62	*HXT2*
YOR134W	2.47	2.88	*BAG7*
YPR194C	-3.01	-3.25	*OPT2*

The *HMS1* constitutive expression mutant had a moderate 11.78-fold increase in expression compared to the wild type control at day two of fermentation. The constitutive expression mutant of *HMS1* resulted in 399 differentially expressed genes. There was a significant gene ontology enrichment for ribosome biogenesis (64 genes). Unsurprisingly, there was a corresponding cellular localization ontology enrichment for the nucleolus (63 genes) and the preribosome (43 genes). There was no molecular function ontology enrichment or pathway enrichment found for the *HMS1* constitutive expression mutant. The mutant resulted in considerable expression changes for several genes; nine genes had greater than 10-fold changes in expression between the mutant and wild type (Table 3.14). The greatest down regulation was for *SIP18* (29.69-fold decrease), a hydrophilin important in desiccation and osmotic stress. The other three strongly downregulated genes were *CTR3* and *PHO84*, encoding copper and phosphate transporters, respectively, and *FET3*, encoding an iron and copper oxidase. Highly upregulated genes included one putative protein of unknown function, YKL070W, which was upregulated 25.65-fold over wild type, as well as *YSR3*, a membrane protein involved in sphingolipid metabolism which was upregulated 10.06-fold. In addition, the high-affinity glucose transporter *HXT4* was upregulated 11.24-fold, and the maltase *MAL12* was upregulated 67.39-fold over wild type. Finally, *PGU1*, encoding a pectolytic enzyme, was upregulated 136.38-fold over wild type. Overall, these highly differentially expressed genes indicated that *HMS1* plays a role in promoting the consumption of alternative carbon sources, while mediating the copper and heavy metal toxicity response.

Table 3.14 HMS1 constitutive expression mutant DE genes with > 10-fold change. The *HMS1* constitutive expression mutant resulted in several genes with large fold-changes in expression, both upregulated and downregulated, compared to wild type on day two of Chardonnay wine fermentation.

Gene	Fold-change (Mutant – WT)	Description
PGU1	136.38	Endo-polygalacturonase; pectolytic enzyme that hydrolyzes the alpha-1,4-glycosidic bonds in the rhamnogalacturonan chains in pectins
MAL12	67.39	Maltase (alpha-D-glucosidase); inducible protein involved in maltose catabolism; encoded in the MAL3 complex locus; functional in genomic reference strain S288C; hydrolyzes the disaccharides maltose, turanose, maltotriose, and sucrose
YKL070W	25.65	Putative protein of unknown function; expression induced in cells treated with mycotoxins patulin or citrinin; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
HXT4	11.24	High-affinity glucose transporter; member of the major facilitator superfamily, expression is induced by low levels of glucose and repressed by high levels of glucose; HXT4 has a paralog, HXT7, that arose from the whole genome duplication
YSR3	10.06	Dihydrosphingosine 1-phosphate phosphatase; membrane protein involved in sphingolipid metabolism; YSR3 has a paralog, LCB3, that arose from the whole genome duplication
CTR3	-11.14	High-affinity copper transporter of the plasma membrane; acts as a trimer; gene is disrupted by a Ty2 transposon insertion in many laboratory strains of S. cerevisiae
FET3	-16.72	Ferro-O2-oxidoreductase; multicopper oxidase that oxidizes ferrous (Fe2+) to ferric iron (Fe3+) for subsequent cellular uptake by transmembrane permease Ftr1p; required for high-affinity iron uptake and involved in mediating resistance to copper ion toxicity, belongs to class of integral membrane multicopper oxidases; protein abundance increases in response to DNA replication stress
PHO84	-21.66	High-affinity inorganic phosphate (Pi) transporter; also low-affinity manganese transporter; regulated by Pho4p and Spt7p; mutation confers resistance to arsenate; exit from the ER during maturation requires Pho86p; cells overexpressing Pho84p accumulate heavy metals but do not develop symptoms of metal toxicity
SIP18	-29.69	Phospholipid-binding hydrophilin; essential to overcome desiccation-rehydration process; expression is induced by osmotic stress; SIP18 has a paralog, GRE1, that arose from the whole genome duplication

There were 31 genes that were found to be differentially expressed in both the *HMS1* null mutant and the *HMS1* constitutive expression mutant (Table 3.15). This list of genes was significantly ($p = 0.00089$) enriched for pentose transmembrane transporter activity, represented by three genes: *HXT2*, *HXT4*, and *HXT7*. All three genes encode high-affinity glucose transporters repressed by high levels of glucose, and all three genes were upregulated in both the constitutive expression and the null mutants. The up regulation was similar for *HXT7* in both mutants (3.50-fold in null, 3.33-fold in CE), while for *HXT2* the up regulation was slightly stronger in the null mutant (4.19-fold in null, 2.59-fold in CE), and for *HXT4* the up regulation was much stronger in the constitutive expression mutant (2.41-fold in null, 11.24-fold in CE). Similar to the *PHD1* mutants, only four genes were upregulated in one mutant and downregulated in the other; the remaining 27 genes were either upregulated or downregulated in both the null and

constitutive expression mutants. Curiously, while *PGU1* was significantly upregulated in the constitutive expression mutant, *PGU1* does not show up on the list of transcripts for the deletion mutant. *SIP18*, *PHO84* and *HXT4* were the only genes that were highly upregulated in the constitutive expression mutant and also differentially expressed in the null mutant. The *SIP18* gene was downregulated 29.69-fold in the constitutive expression mutant and upregulated 2.11-fold in the null mutant, while *PHO84* was downregulated 21.66-fold in the constitutive expression mutant and also downregulated 2.41-fold in the null mutant. These results highlight the differences between deletion and constitutive expression in altering gene expression. Based on these results, however, it could be postulated that *HMS1* may be involved in the regulation of glucose transport.

Table 3.15 Genes differentially expressed in *HMS1* null and constitutive expression mutants. Microarray analysis of *HMS1* null and constitutive expression mutants fermenting Chardonnay wine and sampled at day two were compared to wild type fermentations, and significant differentially expressed genes from both mutant datasets are reported.

Transcript ID	Log2 Fold-Change vs. Wild Type		Gene Symbol
	Constitutive	Null	
YAR068W	2.43	2.97	YAR068W
YBR240C	-2.2	-2.03	*THI2*
YCR104W	2.96	2.22	*PAU15*
YDL214C	4.21	2.86	*PRR2*
YDR034W-B	-2.73	2.11	YDR034W-B
YDR258C	2.19	2.26	*HSP78*
YDR281C	-2.74	-2.43	*PHM6*
YDR342C	3.5	3.34	*HXT6*
YER103W	4.82	4.22	*SSA4*
YER127W	-2.04	-2.07	*LCP5*
YFR015C	5.74	2.90	*GSY1*
YGR043C	-2.94	2.66	*NQM1*
YGR109C	-2.4	-3.94	*CLB6*
YGR142W	2.12	2.58	*BTN2*
YHL021C	3.14	2.42	*AIM17*
YHR092C	11.24	2.41	*HXT4*
YJL122W	-2.35	-2.50	*ALB1*
YJR115W	2.62	2.41	YJR115W
YJR154W	-5.82	-2.66	YJR154W
YKR077W	-2.94	-2.48	*MSA2*
YLR327C	9.59	3.99	*TMA10*
YML027W	-2.72	-3.77	*YOX1*

Transcript ID	Log2 Fold-Change vs. Wild Type		Gene Symbol
	Constitutive	Null	
YML123C	-21.66	-2.41	*PHO84*
YMR011W	2.59	4.19	*HXT2*
YMR175W	-29.69	2.11	*SIP18*
YOL016C	3.54	2.19	*CMK2*
YOR032C	11.78	-187.71	*HMS1*
YOR100C	2.8	2.29	*CRC1*
YOR134W	2.02	3.85	*BAG7*
YOR359W	-2.02	-2.25	*VTS1*
YPL036W	-5.29	2.32	*PMA2*

3.3.6 Deletion of *HXK2* and *GEP5* both result in extensive transcriptomic effects.

Microarray analysis of RNA transcript abundance was conducted for *GEP5* and *HXK2* null mutants sampled on day two and day seven of Chardonnay wine fermentation, respectively. The *HXK2* gene encodes a known hexokinase, which is thought to be the dominant hexokinase during growth on 2% w/v glucose. The *GEP5* gene is classified as encoding a protein of unknown function, but is also known to be required for the maintenance of the mitochondrial genome (Merz and Westermann, 2009). Null mutants of each gene were compared against an Enoferm M2 control fermentation.

The *HXK2* null mutant resulted in 86 differentially expressed genes on day seven of the wine fermentation. This number includes *HXK2*, which was itself downregulated 43-fold, as would be expected for a null mutant. The differentially expressed genes were significantly enriched for the cellular localization ontology of "intrinsic component of the plasma membrane", represented by eleven genes: *SUL1*, *LPP1*, *STL1*, *HXT4*, *GAP1*, *NCW2*, *PHO84*, *BIO5*, *PUT4*, *PMA2* and *AQY1* (Figure 3.12). The vast majority of the differentially expressed genes, 67 out of the total 87, were downregulated in the null mutant. Notably, the proline permease *PUT4* was downregulated 10.84-fold and the proline oxidase *PUT1* was downregulated 4.26-fold. Cytochrome b2 and cytochrome c were also both downregulated in the *HXK2* null mutant, 8.58-fold and 4.55-fold respectively. Ten of the 20 upregulated genes encode proteins or putative proteins of unknown function. Of note was the absence of *HXK1* from the list of upregulated genes; despite the deletion of *HXK2* there appeared to be no compensatory action being taken

by the other well-known hexokinase, *HXK1*. It is possible that *HXK1* was being constitutively expressed at the same level in both wild type and the *HXK2* null mutant, but this would be surprising given previous reports of *HXK1* expression being controlled directly by *HXK2* (Rodriguez *et al.*, 2001).

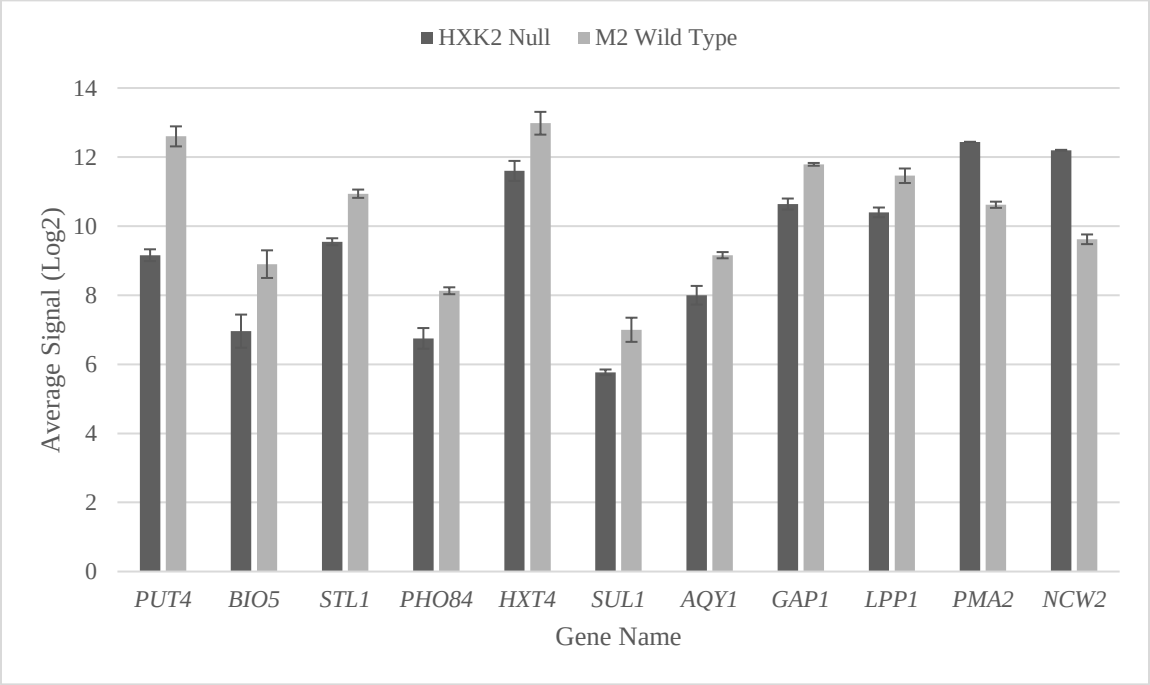


Figure 3.12 Expression of "intrinsic component of membrane" genes in *HXK2* null mutant. Eleven genes that were differentially expressed in the *HXK2* null mutant were annotated with the "intrinsic component of membrane" gene ontology. Expression is reported as the average log signal for each genotype, with error bars representing one standard deviation.

The *GEP5* null mutant resulted in 270 differentially expressed genes at day two of Chardonnay wine fermentation; *GEP5* itself was downregulated 28.56-fold, as was expected for a null mutant. The most highly downregulated gene, however, was YEL073C, a putative protein located in the subtelomeric region of chromosome V. This small 324 bp ORF was expressed 40.75-fold less in the *GEP5* null mutant than in wild type. The third most highly downregulated gene (18.22-fold less than wild type), YGL138C, also encodes a putative protein of unknown function. Neither YEL073C nor YGL138C overlap with any other ORFs. The *GEP5* null mutant demonstrated a pathway enrichment for genes involved in the sulphate assimilation pathway; *MET3*, *MET5*, *MET10*, *MET14* and *MET16* were all upregulated in the null mutant, from 2.13-fold (*MET10*) to 3.40-fold (*MET16*) compared to wild type (Figure 3.13). In addition to sulphate metabolism, the MET genes are also involved in methionine metabolism. The S-

adenosylmethionine synthetase gene, *SAM1*, was upregulated 2.56-fold, and the high-affinity S-methylmethionine permease, *MMP1*, was also upregulated 5.20-fold in the *GEP5* null mutant.

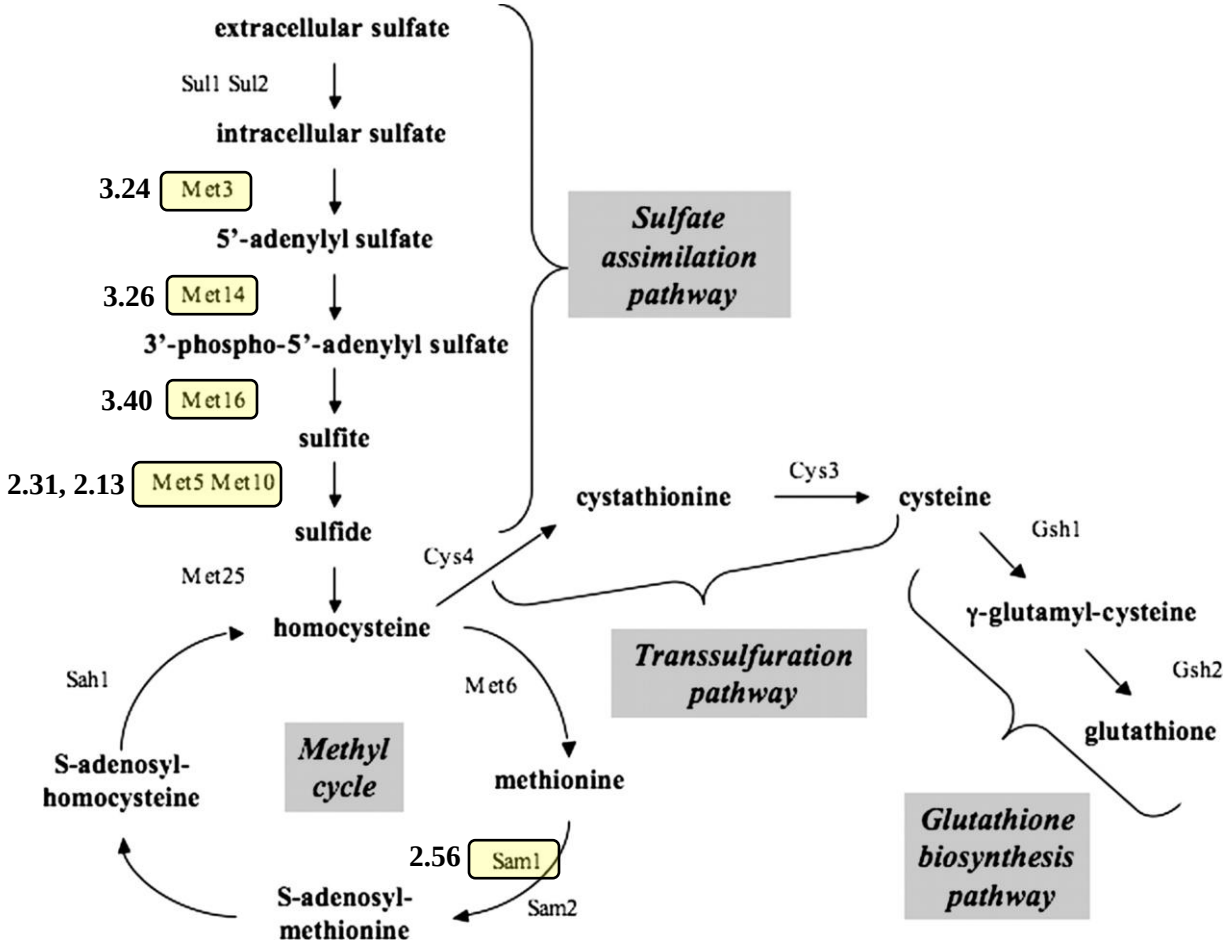


Figure 3.13 The sulphate assimilation pathway. Many of the enzymes involved in the sulphate assimilation pathway were upregulated in the *GEP5* null mutant. Differentially expressed enzymes are highlighted in yellow, with fold-change appearing to the left of the enzyme. Reprinted with permission from Pereira *et al.* (2008).

Differentially expressed genes in the *GEP5* null mutant were also significantly enriched for the biological function gene ontology of "cellular component assembly involved in morphogenesis" (16 genes, p = 7.899e-5) and "ribosome biogenesis" (41 genes, p = 0.005410). There was also a cellular component ontology enrichment for preribosome (30 genes, p = 5.109e-9). The genes involved in the "cellular component assembly for morphogenesis", which govern different aspects of ascospore formation, were all downregulated in the null mutant (Table 3.16). While the most minor change was a 2.2-fold decrease for *AMA1*, several genes had quite strong changes; *OSW1*, *SPS100* and *SMA1* were downregulated 13.8-fold, 9.76-fold and 6.91-fold, respectively. While all involved in sporulation, the

precise function of these genes is not well understood, although both *OSW1* and *SMA1* are known to interact with Spo14p, a phospholipase D involved in membrane assembly during sporulation. The 41 genes involved in ribosome biogenesis were all upregulated in *GEP5* null mutant. Unlike the large changes seen for the ascospore formation genes, the fold-changes seen for ribosome biogenesis were quite minor, ranging from the bare minimum of 2.01-fold to a maximum of 3.09-fold increase over wild type. The contrasting decrease in sporulation and increase in ribosome biogenesis observed in the *GEP5* null mutant could have indicated a general slowing of the yeast life cycle, as an increase in sporulation and a decrease in ribosome biogenesis is typically observed later in the wine fermentation.

Table 3.16 "Cellular component assembly involved in morphogenesis" genes DE in *GEP5* null mutant. The "cellular component assembly involved in morphogenesis" gene ontology was enriched in the *GEP5* null mutant DE genes at day two of wine fermentation, and all genes were downregulated compared to wild type.

Gene Symbol	Identifier	Fold Change	Description
OSW1	YOR255W	-13.8	Protein involved in sporulation; required for the construction of the outer spore wall layers; required for proper localization of Spo14p
SPS100	YHR139C	-9.76	Protein required for spore wall maturation; expressed during sporulation; may be a component of the spore wall; expression also induced in cells treated with the mycotoxin patulin
SMA1	YPL027W	-6.91	Protein of unknown function involved in the assembly of the prospore membrane during sporulation; interacts with Spo14p
SPO19	YPL130W	-4.46	Meiosis-specific prospore protein; required to produce bending force necessary for proper assembly of the prospore membrane during sporulation; identified as a weak high-copy suppressor of the spo1-1 ts mutation
OSW2	YLR054C	-4.34	Protein of unknown function proposed to be involved in the assembly of the spore wall
GAS4	YOL132W	-3.33	1,3-beta-glucanosyltransferase, involved with Gas2p in spore wall assembly; has similarity to Gas1p; localizes to the cell wall
LOH1	YJL038C	-3.23	Protein of unknown function with proposed roles in maintenance of genome integrity and also in spore wall assembly; induced during sporulation; repressed during vegetative growth by Sum1p and Hst1p; sequence similar to *IRC1*
CRR1	YLR213C	-3.10	Putative glycoside hydrolase of the spore wall envelope; required for normal spore wall assembly, possibly for cross-linking between the glucan and chitosan layers; expressed during sporulation
IRC18	YJL037W	-3.07	Putative protein of unknown function; expression induced in respiratory-deficient cells and in carbon-limited chemostat cultures; similar to adjacent ORF, YJL038C; null mutant displays increased levels of spontaneous Rad52p foci
SPS22	YCL048W	-3.05	Protein of unknown function, redundant with Sps2p for the organization of the beta-glucan layer of the spore wall
FUS2	YMR232W	-2.97	Cytoplasmic protein localized to the shmoo tip; required for the alignment of parental nuclei before nuclear fusion during mating
SPO77	YLR341W	-2.62	Meiosis-specific protein of unknown function, required for spore wall formation during sporulation; dispensable for both nuclear divisions during meiosis

Gene Symbol	Identifier	Fold Change	Description
DTR1	YBR180W	-2.44	Putative dityrosine transporter, required for spore wall synthesis; expressed during sporulation; member of the major facilitator superfamily (DHA1 family) of multidrug resistance transporters
RRT8	YOL048C	-2.28	Putative protein of unknown function; identified in a screen for mutants with increased levels of rDNA transcription; green fluorescent protein (GFP)-fusion protein localizes to lipid particles
DON1	YDR273W	-2.24	Meiosis-specific component of the spindle pole body, part of the leading edge protein (LEP) coat, forms a ring-like structure at the leading edge of the prospore membrane during meiosis II
AMA1	YGR225W	-2.20	Activator of meiotic anaphase promoting complex (APC/C); Cdc20p family member; required for initiation of spore wall assembly; required for Clb1p degradation during meiosis

3.4 Proteomic changes during fermentation are both time and genotype dependent

3.4.1 The Enoferm M2 wild type proteome changes throughout the wine fermentation

Enoferm M2 wild type yeast was used as a control for two independent fermentations of Chardonnay wine sampled at three time points. This resulted in two independent datasets each with three biological replicates for the Enoferm M2 proteome over three time points: day two, day four and day seven. The results from each dataset were processed as described in the methods, with statistical testing (Mann-Whitney test, $p < 0.5$, with Benjamini-Hochberg test correction and 1.5-fold change cut off) of the difference from day two – day four and from day four – day seven. The results from the two independent experiments were combined to create a consensus dataset of proteins that were significantly different in abundance.

There was a consensus list of seven differentially expressed proteins between day two and day four of the Chardonnay wine fermentation: Hsp26p, Acs1p, Ino1p, Tkl2p, Ygr201cp, Hsp12p, Nqm1p and Thi4p (Table 3.17). The greatest fold-change was observed for the small heat shock protein Hsp26p, which had an average 2.01-fold up regulation at day four when taking both datasets into account. The ubiquitous heat shock protein Hsp12p was also upregulated at day four. The Tkl2p is a transketolase required for the synthesis of aromatic amino acids and its up regulation could have been associated with the depletion of nitrogen sources in the grape juice. The Ino1p is an inositol 1-phosphate synthase involved in the synthesis of inositol containing phospholipids and could play a role in cell membrane

remodeling as the media conditions change in the wine fermentation. The Acs1p is an acetyl-CoA synthetase known to be expressed under aerobic conditions; its up regulation towards day four of the fermentation was therefore surprising given the displacement of oxygen by carbon dioxide that occurs as the fermentation progresses. One significantly upregulated protein, Ygr201cp, has no known function and is worth pursuing for further annotation. The Ngm1p is a transaldolase that was significantly upregulated, but its function in the diauxic shift remains unknown as well. An enzyme involved in thiamine biosynthesis, Thi4p, was the only protein significantly downregulated at day four.

Table 3.17 Wild type time-course proteomic results. Significantly different proteins during the time course of fermentation, using day two as reference for day four and day four as reference for day seven. * indicates similar trend found in transcriptomic results.

Day	ID	Description	Average Linear Change	Average p-value
4	*HSP26*/ YBR072W	Small heat shock protein with chaperone activity	2.01	< 0.0001
4	*ACS1*/ YAL054C*	Acetyl-coA synthetase isoform expressed during growth on nonfermentable carbon sources and under aerobic conditions	1.88	< 0.0001
4	*INO1*/ YJL153C*	Inositol 1-phosphate synthase, involved in synthesis of inositol phosphates and inositol-containing phospholipids	1.85	< 0.0001
4	*TKL2*/ YBR117C*	Transketolase, catalyzes conversion in the pentose phosphate pathway; needed for synthesis of aromatic amino acids	1.81	< 0.0001
4	YGR201C*	Uncharacterized ORF, Putative protein of unknown function	1.78	0.001115
4	*HSP12*/ YFL014W	Plasma membrane localized protein that protects membranes from desiccation; induced by heat shock, oxidative stress, osmostress, stationary phase entry, glucose depletion, oleate and alcohol	1.70	0.00011
4	*NQM1*/ YGR043C*	Transaldolase of unknown function; transcription is repressed by Mot1p and induced by alpha-factor and during diauxic shift	1.58	< 0.0001
4	*THI4*/ YGR144W	Thiazole synthase, catalyzes formation of a thiazole intermediate during thiamine biosynthesis	-2.05	0.001655
7	*RPS11B*/ YBR048W	Protein component of the small (40S) ribosomal subunit	-1.83	0.00415
7	*THI4*/ YGR144W	Thiazole synthase, catalyzes formation of a thiazole intermediate during thiamine biosynthesis	-1.96	0.000645

There was a consensus list of two differentially expressed proteins between day four and day seven of the Chardonnay wine fermentation, with similar magnitude of changes to the earlier time interval (Table 3.17). The greatest magnitude of change was observed for the thiazole synthase Thi4p, which

decreased an average of 1.96-fold in abundance from day four – day seven, in addition to a similar fold decrease from day two – day four, indicating a consistent decrease as the fermentation progresses. The other protein that was differentially expressed was Rps11bp, which demonstrated a 1.83-fold reduction over this time interval. Rps11bp is a component of the small (40S) ribosomal subunit, and its down regulation was consistent with reduced ribosomal synthesis later in fermentation.

3.4.2 The mutation of novel ORFs affects the yeast proteome during fermentation

Proteomic analysis was conducted for six novel ORF null mutants (M13, M17, M18, M21, M23 and M28) and five constitutive expression mutants (M13, M18, M21, M23 and M28) sampled at three time points (day two, day four and day seven) during a Chardonnay wine fermentation, using three biological replicates for each of the mutants and wild type. The results were processed as described in the methods, with statistical testing (Mann-Whitney test, $p < 0.5$, with Benjamini-Hochberg test correction) of the difference between the mutant and Enoferm M2 wild type at each time point sampled. The deletion or constitutive expression of all novel ORFs except M21 resulted in differential protein expression for at least one time point sampled, for at least one of the mutants (Table 3.18). Unfortunately, the proteomic analysis used for the iTRAQ-generated data did not allow for the identification of novel proteins; only yeast proteins that were present in the pre-established searchable peptide database. Re-analyzing the data with a custom database specific to the Enoferm M2 strain may be able to detect relative abundances of novel proteins present in this strain.

Table 3.18 Proteomic results from the null and constitutive expression of novel ORFs. The proteome was assessed for novel ORF mutants and wild type sampled at three time points during Chardonnay fermentation; genes were determined as significantly different from wild type by Mann-Whitney test with Benjamini-Hochberg correction (*p*-value < 0.05, > 1.5 linear fold-change).

ORF	Mutant type	Day of Fermentation	Short Name	Identifier	Mann-Whitney Test *p*-value	Linear Fold-Change (Mutant / WT)
M13	Constitutive	2	*THI11*	YJR156C	< 0.0001	-1.58
M13	Null	2	*ADH4*	YGL256W	< 0.0001	1.67
M13	Null	7	*UGA1*	YGR019W	0.00063	1.61
M17	Null	2	*THI11*	YJR156C	< 0.0001	-1.80
M17	Null	2	*THI4*	YGR144W	< 0.0001	-1.66

ORF	Mutant type	Day of Fermentation	Short Name	Identifier	Mann-Whitney Test *p*-value	Linear Fold-Change (Mutant / WT)
M18	Null	2	*GRX1*	YCL035C	< 0.0001	1.51
M18	Null	2	*PRD1*	YCL057W	< 0.0001	1.61
M18	Null	2	*RIM1*	YCR028C-A	< 0.0001	1.73
M18	Null	2	*LSB5*	YCL034W	0.00091	1.82
M18	Null	2	*YCP4*	YCR004C	< 0.0001	1.87
M18	Null	4	*SNZ3*	YFL059W	0.0043	-1.68
M18	Null	4	*HBN1*	YCL026C-B	< 0.0001	1.54
M18	Null	4	*RNQ1*	YCL028W	0.00038	1.55
M18	Null	4	*GRX1*	YCL035C	< 0.0001	1.60
M18	Null	4	*ILV6*	YCL009C	< 0.0001	1.61
M18	Null	4	*LSB5*	YCL034W	0.00028	1.62
M18	Null	4	*SYP1*	YCR030C	< 0.0001	1.62
M18	Null	4	*GLK1*	YCL040W	< 0.0001	1.65
M18	Null	4	*RIM1*	YCR028C-A	< 0.0001	1.74
M18	Null	4	*PRD1*	YCL057W	< 0.0001	1.77
M18	Null	4	*YCP4*	YCR004C	< 0.0001	1.82
M23	Constitutive	4	YPL245W	YPL245W	< 0.0001	-2.58
M28	Constitutive	2	*HSP26*	YBR072W	< 0.0001	-1.72
M28	Constitutive	2	*THI11*	YJR156C	< 0.0001	-1.65
M28	Constitutive	2	*RPS11B*	YBR048W	< 0.0001	1.55
M28	Constitutive	2	*RPL36A*	YMR194W	0.0039	1.57

The M13 null mutant resulted in two differentially expressed proteins, and the constitutive expression mutant resulted in one differentially expressed protein; Thi11p, involved in thiamine biosynthesis, was downregulated in the constitutive expression mutant. The zinc-dependent alcohol dehydrogenase, Adh4p, was upregulated at day two of fermentation in the null mutant, while Uga1p, an enzyme involved in 4-aminobutyrate and glutamate degradation, was upregulated at day seven of fermentation.

The M17 null mutant resulted in two differentially expressed proteins, both at day two of the fermentation. These were Thi11p and Thi4p, both proteins involved in the synthesis of thiamine. These results would be worth following up with a specific assay to monitor thiamine levels in the null mutant.

The M18 null mutant resulted in eleven differentially expressed proteins over day two and day four of fermentation. Five of these proteins, encoded by *GRX1*, *PRD1*, *RIM1*, *LSB5* and *YCP4*, were

upregulated at both time points; Ycp4p and Lsb5p are both proteins of unknown function, although Lsb5p is thought to be involved in membrane trafficking. The Prd1p and Rim1p both have some connection to the mitochondria; Prd1p is involved in the degradation of mitochondrial proteins and Rim1p is involved in mitochondrial DNA replication. Glutathione-dependent disulphide oxidoreductase, Grx1p, along with Prd1p, are both known to be upregulated in response to DNA stress.

The M23 constitutive expression mutant resulted in one differentially expressed protein at day four of fermentation; Ypl245wp, the gene product of which is uncharacterized and of unknown function, was upregulated 2.58-fold in the mutant.

The M28 constitutive expression mutant resulted in four differentially expressed proteins at day two of fermentation: Hsp26p, Thi11p, Rps11bp and Rpl36ap. The small heatshock protein Hsp26p, and Thi11p, were both upregulated in the constitutive expression mutant, while Rps11bp and Rpl36ap were both downregulated; Rps11bp and Rpl36ap are protein components of the small (40S) and large (60S) ribosomal subunits, respectively.

Many significantly different proteins were observed in both the null and constitutive expression mutants for the novel ORFs tested in this study, although no mutant resulted in significant changes for both null and constitutive expression mutants. Given the relatively small fold-changes produced by the iTRAQ method, it would be valuable to validate the observed changes by western blotting.

3.4.3 Deletion of putative transcription factors and poorly annotated ORFs produced few changes to the yeast proteome during fermentation

Proteomic analysis was conducted for 17 additional null mutants sampled at day seven during a Chardonnay wine fermentation. Owing to the larger number of null mutants analyzed and the cost involved, only two biological replicates were used and only a single time point was sampled. Due to slow fermentation phenotype, one null mutant (Gep5p/Ylr091wp) was sampled at day ten of the fermentation, when the metabolite profile of the resulting wine was most similar to that of wild type at day seven. Not all proteins were detected in all samples. When a protein was only detected in one replicate the standard

deviation was consequently quite high and this may have resulted in some real differences being reported as non-significant, a difficulty exacerbated by the limited number of biological replicates tested.

Only four of the 17 null mutants resulted in significantly different proteins based on our stringent criteria. Four null mutants (Ybr066cp/Nrg2p, Yfl052wp/Znf1p, Yor032cp/Hms1p and Ygl253wp/Hxk2p) that resulted in no significant differences from wild type are all verified ORFs with some degree of annotation, and had many differentially expressed transcripts determined by previous microarray analysis. This highlights the discrepancy between proteomic and transcriptomic analysis, or could also be a factor of the time point sampled.

The Ybr056wp null mutant resulted in two proteins with significant differential expression compared to wild type, although one was Ybr056wp itself. Given the use of a homozygous null mutant, detection of any Ybr056wp protein was surprising particularly given the lack of overlapping genes. The second differentially expressed protein, Rpl3p, is a protein component of the large (60S) ribosomal subunit and was downregulated 1.60-fold in the null mutant.

The Par32p null mutant resulted in two proteins with significant differential expression compared to wild type; Pdc5p and Thi11p were upregulated 1.92- and 1.77-fold, respectively, in the null mutant. The Thi11p is involved in the synthesis of thiamine, which pathway provides a direction for specific phenotypic tests, while Pdc5p is the minor isoform of pyruvate decarboxylase which is crucial to alcoholic fermentation, and furthermore is known to be repressed by thiamine. Par32p may play some role in mediating the thiamine – pyruvate decarboxylase interaction during alcoholic fermentation.

The Phd1p null mutant resulted in six proteins with significant differential expression; Nqm1p, Ald3p, Tkl2p, Gnd2p and Hbt1p were all upregulated in the null mutant, while Rps5p was downregulated. The four of the five upregulated proteins are all enzymes, although their relation is not abundantly obvious. The Hbt1p is the only non-enzyme; this protein is thought to be important in cell morphogenesis during mating. The Tkl2p and Gnd2p are a transketolase and dehydrogenase, respectively, which are both involved in the pentose phosphate pathway. Nqm1p is a transaldolase of unknown function, although it is induced in response to oxidative stress and has been associated with cell

aging. Ald3p is a cytoplasmic aldehyde dehydrogenase, the expression of which is known to be repressed by glucose and induced by stress. The only downregulated protein, Rps5p, is a component of the small (40S) ribosomal subunit and therefore an essential protein. None of these proteins have been previously reported to have any interaction with Phd1p (SGD).

The Gep5p null mutant resulted in 129 proteins with significant differential expression, 86 of which were upregulated and 43 of which were downregulated. The number of differentially expressed genes allowed them to be entered as a list into the YeastMine tool to analyze for significant gene ontology enrichments. When analyzing the upregulated and downregulated proteins separately, a clear pattern could be observed. The deletion of *GEP5* resulted in the up regulation of enzymes with oxidoreductase activity, important in maintaining the redox balance. These enzymes are largely localized to the mitochondrion. At the same time, constituents of the mitochondrial ribosomes were being downregulated, as well as three subunits (Atp4p, Atp5p, Atp7p) of the mitochondrial ATP synthase required for ATP synthesis (Table 3.19). The protein with the strongest differential expression (5.24-fold increase over wild type), Hsp12p, fell into the oxidative stress response category. While the point of action is not clear, the proteomic results indicated Gep5p deletion acting specifically on the mitochondria and decreasing ribosome generation, translation and ATP synthesis. Enzymes were also activated to improve the redox balance and respond to oxidative stress.

Table 3.19 Gep5p null mutant proteomic results. Gene ontology enrichments found by the YeastMine tool for proteins significantly upregulated or downregulated in the Gep5p deletion mutant at day seven of Chardonnay wine fermentation. Only enrichments with p-value < 0.001 are shown.

	Category	Ontology Name	*p*-value	Number of Genes	Ontology Identifier
Upregulated	Biological process	oxidation-reduction process	4.17445E-06	26	GO:0055114
		small molecule metabolic process	0.000171937	32	GO:0044281
		cellular aldehyde metabolic process	0.000361231	9	GO:0006081
		cellular response to oxidative stress	0.000381084	12	GO:0034599
		protein refolding	0.000449735	6	GO:0042026
		single-organism metabolic process	0.000618879	44	GO:0044710
	Localization	mitochondrial part	9.35E-08	30	GO:0044429
		mitochondrion	1.43E-07	42	GO:0005739
		cytoplasm	1.66425E-05	81	GO:0005737

	Category	Ontology Name	*p*-value	Number of Genes	Ontology Identifier
		mitochondrial envelope	2.35193E-05	22	GO:0005740
		mitochondrial membrane	0.00012178	20	GO:0031966
	Molecular function	aldo-keto reductase (NADP) activity	7.85E-07	6	GO:0004033
		catalytic activity	2.49965E-06	61	GO:0003824
		oxidoreductase activity, acting on CH-OH group of donors	1.10128E-05	12	GO:0016614
		alcohol dehydrogenase (NADP+) activity	4.02841E-05	5	GO:0008106
		oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	5.96993E-05	11	GO:0016616
		oxidoreductase activity	0.000358035	20	GO:0016491
		alditol:NADP+ 1-oxidoreductase activity	0.000655605	4	GO:0004032
Downregulated	Biological process	organonitrogen compound metabolic process	3.83E-10	34	GO:1901564
		organonitrogen compound biosynthetic process	1.50E-09	31	GO:1901566
		organic substance biosynthetic process	2.09E-08	40	GO:1901576
		biosynthetic process	3.53E-08	40	GO:0009058
		cellular biosynthetic process	2.20E-07	39	GO:0044249
		single-organism biosynthetic process	3.92E-07	23	GO:0044711
		small molecule metabolic process	1.04292E-06	24	GO:0044281
		small molecule biosynthetic process	0.000439407	14	GO:0044283
	Localization	ribosome	1.09E-07	15	GO:0005840
		organellar small ribosomal subunit	1.99771E-06	7	GO:0000314
		mitochondrial small ribosomal subunit	1.99771E-06	7	GO:0005763
		ribosomal subunit	4.27668E-06	13	GO:0044391
		mitochondrial proton-transporting ATP synthase, stator stalk	0.000303676	3	GO:0000274
		proton-transporting ATP synthase, stator stalk	0.000303676	3	GO:0045265
	Molecular function	structural constituent of ribosome	0.000128144	1	GO:003735

The proteomic analysis of the putative transcription factor and other ORF null mutants identified four ORFs (YBR056W, *PAR32*, *PHD1*, and *GEP5*) that caused significant differential expression of at least one unique protein compared to wild type at day seven of a wine fermentation. The differentially expressed proteins caused by the deletion of the four ORFs should be validated by Western blotting, or another well-established protein quantification method, and can then be used to strategically design experiments to annotate these genes more specifically.

3.4.4 Adequate sample size is necessary to reduce false positives

Both transcriptomic and proteomic analyses are costly, labour-intensive and time consuming, and because of this it is necessary for researchers to balance the number of biological replicates between what is feasible and what is required for a biologically relevant result.

cDNA microarrays have been available for longer than RNA-seq and iTRAQ technologies, and while the technology has improved somewhat in terms of probe accuracy and detection, the experimental design remains virtually unchanged (Yang and Speed, 2002; Noirel *et al.*, 2011). Work with early microarrays determined that three replicates were sufficient for accurate expression data (Lee *et al.*, 2000), a practice that was imitated in the present study. It should be noted that the study by Lee *et al.* (2000), however, used three technical replicates to measure the reproducibility of results, whereas most modern microarray studies assume low technical variability and use three biological replicates despite biological variation generally being larger than technical variation.

iTRAQ analyses have been noted to inherently under-report the fold-change of proteins, as confirmed by an independent method such as Western blotting. While the direction of the change was generally found to agree, the magnitude was much lower for iTRAQ (Evans *et al.*, 2012). iTRAQ studies have also noted that the application of an arbitrary fold-change cut-off (typically 1.5-fold for iTRAQ) may not be as relevant to biological differences as the application of a statistics-based cut-off. In addition, with earlier iTRAQ methods it was difficult to accurately detect changes in abundance less than two-fold, which compounds the difficulty of detecting subtle changes in the proteome between mutant strains (Evans *et al.*, 2012). The present study noted more significant differences were found for the proteome of mutants tested in triplicate (novel ORFs) versus in duplicate (non-novel ORFs), despite more significant differences reported for the transcriptome of the non-novel ORFs. To date, no comprehensive analysis has been conducted to determine the most appropriate number of biological replicates to use in iTRAQ analyses, as has been conducted for RNA-seq (Conesa *et al.*, 2016), but the number of replicates required will be somewhat dependent on the inherent biological variability of the samples regardless.

3.5 Integration of transcriptomic and proteomic data provides insight into regulation of fermentation-related genes

The parallel transcriptomic and proteomic data sets sampled at three time points (novel ORF mutants) or a single time point (poorly-annotated ORF mutants) during the Chardonnay wine fermentation allowed the determination of transcripts and proteins that have similar or conflicting expression patterns in each analysis. Comparative analysis between the transcriptome and the proteome was performed for both the mutant datasets and the wild type time point analysis. This allows the identification of sites of post-transcriptional and post-translational regulation that may have resulted in disagreements between the data sets.

3.5.1 Five genes shared a regulation pattern between the mRNA transcript and the protein

In the time point analysis of wild type Enoferm M2 during Chardonnay wine fermentation, five genes had similar expression patterns for both RNA transcript and the corresponding protein. These genes were all upregulated at day four of the fermentation and had no significant change at day seven of the fermentation. The magnitude of the fold-change was roughly double for the RNA transcript compared to the encoded protein (Table 3.20). Descriptions of the upregulated genes have been briefly touched on while discussing the transcriptomic results, but will be summarized here as well.

Table 3.20 Genes with significant differences between time points in proteomic and transcriptomic wild type datasets. Corresponding protein and transcript fold-changes, as measured by proteomic and transcriptomic analysis, during the time course of fermentation, using day two as reference for day four and day four as reference for day seven. Significance is based on an ANOVA p-value < 0.05 and linear fold-change > 2 for transcriptomics, and a Mann-Whitney p-value with Benjamini-Hochberg correction < 0.05 and a linear fold-change > 1.5 for proteomics.

Day	ID	Linear Change (Transcript)	Average Linear Change (Protein)
4	*ACS1*/ YAL054C	3.62	1.88
4	*INO1*/ YJL153C	4.66	1.85
4	*TKL2*/ YBR117C	2.54	1.81
4	YGR201C	2.43	1.78
4	*NQM1*/ YGR043C	3.17	1.58

The *TKL2* gene encodes a transketolase required for the synthesis of aromatic amino acids and the non-oxidative pentose phosphate pathway; *TKL2* is complemented in yeast by the gene *TKL1*, which

encodes the major isoform of the gene, and both genes must be deleted to create a mutant with an aromatic amino acid auxotrophy (Schaaff-Gerstenschläger *et al.*, 1993). The transketolases are also important in linking the two main metabolic pathways in yeast, the pentose-phosphate pathway and glycolysis, which could be important as the wine fermentation becomes depleted in oxygen. The *TKL2* gene has not been previously associated with stress response, but its up regulation could be associated with the depletion of nitrogen sources and oxygen in the grape juice at day four of the fermentation.

The *INO1* gene encodes an inositol 1-phosphate synthase involved in the synthesis of inositol containing phospholipids such as phosphatidylinositol, and is positively regulated by *INO2* and *INO4* (Hirsch and Henry, 1986; Graves and Henry, 2000). Phosphatidylinositol is an essential component of the cell membrane and *INO1* has been previously identified as a component of the fermentation stress response (Marks *et al.*, 2008); *INO1* could therefore play a role in cell membrane remodeling as the media conditions change during the wine fermentation.

The *ACS1* gene encodes an acetyl-CoA synthetase known to be expressed under aerobic conditions, so its up regulation towards day four of the fermentation was surprising given the displacement of oxygen by carbon dioxide that occurred as the fermentation progresses (van den Berg *et al.*, 1996). In yeast, however, acetyl-CoA synthetases can act as donors for histone acetylation and therefore influence chromatin regulation and gene expression (Takahashi *et al.*, 2006). *ACS1* expression is also known to be carbon-source dependent, with strong glucose repression and derepression when grown in the presence of ethanol (Kratzer and Schüller, 1995). The second isoform of acetyl-CoA synthetase in yeast, Acs2p, is not subject to glucose repression, and furthermore it appears that Acs2p may be involved in the regulation of Acs1p as the deletion of *ACS2* derepresses the expression of *ACS1* when cells are grown in glucose media (van den Berg *et al.*, 1996; De Jong-Gubbels *et al.*, 1997). The *ACS1* gene has previously been categorized as a fermentation stress response gene (Marks *et al.*, 2008). It appears that despite the oxygen-limited growth conditions of wine fermentation, the "aerobic" *ACS1* is up-regulated at day four of fermentation and could play a role either in histone acetylation or synthesis of lipids for membrane remodelling.

The *NQM1* gene encodes a transaldolase responsible for the conversion of fructose 6-phosphate and erythrose 4-phosphate to sedoheptulose 7-phosphate and glyceraldehyde 3-phosphate in the pentose phosphate pathway. The Nqm1p is the secondary transaldolase in yeast (*TAL1* encodes the major isoform), just as Tkl2p is the secondary transketolase (Huang *et al.*, 2008; Matsushika *et al.*, 2012). *NQM1* has been found to be induced under various stress conditions, such as osmotic and oxidative stress, but has not been implicated in the fermentation stress response (Marks *et al.*, 2008; Michel *et al.*, 2015). While recent work indicates that Nqm1p may not function as a transaldolase in the pentose phosphate pathway *in vivo*, there is evidence that it may confer stress resistance during stationary phase (Michel *et al.*, 2015), which could account for its induction at day four of the wine fermentation.

One significantly upregulated protein, Ygr201cp, is a putative protein of unknown function and is worth pursuing for further annotation. This ORF was identified as a multi-copy repeat in two "wild" wine strains, which could indicate a specific role for this ORF in wine fermentation (Dunn *et al.*, 2012).

While it was promising to see some agreement between the datasets, it was also interesting to investigate the transcripts with major changes (> 10-fold linear change) to see why they are not present in the proteomic data set. A manual search of the 41 genes that had a greater than 10-fold change in their transcript abundance at either day four or day seven against the proteomic results found that most of the encoded proteins were not present in the proteomics data set. Only nine encoded proteins were found in the proteomics data set and the majority of these were not statistically significantly different because they were missing values in at least one sample. This is entirely possible given that the consensus proteomics data set only contained 1858 proteins, whereas the Yeast 2.0 microarray has spots for nearly 6000 different genes.

3.5.2 Thirteen genes were DE at both the transcript and protein level between day two and day seven of fermentation

Differential expression of proteins in Enoferm M2 wild type between day two and day seven of fermentation was considered and compared with the equivalent transcriptomic data set. Twenty-seven

proteins were differentially expressed between day two and day seven of fermentation, but only thirteen had transcripts that demonstrated the same expression pattern (Table 3.21). Of the remaining fourteen genes, twelve were not differentially expressed in the transcriptomic data set, and two (Rtn2p and Ura10p) were regulated in the opposite fashion, albeit with small fold-changes ≤ 2.1.

Table 3.21 Transcriptomic and proteomic DE genes between day two and day seven of fermentation. Transcriptomic (microarray) and proteomic (iTRAQ) differences between Enoferm M2 wild type cells sampled from day two and day seven of Chardonnay wine fermentation were compared and genes/gene products present in both data sets are reported. NA indicates the gene was not differentially expressed in the transcriptome. Fold change (FC) is linear and based on day two as reference.

Protein	Identifier	Protein Mean FC	Transcript FC	Description
ACS1	YAL054C	2.64	5.76	Acetyl-coA synthetase isoform which, along with Acs2p, is the nuclear source of acetyl-coA for histone acetylation; expressed during growth on nonfermentable carbon sources and under aerobic conditions
HSP30	YCR021C	3.60	2.57	Hydrophobic plasma membrane localized, stress-responsive protein that negatively regulates the H(+)-ATPase Pma1p; induced by heat shock, ethanol treatment, weak organic acid, glucose limitation, and entry into stationary phase
RTN2	YDL204W	1.73	-2.07	Protein of unknown function; has similarity to mammalian reticulon proteins; member of the RTNLA (reticulon-like A) subfamily
YGL039W	YGL039W	-1.61	-7.6	Oxidoreductase shown to reduce carbonyl compounds to chiral alcohols
SDT1	YGL224C	-2.23	-2.11	Pyrimidine nucleotidase; overexpression suppresses the 6-AU sensitivity of transcription elongation factor S-II, as well as resistance to other pyrimidine derivatives
NQM1	YGR043C	2.43	3.41	Transaldolase of unknown function; transcription is repressed by Mot1p and induced by alpha-factor and during diauxic shift
YGR201C	YGR201C	2.36	2.2	Putative protein of unknown function
HXT1	YHR094C	-2.17	-6.54	Low-affinity glucose transporter of the major facilitator superfamily, expression is induced by Hxk2p in the presence of glucose and repressed by Rgt1p when glucose is limiting
MPM1	YJL066C	1.62	2.04	Mitochondrial membrane protein of unknown function, contains no hydrophobic stretches
INO1	YJL153C	3.17	5.71	Inositol 1-phosphate synthase, involved in synthesis of inositol phosphates and inositol-containing phospholipids; transcription is coregulated with other phospholipid biosynthetic genes by Ino2p and Ino4p, which bind the UASINO DNA element
YJR096W	YJR096W	1.84	2.36	Putative xylose and arabinose reductase; member of the aldo-keto reductase (AKR) family; GFP-fusion protein is induced in response to the DNA-damaging agent MMS

Protein	Identifier	Protein Mean FC	Transcript FC	Description
ALD3	YMR169C	1.99	2.89	Cytoplasmic aldehyde dehydrogenase, involved in beta-alanine synthesis; uses NAD+ as the preferred coenzyme; very similar to Ald2p; expression is induced by stress and repressed by glucose
URA10	YMR271C	1.88	-2.1	Minor orotate phosphoribosyltransferase (OPRTase) isozyme that catalyzes the fifth enzymatic step in the de novo biosynthesis of pyrimidines, converting orotate into orotidine-5'-phosphate; major OPRTase encoded by URA5
DSE4	YNR067C	-1.68	-3.34	Daughter cell-specific secreted protein with similarity to glucanases, degrades cell wall from the daughter side causing daughter to separate from mother
FAA1	YOR317W	-1.55	-4.11	Long chain fatty acyl-CoA synthetase with a preference for C12:0-C16:0 fatty acids; involved in the activation of imported fatty acids; localized to both lipid particles and mitochondrial outer membrane; essential for stationary phase
HSP26	YBR072W	2.45	NA	Small heat shock protein (sHSP) with chaperone activity; forms hollow, sphere-shaped oligomers that suppress unfolded proteins aggregation; oligomer activation requires a heat-induced conformational change; not expressed in unstressed cells
TKL2	YBR117C	2.17	NA	Transketolase, similar to Tkl1p; catalyzes conversion of xylulose-5-phosphate and ribose-5-phosphate to sedoheptulose-7-phosphate and glyceraldehyde-3-phosphate in the pentose phosphate pathway; needed for synthesis of aromatic amino acids
YDL124W	YDL124W	1.72	NA	NADPH-dependent alpha-keto amide reductase; reduces aromatic alpha-keto amides, aliphatic alpha-keto esters, and aromatic alpha-keto esters; member of the aldo-keto reductase (AKR) family
GLC3	YEL011W	1.83	NA	Glycogen branching enzyme, involved in glycogen accumulation; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm in a punctate pattern
HSP12	YFL014W	1.68	NA	Plasma membrane localized protein that protects membranes from desiccation; induced by heat shock, oxidative stress, osmostress, stationary phase entry, glucose depletion, oleate and alcohol; regulated by the HOG and Ras-Pka pathways
THI4	YGR144W	-3.48	NA	Thiazole synthase, catalyzes formation of a thiazole intermediate during thiamine biosynthesis; required for mitochondrial genome stability in response to DNA damaging agents
GND2	YGR256W	2.18	NA	6-phosphogluconate dehydrogenase (decarboxylating), catalyzes an NADPH regenerating reaction in the pentose phosphate pathway; required for growth on D-glucono-delta-lactone
YLL058W	YLL058W	-2.39	NA	Putative protein of unknown function with similarity to Str2p, which is a cystathionine gamma-synthase important in sulphur metabolism; YLL058W is not an essential gene

Protein	Identifier	Protein Mean FC	Transcript FC	Description
YGP1	YNL160W	2.17	NA	Cell wall-related secretory glycoprotein; induced by nutrient deprivation-associated growth arrest and upon entry into stationary phase; may be involved in adaptation prior to stationary phase entry; has similarity to Sps100p
YNL200C	YNL200C	1.76	NA	Putative protein of unknown function; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
GAS5	YOL030W	-1.89	NA	1,3-beta-glucanosyltransferase, has similarity to Gas1p; localizes to the cell wall
YPR127W	YPR127W	1.72	NA	Protein of unknown function, differentially expressed during alcoholic fermentation; expression activated by transcription factor YRM1/YOR172W; green fluorescent protein (GFP)-fusion protein localizes to both the cytoplasm and the nucleus

The comparison of differentially expressed genes in the transcriptomic and proteomic datasets between day two and day seven identified an additional nine genes (*HSP30*, YGL039W, *SDT1*, *HXT1*, *MPM1*, YJR096W, *ALD3*, *DSE4*, and *FAA1*) that were similarly regulated, that were not identified when day four was also considered as an intermediate step. This is likely due to the systematically underreported fold-change of the iTRAQ method used for the proteomic analyses (Evans *et al.*, 2012). Proteins that change incrementally from day two – day four and again from day four – day seven may have an additive change greater than 1.5, but each increment is below the cut off.

The Hsp30p had the greatest fold-change of any protein between day two and day seven and was one of the few cases where the protein change was greater than the transcript change. This could be due to long-lived transcripts able to be translated many times into the same protein, or very slow turnover rate of Hsp30p. The Hsp30p is induced by many cellular stresses, and is upregulated during the transition to stationary phase (Piper *et al.*, 1994), which would explain the increase in this protein towards day seven of fermentation. The Faa1p is also known to be required for stationary phase (Black and DiRusso, 2007), but both the transcript and protein levels of Faa1p declined towards day seven of fermentation. This could be because the acyl-CoA synthetase Faa1p is required for fatty acid synthesis as the cells approach stationary phase (around day two of wine fermentation), and Faa1p declines thereafter; Hsp30p on the other hand is further upregulated towards day seven because of its role in stress tolerance, and not because of its role in stationary phase.

The Dse4p, Hxt1p, and Ald3p all had expression patterns that were to be expected based on what is known about these proteins. The low-affinity glucose transporter, Hxt1p, is induced by glucose and repressed when glucose is limiting (Luyten, Riou and Blondin, 2002), which was validated by this protein's steady decline towards day seven of wine fermentation when glucose content was < 2% v/v. The cytoplasmic aldehyde dehydrogenase, Ald3p, is known to be stress-induced and glucose repressed (Boubekeur *et al.*, 2001), which is in keeping with this protein's up regulation towards day seven of fermentation. Finally, Dse4p, a secreted protein which degrades the cell wall as daughter cells separate from mother cells, declined from day two – day seven, which could be because cells are no longer actively dividing at this point in the wine fermentation.

Relatively little is known about the function of Mpm1p, Ygl039wp and Yjr096wp. The Mpm1p is only known to be a membrane protein of the mitochondria, although it contains no hydrophobic domain (Inadome *et al.*, 2001); Mpm1p was upregulated towards day seven of wine fermentation. The Ygl039wp demonstrates aldehyde reductase activity *in vitro* (Moon and Liu, 2015), but does not yet have an established purpose *in vivo*; Ygl039wp was downregulated towards day seven of fermentation, the opposite of what was observed for the cytoplasmic aldehyde dehydrogenase *ALD3*. The Ygr096wp is a putative aldehyde reductase which could function in the metabolism of pentose sugars in yeast (Chang *et al.*, 2007); Ygr096wp was upregulated by day seven of fermentation, and could be aiding in the metabolism of alternative carbon sources.

3.5.3 There was no overlap between proteomic and transcriptomic results for novel ORF mutants

Despite robust datasets created using similar cut-off values (1.5- or 2-fold change, ANOVA *p*-value < 0.05) for both the yeast transcriptome and proteome at three time points during the wine fermentation, there was no overlap between significant differentially expressed transcripts and differentially expressed proteins observed in the novel ORF mutants. While the data sets were created using parallel conditions, with the same grape juice media, fermentation temperature and sampling time, the extraction techniques

and quantification were quite different for RNA microarray and iTRAQ proteomics. Possible reasons for these discrepancies will be discussed in detail in section 3.5.5.

3.5.4 *PHD1* and *GEP5* null mutants had some overlap between proteomic and transcriptomic results

While fewer of the poorly annotated null mutants resulted in differentially expressed proteins compared to the novel ORF mutants, there were still two null mutants that resulted in expression changes that were common between the transcriptome and the proteome. It should be noted that for these analyses the transcriptome was observed at day two of the fermentation, while the proteome was observed at day seven of the fermentation. Thus, any overlap between the two datasets may be interpreted as a sustained, rather than transient, response to the elimination of the particular gene.

The *PHD1* null mutant resulted in only one gene that was upregulated in both the proteome and transcriptome, *ALD3*. As previously mentioned, *ALD3* encodes a cytoplasmic aldehyde dehydrogenase, the expression of which is known to be repressed by glucose and induced by stress. This enzyme is involved in polyamine catabolism and beta-alanine synthesis, and most importantly aldehyde dehydrogenases break down toxic acetaldehyde during growth on non-fermentable carbon sources for yeast. This enzyme is non-essential, however, as it is highly similar to the other cytoplasmic aldehyde dehydrogenase Ald2p. Yeast also has two mitochondrial acetaldehyde dehydrogenases encoded by *ALD4* and *ALD5*. No interaction has been previously reported between *PHD1* and *ALD3*, but the deletion of *PHD1* has also never before been tested in the context of wine fermentation.

The *GEP5* null mutant resulted in the largest number of transcripts and proteins with significantly differential expression. To narrow this list, differentially expressed proteins were cross-referenced with differentially expressed transcripts (from day two of fermentation) and twelve genes were identified that overlapped (Figure 3.14). Of these, only three were regulated in the same manner in both datasets: *INO1*, *BNA5* and *CAR2*. This was not entirely surprising however; the microarray data is from earlier in the fermentation than the proteomic data, and so secondary interactions between Gep5p and these genes may

be time-point dependent. The three genes that have similar regulation patterns are the most likely targets for a direct interaction. *INO1* has been previously described. *BNA5* encodes a kynoreninase required for synthesis of tryptophan, and *CAR2* encodes a L-ornithine transaminase involved in arginine degradation, which is upregulated in response to DNA replication stress. Gep5p is thought to potentially play a role in the phospholipid content of the mitochondria, so the strong changes seen in *INO1* could be a mechanism by which Gep5p affects phospholipid synthesis.

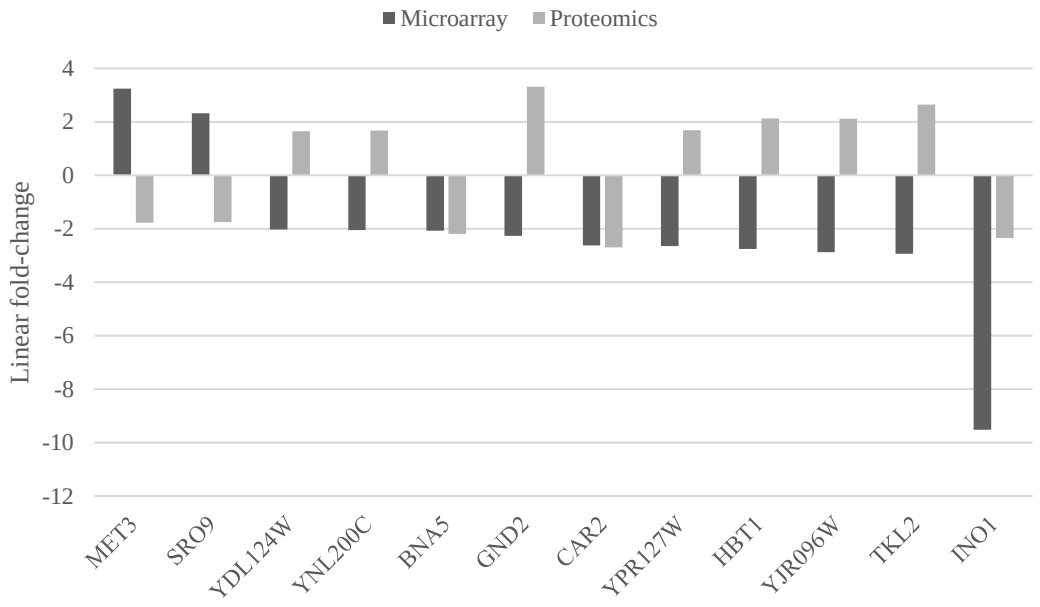


Figure 3.14 Differential expression of transcripts and proteins in the *GEP5* null mutant. The transcriptomic and proteomic datasets for the *GEP5* null mutant were cross-referenced and differentially expressed genes were identified from both datasets. Only twelve genes overlapped, and the expression pattern in the proteomic and transcriptomic datasets was often conflicting.

3.5.5 There is little correlation between the transcriptome and proteome of fermenting wine yeast

Early work correlating the protein and mRNA abundance in yeast under standard laboratory growth conditions found little evidence for transcript levels being a good predictor of protein abundance. Extreme examples from this study found up to 30-fold variation in mRNA levels of different genes resulting in equivalent protein abundances (Gygi *et al.*, 1999). The technologies used to quantify the transcripts and proteins in this study yielded only a small data set of 150 genes, however, and have since been replaced by more accurate and comprehensive methods. More recent studies have focused on the relative levels of transcripts and proteins in both different growth stages or conditions (de Groot *et al.*,

2007; Rossignol *et al.*, 2009), or between different strains of yeast (Rossouw *et al.*, 2010). Despite producing one of the most robust proteomic and transcriptomic data sets available for the difference between aerobic and anaerobic yeast cultures, de Groot *et al.* (2007) choose not to look at direct correlations between the 474 quantified proteins and their respective transcripts. They chose instead to focus on functional categories and found that the correlation between mRNAs and proteins was highly dependent upon the functional category, with ribosome biogenesis showing the best agreement while aminoacyl-tRNA synthetases had no overlap at all. The broad categories of metabolism and energy, however were highly deviant and inconclusive (de Groot *et al.*, 2007). The results were supported by findings using a wine model, which also found weak correlation between mRNA and protein abundance with a strong dependence on functional category (Rossignol *et al.*, 2009). An inter-strain study found that the transcriptome and proteome for a given strain correlated well at a given time point, but that intrastrain comparisons between time points were once again strongly dependent on the functional category of the gene (Rossouw *et al.*, 2010).

While various mechanisms of post-transcriptional and post-translational control have been postulated as being potential causes of the discrepancies between mRNA and protein abundance in yeast cells, much has yet to be discovered about these regulatory processes. The correlation between mRNA and protein abundance appears to be highly dependent on the gene functional category, but also influenced by growth condition, yeast strain and time point. A highly comprehensive study on the deletion of the RNA-binding Puf3p protein including transcriptomics and proteomics as well as polysome profiling and RNA immunoprecipitation and sequencing (RIP-seq) found a role for this protein in the destabilization of mRNA leading to the downregulation of downstream targets. The caveat, however, is that while this protein has been found to bind over a thousand mRNAs, only a handful of these are regulated by Puf3p *in vivo* (Kershaw *et al.*, 2015).

Autophagy is an important process for the catabolism of proteins and recycling of amino acids upon nitrogen starvation (Mizushima, 2005). While this process is active later in fermentation, it is unclear how this may affect the turnover of proteins and the transcriptome-proteome correlation. Furthermore,

yeast cells are generally no longer actively dividing after 48 hours of fermentation (Trabalzini *et al.*, 2003), meaning that our analysis takes place while the yeast are in stationary phase. There has been no comprehensive study on the effect of active cellular division on the translation of transcripts, but this could account for some of the discrepancies between the transcriptome and proteome. Our study adds substantially to the body of knowledge regarding the transcriptome and proteome of actively fermenting yeast and may help inform future studies to pinpoint regulatory mechanisms.

3.6 Mutants with growth phenotypes display altered fermentation performance

The null mutants for *HXK2* and *GEP5*, which have both been characterized as having a slow growth phenotype in rich media, also display a severely altered metabolic profile during wine fermentation. Both null mutants resulted in decreased ethanol production and decreased glucose consumption compared to wild type. Given these metabolic changes, it was highly surprisingly to find that there were no significant proteomic changes between the *HXK2* null mutant and wild type at day seven of fermentation. An additional null mutant for *PHD1* resulted in improved fermentation performance, with greater glucose consumption, ethanol production and glycerol production compared to wild type. The *PHD1* and *GEP5* null mutants, however, resulted in some transcriptomic and proteomic changes that could provide some insight into their observed fermentation phenotype.

3.6.1 *HXK2* deletion shows no compensation by *HXK1*

The Hxk2p has been reported to repress two other proteins of the glycolytic pathway: its paralog Hxk1p and the glucokinase Glk1p (Rodriguez *et al.*, 2001). Notably, however, both the protein and transcripts abundances of *HXK1* and *GLK1* remained completely unchanged in the *HXK2* null mutant. Given the reported regulation of *HXK1* and *GLK1* by *HXK2*, it is expected that these proteins would be differentially regulated in the *HXK2* null mutant. *HXK2* is also known to be an active player in glucose repression; glucose repression is abolished in *HXK2* null mutants (De Winde *et al.*, 1996). The relatively low glucose concentration at day seven of the fermentation could account for the lack of proteomic response in the *HXK2* null mutant, but would still not explain the lack of differential expression of *GLK1*

and *HXK1* at day two of the fermentation. *HXK2* has previously been identified as the key hexokinase enzyme during high-glucose fermentation (Moreno, 2002) and is highly downregulated throughout the wine fermentation (Marks *et al.*, 2008). The current results indicate that *HXK2* may behave differently than expected in the Enoferm M2 genetic background, and warrants follow-up at multiple time points.

3.6.2 *PHD1* may antagonize alcoholic fermentation

The *PHD1* ORF was one of the few genes investigated in this study that produced an altered metabolite profile during wine fermentation, as well as a transcriptomic and proteomic deviation from wild type. Phd1p is considered to be a determinant of pseudohyphal growth, but this gene has been found to have a sustained induction throughout wine fermentation, with a magnitude of the induction too small to be considered a component of the fermentation stress response (Gimeno and Fink, 1994; Marks *et al.*, 2008); *PHD1* also affects cell-cell adhesion and flocculation via *FLO11* regulation, which is an important consideration for wine fermentations (Pan and Heitman, 2000). The *PHD1* null mutant resulted in a slight up regulation of the high-affinity hexose transporter *HXT2* transcript, which could have resulted in slightly faster glucose consumption, as *HXT2* is involved in growth initiation during wine fermentation (Luyten, Riou and Blondin, 2002). The *PHD1* null mutant also increased the transcription of *HSP26*, a known fermentation stress response gene, the converse of which was observed when *PHD1* was constitutively expressed. The constitutive expression of *PHD1* also resulted in the down regulation of ten PAU genes, which are known to be beneficial to wine fermentation. Despite evidence pointing towards *PHD1* being antagonistic to fermentation, when the constitutive expression strain was used to ferment Chardonnay wine it performed slightly better than wild type (data not shown). The expression of neither Hxt2p nor Hsp26p was affected in the *PHD1* null mutant proteome.

The deletion of *PHD1* resulted in the up regulation of *NQM1*, *TKL2*, and *GND2* which encode a transaldolase, transketolase, and dehydrogenase, respectively, of the pentose phosphate pathway. Both *NQM1* and *TKL2* were identified in this study as significantly upregulated at day four of fermentation in both the transcriptome and proteome. While Nqm1p and Tkl2p are considered the minor isoforms of

their respective enzymes, these enzymes may play a larger role in the pentose phosphate pathway during fermentation, and their upregulation could explain improved fermentation performance. The *ALD3* gene was the only gene consistently upregulated in both the proteome and transcriptome of the *PHD1* null mutant; Ald3p is not a component of the fermentation stress response and has been previously noted to have unchanged expression under ethanol stress (Marks *et al.*, 2008; Ma and Liu, 2010). The volatiles in the headspace grape must fermented with the *PHD1* mutant were not tested in this study, but would be worthwhile to analyze as changes in *ALD3* expression may have effects on volatile secondary metabolite production.

3.6.3 *GEP5* deletion produces multiple cellular responses

In addition to a slow growth phenotype, *GEP5* deletion resulted in responses at the metabolic, transcriptomic and proteomic levels. The metabolic changes were large enough that in order to provide a better basis for comparison, the point of fermentation (based on ethanol production) was taken rather than day of fermentation when comparing the proteome between the *GEP5* null mutant and wild type.

The large number of differentially expressed proteins in the *GEP5* null mutant localized to the mitochondrion and involved in redox balance supports the hypothesis that Gep5p is important for mitochondrial maintenance. The combined proteomic and transcriptomic results from the *GEP5* null mutant indicate that its deletion may be inhibiting the mutant's ability to process inositol through Ino1p. No interaction, either genetic or physical, has between reported for these two genes. The primary regulation of *INO1* expression is controlled by the binding of Ino2p and Ino4p to the promoter region of *INO1* (Ambroziak and Henry, 1994). Furthermore, Ino1p has only been known to localize to the cytoplasm in standard growth conditions, whereas Gep5p has been shown to localize to the mitochondria (Huh *et al.*, 2003). Gep5p has been implicated in the maintenance of phosphatidylethanolamine in the mitochondrial membrane (Tamura *et al.*, 2012), and there is a possibility that its actions may extend to other phospholipids.

3.7 Few direct links between altered metabolite profile and transcriptomic/proteomic changes

While altered metabolite profiles and transcriptomic/proteomic changes were observed for some of the null and constitutive expression mutants, it was difficult to ascertain a direct connection between the metabolic change and the change in gene expression.

The M21 constitutive expression mutant resulted in slight increases in ethanol production and acetic acid production, and the M18 constitutive expression mutant resulted in a slight increase in glucose and fructose consumption during fermentation. Both M21 and M18 constitutive expression mutants did not, however, result in any differentially expressed transcripts or proteins at the same time points tested. This could be because the benefits of constitutive expression of these two genes cause expression changes that occurred prior to day two of fermentation, or that the magnitude of the changes is sufficient for small metabolic changes, but does not meet the significance cut-offs for transcriptomics and proteomics applied in this study.

The M15 null mutant resulted in two related fatty acid ethyl esters, ethyl octanoate and ethyl decanoate, being increased compared to wild type in the final wine produced. While the M15 null mutant did not result in any proteomic changes, it did cause transcriptomic changes at day two and day four of the fermentation; *ADH4* was upregulated in the M15 null mutant at day four of fermentation, but this gene is involved in ethanol metabolism (Hazelwood *et al.*, 2008) and has not been linked to ethyl ester production. The two other differentially expressed genes, *TOS6* at day two and YLR154W-E at day four, have no known function and could play a role in ethyl ester synthesis.

The M28 null mutant resulted in significant increases in two unrelated volatile compounds (2,3-butanediol and *p*-ethylacetophenone) compared to wild type, while the M28 constitutive expression mutant resulted in a slight increase in acetic acid production compared to wild type. The M28 null mutant also produced the largest number of transcriptomic changes compared to wild type at all three time points sampled, although it is not immediately clear how these changes may have affected the volatile metabolite production. A number of PAU genes and other genes of unknown function were differentially expressed at days two and four of fermentation, which could be affecting secondary metabolites by a hitherto

unknown mechanism. The M28 constitutive expression mutant resulted in the upregulation of Hsp26p, a heat shock protein well documented to be highly upregulated in stressed yeast cells in stationary phase (Trabalzini *et al.*, 2003; Varela *et al.*, 2005; Marks *et al.*, 2008; Deed, Deed and Gardner, 2015), at day two of the wine fermentation. If Hsp26p is upregulated in response to oxidative stress from a redox imbalance produced by the constitutive expression of M28, this could be a possible mechanism by which M28 constitutive expression increases acetic acid production, as the production of acetic acid can be a redox correction mechanism in yeast (Albers *et al.*, 1996).

The M23 null mutant resulted in significant increases in four related volatile compounds (ethyl acetate, ethyl butyrate, isoamyl acetate, isobutyl acetate) compared to wild type, while the M23 constitutive expression mutant resulted in a slight increase in fructose consumption compared to wild type. The four related volatile compounds are all esters of alcohols and organic acids, and are important aroma compounds in wine. While the M23 null mutant did not have a direct affect on genes involved in ester synthesis, at day two it altered the expression of two genes related to nitrogen usage, the amino acid permease *BAP3* and the proline oxidase *PUT1*, and nitrogen availability can influence the production of esters (Hirst and Richter, 2016). In the M23 null mutant *MIG1* and *MIG2*, two glucose mediated transcriptional repressors, were downregulated at day four of the fermentation; these transcriptional repressors in turn affect the expression of numerous genes under glucose repression (Lutfiyya *et al.*, 1998). It is possible that small changes in expression could have resulted in the altered production of the four volatile esters in the M23 null mutant. The M23 constitutive expression mutant resulted in one downregulated protein at day four of the fermentation, YPL245W, which is uncharacterized and of unknown function. It is possible that M23 is a hexose transporter, and its constitutive expression resulted in the slight increase in fructose consumption observed in the M23 constitutive expression mutant. It is unclear whether the increased fructose consumption or the constitutive expression of M23 was directly responsible for the down regulation of YPL245W protein at day four of the wine fermentation.

Of the allantoin regulating genes, YLL054C, YKL222C, *PHD1* and *HMS1*, only the deletion of YLL054C and *PHD1* resulted in detectable metabolic changes during the wine fermentation. The

deletion of YLL054C resulted in a slight increase in glucose consumption, while the deletion of *PHD1* resulted in increases in glucose and fructose consumption, as well as increases in ethanol and glycerol production. The proteomic analysis of the *PHD1* null mutant found the upregulation of two enzymes, Tkl2p and Gnd2p, of the pentose phosphate pathway, which could explain the improved metabolic rate of the *PHD1* null mutant. Alternatively, the *PHD1* null mutant also resulted in the upregulation of *HXT2, HXT6* and *HXT7* which could be driving the improved sugar metabolism. The YLL054C null mutant could be increasing glucose consumption by upregulating the PAU genes and improving overall stress tolerance of the cell to winemaking conditions.

The *GEP5* null mutant resulted in significant changes in glucose and fructose consumption (decreased), as well as ethanol (decreased), glycerol (decreased) and acetic acid (increased) production. While the slow sugar consumption and ethanol production are symptomatic of a slow growth phenotype, the changes in glycerol and acetic acid production could be directly related to a redox imbalance in the mitochondria. The proteomic analysis of the *GEP5* null mutant identified the upregulation of many proteins involved in the oxidation-reduction process, with specific localization to the mitochondria. It is also possible that the upregulation of the sulphur assimilation pathway in the *GEP5* null mutant is creating an abundance of toxic intermediates such as sulphite (Schimz, 1980; Donalies and Stahl, 2002), and making the cell sick.

4 SUMMARY

The release of the complete sequence of the *S. cerevisiae* genome in 1996 marked a major leap forward for the understanding of genes in a whole genome context. The collaborative nature of the genome sequencing project required the use of a standard yeast strain by all collaborators, and the S288C laboratory strain was a strong candidate based on its extensive use in prior genetic mapping (Mortimer and Schild, 1980; Mortimer, Contopoulou and King, 1992; Goffeau *et al.*, 1996). The complete genome sequence of S288C rapidly accelerated gene annotation efforts in several ways. Firstly, the genome sequence aided the creation of proteomic and transcriptomic assays whereby the cellular expression could be monitored at multiple levels in response to various cellular stresses and growth conditions (Velculescu *et al.*, 1997; Gygi *et al.*, 1999; Marks *et al.*, 2008; Rossignol *et al.*, 2009; Rossouw *et al.*, 2010; Deed, Deed and Gardner, 2015). Secondly, the genome sequence enabled the creation of systematic deletion and overexpression mutants of every known ORF, which in turn could be used to determine phenotypes associated with particular ORFs (Giaever *et al.*, 2002; Hillenmeyer *et al.*, 2008; Ho *et al.*, 2009; Costanzo *et al.*, 2016). Systematic gene annotation efforts have identified a function for ca. 5900 ORFs, but 791 remain uncharacterized. Finally, the S288C genome provided a base upon which the genomes of other *S. cerevisiae* strains could be scaffolded and aided the sequencing of hundreds of new strains for intraspecific comparisons of indels, SNPs and CNVs (Dunn *et al.*, 2012; Borneman, Pretorius and Chambers, 2013; Borneman *et al.*, 2016; Gallone *et al.*, 2016). The sequencing of *S. cerevisiae* strains isolated from various environments has also identified numerous ORFs that do not map to the S288C reference genome, which could be responsible for conferring traits beneficial to particular industries (Du and Takagi, 2007; Borneman *et al.*, 2011).

Previous work by the van Vuuren laboratory re-sequenced and assembled the genome of the commercial wine strain Enoferm M2 (unpublished data), and identified twenty-six ORFs that are not present in the S288C reference genome, many of which have high sequence identity to ORFs found in other commercial strains of *S. cerevisiae* (Iwashita, 2016). To improve the annotation of novel ORFs

from a commercial wine strain (Enoferm M2) of *S. cerevisiae*, in the present study systematic deletion of fifteen novel ORFs was performed, followed by fermentation of grape must into wine using the mutant strains. Concurrently, a number of poorly annotated ORFs were assessed in the Enoferm M2 genetic background using the same methodologies. Null mutants of both novel and poorly annotated ORFs were analyzed by a systems biology approach, monitoring the metabolome, transcriptome and proteome to assess any differences between the mutant and wild type Enoferm M2.

The present study found that fourteen of the fifteen novel ORFs were expressed under enological conditions, but that many of the novel ORFs were not expressed at detectable levels when yeast were grown in YPD at optimal growth conditions. The deletion or constitutive expression of the novel ORFs did not result in any observable growth phenotype and only some novel ORFs resulted in minor changes in the primary and secondary metabolite profiles of the wines during and post- fermentation. The putative transcription factors and poorly annotated ORFs, analyzed in parallel with the novel ORFs resulted in more dramatic changes in the metabolite profiles of the wine during fermentation. Only a limited number of metabolites were measured, however, and as previous high-throughput efforts to annotate deletion strains used up to 1144 assays to attempt to identify a phenotype for every ORF of the S288C reference genome (Hillenmeyer *et al.*, 2008), there may yet be other assays that will reveal further phenotypes for the novel ORFs studied here. There are a number of other technical parameters relevant to wine yeast that were not specifically investigated in this study, such as desiccation tolerance, post-fermentation viability, sporulation, and growth on molasses, all of which are relevant to the growth, storage and distribution of commercial yeast. Furthermore, as Enoferm M2 and the closely related AWRI796 are both South African vineyard isolates (Borneman *et al.*, 2016), there may be conditions and life history directly related to this geographic location that are responsible for the genotypes of these yeast.

Time-course transcriptomic analysis of Enoferm M2 wild type during wine fermentation supported previously observed changes in gene expression (Marks *et al.*, 2008; Rossouw and Bauer, 2009). The transcriptomic analysis found that the majority (65%) of genes are not differentially expressed from day two - day seven of fermentation, and that these “maintenance” genes are involved in primary metabolic

processes and cellular organization. Throughout the fermentation, however, there was a consistent up regulation of genes involved in oxidative stress (*FRM2*, *XBP1*, *SRX1*, *ECM5*, and *MEK1*) and consistent down regulation of genes involved in ribosomal activity (*RPS9B*, *RPS29B*, *RPL13A*, *NHP2*, *RPL12A*, *TMA20*, *RPL30*, *RPS20*, *RPL8A*, *ARD1*, *RPL39*, *ANB1*, *RPL8B*, *RPS0B*, *RPS29A*, *RPL18A*, *RPS7A*, *NAT5*, and *RPL11A*) and tRNA aminoacylation for protein synthesis (*SES1*, *ARC1*, *VAS1*, *DED81*, YHR020W, *THS1*, *DPS1*, YNL247W, and *HTS1*).

The novel ORF null mutants caused several differentially expressed genes at the different time points sampled, and a number of the differentially expressed genes caused by the novel ORF mutants during fermentation had no known function and are strong candidates for follow-up as they may play specific roles in fermentation-related processes. The changes were dependent upon the time point of the sampling, which could be related to the differential expression of the novel ORFs in wild type at different time points of the fermentation. While there is no exact parallel method to the one used in this study available for comparison, transcriptomic studies using multiple strains and sampling multiple time points during a fermentation have noted the strain-specific responses are also time-point dependent (Rossouw and Bauer, 2009; Rossouw *et al.*, 2009). In addition to the time point of sampling, the genetic background (in this case Enoferm M2) may also affect the yeast's response to the mutation of the novel ORFs. Due to the novel nature of these ORFs, deletion is not possible in most strains (except perhaps AWRI796), but the constitutive expression of these novel ORFs in other *S. cerevisiae* genetic backgrounds may aid in the identification of further responses caused by the novel ORFs.

Many of the putative transcription factors and poorly annotated ORFs resulted in extensive transcriptomic remodelling observed at the single time point sampled for these mutants. Of note, genes encoding four putative transcription factors (YLL054C, YKL222C, *PHD1* and *HMS1*) appear to positively regulate the final steps of allantoin metabolism, as the allantoin pathway genes *DAL2*, *DAL3* and *DAL7* were downregulated 2.49-4.30 fold when these ORFs were deleted. Furthermore, the deletion of YLL054C and YKL222C coupled the downregulation of allantoin metabolism with the upregulation of

PAU genes, while the deletion of *PHD1* and *HMS1* resulted in the downregulation of allantoin metabolism along with changes in a number of membrane transport proteins (*HXT2*, *HXT6*/*HXT7*, *OPT2*, *UGA4*, *BIO5*, *PHO84*). The function of *HMS1* in the regulation of transmembrane transport was supported by the *HMS1* constitutive expression mutant resulting in large fold changes in the transcripts of metal and hexose transporters *HXT4* (11.24-fold), *CTR3* (-11.14-fold), *FET3* (-16.72-fold) and *PHO84* (-21.66-fold). The poorly annotated ORF *GEP5* produced significant transcriptomic changes when deleted during wine fermentation, and notably caused a significant upregulation of five sequential MET genes in the sulphur assimilation pathway, *MET3* (3.24-fold), *MET14* (3.26-fold), *MET16* (3.40-fold), *MET5* (2.31-fold), and *MET10* (2.13-fold), which has not been previously reported. In addition, the *GEP5* null mutant also resulted in a decrease in sporulation related genes (*OSW2*, 13.8-fold; *SPS100*, 9.76-fold; *SMA1*, 6.91-fold; *SPO19*, 4.46-fold; *OSW2*, 4.34-fold; *GAS4*, 3.33-fold; *LOH1*, 3.23-fold; *CRR1*, 3.10-fold; *IRC18*, 3.07-fold; *SPS22*, 3.05-fold; *FUS2*, 2.97-fold; *SPO77*, 2.62-fold; *DTR1*, 2.44-fold; *RRT8*, 2.28-fold; *DON1*, 2.24-fold; *AMA1*, 2.20-fold) and an increase in ribosome biogenesis genes (41 genes, $p = 0.005410$), which could be related to the slow-growth phenotype caused by the *GEP5* deletion. The significant results found in this study identified deletions of some ORFs that result in previously unreported effects and interactions.

The proteomic analysis of Enoferm M2 wild type over the course of fermentation only identified nine proteins that were differentially expressed between the three time points sampled, although 27 proteins were differentially expressed when only the larger time gap between day two and day seven was considered. While the transcriptomic analyses found many differentially expressed genes as a result of the null mutants, the proteomic analyses found only a few significant results. The *PHD1* null mutant affected the expression of the transketolase Tkl2p (1.60-fold) and transaldolase Gnd2p (1.67-fold) of the pentose phosphate pathway. In addition, the *GEP5* null mutant resulted in the upregulation of many mitochondrial enzymes involved in oxidation-reduction processes.

Many of the transcripts with large fold-changes between time points were not reflected in a differential protein expression, a phenomenon that has been observed in other studies (Gygi *et al.*, 1999; de Groot *et al.*, 2007). Furthermore, the proteomic results for the mutant strains rarely correlated with the transcriptomic results. The poor correlation could reflect differences in regulatory mechanisms governing the stability of mRNA transcripts and proteins, as well as differences in the sensitivity and accuracy of the microarray and iTRAQ methods employed. While disappointing, the lack of agreement between the proteomic and transcriptomic results corroborates previously published work (Gygi *et al.*, 1999; Rossignol *et al.*, 2009), and furthermore contributes to our understanding of the relation between transcription and translation in a wine fermentation setting.

5 CONCLUSIONS

Despite over a hundred non-reference ORFs being found in a wide variety of yeast strains, very few of these ORFs have been experimentally verified as being transcribed *in vivo*. Studies that have identified non-reference genetic material through comparative genomics have done little to advance the annotation of ORFs found in these regions (Borneman, Pretorius and Chambers, 2013). Some studies have been able to systematically identify a function for novel ORFs; the cross of S288C and Σ1278b strains allowed the identification of the MPR genes, which are responsible for conferring ethanol resistance in baker's yeast (Takagi *et al.*, 2000; Du and Takagi, 2007). While the 10th anniversary update of the reference genome in 2006 focused solely on the improvements upon the original annotation (Fisk *et al.*, 2006), the 2014 update of *S. cerevisiae* reference genome identified the need to incorporate non-S288C genomic data into a *S. cerevisiae* pan-genome (Engel *et al.*, 2014).

The current study confirmed the expression of fourteen novel ORFs present in the genome sequence of Enoferm M2, and made use of two well-established 'omics technologies, RNA microarrays and iTRAQ, in conjunction with the systematic deletion or constitutive expression of novel ORFs to attempt to annotate these novel ORFs in wine fermentation context. While the current approach was able to identify differentially expressed transcripts and proteins in the mutant strains as the strains completed a wine fermentation, there was no agreement between the transcriptomic and proteomic data. Furthermore, it was difficult to relate the observed changes in expression back to a quantifiable change in the yeast metabolic processes or stress responses during the fermentation. We were, however, able to identify some candidate genes that may be interacting with the novel ORFs, and these interactions could be validated with an independent method such as qRT-PCR for transcripts or western blotting for proteins. Some of the novel ORFs, such as the putative hexose transporter M23, could also have their function validated with function specific assays, such as a hexose transport assay in the case of M23.

The parallel analysis of non-novel ORFs also identified several genetic interactions that have not been previously reported and make strong candidates for follow-up studies. The deletion of *GEP5*,

previously implicated in the maintenance of the mitochondrial lipid bilayer and mitochondrial genome, was found to upregulate the sulphur assimilation pathway during wine fermentation. Given the importance of sulphur metabolism during winemaking, both for tolerance to sulphites used as antimicrobials and for the production of sulphur based off-flavours such as hydrogen sulphide, the role of *GEP5* in sulphur metabolism could have serious implications for wine yeast. This study also identified four genes (YLL054C, YKL222C, *PHD1* and *HMS1*) involved as positive regulators of the allantoin metabolic pathway, in addition to having further roles regulating cell wall remodeling and transmembrane transport. The deletion of *PHD1* in the Enoferm M2 background caused a significant improvement in the fermentation performance of the mutant strain, and further study is needed to identify the precise mechanism(s) involved.

Should a similar project be conducted in the future, RNA-seq technology would be better suited to the transcriptomic analyses, as it would be possible to the assess dynamic changes for the transcripts of the novel ORFs present in a given strain without having to use an independent method such as qRT-PCR (Ibáñez *et al.*, 2017). A label-free proteomic method may be better suited to identifying small fold-change differences that may be minimized by iTRAQ analyses (Evans *et al.*, 2012). The creation of custom peptide database including the novel ORFs of interest would also allow the quantification of dynamic changes in the expression of the proteins encoded by novel ORFs along with the rest of the proteome.

The reasonably high cost of microarray and iTRAQ analyses, as well as the number of biological replicates and sampling points required for meaningful analyses, may make the current methodology unfeasible to apply to the remainder of the novel ORFs identified in other commercial wine strains. The significant results found for the putative transcription factors YLR054C, YKL222C, *PHD1* and *HMS1* and the poorly annotated ORF *GEP5*, however, highlight the power of using a systems biology approach in conjunction with relevant industrial stress conditions such as wine fermentation to discover new interactions for yeast genes. Given that many yeast genes are multifunctional (Costanzo *et al.*, 2016), there may yet be undiscovered functions for genes that have already been studied for years.

LITERATURE CITED

Albers, E. *et al.* (1996) 'Influence of the nitrogen source on Saccharomyces cerevisiae anaerobic growth and product formation.', *Applied and Environmental Microbiology*, 62(9), pp. 3187–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8795209.

Albertyn, J. *et al.* (1994) 'GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in Saccharomyces cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway.', *Molecular and Cellular Biology*, 14(6), pp. 4135–4144. doi: 10.1128/MCB.14.6.4135.

Alexandre, H. *et al.* (2001) 'Global gene expression during short-term ethanol stress in Saccharomyces cerevisiae', *FEBS Letters*, 498(1), pp. 98–103. doi: 10.1016/S0014-5793(01)02503-0.

Ambroziak, J. and Henry, S. A. (1994) 'INO2 and INO4 gene products, positive regulators of phospholipid biosynthesis in Saccharomyces cerevisiae, form a complex that binds to the INO1 promoter', *Journal of Biological Chemistry*, 269(21), pp. 15344–15349.

Anderson, M. J. *et al.* (2012) 'Identification of RCN1 and RSA3 as ethanol-tolerant genes in Saccharomyces cerevisiae using a high copy barcoded library', *FEMS Yeast Research*, 12(1), pp. 48–60. doi: 10.1111/j.1567-1364.2011.00762.x.

Argueso, J. J. L. *et al.* (2009) 'Genome structure of a Saccharomyces cerevisiae strain widely used in bioethanol production', *Genome Research*, 19(Gura), pp. 2258–2270. doi: 10.1101/gr.091777.109.

Bąkowska-Żywicka, K., Kasprzyk, M. and Twardowski, T. (2016) 'tRNA-derived short RNAs bind to Saccharomyces cerevisiae ribosomes in a stress-dependent manner and inhibit protein synthesis in vitro', *FEMS Yeast Research*. Edited by C. Mazzoni, 16(6), p. fow077. doi: 10.1093/femsyr/fow077.

Barnett, J. A., Payne, R. W. and Yarrow, D. (1990) 'Yeasts: characteristics and identification'. Cambridge;New York; Cambridge University Press.

Barros de Souza, R. *et al.* (2016) 'Magnesium ions in yeast: setting free the metabolism from glucose catabolite repression', *Metallomics*, 8(11), pp. 1193–1203. doi: 10.1039/C6MT00157B.

Beltran, G. *et al.* (2008) 'Effect of fermentation temperature and culture media on the yeast lipid composition and wine volatile compounds', *International Journal of Food Microbiology*, 121(2), pp. 169–177. doi: 10.1016/j.ijfoodmicro.2007.11.030.

Bely, M., Sablayrolles, J. M. and Barre, P. (1990) 'Description of alcoholic fermentation kinetics: its variability and significance', *American Journal of Enology and Viticulture*, 41(4), pp. 319–324.

van den Berg, M. a *et al.* (1996) 'The Two Acetyl-coenzyme A Synthetases of Saccharomyces cerevisiae Differ with Respect to Kinetic Properties and Transcriptional Regulation', *Journal of Biological*

Chemistry, 271(46), pp. 28953–28959. doi: 10.1074/jbc.271.46.28953.

Bessonov, K. *et al.* (2013) 'Functional Analyses of NSF1 in Wine Yeast Using Interconnected Correlation Clustering and Molecular Analyses', *PLoS ONE*, 8(10), pp. 1–15. doi: 10.1371/journal.pone.0077192.

Bisson, L. F. (2012) 'Geographic origin and diversity of wine strains of saccharomyces', *American Journal of Enology and Viticulture*, 63(2), pp. 165–176. doi: 10.5344/ajev.2012.11083.

Biswas, C. *et al.* (2013) 'Functional characterization of the hexose transporter Hxt13p: An efflux pump that mediates resistance to miltefosine in yeast', *Fungal Genetics and Biology*. Elsevier Inc., 61, pp. 23–32. doi: 10.1016/j.fgb.2013.09.005.

Black, P. N. and DiRusso, C. C. (2007) 'Yeast acyl-CoA synthetases at the crossroads of fatty acid metabolism and regulation', *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1771(3), pp. 286–298. doi: 10.1016/j.bbalip.2006.05.003.

Boratyn, G. M. *et al.* (2013) 'BLAST: a more efficient report with usability improvements', *Nucleic Acids Research*, 41(W1), pp. W29–W33. doi: 10.1093/nar/gkt282.

Borneman, A. R. *et al.* (2008) 'Comparative genome analysis of a Saccharomyces cerevisiae wine strain', *FEMS Yeast Research*, 8(7), pp. 1185–1195. doi: 10.1111/j.1567-1364.2008.00434.x.

Borneman, A. R. *et al.* (2011) 'Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of Saccharomyces cerevisiae', *PLoS Genetics*. Edited by G. Sherlock, 7(2), p. e1001287. doi: 10.1371/journal.pgen.1001287.

Borneman, A. R. *et al.* (2016) 'Whole Genome Comparison Reveals High Levels of Inbreeding and Strain Redundancy Across the Spectrum of Commercial Wine Strains of Saccharomyces cerevisiae.', *G3; Genes|Genomes|Genetics*, 6(4), pp. 957–71. doi: 10.1534/g3.115.025692.

Borneman, A. R. and Pretorius, I. S. (2015) 'Genomic insights into the Saccharomyces sensu stricto complex.', *Genetics*, 199(2), pp. 281–91. doi: 10.1534/genetics.114.173633.

Borneman, A. R., Pretorius, I. S. and Chambers, P. J. (2013) 'Comparative genomics: A revolutionary tool for wine yeast strain development', *Current Opinion in Biotechnology*. Elsevier Ltd, 24(2), pp. 192–199. doi: 10.1016/j.copbio.2012.08.006.

Boubekeur, S. *et al.* (2001) 'Participation of acetaldehyde dehydrogenases in ethanol and pyruvate metabolism of the yeast Saccharomyces cerevisiae', *European Journal of Biochemistry*, 268(19), pp. 5057–5065. doi: 10.1046/j.1432-1327.2001.02418.x.

Bradbury, J. E. *et al.* (2006) 'A homozygous diploid subset of commercial wine yeast strains', *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 89(1), pp. 27–38. doi: 10.1007/s10482-005-9006-1.

Brar, G. a. *et al.* (2012) 'High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling', *Science*, 335(6068), pp. 552–557. doi: 10.1126/science.1215110.

Brem, R. B. (2002) 'Genetic Dissection of Transcriptional Regulation in Budding Yeast', *Science*, 296(5568), pp. 752–755. doi: 10.1126/science.1069516.

Brown, C. A., Murray, A. W. and Verstrepen, K. J. (2010) 'Rapid Expansion and Functional Divergence of Subtelomeric Gene Families in Yeasts', *Current Biology*. Elsevier Ltd, 20(10), pp. 895–903. doi: 10.1016/j.cub.2010.04.027.

Çakar, Z. P. *et al.* (2005) 'Evolutionary engineering of multiple-stress resistant Saccharomyces cerevisiae', *FEMS Yeast Research*, 5(6–7), pp. 569–578. doi: 10.1016/j.femsyr.2004.10.010.

Chang, Q. *et al.* (2007) 'Functional studies of aldo-keto reductases in Saccharomyces cerevisiae', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1773(3), pp. 321–329. doi: 10.1016/j.bbamcr.2006.10.009.

Conesa, A. *et al.* (2016) 'A survey of best practices for RNA-seq data analysis', *Genome Biology*, 17(1), p. 13. doi: 10.1186/s13059-016-0881-8.

Costanzo, M. *et al.* (2016) 'A global genetic interaction network maps a wiring diagram of cellular function', *Science*, 353(6306), p. aaf1420-aaf1420. doi: 10.1126/science.aaf1420.

Cromie, G. A. *et al.* (2013) 'Genomic Sequence Diversity and Population Structure of Saccharomyces cerevisiae Assessed by RAD-seq', *G3; Genes|Genomes|Genetics*, 3(12), pp. 2163–2171. doi: 10.1534/g3.113.007492.

Dahabieh, M. S., Husnik, J. I. and van Vuuren, H. J. J. (2009) 'Functional expression of the DUR3 gene in a wine yeast strain to minimize ethyl carbamate in chardonnay wine', *American Journal of Enology and Viticulture*, 60(4), pp. 537–541.

Dahabieh, M. S., Husnik, J. I. and Van Vuuren, H. J. J. (2010) 'Functional enhancement of Sake yeast strains to minimize the production of ethyl carbamate in Sake wine', *Journal of Applied Microbiology*, 109(3), pp. 963–973. doi: 10.1111/j.1365-2672.2010.04723.x.

David, F. and Siewers, V. (2015) 'Advances in yeast genome engineering', *FEMS Yeast Research*, 15(1), pp. 1–14. doi: 10.1111/1567-1364.12200.

Deed, N. K., Van Vuuren, H. J. J. and Gardner, R. C. (2011) 'Effects of nitrogen catabolite repression and di-ammonium phosphate addition during wine fermentation by a commercial strain of S. cerevisiae', *Applied Microbiology and Biotechnology*, 89(5), pp. 1537–1549. doi: 10.1007/s00253-011-3084-y.

Deed, R. C., Deed, N. K. and Gardner, R. C. (2015) 'Transcriptional response of Saccharomyces cerevisiae to low temperature during wine fermentation.', *Antonie van Leeuwenhoek*, 107(4), pp. 1029–48. doi: 10.1007/s10482-015-0395-5.

DePristo, M. A. *et al.* (2011) ‘A framework for variation discovery and genotyping using next-generation DNA sequencing data’, *Nature Genetics*, 43(5), pp. 491–498. doi: 10.1038/ng.806.

DeRisi, J. L., Iyer, V. R. and Brown, P. O. (1997) ‘Exploring the metabolic and genetic control of gene expression on a genomic scale.’, *Science (New York, N.Y.)*, 278(5338), pp. 680–6. doi: 10.1126/science.278.5338.680.

Ding, M. Z. *et al.* (2009) ‘Comparative metabolomic analysis on industrial continuous and batch ethanol fermentation processes by GC-TOF-MS’, *Metabolomics*, 5(2), pp. 229–238. doi: 10.1007/s11306-008-0145-z.

Donalies, U. E. B. and Stahl, U. (2002) ‘Increasing sulphite formation inSaccharomyces cerevisiae by overexpression ofMET14 andSSU1’, *Yeast*, 19(6), pp. 475–484. doi: 10.1002/yea.849.

Du, X. and Takagi, H. (2007) ‘N-Acetyltransferase Mpr1 confers ethanol tolerance on Saccharomyces cerevisiae by reducing reactive oxygen species’, *Applied Microbiology and Biotechnology*, 75(6), pp. 1343–1351. doi: 10.1007/s00253-007-0940-x.

Dulermo, R. *et al.* (2016) ‘Truncation of Gal4p explains the inactivation of the GAL/MEL regulon in both Saccharomyces bayanus and some Saccharomyces cerevisiae wine strains’, *FEMS Yeast Research*. Edited by P. Daran-Lapujade, 16(6), p. fow070. doi: 10.1093/femsyr/fow070.

Dunn, B. *et al.* (2012) ‘Analysis of the Saccharomyces cerevisiae pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments’, *Genome Research*, 22(5), pp. 908–924. doi: 10.1101/gr.130310.111.

Dunn, B., Levine, R. P. and Sherlock, G. (2005) ‘Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures’, *BMC Genomics*, 6(1), p. 53. doi: 10.1186/1471-2164-6-53.

Engel, S. R. *et al.* (2014) ‘The reference genome sequence of Saccharomyces cerevisiae: then and now.’, *G3; Genes|Genomes|Genetics*, 4(3), pp. 389–98. doi: 10.1534/g3.113.008995.

Evans, C. *et al.* (2012) ‘An insight into iTRAQ: Where do we stand now?’, *Analytical and Bioanalytical Chemistry*, 404(4), pp. 1011–1027. doi: 10.1007/s00216-012-5918-6.

Fernández-González, M., Úbeda, J. F. and Briones, A. I. (2015) ‘Study of Saccharomyces cerevisiae Wine Strains for Breeding Through Fermentation Efficiency and Tetrad Analysis’, *Current Microbiology*, 70(3), pp. 441–449. doi: 10.1007/s00284-014-0741-2.

Finn, R. D. *et al.* (2015) ‘HMMER web server: 2015 Update’, *Nucleic Acids Research*, 43(W1), pp. W30–W38. doi: 10.1093/nar/gkv397.

Fisk, D. G. *et al.* (2006) ‘Saccharomyces cerevisiae S288C genome annotation: a working hypothesis’, *Yeast*, 23(12), pp. 857–865. doi: 10.1002/yea.1400.

França, M. B., Panek, A. D. and Eleutherio, E. C. A. (2007) 'Oxidative stress and its effects during dehydration', *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 146(4), pp. 621–631. doi: 10.1016/j.cbpa.2006.02.030.

Gallone, B. *et al.* (2016) 'Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts', *Cell*. Elsevier, 166(6), p. 1397–1410.e16. doi: 10.1016/j.cell.2016.08.020.

Garay-Arroyo, A. and Covarrubias, A. A. (1999) 'Three genes whose expression is induced by stress inSaccharomyces cerevisiae', *Yeast*, 15(10A), pp. 879–892. doi: 10.1002/(SICI)1097-0061(199907)15:10A<879::AID-YEA428>3.0.CO;2-Q.

García-Ríos, E. *et al.* (2017) 'The genetic architecture of low-temperature adaptation in the wine yeast Saccharomyces cerevisiae', *BMC Genomics*, 18(1), p. 159. doi: 10.1186/s12864-017-3572-2.

García-Ríos, E., López-Malo, M. and Guillamón, J. M. (2014) 'Global phenotypic and genomic comparison of two Saccharomyces cerevisiae wine strains reveals a novel role of the sulfur assimilation pathway in adaptation at low temperature fermentations.', *BMC genomics*, 15, p. 1059. doi: 10.1186/1471-2164-15-1059.

Gavin, A.-C. *et al.* (2006) 'Proteome survey reveals modularity of the yeast cell machinery', *Nature*, 440(7084), pp. 631–636. doi: 10.1038/nature04532.

Giaever, G. *et al.* (2002) 'Functional profiling of the Saccharomyces cerevisiae genome.', *Nature*, 418(6896), pp. 387–391. doi: 10.1038/nature00935.

Gietz, R. D. and Schiestl, R. H. (2007) 'Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method', *Nature Protocols*, 2(1), pp. 35–37. doi: 10.1038/nprot.2007.14.

Gimeno, C. J. and Fink, G. R. (1994) 'Induction of pseudohyphal growth by overexpression of PHD1, a Saccharomyces cerevisiae gene related to transcriptional regulators of fungal development.', *Molecular and Cellular Biology*, 14(3), pp. 2100–12. doi: 10.1128/MCB.14.3.2100.Updated.

Goffeau, A. A. *et al.* (1996) 'Life with 6000 Genes', *Science*, 274(5287), pp. 1187–1191.

Gonçalves, M. *et al.* (2016) 'Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts', *Current Biology*, 26(20), pp. 2750–2761. doi: 10.1016/j.cub.2016.08.040.

Gonzalez, R. *et al.* (2016) 'New genes involved in osmotic stress tolerance in saccharomyces cerevisiae', *Frontiers in Microbiology*, 7(SEP), pp. 1–12. doi: 10.3389/fmicb.2016.01545.

Graves, J. A. and Henry, S. A. (2000) 'Regulation of the yeast INO1 gene. The products of the INO2, INO4 and OPI1 regulatory genes are not required for repression in response to inositol.', *Genetics*, 154(4), pp. 1485–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10747047.

de Groot, M. J. L. *et al.* (2007) 'Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes', *Microbiology*, 153(11),

pp. 3864–3878. doi: 10.1099/mic.0.2007/009969-0.

Gu, Z. *et al.* (2003) 'Role of duplicate genes in genetic robustness against null mutations.', *Nature*, 421(6918), pp. 63–6. doi: 10.1038/nature01198.

Gueldener, U. *et al.* (2002) 'A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast.', *Nucleic Acids Research*, 30(6), p. e23. doi: 10.1093/nar/30.6.e23.

Gygi, S. P. *et al.* (1999) 'Correlation between Protein and mRNA Abundance in Yeast', *Molecular and Cellular Biology*, 19(3), pp. 1720–1730. doi: 10.1128/MCB.19.3.1720.

Hazelwood, L. A. *et al.* (2008) 'The Ehrlich Pathway for Fusel Alcohol Production: a Century of Research on Saccharomyces cerevisiae Metabolism', *Applied and Environmental Microbiology*, 74(12), pp. 3920–3920. doi: 10.1128/AEM.00934-08.

Hebert, A. S. *et al.* (2014) 'The One Hour Yeast Proteome', *Molecular & Cellular Proteomics*, 13(1), pp. 339–347. doi: 10.1074/mcp.M113.034769.

Hillenmeyer, M. E. *et al.* (2008) 'The chemical genomic portrait of yeast: uncovering a phenotype for all genes.', *Science*, 320(5874), pp. 362–365. doi: 10.1126/science.1150021.

Hirsch, J. P. and Henry, S. A. (1986) 'Expression of the Saccharomyces cerevisiae inositol-1-phosphate synthase (INO1) gene is regulated by factors that affect phospholipid synthesis.', *Molecular and Cellular Biology*, 6(10), pp. 3320–8. doi: 10.1128/MCB.6.10.3320.

Hirst, M. B. and Richter, C. L. (2016) 'Review of Aroma Formation through Metabolic Pathways of Saccharomyces cerevisiae in Beverage Fermentations', *American Journal of Enology and Viticulture*, 67(4), pp. 361–370. doi: 10.5344/ajev.2016.15098.

Ho, C. H. *et al.* (2009) 'A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds', *Nature Biotechnology*, 27(4), pp. 369–377. doi: 10.1038/nbt.1534.

Hontz, R. D. *et al.* (2009) 'Genetic identification of factors that modulate ribosomal DNA transcription in Saccharomyces cerevisiae', *Genetics*, 182(1), pp. 105–119. doi: 10.1534/genetics.108.100313.

Huang, C., Roncoroni, M. and Gardner, R. C. (2014) 'MET2 affects production of hydrogen sulfide during wine fermentation', *Applied Microbiology and Biotechnology*, 98(16), pp. 7125–7135. doi: 10.1007/s00253-014-5789-1.

Huang, H. *et al.* (2008) 'The crystal structure and identification of NQM1/YGR043C, a transaldolase from Saccharomyces cerevisiae', *Proteins: Structure, Function, and Bioinformatics*, 73(4), pp. 1076–1081. doi: 10.1002/prot.22237.

Huh, W.-K. *et al.* (2003) 'Global analysis of protein localization in budding yeast', *Nature*, 425(6959), pp. 686–691. doi: 10.1038/nature02026.

Husnik, J. I. *et al.* (2007) 'Functional analyses of the malolactic wine yeast ML01', *American Journal of*

Enology and Viticulture, 58(1), pp. 42–52.

Hyma, K. E. *et al.* (2011) 'Divergence in wine characteristics produced by wild and domesticated strains of Saccharomyces cerevisiae', *FEMS Yeast Research*, 11(7), pp. 540–551. doi: 10.1111/j.1567-1364.2011.00746.x.

Ibáñez, C. *et al.* (2017) 'RNAseq-based transcriptome comparison of Saccharomyces cerevisiae strains isolated from diverse fermentative environments', *International Journal of Food Microbiology*, 257(July), pp. 262–270. doi: 10.1016/j.ijfoodmicro.2017.07.001.

Illumina (2010) 'Technology Spotlight: Illumina Sequencing', pp. 383–388. doi: 10.1016/S0167-7799(03)00189-6.

Inadome, H. *et al.* (2001) 'A novel protein, Mpm1, of the mitochondria of the yeast Saccharomyces cerevisiae.', *Bioscience, Biotechnology, and Biochemistry*, 65(11), pp. 2577–80. doi: 10.1271/bbb.65.2577.

Ingolia, N. T. *et al.* (2009) 'Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.', *Science*, 324(5924), pp. 218–23. doi: 10.1126/science.1168978.

Ivorra, C., Pérez-ortín, J. and del Olmo, M. (1999) 'An inverse correlation between stress resistance and stuck fermentations in wine yeast. A molecular sudy.', *Biotechnology and Bioengineering*, 64(6), pp. 698–708.

Iwashita, M. (2016) *Functional analysis of non-annotated genes in Saccharomyces cerevisiae using wine yeast and Chardonnay fermentation*. University of British Columbia.

De Jong-Gubbels, P. *et al.* (1997) 'The Saccharomyces cerevisiae acetyl-coenzyme a synthetase encoded by the a ACS1 gene, but not the a ACS2-encoded enzyme, is subject to glucose catabolite inactivation', *FEMS Microbiology Letters*, 153(1), pp. 75–81. doi: 10.1016/S0378-1097(97)00236-X.

Jordan, P. *et al.* (2016) 'Hxt13, Hxt15, Hxt16 and Hxt17 from Saccharomyces cerevisiae represent a novel type of polyol transporters', *Scientific Reports*. Nature Publishing Group, 6(1), p. 23502. doi: 10.1038/srep23502.

Karim, A. S., Curran, K. A. and Alper, H. S. (2013) 'Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications', *FEMS Yeast Research*, 13(1), pp. 107–116. doi: 10.1111/1567-1364.12016.

Kellis, M. *et al.* (2003) 'Sequencing and comparison of yeast species to identify genes and regulatory elements.', *Nature*, 423(6937), pp. 241–54. doi: 10.1038/nature01644.

Kershaw, C. J. *et al.* (2015) 'Integrated multi-omics analyses reveal the pleiotropic nature of the control of gene expression by Puf3p', *Scientific Reports*. Nature Publishing Group, 5(1), p. 15518. doi: 10.1038/srep15518.

Kowalczuk, M. *et al.* (1999) 'Total number of coding open reading frames in the yeast genome.', *Yeast*, 15(11), pp. 1031–4. doi: 10.1002/(SICI)1097-0061(199908)15:11<1031::AID-YEA431>3.0.CO;2-G.

Kratzer, S. and Schüller, H. J. (1995) 'Carbon source-dependent regulation of the acetyl-coenzyme A synthetase-encoding gene ACSI from saccharomyces cerevisiae', *Gene*, 161(1), pp. 75–79. doi: 10.1016/0378-1119(95)00289-I.

Krogan, N. J. *et al.* (2006) 'Global landscape of protein complexes in the yeast Saccharomyces cerevisiae', *Nature*, 440(7084), pp. 637–643. doi: 10.1038/nature04670.

Kurtzman, C. P. and Robnett, C. J. (1998) 'Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences', *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 73(4), pp. 331–371. doi: 10.1023/A:1001761008817.

Kvitek, D. J., Will, J. L. and Gasch, A. P. (2008) 'Variations in stress sensitivity and genomic expression in diverse S. cerevisiae isolates', *PLoS Genetics*, 4(10), pp. 31–35. doi: 10.1371/journal.pgen.1000223.

Lallemand (2015) *Enoferm M2, Australia & New Zealand Technical Data Sheet*. Available at: http://winequip.com.au/wp-content/uploads/2015/01/M2.pdf (Accessed: 25 October 2015).

Lashkari, D. A. *et al.* (1997) 'Yeast microarrays for genome wide parallel genetic and gene expression analysis', *Proceedings of the National Academy of Sciences*, 94(24), pp. 13057–13062. doi: 10.1073/pnas.94.24.13057.

Lee, M.-L. L. *et al.* (2000) 'Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.', *Proceedings of the National Academy of Sciences*, 97(18), pp. 9834–9. doi: VL - 97.

Legras, J. L. *et al.* (2007) 'Bread, beer and wine: Saccharomyces cerevisiae diversity reflects human history', *Molecular Ecology*, 16(10), pp. 2091–2102. doi: 10.1111/j.1365-294X.2007.03266.x.

Lertwattanasakul, N. *et al.* (2015) 'Genetic basis of the highly efficient yeast Kluyveromyces marxianus: complete genome sequence and transcriptome analyses', *Biotechnology for Biofuels*, 8(1), p. 47. doi: 10.1186/s13068-015-0227-x.

Liti, G. *et al.* (2009) 'Population genomics of domestic and wild yeasts.', *Nature*. Nature Publishing Group, 458(7236), pp. 337–41. doi: 10.1038/nature07743.

Luo, Z. and van Vuuren, H. J. J. (2009) 'Functional analyses of PAU genes in Saccharomyces cerevisiae', *Microbiology*, 155(12), pp. 4036–4049. doi: 10.1099/mic.0.030726-0.

Lutfiyya, L. L. *et al.* (1998) 'Characterization of three related glucose repressors and genes they regulate in Saccharomyces cerevisiae', *Genetics*, 150(4), pp. 1377–1391.

Luyten, K., Riou, C. and Blondin, B. (2002) 'The hexose transporters of Saccharomyces cerevisiae play

different roles during enological fermentation', *Yeast*, 19(8), pp. 713–726. doi: 10.1002/yea.869.

Ma, M. and Liu, Z. L. (2010) 'Mechanisms of ethanol tolerance in saccharomyces cerevisiae', *Applied Microbiology and Biotechnology*, 87(3), pp. 829–845. doi: 10.1007/s00253-010-2594-3.

Marks, V. D. *et al.* (2008) 'Dynamics of the yeast transcriptome during wine fermentation reveals a novel fermentation stress response', *FEMS Yeast Research*, 8(1), pp. 35–52. doi: 10.1111/j.1567-1364.2007.00338.x.

Martínez-Montañés, F. *et al.* (2013) 'Activator and repressor functions of the Mot3 transcription factor in the osmostress response of Saccharomyces cerevisiae', *Eukaryotic Cell*, 12(5), pp. 636–647. doi: 10.1128/EC.00037-13.

Martiniuk, J. T. *et al.* (2016) 'Impact of Commercial Strain Use on Saccharomyces cerevisiae Population Structure and Dynamics in Pinot Noir Vineyards and Spontaneous Fermentations of a Canadian Winery', *PLOS ONE*. Edited by J. Schacherer, 11(8), p. e0160259. doi: 10.1371/journal.pone.0160259.

Marullo, P. *et al.* (2009) 'Genetic improvement of thermo-tolerance in wine Saccharomyces cerevisiae strains by a backcross approach', *FEMS Yeast Research*, 9(8), pp. 1148–1160. doi: 10.1111/j.1567-1364.2009.00550.x.

Matsushika, A. *et al.* (2012) 'Characterization of non-oxidative transaldolase and transketolase enzymes in the pentose phosphate pathway with regard to xylose utilization by recombinant Saccharomyces cerevisiae', *Enzyme and Microbial Technology*, 51(1), pp. 16–25. doi: 10.1016/j.enzmictec.2012.03.008.

McIlwain, S. J. *et al.* (2016) 'Genome Sequence and Analysis of a Stress-Tolerant, Wild-Derived Strain of Saccharomyces cerevisiae Used in Biofuels Research.', *G3; Genes|Genomes|Genetics*, 6(6), pp. 1757–66. doi: 10.1534/g3.116.029389.

Merz, S. and Westermann, B. (2009) 'Genome-wide deletion mutant analysis reveals genes required for respiratory growth, mitochondrial genome maintenance and mitochondrial protein synthesis in Saccharomyces cerevisiae', *Genome Biology*, 10(9), p. R95. doi: 10.1186/gb-2009-10-9-r95.

Michel, S. *et al.* (2015) 'A haploproficient interaction of the transaldolase paralogue NQM1 with the transcription factor VHR1 affects stationary phase survival and oxidative stress resistance.', *BMC genetics*, 16(1), p. 13. doi: 10.1186/s12863-015-0171-6.

Mizushima, N. (2005) 'The pleiotropic role of autophagy: from protein metabolism to bactericide', *Cell Death and Differentiation*, 12, pp. 1535–1541. doi: 10.1038/sj.cdd.4401728.

Moon, J. and Liu, Z. L. (2015) 'Direct enzyme assay evidence confirms aldehyde reductase function of Ydr541cp and Ygl039wp from Saccharomyces cerevisiae', *Yeast*, 32(4), pp. 399–407. doi: 10.1002/yea.3067.

Moreno, F. (2002) 'The hexokinase 2-dependent glucose signal transduction pathway of Saccharomyces

cerevisiae', *FEMS Microbiology Reviews*, 26(1), pp. 83–90. doi: 10.1016/S0168-6445(01)00077-8.

Mortimer, R. K. (2000) 'Evolution and variation of the yeast (Saccharomyces) genome', *Genome Research*, 10(4), pp. 403–409. doi: 10.1101/gr.10.4.403.

Mortimer, R. K., Contopoulou, C. R. and King, J. S. (1992) 'Genetic and physical maps of Saccharomyces cerevisiae, Edition 11.', *Yeast*, 8(10), pp. 817–902. doi: 10.1002/yea.320081002.

Mortimer, R. K. and Johnston, J. R. (1986) 'Genealogy of principal strains of the yeast genetic stock center', *Genetics*, 113(1), pp. 35–43.

Mortimer, R. K. and Schild, D. (1980) 'Genetic map of Saccharomyces cerevisiae.', *Microbiological Reviews*, 44(4), pp. 519–71.

Nakagawa, Y. *et al.* (2017) 'Development of intra-strain self-cloning procedure for breeding baker's yeast strains', *Journal of Bioscience and Bioengineering*, 123(3), pp. 319–326. doi: 10.1016/j.jbiosc.2016.10.008.

Noirel, J. *et al.* (2011) 'Methods in Quantitative Proteomics: Setting iTRAQ on the Right Track', *Current Proteomics*, 8(1), pp. 17–30. doi: 10.2174/157016411794697408.

Novo, M. *et al.* (2009) 'Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118.', *Proceedings of the National Academy of Sciences*, 106(38), pp. 16333–16338. doi: 10.1073/pnas.0904673106.

Novo, M. *et al.* (2013) 'Genome-Wide Study of the Adaptation of Saccharomyces cerevisiae to the Early Stages of Wine Fermentation', *PLoS ONE*, 8(9), pp. 1–15. doi: 10.1371/journal.pone.0074086.

Pan, X. and Heitman, J. (2000) 'Sok2 regulates yeast pseudohyphal differentiation via a transcription factor cascade that regulates cell-cell adhesion.', *Molecular and Cellular Biology*, 20(22), pp. 8364–8372. doi: 10.1128/MCB.20.22.8364-8372.2000.

Pelet, S. *et al.* (2011) 'Transient Activation of the HOG MAPK Pathway Regulates Bimodal Gene Expression', *Science*, 332(6030), pp. 732–735. doi: 10.1126/science.1198851.

Pereira, Y. *et al.* (2008) 'Chromate causes sulfur starvation in yeast', *Toxicological Sciences*, 106(2), pp. 400–412. doi: 10.1093/toxsci/kfn193.

Piggott, N. *et al.* (2011) 'Genome-wide Fitness Profiles Reveal a Requirement for Autophagy During Yeast Fermentation', *G3; Genes|Genomes|Genetics*. Edited by C. Nislow, 1(5), pp. 353–367. doi: 10.1534/g3.111.000836.

Piper, P. W. *et al.* (1994) 'Induction of major heat-shock proteins of Saccharomyces cere visiae , i nc I u d i ng p I as ma membrane HspSO , by ethanol levels above a critical threshold', *Microbiology*, 140(1 994), pp. 3031–3038. doi: 10.1099/13500872-140-11-3031.

Piper, P. W. *et al.* (1997) 'Hsp30, the integral plasma membrane heat shock protein of Saccharomyces

cerevisiae, is a stress-inducible regulator of plasma membrane H(+)-ATPase', *Cell Stress Chaperones*, pp. 12–24. doi: 10.1379/1466-1268(1997)002<0012:HTIPMH>2.3.CO;2.

Pretorius, I. S. (2000) 'Tailoring wine yeast for the new millennium: Novel approaches to the ancient art of winemaking', *Yeast*, 16(8), pp. 675–729. doi: 10.1002/1097-0061(20000615)16:8<675::AID-YEA585>3.0.CO;2-B.

Puig, S. and Pérez-Ortín, J. E. (2000) 'Stress response and expression patterns in wine fermentations of yeast genes induced at the diauxic shift', *Yeast*, 16(2), pp. 139–148. doi: 10.1002/(SICI)1097-0061(20000130)16:2<139::AID-YEA512>3.0.CO;2-J.

Rachidi, N. *et al.* (2000) 'Saccharomyces cerevisiae PAU genes are induced by anaerobiosis', *Molecular Microbiology*, 35(6), pp. 1421–1430. doi: 10.1046/j.1365-2958.2000.01807.x.

R Development Core Team (2017) 'R: A language and environment for statistical computing', *R Development Core Team*. Vienna, Austria: R Foundation for Statistical Computing, p. 63. doi: 3-900051-14-3.

Ramírez, M., Peréz, F. and Regodón, J. A. (1998) 'A simple and reliable method for hybridization of homothallic wine strains of Saccharomyces cerevisiae', *Applied and Environmental Microbiology*, 64(12), pp. 5039–5041.

Reed, G. and Chen, S. L. (1978) 'Evaluating Commercial Active Dry Wine Yeasts by Fermentation Activity', *American Journal of Enology and Viticulture*, 29(3), pp. 165–168.

Regenberg, B. *et al.* (1998) 'Dip5p mediates high-affinity and high-capacity transport of L-glutamate and L-aspartate in Saccharomyces cerevisiae', *Current Genetics*, 33(3), pp. 171–177. doi: 10.1007/s002940050324.

Regenberg, B. *et al.* (2006) 'Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae.', *Genome biology*, 7(11), p. R107. doi: 10.1186/gb-2006-7-11-r107.

Riou, C. *et al.* (1997) 'Stationary-phase gene expression in Saccharomyces cerevisiae during wine fermentation', *Yeast*, 13(10), p. 903–15. doi: 10.1002/(SICI)1097-0061(199708)13:10<903::AID-YEA145>3.0.CO;2-1 [pii]\r10.1002/(SICI)1097-0061(199708)13:10<903::AID-YEA145>3.0.CO;2-1.

Rodriguez, A. *et al.* (2001) 'The hexokinase 2 protein regulates the expression of the GLK1, HXK1 and HXK2 genes of Saccharomyces cerevisiae', *Biochemical Journal*, 355(3), pp. 625–631. doi: 10.1042/bj3550625.

Romano, P. *et al.* (1985) 'Improvement of a Wine Saccharomyces cerevisiae Strain by a Breeding Program.', *Applied and Environmental Microbiology*, 50(4), pp. 1064–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16346903.

Ross, P. L. *et al.* (2004) 'Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents', *Molecular & Cellular Proteomics*, 3(12), pp. 1154–1169. doi: 10.1074/mcp.M400129-MCP200.

Rossignol, T. *et al.* (2003) 'Genome-wide monitoring of wine yeast gene expression during alcoholic fermentation', *Yeast*, 20(16), pp. 1369–1385. doi: 10.1002/yea.1046.

Rossignol, T. *et al.* (2009) 'The proteome of a wine yeast strain during fermentation, correlation with the transcriptome', *Journal of Applied Microbiology*, 107(1), pp. 47–55. doi: 10.1111/j.1365-2672.2009.04156.x.

Rossouw, D. *et al.* (2009) 'Comparative transcriptomic approach to investigate differences in wine yeast physiology and metabolism during fermentation', *Applied and Environmental Microbiology*, 75(20), pp. 6600–6612. doi: 10.1128/AEM.01251-09.

Rossouw, D. *et al.* (2010) 'Comparative transcriptomic and proteomic profiling of industrial wine yeast strains', *Applied and Environmental Microbiology*, 76(12), pp. 3911–3923. doi: 10.1128/AEM.00586-10.

Rossouw, D. and Bauer, F. F. (2009) 'Comparing the transcriptomes of wine yeast strains: Toward understanding the interaction between environment and transcriptome during fermentation', *Applied Microbiology and Biotechnology*, 84(5), pp. 937–954. doi: 10.1007/s00253-009-2204-4.

Salmon, J. M. and Barre, P. (1998) 'Improvement of nitrogen assimilation and fermentation kinetics under enological conditions by derepression of alternative nitrogen-assimilatory pathways in an industrial Saccharomyces cerevisiae strain.', *Applied and Environmental Microbiology*, 64(10), pp. 3831–7.

Schaaff-Gerstenschläger, I. *et al.* (1993) 'TKL2, a second transketolase gene of Saccharomyces cerevisiae', *European Journal of Biochemistry*, 217, pp. 487–492. doi: 10.1111/j.1432-1033.1993.tb18268.x.

Schacherer, J. *et al.* (2007) 'Genome-wide analysis of nucleotide-level variation in commonly used Saccharomyces cerevisiae strains', *PLoS ONE*, 2(3), pp. 1–7. doi: 10.1371/journal.pone.0000322.

Schacherer, J. *et al.* (2009) 'Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae.', *Nature*, 458(7236), pp. 342–5. doi: 10.1038/nature07670.

Schimz, K. L. (1980) 'The effect of sulfite on the yeast Saccharomyces cerevisiae.', *Archives of Microbiology*, 125(1–2), pp. 89–95. doi: 10.1007/BF00435682.

Schoch, C. L. *et al.* (2012) 'Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi', *Proceedings of the National Academy of Sciences*, 109(16), pp. 1–6. doi: 10.1073/pnas.1117018109.

Senko, M. W. *et al.* (2013) 'Novel parallelized quadrupole/linear ion trap/orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates', *Analytical Chemistry*,

85(24), pp. 11710–11714. doi: 10.1021/ac403115c.

Da Silva, T. *et al.* (2015) 'Hybridization within Saccharomyces genus results in homoeostasis and phenotypic novelty in winemaking conditions', *PLoS ONE*, 10(5), pp. 1–24. doi: 10.1371/journal.pone.0123834.

Skelly, D. A. *et al.* (2013) 'Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast', *Genome Research*, 23(9), pp. 1496–1504. doi: 10.1101/gr.155762.113.

Solis-Escalante, D. *et al.* (2015) 'The genome sequence of the popular hexose-transport-deficient Saccharomyces cerevisiae strain EBY.VW4000 reveals LoxP/Cre-induced translocations and gene loss', *FEMS Yeast Research*, 15(2), pp. 1–12. doi: 10.1093/femsyr/fou004.

Steensels, J. *et al.* (2014) 'Improving industrial yeast strains: Exploiting natural and artificial diversity', *FEMS Microbiology Reviews*, 38(5), pp. 947–995. doi: 10.1111/1574-6976.12073.

Strope, P. K. *et al.* (2015) 'The 100-genomes strains, an S. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen', *Genome Research*, 125(5), pp. 762–774. doi: 10.1101/gr.185538.114.

Takagi, H. *et al.* (2000) 'Saccharomyces cerevisiae Σ1278b has novel genes of the N- acetyltransferase gene superfamily required for L-proline analogue resistance', *Journal of Bacteriology*, 182(15), pp. 4249–4256. doi: 10.1128/JB.182.15.4249-4256.2000.

Takahashi, H. *et al.* (2006) 'Nucleocytosolic Acetyl-Coenzyme A Synthetase Is Required for Histone Acetylation and Global Transcription', *Molecular Cell*, 23(2), pp. 207–217. doi: 10.1016/j.molcel.2006.05.040.

Tamura, Y. *et al.* (2012) 'Phosphatidylethanolamine biosynthesis in mitochondria: Phosphatidylserine (PS) trafficking is independent of a PS decarboxylase and intermembrane space proteins Ups1p and Ups2p', *Journal of Biological Chemistry*, 287(52), pp. 43961–43971. doi: 10.1074/jbc.M112.390997.

Terrell, E., Cliff, M. A. and Van Vuuren, H. J. J. (2015) 'Functional characterization of individual- and mixed-Burgundian Saccharomyces cerevisiae isolates for fermentation of pinot noir', *Molecules*, 20(3), pp. 5112–5136. doi: 10.3390/molecules20035112.

Thompson, D. A. and Cubillos, F. A. (2017) 'Natural gene expression variation studies in yeast', *Yeast*, 34(1), pp. 3–17. doi: 10.1002/yea.3210.

Trabalzini, L. *et al.* (2003) 'Proteomic response to physiological fermentation stresses in a wild-type wine strain of Saccharomyces cerevisiae.', *The Biochemical Journal*, 370(Pt 1), pp. 35–46. doi: 10.1042/BJ20020140.

Treu, L., Campanaro, S., *et al.* (2014) 'Oxidative stress response and nitrogen utilization are strongly variable in Saccharomyces cerevisiae wine strains with different fermentation performances', *Applied*

Microbiology and Biotechnology, 98(9), pp. 4119–4135. doi: 10.1007/s00253-014-5679-6.

Treu, L., Toniolo, C., *et al.* (2014) 'The impact of genomic variability on gene expression in environmental Saccharomyces cerevisiae strains', *Environmental Microbiology*, 16(5), pp. 1378–1397. doi: 10.1111/1462-2920.12327.

Tronchoni, J. *et al.* (2017) 'Early transcriptional response to biotic stress in mixed starter fermentations involving Saccharomyces cerevisiae and Torulaspora delbrueckii', *International Journal of Food Microbiology*, 241, pp. 60–68. doi: 10.1016/j.ijfoodmicro.2016.10.017.

Varela, C. *et al.* (2005) 'Quantitative analysis of wine yeast gene expression profiles under winemaking conditions', *Yeast*, 22(5), pp. 369–383. doi: 10.1002/yea.1217.

Varela, C., Pizarro, F. and Agosin, E. (2004) 'Biomass content governs fermentation rate in nitrogen-deficient wine musts', *Applied and Environmental Microbiology*, 70(6), pp. 3392–3400. doi: 10.1128/AEM.70.6.3392-3400.2004.

Velculescu, V. E. *et al.* (1995) 'Serial Analysis of Gene Expression', *Science*, 270(5235), pp. 484–487. doi: 10.1126/science.270.5235.484.

Velculescu, V. E. *et al.* (1997) 'Characterization of the yeast transcriptome', *Cell*, 88(2), pp. 243–251. doi: 10.1016/S0092-8674(00)81845-0.

Verstrepen, K. J., Chambers, P. J. and Pretorius, I. S. (2006) 'The Development of Superior Yeast Strains for the Food and Beverage Industries: Challenges, Opportunities and Potential Benefits', in *Yeasts in Food and Beverages*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 399–444. doi: 10.1007/978-3-540-28398-0_13.

Walkey, C. J. *et al.* (2011) 'The Saccharomyces cerevisiae fermentation stress response protein Igd1p/Yfr017p regulates glycogen levels by inhibiting the glycogen debranching enzyme', *FEMS Yeast Research*, 11(6), pp. 499–508. doi: 10.1111/j.1567-1364.2011.00740.x.

Walkey, C. J. *et al.* (2012) 'The Fermentation Stress Response Protein Aaf1p/Yml081Wp Regulates Acetate Production in Saccharomyces cerevisiae', *PLoS ONE*, 7(12), pp. 1–8. doi: 10.1371/journal.pone.0051551.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature Reviews Genetics*, 10(1), pp. 57–63. doi: 10.1038/nrg2484.

Warringer, J. *et al.* (2011) 'Trait variation in yeast is defined by population history', *PLoS Genetics*, 7(6). doi: 10.1371/journal.pgen.1002111.

Wieczorke, R. *et al.* (1999) 'Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in Saccharomyces cerevisiae', *FEBS Letters*, 464(3), pp. 123–128. doi: 10.1016/S0014-5793(99)01698-1.

Wiens, F. *et al.* (2008) ‘Chronic intake of fermented floral nectar by wild treeshrews.’, *Proceedings of the National Academy of Sciences*, 105(30), pp. 10426–31. doi: 10.1073/pnas.0801628105.

De Winde, J. H. *et al.* (1996) ‘Differential requirement of the yeast sugar kinases for sugar sensing in establishing the catabolite-repressed state’, *European Journal of Biochemistry*, 241(2), pp. 633–643. doi: 10.1111/j.1432-1033.1996.00633.x.

Wodicka, L. *et al.* (1997) ‘Genome-wide expression monitoring in Saccharomyces cerevisiae’, *Nature Biotechnology*, 15(13), pp. 1359–1367. doi: 10.1038/nbt1297-1359.

Wolfe, K. H. and Shields, D. C. (1997) ‘Molecular evidence for an ancient duplication of the entire yeast genome’, *Nature*, 387(6634), pp. 708–713. doi: 10.1038/42711.

Wong, S. and Wolfe, K. H. (2005) ‘Birth of a metabolic gene cluster in yeast by adaptive gene relocation’, *Nature Genetics*, 37(7), pp. 777–782. doi: 10.1038/ng1584.

Xia, K. *et al.* (2016) ‘New insights into the mechanisms of acetic acid resistance in Acetobacter pasteurianus using iTRAQ-dependent quantitative proteomic analysis’, *International Journal of Food Microbiology*, 238, pp. 241–251. doi: 10.1016/j.ijfoodmicro.2016.09.016.

Yagoub, D. *et al.* (2015) ‘Proteogenomic Discovery of a Small, Novel Protein in Yeast Reveals a Strategy for the Detection of Unannotated Short Open Reading Frames’, *Journal of Proteome Research*, 14(12), pp. 5038–5047. doi: 10.1021/acs.jproteome.5b00734.

Yang, Y. H. and Speed, T. (2002) ‘Design issues for cDNA microarray experiments.’, *Nature Reviews Genetics*, 3(8), pp. 579–88. doi: 10.1038/nrg863.

APPENDICES

A. Currently available *S. cerevisiae* genomes

Table A.1 Currently available *S. cerevisiae* genomes. The complete list of *S. cerevisiae* genome assemblies available for download from the NCBI Genome Database (https://www.ncbi.nlm.nih.gov/genome/15) as of November 1st, 2017.

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
S288C	12.1571	GCA_000146045.2	17	6445	6002	1999-11-08	Complete Genome
ySR127	12.0863	GCA_001051215.1	17	-	-	2015-07-09	Complete Genome
YJM993	12.502	GCA_000662435.1	18	6961	5471	2014-05-01	Chromosome
YJM195	12.7855	GCA_000975585.1	17	7079	5373	2015-02-23	Chromosome
YJM244	12.9872	GCA_000975615.1	18	7376	5329	2015-02-23	Chromosome
YJM1078	11.8467	GCA_000975645.2	18	6152	5424	2015-02-25	Chromosome
YJM1083	12.5085	GCA_000975675.1	18	6756	5404	2015-02-23	Chromosome
YJM1129	13.0872	GCA_000975705.1	18	7642	5475	2015-02-23	Chromosome
YJM189	12.2707	GCA_000975735.3	18	6564	5437	2015-02-23	Chromosome
YJM193	13.0573	GCA_000975765.1	18	7498	5411	2015-02-23	Chromosome
YJM248	12.0749	GCA_000975795.1	18	6357	5391	2015-02-23	Chromosome
YJM270	12.3693	GCA_000975825.1	17	6844	5427	2015-02-23	Chromosome
YJM271	13.5334	GCA_000975855.2	18	8054	5367	2015-02-23	Chromosome
YJM320	13.5229	GCA_000975885.1	18	8021	5387	2015-02-23	Chromosome
YJM326	12.4289	GCA_000975915.1	18	6658	5416	2015-02-23	Chromosome
YJM428	12.3275	GCA_000975945.1	18	6533	5402	2015-02-23	Chromosome
YJM450	12.4178	GCA_000975975.2	18	6712	5435	2015-02-23	Chromosome
YJM451	12.3618	GCA_000976005.1	18	6632	5405	2015-02-23	Chromosome
YJM453	12.6829	GCA_000976035.1	18	7121	5393	2015-02-23	Chromosome
YJM456	12.3959	GCA_000976065.1	18	6574	5476	2015-02-23	Chromosome
YJM470	12.5987	GCA_000976095.1	17	6927	5395	2015-02-23	Chromosome
YJM541	12.9961	GCA_000976125.1	18	7389	5398	2015-02-23	Chromosome
YJM554	13.052	GCA_000976155.3	18	7409	5363	2015-02-23	Chromosome
YJM555	13.3722	GCA_000976185.1	18	7802	5460	2015-02-23	Chromosome
YJM627	12.6266	GCA_000976215.1	18	6936	5400	2015-02-23	Chromosome
YJM681	12.3273	GCA_000976245.1	18	6584	5398	2015-02-23	Chromosome
YJM682	12.4194	GCA_000976275.3	18	6628	5442	2015-02-23	Chromosome
YJM683	12.5361	GCA_000976305.1	17	6815	5464	2015-02-23	Chromosome
YJM689	12.2975	GCA_000976335.1	17	6576	5393	2015-02-23	Chromosome
YJM693	13.1562	GCA_000976365.1	17	7541	5372	2015-02-23	Chromosome
YJM969	12.3221	GCA_000976395.1	18	6768	5455	2015-02-23	Chromosome
YJM972	12.972	GCA_000976425.1	18	7492	5441	2015-02-23	Chromosome
YJM975	12.2374	GCA_000976455.2	18	6608	5438	2015-02-23	Chromosome
YJM978	12.483	GCA_000976485.1	18	6963	5448	2015-02-23	Chromosome
YJM981	12.9198	GCA_000976515.1	18	7462	5479	2015-02-23	Chromosome

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
YJM984	13.0939	GCA_000976545.1	18	7602	5424	2015-02-23	Chromosome
YJM987	12.9619	GCA_000976575.3	18	7532	5482	2015-02-23	Chromosome
YJM990	12.4481	GCA_000976605.2	18	6794	5386	2015-02-23	Chromosome
YJM996	12.7062	GCA_000976665.1	18	7258	5476	2015-02-23	Chromosome
YJM1133	12.433	GCA_000976695.2	18	6607	5338	2015-02-23	Chromosome
YJM1190	12.9423	GCA_000976725.1	18	7380	5441	2015-02-23	Chromosome
YJM1199	12.3567	GCA_000976755.1	18	6662	5471	2015-02-23	Chromosome
YJM1202	12.5711	GCA_000976785.1	18	6785	5403	2015-02-23	Chromosome
YJM1208	13.1366	GCA_000976815.2	18	7446	5395	2015-02-23	Chromosome
YJM1242	12.1831	GCA_000976845.2	18	6542	5402	2015-02-23	Chromosome
YJM1244	12.7286	GCA_000976875.2	18	7136	5363	2015-02-23	Chromosome
YJM1248	12.6069	GCA_000976905.1	17	6848	5319	2015-02-23	Chromosome
YJM1250	12.3537	GCA_000976935.1	18	6525	5328	2015-02-23	Chromosome
YJM1252	12.6455	GCA_000976965.2	17	7019	5392	2015-02-23	Chromosome
YJM1273	12.6094	GCA_000976995.1	17	6904	5471	2015-02-23	Chromosome
YJM1304	12.4187	GCA_000977025.1	18	6661	5381	2015-02-23	Chromosome
YJM1307	13.071	GCA_000977055.2	18	7510	5414	2015-02-23	Chromosome
YJM1311	12.5773	GCA_000977085.1	18	6880	5438	2015-02-23	Chromosome
YJM1326	12.444	GCA_000977115.2	18	6661	5431	2015-02-23	Chromosome
YJM1332	12.3026	GCA_000977145.1	18	6711	5463	2015-02-23	Chromosome
YJM1336	12.8564	GCA_000977175.2	18	7322	5422	2015-02-24	Chromosome
YJM1338	12.2815	GCA_000977205.1	18	6543	5423	2015-02-24	Chromosome
YJM1341	12.5081	GCA_000977235.1	18	6931	5445	2015-02-24	Chromosome
YJM1342	12.6892	GCA_000977265.2	17	6912	5409	2015-02-24	Chromosome
YJM1355	12.9034	GCA_000977295.1	18	7180	5371	2015-02-24	Chromosome
YJM1356	12.6028	GCA_000977325.1	18	7038	5465	2015-02-24	Chromosome
YJM1381	12.4563	GCA_000977355.2	18	6806	5470	2015-02-24	Chromosome
YJM1383	12.6305	GCA_000977385.1	18	6885	5427	2015-02-24	Chromosome
YJM1385	13.0244	GCA_000977415.1	17	7280	5442	2015-02-24	Chromosome
YJM1386	12.6431	GCA_000977445.1	18	6875	5378	2015-02-24	Chromosome
YJM1387	12.773	GCA_000977475.2	17	7218	5453	2015-02-24	Chromosome
YJM1388	12.5426	GCA_000977505.1	17	6935	5398	2015-02-24	Chromosome
YJM1389	12.2761	GCA_000977535.1	17	6591	5386	2015-02-24	Chromosome
YJM1399	12.4163	GCA_000977565.1	17	6488	5228	2015-02-24	Chromosome
YJM1400	12.4168	GCA_000977595.1	18	6623	5415	2015-02-24	Chromosome
YJM1401	12.2356	GCA_000977625.1	18	6386	5362	2015-02-24	Chromosome
YJM1402	12.6222	GCA_000977655.1	17	6924	5457	2015-02-24	Chromosome
YJM1415	12.8709	GCA_000977685.1	18	7422	5469	2015-02-24	Chromosome
YJM1417	12.7121	GCA_000977715.3	18	7207	5441	2015-02-24	Chromosome
YJM1418	12.5052	GCA_000977745.1	17	6772	5426	2015-02-24	Chromosome
YJM1419	13.6073	GCA_000977775.2	18	8081	5412	2015-02-24	Chromosome

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
YJM1433	12.5725	GCA_000977805.1	18	6920	5386	2015-02-24	Chromosome
YJM1434	12.5942	GCA_000977835.1	17	6865	5470	2015-02-24	Chromosome
YJM1439	12.9559	GCA_000977865.1	17	7351	5377	2015-02-24	Chromosome
YJM1443	12.2603	GCA_000977895.1	17	6397	5289	2015-02-24	Chromosome
YJM1444	12.4269	GCA_000977925.1	17	6581	5366	2015-02-24	Chromosome
YJM1447	12.1159	GCA_000977955.1	17	6302	5227	2015-02-24	Chromosome
YJM1450	13.0549	GCA_000977985.1	18	7373	5378	2015-02-24	Chromosome
YJM1460	12.785	GCA_000978015.1	17	7166	5382	2015-02-24	Chromosome
YJM1463	13.4375	GCA_000978045.1	18	7961	5372	2015-02-24	Chromosome
YJM1477	12.1914	GCA_000978075.2	18	6607	5429	2015-02-24	Chromosome
YJM1478	12.6776	GCA_000978105.1	18	6997	5414	2015-02-24	Chromosome
YJM1479	12.4809	GCA_000978135.1	18	6709	5413	2015-02-24	Chromosome
YJM1526	12.5382	GCA_000978165.1	18	6862	5363	2015-02-24	Chromosome
YJM1527	12.7997	GCA_000978195.1	18	7112	5402	2015-02-24	Chromosome
YJM1549	13.2595	GCA_000978225.1	17	7739	5452	2015-02-24	Chromosome
YJM1573	12.301	GCA_000978255.1	17	6531	5458	2015-02-24	Chromosome
YJM1574	12.4702	GCA_000978285.1	18	6738	5359	2015-02-24	Chromosome
YJM1592	12.3806	GCA_000978315.1	17	6742	5393	2015-02-24	Chromosome
YJM1615	12.651	GCA_000978345.1	18	6940	5412	2015-02-24	Chromosome
NCIM3186	12.1509	GCA_001029075.1	17	-	-	2015-06-17	Chromosome
BSPX042	12.0719	GCA_001592655.1	17	-	-	2016-03-15	Chromosome
HPRMAwf_D10	12.1175	GCA_001669875.1	17	-	-	2016-06-20	Chromosome
T.52_3C	12.1311	GCA_001669905.1	17	-	-	2016-06-20	Chromosome
T.52_3A	12.1226	GCA_001669935.1	17	-	-	2016-06-20	Chromosome
T.52_2H	12.1119	GCA_001669965.1	17	-	-	2016-06-20	Chromosome
Sol7-2	12.1145	GCA_001669995.1	17	-	-	2016-06-20	Chromosome
HB_C_TUKITUKI2_4	12.1255	GCA_001670025.1	17	-	-	2016-06-20	Chromosome
NSERVsf_F8	12.1228	GCA_001670055.1	17	-	-	2016-06-20	Chromosome
HCNTHsf_F8	12.1359	GCA_001670085.1	17	-	-	2016-06-20	Chromosome
T78	12.1253	GCA_001670115.1	17	-	-	2016-06-20	Chromosome
HCNTHsf_C5	12.1368	GCA_001670145.1	17	-	-	2016-06-20	Chromosome
TNPLST-4-S-2	12.1236	GCA_001670175.1	17	-	-	2016-06-20	Chromosome
HB_S_GIMBLETTROAD_9	12.1266	GCA_001670205.1	17	-	-	2016-06-20	Chromosome
T8	12.1217	GCA_001670235.1	17	-	-	2016-06-20	Chromosome
MTKSKsf_E2	12.1174	GCA_001670265.1	17	-	-	2016-06-20	Chromosome
CRIRIwf_A11	12.1231	GCA_001670295.1	17	-	-	2016-06-20	Chromosome
CDRDR_sf_H	12.131	GCA_001670325.1	17	-	-	2016-06-20	Chromosome
Soil7-1	12.0855	GCA_001670355.1	17	-	-	2016-06-20	Chromosome
HPRMTsf_H7	12.1194	GCA_001670385.1	17	-	-	2016-06-20	Chromosome
WI_S_OAKURA_4	12.102	GCA_001670415.1	17	-	-	2016-06-20	Chromosome
HB_S_BILANCHER_12	12.1284	GCA_001670845.1	17	-	-	2016-06-20	Chromosome

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
HB_S_GIMBLETTROAD_22	12.129	GCA_001670935.1	17	-	-	2016-06-20	Chromosome
HB_C_OMARUNUI_7	12.1142	GCA_001670965.1	17	-	-	2016-06-20	Chromosome
HB_C_TUKITUKI1_16	12.1245	GCA_001670995.1	17	-	-	2016-06-20	Chromosome
WI_S_JASA_5	12.1114	GCA_001671025.1	17	-	-	2016-06-20	Chromosome
T52	12.1209	GCA_001671055.1	17	-	-	2016-06-20	Chromosome
HB_C_KOROKIPO_3	12.1262	GCA_001671085.1	17	-	-	2016-06-20	Chromosome
HB_C_OMARUNUI_14	12.1253	GCA_001671115.1	17	-	-	2016-06-20	Chromosome
HB_C_OMARUNUI_6	12.1201	GCA_001671165.1	17	-	-	2016-06-20	Chromosome
WA_C_WAITAKEREROAD_7	12.1212	GCA_001671195.1	17	-	-	2016-06-20	Chromosome
WI_C_MB95MBMZ_4	12.091	GCA_001671225.1	17	-	-	2016-06-20	Chromosome
WI_C_MBSP_15	12.1001	GCA_001671255.1	17	-	-	2016-06-20	Chromosome
WI_C_MBSP_4	12.1228	GCA_001671285.1	17	-	-	2016-06-20	Chromosome
T.52_5E	12.125	GCA_001671315.1	17	-	-	2016-06-20	Chromosome
WA_C_KINGSMILL_10	12.1151	GCA_001671345.1	17	-	-	2016-06-20	Chromosome
WA_C_MATES_10	12.1212	GCA_001671375.1	17	-	-	2016-06-20	Chromosome
WSERCsf_G4	12.0992	GCA_001671405.1	17	-	-	2016-06-20	Chromosome
HB_S_GIMBLETTROAD_5	12.1329	GCA_001671435.1	17	-	-	2016-06-20	Chromosome
HCNKIsf_G7	12.1174	GCA_001671565.1	17	-	-	2016-06-20	Chromosome
WSETAwf_B1	12.1187	GCA_001671595.1	17	-	-	2016-06-20	Chromosome
NSEBRsf_A9	12.1358	GCA_001671625.1	17	-	-	2016-06-20	Chromosome
MARARsf_A10	12.1171	GCA_001671655.1	17	-	-	2016-06-20	Chromosome
T16	12.1188	GCA_001671685.1	17	-	-	2016-06-20	Chromosome
WA_C_CODDINGTON_2	12.1213	GCA_001671715.1	17	-	-	2016-06-20	Chromosome
T.52_5A	12.113	GCA_001671745.1	17	-	-	2016-06-20	Chromosome
WA_C_MATES_13	12.1248	GCA_001671775.1	17	-	-	2016-06-20	Chromosome
HB_C_KOROKIPO_12	12.1114	GCA_001671805.1	17	-	-	2016-06-20	Chromosome
WI_S_JASA_13	12.1111	GCA_001671835.1	17	-	-	2016-06-20	Chromosome
HB_C_TUKITUKI2_10	12.0983	GCA_001671865.1	17	-	-	2016-06-20	Chromosome
HB_S_GIMBLETTROAD_14	12.1193	GCA_001671895.1	17	-	-	2016-06-20	Chromosome
T63	12.1159	GCA_001671925.1	17	-	-	2016-06-20	Chromosome
HB_S_BILANCHER_6	12.1269	GCA_001671955.1	17	-	-	2016-06-20	Chromosome
HB_S_GIMBLETTROAD_16	12.134	GCA_001671985.1	17	-	-	2016-06-20	Chromosome
S288c	12.2429	GCA_002057635.1	17	-	-	2017-03-21	Chromosome
DBVPG6765	11.8949	GCA_002057805.1	17	-	-	2017-03-21	Chromosome
SK1	12.1479	GCA_002057885.1	17	-	-	2017-03-21	Chromosome
YPS128	11.988	GCA_002057995.1	17	-	-	2017-03-21	Chromosome
UWOPS03-461.4	11.8215	GCA_002058095.1	17	-	-	2017-03-21	Chromosome
Y12	11.964	GCA_002058645.1	17	-	-	2017-03-21	Chromosome
DBVPG6044	12.0391	GCA_002079025.1	17	-	-	2017-03-21	Chromosome
Sigma1278b	11.9459	GCA_000151485.1	16	-	-	2010-04-29	Chromosome
AWRI796	11.7417	GCA_000190195.1	89	5267	3786	2011-02-11	Chromosome

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
Vin13	11.6675	GCA_000190215.1	107	5250	3930	2011-02-11	Chromosome
VL3	11.6622	GCA_000190235.1	84	5295	4025	2011-02-11	Chromosome
FostersB	11.5146	GCA_000190255.1	90	5187	3689	2011-02-11	Chromosome
CEN.PK113-7D	11.9991	GCA_000269885.1	70	5451	5451	2012-06-18	Chromosome
YJSH1	11.5505	GCA_000275665.1	17	-	-	2012-07-02	Chromosome
W303	12.1532	GCA_000292815.1	18	-	-	2012-08-24	Chromosome
R008	11.6014	GCA_000568005.1	32	6074	5222	2013-12-31	Chromosome
P301	11.4871	GCA_000568055.1	41	6075	4767	2013-12-31	Chromosome
P283	11.4081	GCA_000568295.1	34	5266	4400	2013-12-31	Chromosome
R103	11.4897	GCA_000568365.1	73	6025	5045	2013-12-31	Chromosome
UFMG A-905	11.4304	GCA_000733235.3	210	-	-	2014-07-24	Chromosome
GLBRCY22-3	12.3772	GCA_001634645.1	32	6967	6319	2016-04-28	Chromosome
wild007	12.8114	GCA_001735825.1	972	-	-	2016-09-12	Chromosome
wild003	12.0786	GCA_001737145.2	156	-	-	2016-09-12	Chromosome
wild005	12.9322	GCA_001737155.1	1156	-	-	2016-09-12	Chromosome
wild004	11.8841	GCA_001737165.1	64	-	-	2016-09-12	Chromosome
wild006	12.8221	GCA_001737175.1	1050	-	-	2016-09-12	Chromosome
wild002	12.0095	GCA_001737265.1	68	-	-	2016-09-12	Chromosome
wild001	12.1491	GCA_001737275.1	185	-	-	2016-09-12	Chromosome
wine019	12.2035	GCA_001737285.1	216	-	-	2016-09-12	Chromosome
wine018	12.168	GCA_001737295.2	186	-	-	2016-09-12	Chromosome
wine017	12.0749	GCA_001737385.2	148	-	-	2016-09-12	Chromosome
wine016	12.5063	GCA_001737395.2	454	-	-	2016-09-12	Chromosome
wine015	11.8712	GCA_001737415.2	57	-	-	2016-09-12	Chromosome
wine014	11.9019	GCA_001737445.2	81	-	-	2016-09-12	Chromosome
wine013	12.1041	GCA_001737505.2	169	-	-	2016-09-12	Chromosome
wine012	12.1565	GCA_001737515.2	302	-	-	2016-09-12	Chromosome
wine011	12.1541	GCA_001737535.2	103	-	-	2016-09-12	Chromosome
wine010	12.3265	GCA_001737545.2	220	-	-	2016-09-12	Chromosome
wine009	12.2287	GCA_001737625.2	98	-	-	2016-09-12	Chromosome
wine008	11.8436	GCA_001737635.2	200	-	-	2016-09-12	Chromosome
wine007	12.081	GCA_001737645.2	190	-	-	2016-09-12	Chromosome
wine006	12.0275	GCA_001737655.2	100	-	-	2016-09-12	Chromosome
wine005	11.891	GCA_001737745.2	112	-	-	2016-09-12	Chromosome
wine004	12.0356	GCA_001737755.2	75	-	-	2016-09-12	Chromosome
wine003	12.2394	GCA_001737765.2	122	-	-	2016-09-12	Chromosome
wine002	12.0617	GCA_001737785.2	58	-	-	2016-09-12	Chromosome
wine001	11.9942	GCA_001737865.2	86	-	-	2016-09-12	Chromosome
spirits011	12.0634	GCA_001737875.1	177	-	-	2016-09-12	Chromosome
spirits010	12.3951	GCA_001737885.1	500	-	-	2016-09-12	Chromosome
spirits009	12.3136	GCA_001737895.1	485	-	-	2016-09-12	Chromosome

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
spirits008	12.4098	GCA_001737985.1	587	-	-	2016-09-12	Chromosome
spirits007	12.3584	GCA_001737995.1	325	-	-	2016-09-12	Chromosome
spirits006	12.0773	GCA_001738005.1	112	-	-	2016-09-12	Chromosome
spirits005	12.0994	GCA_001738015.1	263	-	-	2016-09-12	Chromosome
spirits004	12.2543	GCA_001738105.1	268	-	-	2016-09-12	Chromosome
spirits003	12.2977	GCA_001738115.1	415	-	-	2016-09-12	Chromosome
spirits002	12.0836	GCA_001738125.1	84	-	-	2016-09-12	Chromosome
spirits001	13.2505	GCA_001738135.2	1476	-	-	2016-09-12	Chromosome
sake007	11.8735	GCA_001738225.1	123	-	-	2016-09-12	Chromosome
sake006	11.789	GCA_001738235.1	93	-	-	2016-09-12	Chromosome
sake005	11.8583	GCA_001738255.1	87	-	-	2016-09-12	Chromosome
sake004	11.8898	GCA_001738265.1	117	-	-	2016-09-12	Chromosome
sake003	11.9472	GCA_001738345.1	85	-	-	2016-09-12	Chromosome
sake002	12.0684	GCA_001738355.1	89	-	-	2016-09-12	Chromosome
sake001	11.8606	GCA_001738375.1	115	-	-	2016-09-12	Chromosome
laboratory002	12.2386	GCA_001738395.1	76	-	-	2016-09-12	Chromosome
laboratory001	12.0918	GCA_001738465.1	91	-	-	2016-09-12	Chromosome
bread004	12.3816	GCA_001738485.1	358	-	-	2016-09-12	Chromosome
bread003	12.3435	GCA_001738495.1	360	-	-	2016-09-12	Chromosome
bread002	12.3189	GCA_001738515.1	384	-	-	2016-09-12	Chromosome
bread001	12.3625	GCA_001738585.1	400	-	-	2016-09-12	Chromosome
bioethanol005	12.5131	GCA_001738595.1	299	-	-	2016-09-12	Chromosome
bioethanol004	11.9136	GCA_001738605.1	124	-	-	2016-09-12	Chromosome
bioethanol003	11.8691	GCA_001738615.1	93	-	-	2016-09-12	Chromosome
bioethanol002	12.4993	GCA_001738705.1	414	-	-	2016-09-12	Chromosome
bioethanol001	11.8844	GCA_001738715.1	124	-	-	2016-09-12	Chromosome
beer102	12.168	GCA_001738725.1	395	-	-	2016-09-12	Chromosome
beer101	12.3586	GCA_001738755.1	588	-	-	2016-09-12	Chromosome
beer100	12.3493	GCA_001738825.1	638	-	-	2016-09-12	Chromosome
beer099	12.4921	GCA_001738835.1	725	-	-	2016-09-12	Chromosome
beer098	12.1685	GCA_001738845.1	477	-	-	2016-09-12	Chromosome
beer097	12.8081	GCA_001738855.1	898	-	-	2016-09-12	Chromosome
beer096	12.9195	GCA_001738945.1	1012	-	-	2016-09-12	Chromosome
beer095	12.7425	GCA_001738955.1	787	-	-	2016-09-12	Chromosome
beer094	12.3048	GCA_001738965.1	517	-	-	2016-09-12	Chromosome
beer093	12.1413	GCA_001738985.1	347	-	-	2016-09-12	Chromosome
beer092	12.3363	GCA_001739065.1	385	-	-	2016-09-12	Chromosome
beer091	12.4008	GCA_001739075.1	501	-	-	2016-09-12	Chromosome
beer090	12.2899	GCA_001739095.1	482	-	-	2016-09-12	Chromosome
beer089	12.2866	GCA_001739105.1	505	-	-	2016-09-12	Chromosome
beer088	12.2237	GCA_001739185.1	214	-	-	2016-09-12	Chromosome

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
beer087	12.293	GCA_001739215.1	490	-	-	2016-09-12	Chromosome
beer086	12.3623	GCA_001739225.1	563	-	-	2016-09-12	Chromosome
beer085	12.4792	GCA_001739235.1	533	-	-	2016-09-12	Chromosome
beer084	12.1445	GCA_001739305.1	228	-	-	2016-09-12	Chromosome
beer083	12.1052	GCA_001739335.1	182	-	-	2016-09-12	Chromosome
beer082	12.1824	GCA_001739345.1	468	-	-	2016-09-12	Chromosome
beer081	12.2596	GCA_001739355.1	537	-	-	2016-09-12	Chromosome
beer080	12.3427	GCA_001739425.1	497	-	-	2016-09-12	Chromosome
beer079	12.2055	GCA_001739455.1	406	-	-	2016-09-12	Chromosome
beer078	12.1924	GCA_001739475.1	402	-	-	2016-09-12	Chromosome
beer077	12.1545	GCA_001739495.1	389	-	-	2016-09-12	Chromosome
beer076	12.1878	GCA_001739545.1	400	-	-	2016-09-12	Chromosome
beer075	12.2068	GCA_001739575.1	430	-	-	2016-09-12	Chromosome
beer074	12.1134	GCA_001739595.1	334	-	-	2016-09-12	Chromosome
beer073	12.3263	GCA_001739625.1	542	-	-	2016-09-12	Chromosome
beer072	12.0216	GCA_001739665.1	317	-	-	2016-09-12	Chromosome
beer071	12.0274	GCA_001739695.1	331	-	-	2016-09-12	Chromosome
beer070	11.993	GCA_001739705.1	178	-	-	2016-09-12	Chromosome
beer069	12.2686	GCA_001739735.1	478	-	-	2016-09-12	Chromosome
beer068	11.9066	GCA_001739775.1	157	-	-	2016-09-12	Chromosome
beer067	12.1821	GCA_001739815.1	438	-	-	2016-09-12	Chromosome
beer066	12.174	GCA_001739825.1	443	-	-	2016-09-12	Chromosome
beer065	12.0982	GCA_001739845.1	402	-	-	2016-09-12	Chromosome
beer064	12.2899	GCA_001739865.1	575	-	-	2016-09-12	Chromosome
beer063	12.3228	GCA_001739935.1	422	-	-	2016-09-12	Chromosome
beer062	12.3958	GCA_001739945.1	547	-	-	2016-09-12	Chromosome
beer061	12.2688	GCA_001739965.1	416	-	-	2016-09-12	Chromosome
beer060	12.2727	GCA_001739995.1	495	-	-	2016-09-12	Chromosome
beer059	12.3837	GCA_001740055.1	398	-	-	2016-09-12	Chromosome
beer058	12.1529	GCA_001740075.1	395	-	-	2016-09-12	Chromosome
beer057	12.342	GCA_001740095.1	566	-	-	2016-09-12	Chromosome
beer056	12.3147	GCA_001740115.1	525	-	-	2016-09-12	Chromosome
beer055	12.2529	GCA_001740175.1	497	-	-	2016-09-12	Chromosome
beer054	12.2005	GCA_001740185.1	437	-	-	2016-09-12	Chromosome
beer053	12.1577	GCA_001740205.1	430	-	-	2016-09-12	Chromosome
beer052	12.572	GCA_001740225.1	599	-	-	2016-09-12	Chromosome
beer051	12.1447	GCA_001740295.1	370	-	-	2016-09-12	Chromosome
beer050	12.3103	GCA_001740305.1	536	-	-	2016-09-12	Chromosome
beer049	12.2624	GCA_001740325.1	488	-	-	2016-09-12	Chromosome
beer048	12.2448	GCA_001740345.1	476	-	-	2016-09-12	Chromosome
beer047	12.2528	GCA_001740415.1	466	-	-	2016-09-12	Chromosome

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
beer046	12.2566	GCA_001740425.1	489	-	-	2016-09-12	Chromosome
beer045	12.2869	GCA_001740445.1	517	-	-	2016-09-12	Chromosome
beer044	11.9934	GCA_001740455.1	129	-	-	2016-09-12	Chromosome
beer043	12.1271	GCA_001740535.1	283	-	-	2016-09-12	Chromosome
beer042	12.0086	GCA_001740545.1	192	-	-	2016-09-12	Chromosome
beer041	12.0608	GCA_001740565.1	225	-	-	2016-09-12	Chromosome
beer040	12.0853	GCA_001740575.1	208	-	-	2016-09-12	Chromosome
beer039	12.1854	GCA_001740655.1	261	-	-	2016-09-12	Chromosome
beer038	12.2935	GCA_001740665.1	380	-	-	2016-09-12	Chromosome
beer037	12.1041	GCA_001740675.1	287	-	-	2016-09-12	Chromosome
beer036	12.0924	GCA_001740705.1	267	-	-	2016-09-12	Chromosome
beer035	12.1229	GCA_001740775.1	309	-	-	2016-09-12	Chromosome
beer034	12.0774	GCA_001740785.1	77	-	-	2016-09-12	Chromosome
beer033	11.9716	GCA_001740795.1	113	-	-	2016-09-12	Chromosome
beer032	12.1895	GCA_001740805.1	249	-	-	2016-09-12	Chromosome
beer031	11.992	GCA_001740895.1	212	-	-	2016-09-12	Chromosome
beer030	12.0044	GCA_001740905.1	219	-	-	2016-09-12	Chromosome
beer029	12.3566	GCA_001740915.1	392	-	-	2016-09-12	Chromosome
beer028	12.3636	GCA_001740925.1	373	-	-	2016-09-12	Chromosome
beer027	12.1418	GCA_001741015.1	217	-	-	2016-09-12	Chromosome
beer026	12.1174	GCA_001741025.1	257	-	-	2016-09-12	Chromosome
beer025	12.4194	GCA_001741035.1	432	-	-	2016-09-12	Chromosome
beer024	12.0957	GCA_001741045.1	93	-	-	2016-09-12	Chromosome
beer023	12.3957	GCA_001741135.1	378	-	-	2016-09-12	Chromosome
beer022	12.0439	GCA_001741145.1	228	-	-	2016-09-12	Chromosome
beer021	12.2134	GCA_001741165.1	263	-	-	2016-09-12	Chromosome
beer020	11.9086	GCA_001741195.1	86	-	-	2016-09-12	Chromosome
beer019	12.0947	GCA_001741255.1	221	-	-	2016-09-12	Chromosome
beer018	12.1726	GCA_001741285.1	45	-	-	2016-09-12	Chromosome
beer017	12.0459	GCA_001741295.1	191	-	-	2016-09-12	Chromosome
beer016	12.0556	GCA_001741305.1	177	-	-	2016-09-12	Chromosome
beer015	12.0006	GCA_001741375.1	181	-	-	2016-09-12	Chromosome
beer014	11.9323	GCA_001741405.1	126	-	-	2016-09-12	Chromosome
beer013	12.2893	GCA_001741415.1	298	-	-	2016-09-12	Chromosome
beer012	12.1148	GCA_001741425.1	244	-	-	2016-09-12	Chromosome
beer011	12.1535	GCA_001741495.1	158	-	-	2016-09-12	Chromosome
beer010	12.2331	GCA_001741525.1	284	-	-	2016-09-12	Chromosome
beer009	12.0247	GCA_001741535.1	295	-	-	2016-09-12	Chromosome
beer008	12.1412	GCA_001741555.1	263	-	-	2016-09-12	Chromosome
beer007	11.917	GCA_001741605.1	218	-	-	2016-09-12	Chromosome
beer006	12.5078	GCA_001741685.1	478	-	-	2016-09-12	Chromosome

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
beer005	12.3614	GCA_001741695.1	341	-	-	2016-09-12	Chromosome
beer004	12.2043	GCA_001741705.1	276	-	-	2016-09-12	Chromosome
beer003	12.0585	GCA_001741725.1	92	-	-	2016-09-12	Chromosome
beer001	11.9587	GCA_001741805.1	247	-	-	2016-09-12	Chromosome
beer002	12.1359	GCA_001741815.1	195	-	-	2016-09-12	Chromosome
W303	12.4334	GCA_002163515.1	21	-	-	2017-06-01	Chromosome
DBVPG6044	11.7774	GCA_002271905.1	23	-	-	2017-08-11	Chromosome
Y12	11.688	GCA_002271945.1	23	-	-	2017-08-11	Chromosome
YJM789	11.991	GCA_000181435.1	258	5903	5902	2005-01-06	Scaffold
JAY291	11.5381	GCA_000182315.2	453	5594	5197	2009-08-20	Scaffold
Kyokai no. 7	12.3892	GCA_000260735.1	522	6592	5795	2011-09-22	Scaffold
Lalvin QA23	11.8041	GCA_000325965.1	185	5336	4022	2011-02-11	Scaffold
RM11-1a	11.737	GCA_000149365.1	17	5960	5377	2005-03-16	Scaffold
Y10	14.2673	GCA_000192375.1	1157	-	-	2011-03-11	Scaffold
CBS 7960	12.9079	GCA_000192455.1	1402	-	-	2011-03-11	Scaffold
CLIB324	13.4362	GCA_000192495.1	998	-	-	2011-03-11	Scaffold
YJM269	12.4093	GCA_000192515.1	451	-	-	2011-03-11	Scaffold
FL100	12.4154	GCA_000192535.1	926	-	-	2011-03-11	Scaffold
CLIB215	12.4385	GCA_000192555.1	583	-	-	2011-03-11	Scaffold
PW5	11.8024	GCA_000209265.1	461	-	-	2011-04-25	Scaffold
UC5	11.7165	GCA_000209285.1	506	-	-	2011-04-25	Scaffold
T73	12.4352	GCA_000209305.1	2689	-	-	2011-04-25	Scaffold
CLIB382	10.2559	GCA_000209345.1	10262	-	-	2011-04-25	Scaffold
T7	11.7814	GCA_000209365.1	212	-	-	2011-04-25	Scaffold
EC9-8	12.2998	GCA_000234495.1	96	-	-	2011-11-04	Scaffold
M3707	11.5097	GCA_000365045.1	42	-	-	2013-04-15	Scaffold
M3836	11.5327	GCA_000365065.1	52	-	-	2013-04-15	Scaffold
M3837	11.5303	GCA_000365085.1	45	-	-	2013-04-15	Scaffold
M3838	11.5716	GCA_000365105.1	34	-	-	2013-04-15	Scaffold
M3839	11.5502	GCA_000365125.1	44	-	-	2013-04-15	Scaffold
NAM34-4C	11.5635	GCA_000508805.2	56	-	-	2013-12-11	Scaffold
IR-2	11.4648	GCA_000508825.2	90	-	-	2013-12-11	Scaffold
M22	10.7768	GCA_000182075.1	6145	-	-	2008-06-16	Contig
YPS163	10.7188	GCA_000182095.1	4725	-	-	2008-06-16	Contig
AWRI1631	11.1769	GCA_000182175.1	2484	5568	5451	2008-09-30	Contig
ZTW1	11.4148	GCA_000308935.1	33	-	-	2012-10-31	Contig
FostersO	11.6849	GCA_000326005.1	222	5207	3575	2011-02-11	Contig
NY1308	11.5142	GCA_000416405.1	35	-	-	2013-06-28	Contig
N85	11.9053	GCA_000723645.1	371	-	-	2014-06-04	Contig
UCD51	11.2113	GCA_000756235.1	6048	-	-	2014-09-18	Contig
M5	11.1042	GCA_000756245.1	7718	-	-	2014-09-18	Contig

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
9464	12.5341	GCA_000773925.1	38	-	-	2014-11-10	Contig
EBY.VW4000	11.3861	GCA_000775555.1	354	-	-	2014-11-12	Contig
GSY2239	11.538	GCA_001029995.1	568	-	-	2015-06-19	Contig
W303	13.4896	GCA_001305865.1	109	-	-	2015-10-02	Contig
74-D694	11.3299	GCA_001578265.1	1513	-	-	2016-03-02	Contig
15V-P4	11.6655	GCA_001578275.1	1162	-	-	2016-03-02	Contig
1B-D1606	11.5669	GCA_001578285.1	479	-	-	2016-03-02	Contig
25-25-2V-P3982	11.6135	GCA_001578295.1	890	-	-	2016-03-02	Contig
6P-33G-D373	10.0128	GCA_001578385.1	3038	-	-	2016-03-02	Contig
MT1	11.6209	GCA_001584535.1	125	-	-	2015-12-07	Contig
Lalvin L2056	13.8543	GCA_001611835.1	4677	-	-	2016-04-06	Contig
131	12.0056	GCA_001983315.1	236	5445	5445	2017-02-02	Contig
SRCM100587	11.9442	GCA_002214855.1	55	-	-	2017-07-05	Contig
YS9	11.6064	GCA_000766165.2	851	-	-	2014-10-12	Scaffold
YPS163	11.6469	GCA_000766175.2	607	-	-	2014-10-12	Scaffold
YJM339	11.6247	GCA_000766185.2	554	-	-	2014-10-12	Scaffold
YPS128	11.5563	GCA_000766195.2	670	-	-	2014-10-12	Scaffold
Y55	11.645	GCA_000766245.2	413	-	-	2014-10-12	Scaffold
g833-1B	11.6172	GCA_000766265.2	451	-	-	2014-10-12	Scaffold
DBVPG6044	11.6177	GCA_000766275.2	624	-	-	2014-10-12	Scaffold
RedStar	11.8442	GCA_000766305.2	577	-	-	2014-10-12	Scaffold
BC187	11.508	GCA_000766315.2	614	-	-	2014-10-12	Scaffold
YPH499	11.668	GCA_000766375.2	344	-	-	2014-10-12	Scaffold
FY1679	11.6368	GCA_000766395.2	388	-	-	2014-10-12	Scaffold
10560-6B	11.5786	GCA_000766415.2	389	-	-	2014-10-12	Scaffold
BY4742	11.6149	GCA_000766435.2	418	-	-	2014-10-12	Scaffold
JK9-3d	11.6031	GCA_000766455.2	431	-	-	2014-10-12	Scaffold
W303	11.6336	GCA_000766475.2	415	-	-	2014-10-12	Scaffold
CEN.PK2-1Ca	11.5904	GCA_000766495.2	389	-	-	2014-10-12	Scaffold
FL100	11.597	GCA_000766515.2	402	-	-	2014-10-12	Scaffold
X2180-1A	11.6271	GCA_000766535.2	409	-	-	2014-10-12	Scaffold
D273-10B	11.6473	GCA_000766555.2	403	-	-	2014-10-12	Scaffold
BY4741	11.6166	GCA_000766575.2	397	-	-	2014-10-12	Scaffold
SEY6210	11.6059	GCA_000766595.2	366	-	-	2014-10-12	Scaffold
L1528	11.6202	GCA_000766615.2	538	-	-	2014-10-12	Scaffold
RM11-1A	11.533	GCA_000766635.2	325	-	-	2014-10-12	Scaffold
K11	11.5047	GCA_000767965.1	480	-	-	2014-10-12	Scaffold
UWOPS05_217_3	11.3773	GCA_000768095.1	1348	-	-	2014-10-12	Scaffold
4124-S60	11.7612	GCA_001006245.1	213	-	-	2015-05-12	Scaffold
ISO12	11.3569	GCA_001078085.1	361	-	-	2015-07-15	Scaffold
Ethanol Red	11.4923	GCA_001078105.1	218	-	-	2015-07-15	Scaffold

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
NRRL Y-12632	11.4186	GCA_001282415.1	236	-	-	2015-09-14	Scaffold
GUJ105	11.5538	GCA_001750485.1	201	-	-	2016-09-30	Scaffold
BG1	11.6932	GCA_001932575.1	213	-	-	2017-01-04	Scaffold
VTT A-81062	11.9265	GCA_001937245.1	40	-	-	2017-01-05	Scaffold
SK1	11.5481	GCA_002250225.1	495	-	-	2017-08-14	Scaffold
	11.3776	GCA_900099165.1	723	-	-	2017-02-08	Scaffold
	11.5978	GCA_900099195.1	2159	-	-	2017-02-08	Scaffold
	11.5611	GCA_900099225.1	1284	-	-	2017-02-08	Scaffold
	11.3267	GCA_900099255.1	570	-	-	2017-02-08	Scaffold
	11.3903	GCA_900099265.1	675	-	-	2017-02-08	Scaffold
	11.5071	GCA_900099275.1	1218	-	-	2017-02-08	Scaffold
	11.4532	GCA_900099285.1	909	-	-	2017-02-08	Scaffold
	11.5283	GCA_900099305.1	1022	-	-	2017-02-08	Scaffold
	11.5435	GCA_900099345.1	1124	-	-	2017-02-08	Scaffold
	11.5678	GCA_900099355.1	1556	-	-	2017-02-08	Scaffold
	11.4172	GCA_900099375.1	604	-	-	2017-02-08	Scaffold
	11.7509	GCA_900099395.1	2954	-	-	2017-02-08	Scaffold
	11.5409	GCA_900099425.1	1080	-	-	2017-02-08	Scaffold
	11.5109	GCA_900099445.1	1117	-	-	2017-02-08	Scaffold
	12.6297	GCA_900177835.1	11496	-	-	2017-04-25	Scaffold
	11.6042	GCA_900177845.1	2975	-	-	2017-04-25	Scaffold
	11.6101	GCA_900177855.1	3479	-	-	2017-04-25	Scaffold
	13.1666	GCA_900177865.1	7513	-	-	2017-04-25	Scaffold
	12.6982	GCA_900177875.1	23151	-	-	2017-04-25	Scaffold
	13.0945	GCA_900177885.1	11280	-	-	2017-04-26	Scaffold
	11.7512	GCA_900177905.1	4892	-	-	2017-04-25	Scaffold
	12.5048	GCA_900177915.1	4708	-	-	2017-04-25	Scaffold
	12.0463	GCA_900177925.1	4783	-	-	2017-04-25	Scaffold
	11.6426	GCA_900177935.1	2167	-	-	2017-04-25	Scaffold
	12.8748	GCA_900177945.1	24998	-	-	2017-04-25	Scaffold
	12.8759	GCA_900177975.1	7075	-	-	2017-04-25	Scaffold
	11.5154	GCA_900177985.1	2056	-	-	2017-04-25	Scaffold
	11.5849	GCA_900177995.1	1509	-	-	2017-04-25	Scaffold
	11.7185	GCA_900178015.1	3001	-	-	2017-04-25	Scaffold
	12.6994	GCA_900178025.1	7290	-	-	2017-04-25	Scaffold
	13.8535	GCA_900178035.1	8158	-	-	2017-04-25	Scaffold
	11.5139	GCA_900178045.1	1441	-	-	2017-04-25	Scaffold
	11.5691	GCA_900178065.1	1144	-	-	2017-04-26	Scaffold
	12.5459	GCA_900178075.1	6246	-	-	2017-04-25	Scaffold
MCN1500_3C	11.6136	GCA_900178105.1	633	-	-	2017-04-21	Scaffold
MCN1500_10C	12.0676	GCA_900178125.1	1926	-	-	2017-04-21	Scaffold

Strain	Size (Mb)	Assembly	Scaffolds	Genes	Proteins	Release Date	Level
MUCL42908	11.7839	GCA_900178135.1	1876	-	-	2017-04-21	Scaffold
CLIB560	12.2678	GCA_900178145.1	2876	-	-	2017-04-21	Scaffold
Lava32_6	11.8578	GCA_900178155.1	1296	-	-	2017-04-21	Scaffold
N15_4	11.8751	GCA_900178165.1	1545	-	-	2017-04-21	Scaffold
NRRY1791	11.97	GCA_900178175.1	1892	-	-	2017-04-21	Scaffold
309	12.8795	GCA_900178185.1	4008	-	-	2017-04-21	Scaffold
CLIB215_3B	12.4981	GCA_900178195.1	2577	-	-	2017-04-21	Scaffold
CLIB219	11.7429	GCA_900178205.1	766	-	-	2017-04-21	Scaffold
ZP611	11.7038	GCA_900178215.1	934	-	-	2017-04-21	Scaffold
Lava38_1	11.761	GCA_900178225.1	645	-	-	2017-04-21	Scaffold
TL229_alpha	11.517	GCA_900178235.1	874	-	-	2017-04-21	Scaffold
390_D2	11.8084	GCA_900178245.1	995	-	-	2017-04-21	Scaffold
YQ5	11.9169	GCA_900178255.1	1675	-	-	2017-04-21	Scaffold
6464	12.5658	GCA_900178265.1	2475	-	-	2017-04-21	Scaffold
Lava32_15	11.6351	GCA_900178275.1	583	-	-	2017-04-21	Scaffold
EDV493	11.7647	GCA_900178285.1	965	-	-	2017-04-21	Scaffold
MJ73	12.1058	GCA_900178295.1	2127	-	-	2017-04-21	Scaffold
CBS7957	12.5576	GCA_900178305.1	3318	-	-	2017-04-21	Scaffold
376	12.3412	GCA_900178315.1	2667	-	-	2017-04-21	Scaffold
OakArd11_2_2	11.7657	GCA_900178325.1	638	-	-	2017-04-21	Scaffold
ZP1050	11.8042	GCA_900178335.1	772	-	-	2017-04-21	Scaffold
OakRom3_2	11.7285	GCA_900178345.1	553	-	-	2017-04-21	Scaffold
CBS7959	11.7604	GCA_900178355.1	1182	-	-	2017-04-21	Scaffold
CLIB215	12.6717	GCA_900178365.1	6295	-	-	2017-04-21	Scaffold
ZP851	11.7214	GCA_900178375.1	515	-	-	2017-04-21	Scaffold
OakBod21_1	11.7583	GCA_900178385.1	464	-	-	2017-04-21	Scaffold
NRRLY1545	12.0029	GCA_900178395.1	2388	-	-	2017-04-21	Scaffold
OakGri7_1	11.795	GCA_900178405.1	759	-	-	2017-04-21	Scaffold
VPDN_Fino	11.5745	GCA_900178415.1	924	-	-	2017-04-21	Scaffold
CBS1171	12.1625	GCA_900178425.1	2430	-	-	2017-04-21	Scaffold
KS11	12.0671	GCA_900178435.1	2718	-	-	2017-04-21	Scaffold
YA3	11.6276	GCA_900178445.1	490	-	-	2017-04-21	Scaffold
460	12.5711	GCA_900178455.1	3297	-	-	2017-04-21	Scaffold
	12.0392	GCA_900178465.1	1925	-	-	2017-04-21	Scaffold
ZP848	11.7086	GCA_900178475.1	590	-	-	2017-04-21	Scaffold
VKMY373	11.8194	GCA_900178485.1	2013	-	-	2017-04-21	Scaffold
245	11.7081	GCA_900178495.1	1028	-	-	2017-04-21	Scaffold
M2ONO800_1A	11.8189	GCA_900178505.1	799	-	-	2017-04-21	Scaffold

B. Sequence of novel ORFs

M6

ATGCCAGCCACTGGGGCAACCATTAGGCATGCTGAGCAATTCATAGACGAGTCAGTTGGTTT
TACATTTGGTTGGATCTCTACATACTCATCACTAATGCCTGGTGAGTTATCGGCAACAGCAGT
TATCATGACTTACTGGACGGATGTAAGCCCCGCAATTTTCATAACTGTATTTGGTGTTCTCTT
TGTGGCGACGAACATTTATACAATTCGATTCTACGGCGAAATTGAATATATATTCGGATGGT
TAAAGGTTTTATTAATCGTTATATTGATCGTCTCAGGATTGGTTATCGATCTTGGCGGAACTA
AAGGACAAGAAAGGTTAGGATTTCATTACTGGAGGGATCCTGGTCCCTTTGCAAATTATTTG
GTGGGAGGACACATAGGGAAGTTTGTTGGATTTTGGGCTGCCATTTCCTCCGTTGTATACTCC
TACTCTGGTATTCAAAATATTGCCATTCTCGCTGGCGAGACCAAAAACAGCAGGCATGCAAT
TTTTCACGGCGCCAAAAATGTCTTTCTACGCATTATAGTTCTATATTTGGTCACAGTATTCAT
ATTAACGTTGATTGTTCCCTACAATGACAAGTTAATTGCTACAGGAACAGGTACTGCCAGGT
CAAGTCCCTTTGTCATTGCGATGAACAGAGCCGGGATCAAAGTCTTACCTCATATTGTCAAT
GCCTTGATCTTAACATCTGCATGGTCTGCAGGTAATTTGGCAATCATTGAAGGTTCTAGGAA
TTTGTTCTGCTTAGCAACAAAAAATCAAGCACCTAAGATATTTTTGAGGACAAGCAAAAGGG
GTATTCCATATGTTGGTGTGATATTCATTTCGAGCTTTCTACCGTTGGCGTACATGTCGTGCT
CCAAATCGTCTGCCACTGTCTTTGGGTGGTTCCAAGAATTGGTATCTTCAAACACTCTGCTAC
GTTGGATTTTGATTTCGGCAAACCATATTCATATGGACAGGGCTTTGAAAGCTCAGGGATAC
AGTAGGTCTGACCTGCCATATTCCACACGCATTGGCCCTTTTGCCGCTTGGTTCTCCGGTATA
ATGTCGTTTATTTTCTTGCTTACTGGGGGCTTTTACAATTTCATACATGGTCATTTTGATATCG
AATCATTTTTCACCAGGTACTTCATCATTCCATTAGCAATTGGATTATTTACTTTCTGGAAGC
TGTTCAAAAAGACTAGATATCTACGTCCGCATGAAGTCGATCTTGAATCCATATTCGAGGAT
ATCAAGGAAAACCCAGAACATATTGAAGAGTCCAAACCAATTTGGGCAAAATTTTCATTAA
AAAAAGGTTTAAAAAACGAAAAAAAAGTCTAA

M13

ATGAAAGCATTTCTTTGGACCCTTTGCTTAAGTTTTTTTGCCGAAGCGTTGCAGGCAATCTCA
GTATCTAAGCAAGCTAAGCTTGGCGACATTTCGTATTATGTTCCTGATGTTCCAGAACTGAC
GATTGATGATAGATTAATTTCGAAGCAATTCGGAAGTTTAGGCCACTCGCTACTCTTTCCATT
GACCGTAATTAACCACTCAGGTAGTCTGGACTATCCAGCGGTTAAATCTATTTCTGATAACTT
TTCTTCGTCTGATGATGTTTACCAGGACTATTTCCTGGAAACTATTTTGGTTCAATCTTCAGG
TGCTAGCGCCACGTTATGCACAAACGCTTCTATTTCTTCCCTCAATGTATCCTCCGTTCTGTC
ATTGGACGGCTCTATTCTACCAAATGGTCCATACTTTGGTTCGTATTTGGATGGTAAATTGAG
TATCTACAAGGCTTACCGTCTATATGCAGACACGTATGCCGCATTCCAATCAGGTATCGTTCC
CTCTGATGAGGACGCTAATTCCTTTATGGTTCTACCAGGAGGTGTTGCTGTTGCGCACGCAC
AGACCATTGCGGTGCCTTCTAAGTTGTACTATACAGTTACGAAAGAACAACCACTAGCGGGT
TACCGTGTTGCTATTAAGGACCTGTATGATATCGCCGGTGTCAAAACTGGTGGATCCTCCAG
GACCTATTATGATGTTTACGGTGAGGCTAATATGACCTGTCCTTCAGTTCAGCGTTTAATTGA
CATGGGTGCTGTTATTGTAGGAAAGCTAAAGCTAACTCAGTTTGCAAATGGTGAGACACCAA
CAGCAGATTATGTCGATTACCATGCACCATTTAATCCCAGAGGTGATGGATATCAGTCCCCT
TCTTCTTCATCATGTGGGTCTGGTGCAGCAGAAGCTGCTTATGATTGGTTGGACTTCACAGTT
GGATCCGATACTGGCTGTTCGGTGAGATGTCCTGCTGGGGCCCAAGGTCTATATGGGTTGAG
ACCTACTTTTGATGCCATCTCCTTAGATGGTATTATTCCGATGAGTGATATCATGGACACTGC
TGGTTACTTTTCAAGAGATCCTGAGTTATTCAGAGTGTTCGGTGAAGCCTGGTACGGCGAGA
ATGAAAACATTTCTAAGAGCTACACTTCGTTCCCCAACACTGTTTACACTTTCGACATCAAG
GAAGAGCAGGTGGGGTTCACTCAGAGTCGTGCTAGTCCTGAAGCACTTGAGATGTTCAACA
AGTTTGTCAACGATGTTGTCAATTTTGTGAACGGTACCAACCCTAAGTTAGATGTTTATTCCA
AATTTGAAGAAGATACGGGTCAAAGCCTTACCGATGTATCCAATAGTACGTGGTCTGGTCTA
GCAGGGTACTACCAATATGTGAACATCTGGCAACCGTTCGCCAAGGACTATCAAGAAGCGTT
CGATGGTGATACTCCCTTTTTAGACCCTATTCCCAAATTCAGATGGGACTGGGCTTACTTCAA

TTTTACTGAAACCGGCTATGAGCGGGCGGTAGCAAACAAACAGTTGTTCGACCAATGGTGG
AACACCAACGTCACCACTCGGGACCCAGAAACGTGTTCGAACTCCTTGTATTTGTTTTTATCA
AGTATTGGCACGGTTAACTACAGAGACACCTATAGATCAGCGCCTTCTGCTTCTTCGATATC
CGGTTTTGCTGATATGATGATTTCATCTTTTGCACGTACGCCTGAAGCTATTGTTCCTCTTGGT
GAAGTGCCATACAATTCTAGAATCACCTTGACAGAAAAGTATTTGCCAGTAACTGCGGCTAT
TGGAGCTGCGCCAGGTTGCGATTTTATGGTCCTGGATTTAATAGACAAACTGAGTGAGGCTG
GAATTATTGGAGGAGTTGCCACAGGTCCAAGGATGTTTCCTTCAAAAAATTGA

M14

ATGTGGGCAATGGTAAAGTCCCCACTTATCATTGGTGCCGACTTGAATCACTTAAAGGCATC
TTCGTACTCGATCTACAGTCAAGCCTCTGTCATCGCAATTAATCAAGATCCAAAGGGTATTC
CAGCCACAAGAGTCTGGAGATATTATGTTTCAGACACCGATGAATATGGACAAGGTGAAAT
TCAAATGTGGAGTGGTCCGCTTGACAATGGTGACCAAGTGGTTGCTTTATTGAATGGAGGAA
GCGTAGCAAGACCAATGAACACGACCTTGGAAGAGATTTTCTTTGACAGCAATTTGGGTTCA
AAGGAACTGACATCGACTTGGGATATTTACGACTTATGGGCCAACAGAGTTGACAACTCTAC
GGCGTCTGCTATCCTTGAACAGAATAAGGCAGCCACCGGTATTCTCTACAATGCTACAGAGC
AGTCTTATAAAGACGGTTTGTCTAAGAATGATACAAGACTGTTTGGCCAGAAAATTGGTAGT
CTTTCTCCAAATGCTATACTTAACACAACTGTTCCAGCTCATGGTATCGCCTTCTATAGGTTG
AGACCCTCGGCTTAA

M15

ATGTTTGCTTTCTACTTTCTCACCGCATGCACCACTTTGAAGGGTGTTTTCGGAGTTTCTCCG
AGTTACAATGGTCTTGGTCTCACCCCACAGATGGGTTGGGACAGCTGGAACACGTTTGCCTG
CGATGTCAGTGAACAGCTACTTCTAGACACTGCTGATAGAATTTCTGACTTGGGGCTAAAGG
ATATGGGTTACAAGTATGTCATCCTAGATGACTGTTGGTCTAGCGGCAGGGATTCCGACGGT
TTCCTCGTTGCAGACAAGCACAAATTTCCCAACGGTATGGGCCATGTTGCAGACCACCTGCA
TAATAACAGCTTTCTTTTCGGTATGTATTCGTCTGCTGGTGAGTACACCTGTGCTGGGTACCC
TGGGTCTCTGGGGCGTGAGGAAGAAGATGCTCAATTCTTTGCAAATAACCGCGTTGACTACT
TGAAGTATGATAATTGTTACAATAAAGGTCAATTTGGTACACCAGACGTTTCTTACCACCGT
TACAAGGCCATGTCAGATGCTTTGAATAAAACTGGTAGGCCTATTTTCTATTCTCTATGTAAC
TGGGGTCAGGATTTGACATTTTACTGGGGCTCTGGTATCGCCAATTCTTGGAGAATGAGCGG
AGATATTACTGCTGAGTTCACCCGTCCAGATAGCAGATGTCCCTGTGACGGTGACGAATATG
ATTGCAAGTACGCCGGTTTCCATTGTTCTATTATGAATATTCTTAACAAGGCAGCTCCAATGG
GGCAAAATGCAGGTGTTGGTGGTTGGAACGATCTGGACAATCTAGAGGTCTGA

M16

ATGGCTTTGGTTAAACAGGTAAAATTCGGCAATGCAGGATTGAAGGTCTCACCAGTCGTTGT
TGGATGTATGTCATATGGTTCAAAGAGTTGGTCTCCTTGGGTTTTAGATGATAAGGAGCAGG
TATTCGAGATTTTGAAGTACTGCTACGACCGAGGTTTACGTACTTTNNNNNNNNNNNNNNNN
NNNCGAGGTTTACGTACTTTTGACACTGCAGATGTGTACAGTAATGGTAAAAGTGAGCGTTT
ATTGGGAGAGTTCTTGAAGCACTACAACATCAACAGGGAGACCGTTGTAATCTTATCGAAAG
TTCGTTTCGGTGTTGATGAATCTTTGGAATTGCCTCTCAGTGCACTATACACAACCGATGAAC
AAACCGCATTGACTCTTGCCAATCAAGGTGGACTTTCTCGTAAGCACATTCTGGATGGTGTT
GCAAAGTCTGTTGAAAGGTTGGGAACATATATTGACGTTCTACAAATTCACAGGGCTGATCC
GGACACTCCAATGGAAGAAACTATGAAGGCTCTGAACGACGTTATTGAGAGAGGTGATGTN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTGTCAGCT
GGCAGTCGTGTTACAACTTGTTGTACAGAGAGGACGAAAGAGAAGTCATTCCGTTTGCCAAG
AAACATAA

M17

ATGTCCTATCCAGAAAAGTTTAAAGGTATTGGTATTTCCAACCCAGAAGATTGGAAACACCC
CACATTGGTGAGTTTCGACCCAAAACCATTTGCTGACCACGATGTTGATATTGAAATCGAAG
CTTGTGGTATCTGTGGGTCTGATTTCCATGTAGCAGCCGGTGATTGGGGCCCAGTTCCTAAA
AATGAAATTCTTGGCCATGAAGTGATTGGCCATGTTGTTAAGGTAGGGCCAAAGTGCCACAC
CGGCATCAAGGTGGGAGATCGTGTTGGTGTTGGTGCTCAGGTTTTGGCCTGCCTTGAATGTG
ATCGCTGTAAGAGTGACAATGAGCAGTATTGTCCTATTGACCGTGTATGGACTATCATGTCC
CCATATAAGGATGGCTATATTGCACAGGGTGGCTTTGCATCTCATATCCGAGTCCATGAGCA
TTTTGCTATTCCCATACCAGAAAATATTCCAAGTCCATTGGCAGCTCCTTTGCTGTGTGGTGG
TATCACTGTATTTTCACCATTATCGAGGAATGGCTGTGGTCCAGGTAAAAAGGTGGGCATAA
TCGGTGTCGGTGGTATTGGGCACATGGGTATTTTATTGGCCAAAGCTATGGGAGCTGAAGTG
TATGCCTTCTCACGTGGCCAATCGAAAAAGGAGGACTCTTTAAAGCTCGGTGCTGATCACTA
TATTGCCACATCAGAAGATGAAAACTGGTCCGAGAAGTACTTTAATACTCTGGACCTAATTG
TCATTTGTTCATCCTCTTTAACAGGGGTTGATCTCGACAAGGTGGTTAACGTTATGAAGATCG
GAGGCTCTATTGTCTCGATTGCTGCTCCAGATGCGAGCGAAAGATTAGTTTTACAACCATTG
GGGCTTATGGGAGTCTCAATTTCTAGTAGCGGTGTTGGTTCCAGGAAGGAAATTGAGGAGCT
ACTAAAATTGGTTTCTGAAAAAGATATCAAGATTTGGGTGGAAGAACTCCCAATTAGTGAAG
AAGGTGTCAATGAGGCTTTCACAAGAATGGAAAAAGGGGATGTCAAGTATAGATTTACTTT
GGTAGACTATGATAAAGAGTTCTAG

M18

ATGGAGAAAAAATTGGTGAAAAAACCCAGACTGACGTTGGTTTGTGTACAGTGTAAAAAAA
GTAAACGAAAGTGTGATAAATTACGACCTGTTTGCTCACGATGTCAGCAAGCTTTGCTGGAG
TGTACCTACGAGAGTGCCGCTGACCGGTGTGTCGATACGGTAACGGAAAGCTCCAAGAACC
ATATATCTCAAGGGCTTGTAAATAATAAACATATAACACCTTATGAGAGGGACGGCTATAAG
TTATCTTTCGATAGTTCTACAAAGGACCTTGTCAATGTTCCATTGTGTTTATGGAATGTGGAA
GATATGTTAGTAATCTTGGGCTCCATGACGTTTATGGATTATCCATTTGCATCACACAGTCTG
GTCGAGTATGACCTATATCTTAGAGCCCTATGTGGATCTTTGCACGGCATGACGCTCGTTGAT
CTAAGCAGTCGCCTCAATGGCCTGCCTTCTCAAGACTCATTGAAGCGAGTTCTGAGCCCTCT
GCCGTTTATAGAGAAGGTGATACAACGACGAATAGAATATTCCAAGGTAAATCGAGTTCAG
CCACCTGCTCTAGGAATGTCGTATGACGGATGTTCAATAGAAGACGACAACCTAGCTGATGT
TCTGCAAGCACTTGTGGTAGAAATCGAAGGTCTGCCGACGCAGAAAAAAGACTGTGACACG
CTTTTGAAGAGTTTCTACCAAAATATTTACCCATTCCATCCATTCATTGATATAAGTAGTTTT
GAAAATGATCTTGCTATACTGTTTGTGGAAGACAACAACCACAACTGGAAGATCAGTACTGA
TGGTAGAGATATTCGCAGAAAAATGGAAACTTTGTCATTACTCGCTATCATCATGGCAATAA
CTCTGAAACACTCAACCCTGGACGTAGATATTCTTTCCATGGTCAAAACAAGTGCTTCTGAA
ACTGCCAAAAAACTATCGTTGTTGTGTCACAGACTATTATGTCTACTGGATGTGTTTCGTTAT
CCGAATGAGAGTACATTTAGTTGTCTATTGTATTTTTACATTTCAGAACATCTAGATCCTGAC
AGCCCTGAAAGCATAATGCTGCCCACTAAATTACTCGGGCTTAACCACCTCACAAATCTATC
CATGATTCTGGGTCTCCAATACGATCCTTGCAAATACAAACGTATTCGAAATCCGCAACATA
TAAGACGTAGACGCCTACTGTGGTTAGGCGTTCAATCATTAAGATTTCAGATTACACTTCCC
GAAGGTAATAGTGATAAGACGAATAACGAGTACATGGAGATGTTCTTGGCAGACAGTGAGT
CTACAAAGAGTTCATCCAACGACTTTGCAAGCGCTATGGATGAATTTTATGTGCAATTTTCTG
ACATAGCGTGGGAAAAGTACCAGTTTCACGTTTTACTGAACAAGCTAATTTCTAGTTATACT
TCAATAGTAGAAAACCCATTGCTTTACATGGTGTTGGAGAATATCAAAAGGTCAGAAGATTT
CATGCTTGAGCATTTTCCTTTAGATCTAATATACAGCCCCTTGAATGACCCAACATTGAATAC
GGTGCCGTTTAGCAAAGGTAGCATACTTAATATTTGCGACGTTAAAAAGACGGAAGTATTCA
TGACAAATATTGTTGGACGTATATGCATCTTCAACACTTTGGACGTACTGTCACTATATTTCG
AGAAGAAATGTATTATCCACTGGGAACAATACGAAAAGAATTATCATTACTCCACATTAAG
AAGTTTCAATGTATACCTAGAAGTTTCGGGTATGATTTCAGATTATCTTGATACCAGATTTGG
AGGAAACATTCCGCAGCAGTATGAATACATTGTCAATAAACAGGTATGCTTCACCCTTATCA
GAATATGGTTATTTCAATGTAGGATTCTTTTAAGACTTTCATACAAACAAGAGACAGAAATG

AAACGATTCTCTCCGAATGGTACACAGGCGGCAAGTAATGTGACAAGGATTTTAGCAAGAC
TTATTCAGCATATTCGCAACCAAATGGCATACTTAATTGATTTAGCCAGTAAAAGACTTCAA
CAGTGCTATTTTGGTACTTTCCAAGCCGTTGAGATGTTTAGGTATATTGTGTATGTAGTGGAT
GCTGGTAAATTGGTATCGGTAACAAATGAGTTCTGGGAAAGAAATGTTGACCGAAGCAGAA
TACCGCAAAATATACAGCAAACAATAGGCTTAAAATGGGGACTGGAAAGCAAGAATTCCAG
CGCGATCAAACGGAAACTGACAAGCGCTCAAACTTTGGAGAGTTTCAATGAATTGTTATTGG
GTCAGATGGAGGGTGCTGTTCTTGGTACCTCTTTTGGGCAAAAATCTAATACTGTAATGTCC
AGTAACCTTCCTGAAGAGGCCTTCAATGTCAATGAGGAGGAAACATTAAGCCAGCTGTTGG
AAAGCAATTTTGATTTATTCTGGGATTTACTGGGTGAGAACATGAGTGATGTATACTCTTTGT
GA

M19

ATGTCCCTCGTGAAGCAAGTTCGATACGGTAATACGGGTTTGAAAATTTCCCCGATTATTGTT
GGATGTATGTCATATGGCTCAAAGAGTTGGTCTCCTTGGGTTTTAGATGATAAGGAGCAGGT
ATTCGAGATTTTGAAGTACTGCTACGACCGAGGTTTACGTACTTTTGACACTGCAGATGTGT
ACAGTAATGGTAAAAGTGAGCGTTTATTGGGAGAGTTCTTGAAGCACTACAACATCAATAG
GGAGACAGTTGTTATACTCTCTAAGGTTCATTTCCCTGTTGATGAGTCATTCGACGTTCCTCA
ACCTGGGTTTAGTGAACTCACCACATGGCAATCACTGGAGTTAGCGAATCAGAGGGGCTTAT
CTCGCAAGCACATTCTGGATGGTGTTGCAAAGTCTGTTGAAAGGTTGGGAACATATATTGAC
GTCCTACAAATTCACAGGGCTGATCCGGACACTCCAATGGAAGAAACTATGAAGGCTCTGA
ACGACGTTATTGAGAGAGGTGATGTTAGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
CACAATTGGTTCAAATTTGTCAGCTGGCAGTCGTGTTACAACTTGTTGTACAGAGAGGACGA
AAGAGAAGTCATTCCGTTTGCCAAGAAACATAA

M20

ATGTCCAATATGTTCGAGCCCGCACCTGAACCTCCTACTGAGTTAGGAAATCTCAGAATTCT
CTCAAAAACTGCTGGTATCAGAGTTTCCCCACTCGTTCTAGGAGGAATGTCGATTGGTGACG
CATGGTCAGAGGTCATGGGGTCAATGGACAAGAAACACGCTTTTCAGTTGCTTGATGCTTTT
TATGATGCAGGTGGGAACTTCGTTGATACTGCAAACAATTATCAAAACGAGCAATCGGAAA
CATGGATTGGTGAGTGGATGGACTCGAGAAAGCTACGTGACCAAATGGTTATTGCTACCAA
ATATACACATGACTATAAGTGGTTTGAGGTAGGCAAAGGGAAGAGTGCCAATTTTTGCGGC
AATCACAAGCGTAGTTTGCATATGAGTGTGAGAGATTCGCTTTGCAAATTGAAAACTGACTG
GATTGATATCCTTTATGTTCACTGGTGGGATTATATGACTTCGATCGAAGAAGTTATGGATA
GTTTACACATTCTTGTCCAACAGGGCAAGGTCCTCTATTTAGGTGTGTCAGATACTCCTGCCT
GGGTTGTCTCTGCGGCAAATTATTACGCTACGTCTCACGGTAAAACTCCTTTCAGCGTCTATC
AAGGTAAGTGGAATATATTGAATAGAGATTTTGAGCGTGAAATAATTCCAATGGCGAGACA
TTTTGGTATGGCACTTGTTCCCTGGGATGTGATCGGAGGCGGTAGATTCCAGAGTAAGAAAG
CTATAGAAGAACGGAAGAAGGCTGGAGAGGGTTTGCGCAGTTTTACTGGTGCTTCTGAACA
GACAGAATTGGAAGTCAAAATTAGTGAAGCATTGTTTAAAGTTGCTGAGGAACATGGTACT
GAATCTGTCACTGCTATTGCTATCGCCTATGTCCGCTCCAAAGCAAAAAATGTTTTCCCATTG
GTTGGAGGAAGAAAAATTGAACACCTCAAACAGAACATTGAGGCTCTAAGTATCAAGTTGA
CGCCAGAACAGATAAAGTTCCTAGAAAGCATCGTTCCATTTGATGTTGGGTTCCCCAGTAAT
TTTATCGGGGAGGACCCTGCAGTTACTAAAGTAGCTCCACCTCTCACAGCAATGTCTGCCAG
GATTGTTTTTGATTGA

M21

ATGGCATATCAAGACGAAGTCAAAGGGCCATCTATAACGTCGCAGAAGCAGATTGACACTT
CGAGCATTTCTATCAATGACCCTATATTATTGTTAGATGATAATGAATTGAAGCAGAAAAAG
AGAGTACTAATGAAAATAGATCTAGTAATACTTCCCTTCCTAACCTGCTGTGTGTTCCTACAA
ATCTTAGATAAGAACTCCCTGTCATACGCTGCAATCTATGATTTGCGTAAAGACTTGCATTTG
GAGGGACAACAATACAGCTGGTTTGCAACGATCTTTTACATTAGTTACTTGCTTATGCAAAT

CCCCGTTTTTTTCCTGCTGCCAAAGATCAAAAATTTGACCGGTTTCATGTCATTATTGCTAAC
GTGTTGGGGTATTCTAGTTATGCTTCTTGCTGCATGCAAGAACTTTGCTGGGTTTGCAACAGT
CAGGTTTTTTCTTGGAATGTTCGAAGCTGCTCTTCAACCTTGTTGTATTATTGTTTCAACAAG
CTGGTACACTAAAAGAGAGCAACCGTTGAGAATGGCAATGTGGTCCAACACCTTTGCAGGT
ATATTCAATGGTATTTTTGGATATGCATTTGGTCACTGGAATGCAAAACTACATATCTGGCA
GTATATGTTCATCGTCTACGGAGCGGTCACAGTTGTGTTCGGCATATTCACATACTTTGTTAT
CCCTGCAGACATCGAGTCTGCTTGGTTTCTGAATAAAGAAGAAAAGAAAATTGCCACTATGA
GAATTGCCACTAATCAGACGGGCATCAAATTCAAGGTAGAAACGCTCAAATGGACTCAAAT
TCTCGAGAGTATTCGTCATGTCAAGTATTGGCTGATAGTCATTTTCATTATCGTTCAGAACTT
TATTAATGCAGGTATCACGAACTTCAATACTTTCATCATAAAAGGCTTCGGATATTCAAACTT
CAGGACTATGCTGTTAGCCACCCCACAGGCAAGTGTCGCAATGGTCGTGTCACTGTTGGCTG
CGGCTGCAACATATTATATCAAAAGAATAAGATGTCTTTTGATATGCATCACAACAGGCGTG
ACAATGACCGGCATAATAATGATATGGAAGATTTCTCCAGAGACACATAGAAACGCCTGTTT
GGCAGCCATTTATATTGCTGGATTTTTCAATACTCCTTACGTTCTGATTCTTTCTTTACTTGCT
TCCAACACCAGTGGAGATACTAAAAAAGCGTTTGCATCCTTCAGTGTTGGTGTTTTCTATGC
ACTAGGAAACTTAATCGGCCCACAATTTTTCCTAAATTATGAAAGTCCAAATACCAAACTGG
TATCAAAGCAATGCTATCAGCCTGTGTTATTATGA

M22

ATGGACGGCTCGACGAAGTATAGTGCGCAGAGCTATGCGGACGATCCCCGTAACGAAAACG
TTCTGATCTCCGTGAATGGAGAGCTTGTTCCTCGAGACCAAGGTTGCGTGTCAGTGTTTGATG
CTGGCTTTGCACTGGGCGACGGTGTGTGGGAAGGTATGCGTCTTATTAATGACAAGATCTTC
GCCAGTTCTGAGCATATAGACCGTCTTTATGCAGGCGCATTGAGTATTCAGCTTAATATTGG
ACTATCAAAGGACGAAATCCTTCGTGAAATCTACAAAACAACAGAGGCAAATGATATGCAT
GATGGTGCGCATATTCGTCTCATGGTCACACGAGGCAAGAAAAAGACTCCTAATCAGGATCC
TCGTTTTGCGTTAGGAGCACCAACTATCGTTATCATTGCTGAATACAAAGCTCCAAACCCAG
CCGTATGGTCCAAAGGTTTACAACTTTTAACATCTACCATTCGCTGCAGCAACCCTGATGTTT
TCGACTTGCATTTGAACTCGCACAGTCGATTGAATTTAATTCAGGCACTTCTTCAGGCCATTA
ATGCGGGCTGTGATGAAGCGTTGATGCTTGATCAACGTGGTTTTGTAGCCAGTTGTAACTCC
ACCAACTTCTTCATTGTATGCAAGGGAGAACTATGGACGTCCACCGGACTGTACAATTTCAA
AGGTATAACTAGAAAAAAAATTATAGAATTGTTCAAGAATAAGGTTGGAGTTGTGCGTGAG
GTAGATTTTACACTTGCAGAGGCTTATTCCGCAGATGAGGCCTTTGTTACAGGAACTTTAGG
CGGTGTCACTCCTGTCACGGTGATAGATGGACGTCCGATAGGGACCGGTCAAGTCGGTGATG
TAACCAGAAAGATCAGCGATATATACAGGGAGTTCATTTCAACTTAG

M23

ATGGACAACTTTAAAATGAATTTTGGGTCTTACAAGCATAGCACTGGAGAGTATTATTTAAG
CAATGTGCGTATGGGTCTTCTAGTGGCAATGTTCAGTATTGGTTGTGCTATAGGAGGTCTTCT
TTTCGCCCGTTTTGCCGATACTTTAGGTAGAAGGCTGGCGATTGTGATCGTGGTGTTGGTATA
TATGGTCGGCGCAATTATTCAGATTAGTTCGAACCATAAATGGTACCAATACTTTATCGGGA
AGATCATTTACGGTCTTGGTGCTGGTGGCTGTTCGGTGTTGTGTCCAATGCTTTTGTCTGAAA
TAGCTCCCAAAGACTTGAGAGGTGGACTTGTTTCATTGTACCAATTGAACATGACCTTTGGT
ATTTTCTTAGGTTACTGTAGTGTTTATGGTACGAGAAAATACGACAACACAGCACAATGGAG
AGTGCCTGTTGGGCTTTGCTTTTTATGGGCCCTAATTATCATCATTGGTATGTTATTGGTGCC
AGAATCCCCAAGATACTTGATTGAACAGGAAAACCATGAGCAGGCACGTATTTCAATCGCC
AAAATCAACAAGGTTTCGGCAGAGGACCCATGGGTGCTTAAACAGGTTGATGAAATCAGCG
TCGGTGTGCTTGCCCAAAAAGAGCTAGGAGATGCTTCTTGGAAGGAACTGTTCTCAGTCAAA
ACAAAGGTCTTCCAACGTTTGATTACAGGTATTCTCGTGCAAACCTTCTTGCAACTTACCGGT
GAAAACTACTTTTTCTTTTACGGGACTACCATTTTCAAATCAGTCGGGCTTACTGATGGTTTT
GAGACCTCGATTATCCTGGGTACAGTGAATTTCTTTTCCACCATTATCGCCGTTATGGTCGTA
GACAAAATTGGCCGTCGTAAATGTCTGTTGTTCGGGGCTGCTGGGATGATGGCTTGTATGGT

TATATTCGCAAGTATCGGGGTGAAAAGCCTTTACCCTCATGGCGACAATGGTCCCTCTTCGA
AAGGTGCAGGTAATGCGATGATTGTATTCACCTGTTTCTACATATTCTGCTTTGCAACGACAT
GGGCTCCTGTTGCATATATTGTGGTTGCCGAGTCGTTCCCCTCGAAGGTCAAATCTAGAGCC
ATGTCGATTTCGACTGCATTCAACTGGTTGTGGCAATTTTTGATTGGTTTCTTCACGCCATTC
ATCACGGGATCTATCCACTTCTACTATGGTTATGTGTTTGTAGGTTGTTTGGTTGCTATGTTTT
TGTACGTCTTCTTCTTCTTACCAGAAACAATTGGTCTATCTTTGGAGGAAATCCAATTACTAT
ATGAAGAAGGTGTAAAGCCATGGAAATCTGCATCTTGGGTTCCTCCTTCGAAGAGAGGAATT
CCTGCTGAGGAAAGAAAGACTGAGAAAAAAGATTGGAAGAAATTTTTGAAGTTCCCAAAGG
GTTCAAATTGA

M25
ATGTTTGCGTCACCCTTTCTGAGGATCCGGCCCTCAATCAGAACGTATTCACCAGCGGACTC
GAAGTGTTCGCAGCTAAATGCTCTTTTGTTTTCATTAGACGTCTGTGGCGCTACCTCCCCCTG
TTGGGTCTGTACTGCTTCTTTCATATCGCCTCCAAATGAGGATGCTGAAGATATGTGCTCCAT
TTTAAAATGTTCACTATTTATGTCTGCGTTTGATAATTCACAGGCTAAAGCATTTGCTTCAAG
GTACAGAATTTCTCATCTTTTATACGTGCTAAACTTTTCCATAGTTTCCTTTTTTGTTGAGCAC
ACTACTACAGTGGAAAACTACAACAAAAGTCTGTTGTTTCCGGGCCACATTCTGGGTCACAC
TATGGAGAAAGACAGAAAGAGAAAAAGCTCCTTGAAAAATTGCTTATTTTTCTTCTACATGC
CTTTGCTATTACTCTTGCTGTTTGAATCCGAGAAAAGAAGGAATAAAAGCATAAATAGTCAT
ACCATCCGGTTGTGGGGGCGGGGGTAA

M26
ATGGAGCACATATCTTCAGCATCCTCATTTGGAGGCGATATGAAAGAAGCAGTACAGACCC
AACAGGGGGAGGTAGCGCCACAGACGTCTAATGAAAACAAAAGAGCATTTAGCTGCGAACA
CTTCGAGTCCGCTGGTGAATACGTTCTGATTGAGGGCCGGATCCTCAGAAAGGGTGACGCAA
ACATCTTCTTAGAAGACCTAACTCCCGCTACACACAAAGAGCTACCAAAGCAAGTTGGATTT
GCCAATCCTCTTCCACTCGGGCTGGCCTCTTTCTCGTTTATGTGTTTAACATTGGGTTTGGTTA
ACGCAAGAGTGCGCGGGGTCACAAATCTTTATTTATTGAACGCATCCTTTATTTTTGGTGGTG
CTGTTGTTTTGCTATCAGGACTTCTCTCCTTCTGTGTCGGAGACACATTTTGTATGACGGTTTT
TGGCTCATTTGGAGGCTTTTGGATTAGTTGGGGGTGCCTAAACATCGAACAATTCGGTGTAA
CTAAGGCATATGCTGATGATCCTCAAGCACTACAAAATATACTGGGATTTTACCTTGCCGGA
TGGACCGTATTCAATTTTCTGGTTATGGTCTGTTCCATGAAAAGTACCTGGGGAATATTCCTG
CTGCTATTTTTCCTGGACCTGACGTTTTTAATGTTATGTATCGGTTCTTTCACACAAAGCGTC
AATGCGTCCATGGCAGGAGGATATTTTGGTATATTGAGCAGCTGTTGCGGTTGGTATTCCTT
ATATTGTTCAATTGCGAATAAAGATAGCTCCTACGTCCCTTTAGTCGCATATCCAATGCCAG
GTTCTCAAATTGTATGA

M28
ATGCTCAATCGCTTTAATAAGTTTCAAGCTGCTTTAGCTTTAGCCCTTTATTCACAAAGTGTA
CTAGGTCAGTACTATCCAAATAGCACTACAATTCCAAGCAACAGCTCATCATCTTCTTTTTCA
TCAACCTCATCAAGTTCCTTTTCTATTAGTAGTTCTATTACTCAATCGACTTCATCAACCCCC
GATGTTTCGAGTTCTCTCACTCAATTAACCTCATCTTCGGATGTTTCGAGCTCTATTGCTTCAT
CGTCCTCTCCTGGTTCAGGAGTGTCAAGTTCTGTTACTCAATCGAGTTCAAGTGGCAGTAGTC
CATCTGGTACTGGTTCCACAGCTACTACTCCGTCTGGCACTACTCCATCTGGTACTGGTTCCA
CAGCTACTACACCATCTGGCACTACACCATCTGGCACCGGTTCCACAGCTACTACACCATCT
GGCACTACACCATCTGGTACTGGTTCCACAGCTACTACACCATCTGGCACTACACCATCTGG
CACTGGTTCCACAGCTACTGGCCCATCTGGCACGGGTTCAACAGCTACTACTCCGTCTGGCA
CTACTCCATCTGGCACCGGTTCAACAGCTACTACTCCATCTGGTACTGGTTCCACAGCTACTA
CACCATCTGGCACTACACCATCTGGCACTGGTTCCACAGCTACTGGCCCATCTGGCACGGGT
TCAACAGCTACTACTCCGTCTGGCACTACTCCATCTGGCACCGGTTCAACAGCTACTACTCCA
TCTGGTACTGGTTCCACAGCTACTGGCCCATCTGGCACGGGTTCAACAGCTACTACTCCATCT

GGTACGGGTTCAACAACCACTAGTCCATCGGGCAGTGGTCCATCTGAATCTGGATCATCCGC
CTCTATCACACAATCGGTCTAA

C. Sequence of plasmids

pUG6

GAACGCGGCCGCCAGCTGAAGCTTCGTACGCTGCAGGTCGACAACCCTTAATATAACTTCGT
ATAATGTATGCTATACGAAGTTATTAGGTCTAGAGATCTGTTTAGCTTGCCTCGTCCCCGCCG
GGTCACCCGGCCAGCGACATGGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAG
CTCAGGGGCATGATGTGACTGTCGCCCGTACATTTAGCCCATACATCCCCATGTATAATCATT
TGCATCCATACATTTTGATGGCCGCACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCAGA
CCTGCGAGCAGGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTG
TAGAGAAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAA
ATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACCATGGGTAAGGAAAA
GACTCACGTTTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAAT
GGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGAT
GCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGAT
GGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTA
CTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGCAAAACAGCATTCCAGGTATTA
GAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTT
GCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGC
GCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCT
GGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTC
GTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGT
ATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTG
CCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCC
TGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGTACTGACAAT
AAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTTCTAT
TTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGCGA
AGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATACTG
CTGTCGATTCGATACTAACGCCGCCATCCAGTGTCGAAAACGAGCTCTCGAGAACCCTTAAT
ATAACTTCGTATAATGTATGCTATACGAAGTTATTAGGTGATATCAGATCCACTAGTGGCCT
ATGCGGCCGCGGATCTGCCGGTCTCCCTATAGTGAGTCGTATTAATTTCGATAAGCCAGGTT
AACCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTC
CGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTC
ACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTG
AGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCAT
AGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACC
CGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCT
CAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGT
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGC
GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
GGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGAT
TACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTC
AGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACC
TAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTC
ATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTG

GCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATA
AACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAAC
GTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGC
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAG
CTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTAT
GGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGA
GTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGT
CAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACG
TTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
CTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAA
CAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCA
TACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTG
CCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCAC
GAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCC
CGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGC
GTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTAC
TGAGAGTGCACCATATGGACATATTGTCGTTAGAACGCGGCTACAATTAATACATAACCTTA
TGTATCATACACATACGATTTAGGTGACACTATA

pCW1

TATAGTGTCACCTAAATCGTATGTGTATGATACATAAGGTTATGTATTAATTGTAGCCGCGTT
CTAACGACAATATGTCCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
AGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGC
ATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGT
CATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTC
ATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC
TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATA
AATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTA
TTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAA
AAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGG
TAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCT
GCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATAC
ACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGC
ATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTT
ACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATC
ATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCG
TGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTAC
TTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCA
CTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGT
GGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGT
GCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGAT
TTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACC
AAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGG
ATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCT
ACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCT
TCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTC
AAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCC
AGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC

AGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACAC
CGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCA
GGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCG
ATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTT
TACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC
TGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCG
AGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCC
CGCGCGTTGGCCGATTCATTAATGCAGGTTAACCTGGCTTATCGAAATTAATACGACTCACT
ATAGGGAGACCGGCAGATCCGCGGCCGCATAGGCCACTAGTGGATCTGATATCACCTAATA
ACTTCGTATAGCATACATTATACGAAGTTATATTAAGGGTTCTCGAGAGCTCGTTTTCGACAC
TGGATGGCGGCGTTAGTATCGAATCGACAGCAGTATAGCGACCAGCATTCACATACGATTGA
CGCATGATATTACTTTCTGCGCACTTAACTTCGCATCTGGGCAGATGATGTCGAGGCGAAAA
AAAATATAAATCACGCTAACATTTGATTAAAATAGAACAACTACAATATAAAAAAACTATA
CAAATGACAAGTTCTTGAAAACAAGAATCTTTTTATTGTCAGTACTGATTAGAAAAACTCAT
CGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAA
AGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCT
GGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAA
AAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAA
AAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATC
ACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGAT
CGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAG
CGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTGCC
GGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTC
GGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGC
AACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGAT
AGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCA
TCCATGTTGGAATTTAATCGCGGCCTCGAAACGTGAGTCTTTTCCTTACCCATGGTTGTTTAT
GTTCGGATGTGATGTGAGAACTGTATCCTAGCAAGATTTTAAAAGGAAGTATATGAAAGAA
GAACCTCAGTGGCAAATCCTAACCTTTTATATTTCTCTACAGGGGCGCGGCGTGGGGACAAT
TCAACGCGTCTGTGAGGGGAGCGTTTCCCTGCTCGCAGGTCTGCAGCGAGGAGCCGTAATTT
TTGCTTCGCGCCGTGCGGCCATCAAAATGTATGGATGCAAATGATTATACATGGGGATGTAT
GGGCTAAATGTACGGGCGACAGTCACATCATGCCCCTGAGCTGCGCACGTCAAGACTGTCA
AGGAGGGTATTCTGGGCCTCCATGTCGCTGGCCGGGTGACCCGGCGGGGACGAGGCAAGCT
AAACAGATCTCTAGACCTAATAACTTCGTATAGCATACATTATACGAAGTTATATTAAGGGT
TGTCGACCCTTAATTTTTATTTTAGATTCCTGACTTCAACTCAAGACGCACAGATATTATAAC
ATCTGCATAATAGGCATTTGCAAGAATTACTCGTGAGTAAGGAAAGAGTGAGGAACTATCG
CATACCTGCATTTAAAGATGCCGATTTGGGCGCGAATCCTTTATTTTGGCTTCACCCTCATAC
TATTATCAGGGCCAGAAAAAGGAAGTGTTTCCCTCCTTCTTGAATTGATGTTACCCTCATAA
AGCACGTGGCCTCTTATCGAGAAAGAAATTACCGTCGCTCGTGATTTGTTTGCAAAAAGAAC
AAAACTGAAAAAACCCAGACACGCTCGACTTCCTGTCTTCCTATTGATTGCAGCTTCCAATTT
CGTCACACAACAAGGTCCTAGCGACGGCTCACAGGTTTTGTAACAAGCAATCGAAGGTTCTG
GAATGGCGGGAAAGGGTTTAGTACCACATGCTATGATGCCCACTGTGATCTCCAGAGCAAA
GTTCGTTCGATCGTACTGTTACTCTCTCTCTTTCAAACAGAATTGTCCGAATCGTGTGACAAC
AACAGCCTGTTCTCACACACTCTTTTCTTCTAACCAAGGGGGTGGTTTAGTTTAGTAGAACCT
CGTGAAACTTACATTTACATATATATAAACTTGCATAAATTGGTCAATGCAAGAAATACATA
TTTGGTCTTTTCTAATTCGTAGTTTTTCAAGTTCTTAGATGCTTTCTTTTTCTCTTTTTTACAGA
TCATCAAGGAAGTAATTATCTACTTTTTTACAACAAATATAAAACATCGACCTGCAGCGTACG
AAGCTTCAGCTGGCGGCCGCGTTC

D. Primers used for creation of mutants

Table A.2 Primers used the creation of null and constitutive expression mutants. Primers with overlapping ends for the ORF of interest were used to amplify either the kanMX cassette from the pUG6 plasmid, or the kanMX-PCK1prom cassette from the pCW1 plasmid, for the creation null and constitutive expression mutants, respectively.

ORF	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Template	Purpose
kanMX		CGTGAGTCTTTTCCTTACCCATG	Null mutant	Confirm insertion of cassette
PGK1 prom	GGTTTTGTAACAAGCAATCGAAGG		Constitutive expression mutant	Confirm insertion of cassette
M6	AACATCTTACCTAAGGTTATGTAATTACCGTATATTTCTATTCTATTAAAATAAAATAACGTAGCTGAAGCTTCGTACGC	CTCTTTGTTGGTATTGTCGTAGGCTGTAACCAAATTGCAATAGCCGAGGTTGCCTCTTTCTAGGCCACTAGTGGATCTG	pUG6	Amplify deletion cassette
M6	GCTTGTACGAGGTGTGTTGTTC	GGCTTTTACAATTTCATACATGGTC	Genomic DNA	Confirm null mutant
M13	TTTCTGGTTTCATTCAGTTCAAATGTTTTTATCCATATCTCCAAGTAAACCTCTGAAGCAAGCTGAAGCTTCGTACGC	GTATACATTGCATTAGCAATATTACATACTAGTGACTTTTGAGAAGTGGCTTTTGGTTTCTAGGCCACTAGTGGATCTG	pUG6	Amplify deletion cassette
M13	GCTAACGTAGGATAACAAGCAC	CCTGGTAAACATCATCAGACGAAG	Genomic DNA	Confirm null mutant
M13	TTTCTGGTTTCATTCAGTTCAAATGTTTTTATCCATATCTCCAAGTAAACCTCTGAAGCACTATAGGGAGACCGGCAGATC	GATTGCCTGCAACGCTTCGGCAAAAAAACTTAAGCAAAGGGTCCAAAGAAATGCTTTCATTGTTTTATATTTGTTGTAAAAAGTAGATAATTACTTCC	pCW1	Amplify PGK1 promoter cassette
M14	CATGAACTTCTTAACAATCATTATTTTTTTTTTCTCGTCCTGCTTTGCTCAACATTGAGCAGCTGAAGCTTCGTACGC	GACAATCTAGAGGTCTGAGTCGGTAATTTGACTGACGATGAGGAAAAGGCCCATTTCTCTTAGGCCACTAGTGGATCTG	pUG6	Amplify deletion cassette
M14	GTTCATTGGTCTTGCTACGC	TGCTGAGTTCACCCGTCC	Genomic DNA	Confirm null mutant
M15	CTTTACCATTGCCCACATAGAGAAATGGGCCTTTTCCTCATCGTCAGTCAAATTACCGACAGCTGAAGCTTCGTACGC	AATGTTTTCGATAATTTCTTATTTGTTTCTAGTAGAGCAACGGTAATAAAAATAACAACGTAGGCCACTAGTGGATCTG	pUG6	Amplify deletion cassette
M15	TCACCTTGTCCATATTCATCGG	CTCTATGTAACTGGGGTCAGG	Genomic DNA	Confirm null mutant
M15	AATGTTTTCGATAATTTCTTATTTGTTTCTAGTAGAGCAACGGTAATAAAAATAACAACGCTATAGGGAGACCGGCAGATC	AGAAACTCCGAAAACACCCTTCAAAGTGGTGCATGCGGTGAGAAAGTAGAAAGCAAACATTGTTTTATATTTGTTGTAAAAAGTAGATAATTACTTCC	pCW1	Amplify PGK1 promoter cassette
M16	TCACCCAGAGGCCTTGTTAGTAAACCTCTTGCATTGGGAGACCAAGGAATCAAGGCGATAAGCTGAAGCTTCGTACGC	TGTTACCGCAGCAATTAAGAAACGGCAGTTAAGAAATAAGAATAAAATAAAAAGAAAACTTAGGCCACTAGTGGATCTG	pUG6	Amplify deletion cassette
M16	CCATTACTGTACACATCTGCAGTG	CCGACAAGGATATCAGGGCTG	Genomic DNA	Confirm null mutant
M17	AAAACACACCTAAGTTTTAAAGTACTACGCAAAATCTGAGAAAAAAAAAACAAAAAAAAAAGCTGAAGCTTCGTACGC	AAAAAAAAAAAAAAAAGAGGAAGGAAAAGTTAAATAAATGCAGAAATGCTACTACTTGTCTAGGCCACTAGTGGATCTG	pUG6	Amplify deletion cassette
M17	GCAATTTTCCGAGCCGCAG	CAAAACCTGAGCACCAACACC	Genomic DNA	Confirm null mutant

ORF	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Template	Purpose
M18	CAAGTGCTTTCCGAAGGTGACA AAAGAGCAAATCAGCTCGATAA TACCGTCAAGAGAGCTGAAGCT TCGTACGC	TTAATTGTTAAAAATAATCTTC ATGAGAGTTATGTTATTTGCTTT TCGGCTAAAGCTTTCTAGGCCA CTAGTGGATCTG	pUG6	Amplify deletion cassette
M18	CGTTGACTTCTGGTAGTCCTCG	GATAACTTATAGCCGTCCCTCT C	Genomic DNA	Confirm null mutant
M18	TCTTCAAGTGCTTTCCGAAGGT GACAAAAGAGCAAATCAGCTC GATAATACCGTCAAGAGCTATA GGGAGACCGGCAGATC	TTTTTTACACTGTACACAAACC AACGTCAGTCTGGGTTTTTTCAC CAATTTTTTCTCCATTGTTTTAT ATTTGTTGTAAAAAGTAGATAA TTACTTCC	pCW1	Amplify PGK1 promoter cassette
M19	CCTTTCCATTTGCTAACCGCAA AATTCTAAGTACAAACAAAATA CAACCATATCTTACAGCTGAAG CTTCGTACGC	CACCCCATGGCCTTGTTAGTAA ACCTCTTGCATTGGGAGACCAA GGAATCAAGGCAATATAGGCCA CTAGTGGATCTG	pUG6	Amplify deletion cassette
M19	ACTAAGCAACGAGATATGGTCG	CCATTACTGTACACATCTGCAG TG	Genomic DNA	Confirm null mutant
M20	GTATTATCTTTCTTTCTTCTGCT TCACTTTTGTAAAACAAAACAA CCATTACTTTCGATAGCTGAAG CTTCGTACGC	GCTAGATAGATACAACCGATCA ATGCACAAACATAAACTTCCCT CATCATCCATACTTAGGCCACT AGTGGATCTG	pUG6	Amplify deletion cassette
M20	TGCTATCGCCTATGTCCGC	GACTTACGGCAGACATAGTCC	Genomic DNA	Confirm null mutant
M21	GTCCCTCTCTCTGTTCTCCTTAA TGCATAGAAGAGCGTACAACAT GGCACAAAAAAAGCTGAAGCTT CGTACGC	GATAAGTACAATGGTTGACAGG ACCTACCTGTGACAAAGGCAGT GTACACTTGAAAACTAGGCCAC TAGTGGATCTG	pUG6	Amplify deletion cassette
M21	AGTTACGCCTTTTACTCCCAG	CAACAGGCGTGACAATGACC	Genomic DNA	Confirm null mutant
M21	TTGATAAGTACAATGGTTGACA GGACCTACCTGTGACAAAGGCA GTGTACACTTGAAAACCTATAG GGAGACCGGCAGATC	AGTGTCAATCTGCTTCTGCGAC GTTATAGATGGCCCTTTGACTTC GTCTTGATATGCCATTGTTTTAT ATTTGTTGTAAAAAGTAGATAA TTACTTCC	pCW1	Amplify PGK1 promoter cassette
M22	CTCAAGCAAACTGAAGAAGAA CTTTGATTGACGAGATGTGCAT AATCAGGAACTCAGCTGAAGCT TCGTACGC	GTAGAAAAATTAACCGGCATTT GAAACAAACCTGTTGAAGGCAT CAGCTAATAAAGTCTAGGCCAC TAGTGGATCTG	pUG6	Amplify deletion cassette
M22	GGAAGTGGGAAATCGACCTTAC	AGGGAGAACTATGGACGTCC	Genomic DNA	Confirm null mutant
M23	CTCCTTTGGGGGGTTTCTGCCTG GTTGGGATAGTGGTATTACAGC AGGTTTTATTAACAGCTGAAGC TTCGTACGC	CGTTAGGTAAAATATAATATAC AATGCTGTCTATAAGGACAAGC GTATGCACTCAATGATAGGCCA CTAGTGGATCTG	pUG6	Amplify deletion cassette
M23	AGATATTCGAGACGCTTCTGG	GACCGTAAATGATCTTCCCG	Genomic DNA	Confirm null mutant
M23	ATCTCCTTTGGGGGGTTTCTGCC TGGTTGGGATAGTGGTATTACA GCAGGTTTTATTAACCTATAGG GAGACCGGCAGATC	TAAATAATACTCTCCAGTGCTA TGCTTGTAAGACCCAAAATTCA TTTTAAAGTTGTCCATTGTTTTA TATTTGTTGTAAAAAGTAGATA ATTACTTCC	pCW1	Amplify PGK1 promoter cassette
M25	GGACCAAAACACAATAATGGAT GGAGAACAATAAATTCAATATT AAGTTTAGGTAATAATAGCTGA AGCTTCGTACGC	CAAATCCAACTTGCTTTGGTAG CTCTTTGTGTGTAGCGGGAGTT AGGTCTTCTAAGAAGTAGGCCA CTAGTGGATCTG	pUG6	Amplify deletion cassette
M25	GGAACGGTATGAGTTGCCTC	GCTTCAAGGTACAGAATTTCTC ATC	Genomic DNA	Confirm null mutant
M26	GAAGCAAATGCTTTAGCCTGTG AATTATCAAACGCAGACATAAA TAGTGAACATTTTAAAAGCTGA AGCTTCGTACGC	CAATAAAAAGGCCCAAATAAA ATAGTAATGCATAGCAAAAGAA GAAATAATAGCGCCTCTAGGCC ACTAGTGGATCTG	pUG6	Amplify deletion cassette

ORF	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Template	Purpose
M26	CTGGGGAATATTCCTGCTGC	GTGGTGTATCAATAATGGCAACC	Genomic DNA	Confirm null mutant
M28	GAAACAGTAGAACCGGAGGAG GTTGATTGGGGAATTGAACTTG AGAGACTAGGAGCGGATAGCT GAAGCTTCGTACGC	CAGCGGCAACAGTATCTTATCG TTGCAATTTTCTGGTAATATCTC GCCTGTGTTGAAGAGTAGGCCA CTAGTGGATCTG	pUG6	Amplify deletion cassette
M28	CTGTTGCAATCACAGAACCTG	GCACTGGTTCCACAGCTACTG	Genomic DNA	Confirm null mutant
M28	CAGCGGCAACAGTATCTTATCG TTGCAATTTTCTGGTAATATCTC GCCTGTGTTGAAGAGCTATAGG GAGACCGGCAGATC	ACTTTGTGAATAAAGGGCTAAA GCTAAAGCAGCTTGAAACTTAT TAAAGCGATTGAGCATTGTTTT ATATTTGTTGTAAAAAGTAGAT AATTACTTCC	pCW1	Amplify PGK1 promoter cassette

E. Primers for qRT-PCR

Table A.3 Primers for the qRT-PCR amplification of novel ORFs. Unique primers corresponding the < 200bp fragments of novel ORFs were used for the relative quantification of novel ORFs expressed in wild type Enoferm M2.

ORF	Forward Primer (5' - 3')	Reverse Primer (5' - 3')
M6	ACACTCTGCTACGTTGGATTT	GCCAATGCGTGTGGAATATG
M13	TGTCATTGGACGGCTCTATTC	CAGAGGGAACGATACCTGATTG
M14	AGCGTAGCAAGACCAATGAA	TGTTGGCCCATAAGTCGTAAA
M15	CGTGAGGAAGAAGATGCTCAA	GCCTTGTAACGGTGGTAAGA
M16	CAAAGTCTGTTGAAAGGTTGGG	GTCGTTCAGAGCCTTCATAGTT
M17	AATCGGTGTCGGTGGTATTG	GTGATCAGCACCGAGCTTTA
M18	TACTGTGGTTAGGCGTTCAATC	ACTGTCTGCCAAGAACATCTC
M19	GGTTCATTTCCCTGTTGATGAG	CAGAATGTGCTTGCGAGATAAG
M20	CTAGGAGGAATGTCGATTGGTG	AGTTCCCACCTGCATCATAAA
M21	CCTGCTGTGTGTTCCTACAA	GCTGTATTGTTGTCCCTCCA
M22	GGCACTTCTTCAGGCCATTA	GTCCATAGTTCTCCCTTGCATAC
M23	ACTTGAGAGGTGGACTTGTTTC	CATTGTGCTGTGTTGTCGTATTT
M25	GCGTTTGATAATTCACAGGCTAAA	CCACTGTAGTAGTGTGCTCAAC
M26	TAACTCCCGCTACACACAAAG	GCGTTAACCAAACCCAATGTTA
M28	CCGATGTTTCGAGTTCTCTCAC	CAGAACTTGACACTCCTGAACC
TAF10	GTAGTGGATGATGGGAGTGAAA	TTACTGCATCGGGAATGATAGG

F. Gene Ontology Enrichments for Enoform M2 Transcriptome

Table A.4 YeastMine gene ontology enrichments for different transcript expression profiles during fermentation. Gene lists were produced by categorizing transcripts by their expression pattern in Enoform M2 wild type over three time points in Chardonnay wine fermentation. These lists were analyzed with the YeastMine tool for ontology enrichment and significant (p-value < 0.05, with Holm-Bonferroni correction) results are reported.

		Biological Process			**Cellular Localization**			**Molecular Function**		
Expression Trend	**Number of Genes**	Ontology	p-value	# of genes	Ontology	p-value	# of genes	Ontology	p-value	# of genes
Down-Down	132	organonitrogen compound biosynthetic process	5.78E-08	59	cytosolic ribosome	0.000263779	6	ligase activity	0.001196321	14
		small molecule metabolic process	6.27E-08	48	cytoplasm	0.000919594	116	aminoacyl-tRNA ligase activity	0.005514938	8
		cellular amino acid metabolic process	7.43E-07	25	intracellular part	0.002698351	126	ligase activity, forming carbon-oxygen bonds	0.005514938	8
		oxoacid metabolic process	1.35292E-06	33	intracellular	0.003687824	126	ligase activity, forming aminoacyl-tRNA and related compounds	0.005514938	8
		organic acid metabolic process	1.43486E-06	33	cytosolic part	0.02349245	16			
		organonitrogen compound metabolic process	1.65381E-06	63	ribosome	0.026529598	17			
		carboxylic acid metabolic process	2.23121E-06	32						
		single-organism biosynthetic process	2.62314E-05	40						
		tRNA aminoacylation for protein translation	0.000443652	9						
		single-organism metabolic process	0.00083372	59						
		amino acid activation	0.000901628	9						
		tRNA aminoacylation	0.000901628	9						
		small molecule biosynthetic process	0.000994311	24						
		cytoplasmic translation	0.002159439	17						
		biosynthetic process	0.005590231	81						
		organic substance biosynthetic process	0.017234101	79						
		organic acid biosynthetic process	0.032816699	16						

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		carboxylic acid biosynthetic process	0.032816699	16						
		cellular amino acid biosynthetic process	0.03489691	13						
Down-Null	457	cytoplasmic translation	2.35E-47	87	cytosolic ribosome	8.83E-47	81	structural constituent of ribosome	2.66E-27	79
		ribosome biogenesis	7.93E-22	104	cytosolic part	3.39E-34	82	structural molecule activity	9.74E-18	87
		organonitrogen compound biosynthetic process	5.00E-20	174	ribosomal subunit	4.82E-34	82	snoRNA binding	0.006221918	11
		ribonucleoprotein complex biogenesis	5.48E-17	107	ribosome	8.16E-31	83	oxidoreductase activity, acting on the CH-NH group of donors	0.018416211	8
		organonitrogen compound metabolic process	2.42E-16	188	cytosolic large ribosomal subunit	2.90E-28	48			
		small molecule biosynthetic process	8.13E-15	74	preribosome	6.12E-26	63			
		ribosome assembly	8.55E-15	34	intracellular ribonucleoprotein complex	6.46E-26	135			
		cellular amino acid metabolic process	8.23E-13	60	ribonucleoprotein complex	6.46E-26	135			
		ribosomal small subunit biogenesis	5.57E-12	43	large ribosomal subunit	1.68E-19	49			
		ribosomal large subunit biogenesis	2.17E-11	38	cytosolic small ribosomal subunit	4.85E-18	33			
		alpha-amino acid metabolic process	5.05E-11	48	90S preribosome	2.62E-13	33			
		organic acid biosynthetic process	5.28E-10	48	non-membrane-bounded organelle	2.68E-13	170			
		carboxylic acid biosynthetic process	5.28E-10	48	intracellular non-membrane-bounded organelle	2.68E-13	170			
		alpha-amino acid biosynthetic process	7.17E-10	37	small ribosomal subunit	6.50E-12	33			
		amide biosynthetic process	9.42E-10	117	preribosome, large subunit precursor	5.57E-11	29			

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		small molecule metabolic process	9.56E-10	116	cytosol	4.07591E-06	126			
		cellular amino acid biosynthetic process	1.11E-09	38	cell part	2.73278E-05	429			
		rRNA-containing ribonucleoprotein complex export from nucleus	3.20E-09	25	cell	3.09445E-05	429			
		ncRNA metabolic process	6.70E-09	91	macromolecular complex	0.00072979	200			
		cellular amide metabolic process	1.04E-08	120	small-subunit processome	0.002239031	15			
		carboxylic acid metabolic process	1.35E-08	74	intracellular part	0.00303027	418			
		organic acid metabolic process	1.35E-08	76	intracellular	0.005961699	418			
		translation	2.35E-08	108	nucleolus	0.007682429	50			
		peptide biosynthetic process	3.31E-08	108						
		oxoacid metabolic process	3.42E-08	75						
		rRNA export from nucleus	6.13E-08	13						
		rRNA transport	6.13E-08	13						
		peptide metabolic process	2.36E-07	109						
		maturation of SSU-rRNA	3.61E-07	32						
		ergosterol biosynthetic process	4.05E-07	16						
		phytosteroid biosynthetic process	4.05E-07	16						
		cellular alcohol biosynthetic process	4.05E-07	16						
		cellular lipid biosynthetic process	4.05E-07	16						
		rRNA processing	4.18E-07	62						
		maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	6.04E-07	30						
		secondary alcohol biosynthetic process	7.54E-07	16						
		sterol biosynthetic process	8.04E-07	18						

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		steroid biosynthetic process	1.31061E-06	18						
		ribosomal small subunit assembly	1.55002E-06	15						
		ergosterol metabolic process	2.40193E-06	16						
		phytosteroid metabolic process	2.40193E-06	16						
		rRNA metabolic process	3.09376E-06	62						
		organic hydroxy compound biosynthetic process	3.43856E-06	25						
		secondary alcohol metabolic process	4.12125E-06	16						
		single-organism biosynthetic process	4.5667E-06	96						
		ncRNA processing	4.90424E-06	72						
		organelle assembly	5.34156E-06	38						
		ribonucleoprotein complex subunit organization	8.39009E-06	41						
		cellular alcohol metabolic process	1.13464E-05	16						
		ribosomal large subunit assembly	1.22018E-05	17						
		ribonucleoprotein complex assembly	1.52083E-05	39						
		alcohol biosynthetic process	3.57475E-05	20						
		ribonucleoprotein complex export from nucleus	5.13451E-05	30						
		organic substance biosynthetic process	5.21744E-05	239						
		cellular biosynthetic process	7.19365E-05	237						
		ribonucleoprotein complex localization	7.54794E-05	30						
		sterol metabolic process	0.000109955	18						
		biosynthetic process	0.000124318	240						

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		steroid metabolic process	0.000153665	18						
		nuclear export	0.000227454	30						
		organic hydroxy compound metabolic process	0.000995237	29						
		maturation of 5.8S rRNA	0.001876329	22						
		nitrogen compound metabolic process	0.002818974	272						
		metabolic process	0.004092092	355						
		maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	0.006501848	21						
		maturation of LSU-rRNA	0.007393915	16						
		ribosomal subunit export from nucleus	0.03281662	14						
		establishment of ribosome localization	0.03281662	14						
		alcohol metabolic process	0.034567782	21						
		single-organism metabolic process	0.042533219	149						
		ribosome localization	0.042821774	14						
		RNA export from nucleus	0.048475977	21						
		cellular metabolic process	0.049211802	336						
Down-Up	102	none			none			none		
Null-Down	366				cytoplasm	0.004497722	278	catalytic activity	0.000120589	165
					cell part	0.018955614	320			
					cell	0.020636387	320			
Null-Null	3825	cellular component organization	6.07E-39	1464	cell	3.73E-93	3472	protein binding	7.57E-17	582
		biological regulation	1.11E-28	1403	cell part	9.87E-93	3471	ion binding	0.000106491	1003
		organelle organization	5.95E-26	1050	membrane-bounded organelle	4.49E-71	2841	enzyme binding	0.001788065	130
		regulation of biological process	1.72E-20	1161	intracellular organelle	1.59E-68	2985	molecular function regulator	0.002602179	212

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		cellular component organization or biogenesis	6.61E-19	1623	organelle	2.70E-68	2985	catalytic activity	0.004849303	1353
		regulation of cellular process	2.46E-18	1117	intracellular membrane-bounded organelle	5.53E-66	2800	transferase activity	0.012487111	532
		cellular protein modification process	1.83E-15	591	intracellular	9.81E-66	3384	protein kinase activity	0.01906804	99
		protein modification process	1.83E-15	591	intracellular part	1.88E-65	3378	ubiquitin-like protein transferase activity	0.022176556	55
		localization	3.30E-15	1068	organelle part	1.35E-36	1989	enzyme regulator activity	0.031208467	184
		macromolecule localization	8.31E-14	649	intracellular organelle part	3.01E-36	1985	ubiquitin-protein transferase activity	0.036988478	49
		establishment of localization	1.70E-13	967	organelle membrane	5.69E-30	723			
		chromosome organization	1.05E-12	455	protein complex	1.35E-29	847			
		transport	2.42E-12	928	macromolecular complex	5.73E-22	1424			
		protein localization	3.03E-12	543	bounding membrane of organelle	6.75E-21	461			
		regulation of primary metabolic process	4.37E-12	787	endomembrane system	6.52E-18	740			
		regulation of metabolic process	9.90E-12	808	cytoplasm	1.92E-17	2803			
		cellular macromolecule localization	1.05E-11	431	whole membrane	5.36E-15	396			
		regulation of cellular metabolic process	1.62E-11	789	mitochondrial part	8.66E-14	447			
		cellular localization	2.38E-11	626	transferase complex	5.47E-13	288			
		regulation of biosynthetic process	6.28E-11	668	catalytic complex	4.83E-12	480			
		cellular protein localization	7.72E-11	410	nucleoplasm part	9.36E-12	197			
		regulation of cellular biosynthetic process	9.59E-11	666	Golgi apparatus	1.61E-10	217			
		regulation of macromolecule metabolic process	1.08E-10	768	nucleus	2.51E-10	1498			
		cellular component assembly	1.78E-10	596	mitochondrion	4.06E-10	793			

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		establishment of protein localization	2.04E-10	449	nucleoplasm	1.48E-09	216			
		vesicle-mediated transport	4.70E-10	304	membrane-enclosed lumen	7.76E-09	789			
		mitochondrion organization	6.79E-10	221	organelle lumen	7.76E-09	789			
		response to stimulus	1.26E-09	839	intracellular organelle lumen	7.76E-09	789			
		protein transport	2.06E-09	432	organelle envelope	9.78E-09	400			
		regulation of macromolecule biosynthetic process	6.49E-09	637	envelope	9.78E-09	400			
		cellular macromolecule catabolic process	9.23E-09	307	vesicle	6.90E-08	200			
		chromatin organization	1.66E-08	261	cytoplasmic vesicle	7.17E-08	198			
		regulation of nitrogen compound metabolic process	1.94E-08	670	intracellular vesicle	7.17E-08	198			
		cellular response to stimulus	2.15E-08	734	Golgi apparatus part	1.55E-07	163			
		establishment of localization in cell	3.57E-08	491	organellar ribosome	2.07E-07	76			
		macromolecule modification	4.83E-08	700	mitochondrial ribosome	2.07E-07	76			
		regulation of cellular macromolecule biosynthetic process	5.95E-08	624	mitochondrial membrane	2.10E-07	283			
		macromolecule catabolic process	1.06E-07	332	mitochondrial envelope	2.89E-07	307			
		RNA biosynthetic process	1.19E-07	551	chromosomal part	3.06E-07	290			
		proteolysis involved in cellular protein catabolic process	1.51E-07	195	chromosome	7.64E-07	308			
		intracellular transport	1.53E-07	456	cytoplasmic vesicle part	7.98E-07	128			
		nucleic acid-templated transcription	1.53E-07	548	mitochondrial matrix	9.28E-07	169			
		regulation of nucleobase-containing compound metabolic process	2.10E-07	577	mitochondrial membrane part	2.37006E-06	139			

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		intracellular protein transport	2.63E-07	296	vacuolar membrane	3.16324E-06	198			
		cellular protein catabolic process	2.65E-07	208	vacuolar part	3.49238E-06	203			
		covalent chromatin modification	4.90E-07	149	Golgi membrane	4.34191E-06	117			
		transcription, DNA-templated	4.97E-07	542	nuclear chromosome part	6.46582E-06	199			
		protein localization to organelle	6.35E-07	302	cellular bud	1.04439E-05	183			
		regulation of gene expression	6.46E-07	608	intrinsic component of organelle membrane	1.47881E-05	101			
		regulation of RNA metabolic process	9.01E-07	527	endoplasmic reticulum	1.5091E-05	444			
		protein catabolic process	9.72E-07	220	nuclear chromosome	2.14439E-05	212			
		regulation of nucleic acid-templated transcription	1.41898E-06	511	endosome membrane	3.86384E-05	74			
		regulation of RNA biosynthetic process	1.41898E-06	511	organellar large ribosomal subunit	5.66398E-05	44			
		protein complex biogenesis	2.16634E-06	247	mitochondrial large ribosomal subunit	5.66398E-05	44			
		organic substance transport	3.16676E-06	629	nuclear chromatin	6.73303E-05	103			
		ubiquitin-dependent protein catabolic process	4.89828E-06	177	membrane protein complex	6.81325E-05	178			
		modification-dependent protein catabolic process	4.89828E-06	177	SWI/SNF superfamily-type complex	8.40784E-05	52			
		cellular response to stress	4.90639E-06	496	integral component of organelle membrane	8.9338E-05	96			
		cell cycle	5.02543E-06	555	organelle inner membrane	0.000137767	191			
		regulation of transcription, DNA-templated	5.40316E-06	504	endosome	0.000143055	130			
		macromolecular complex subunit organization	5.48541E-06	492	mitochondrial inner membrane	0.000182737	183			
		single-organism organelle organization	7.45379E-06	388	endosomal part	0.000226954	85			
		protein complex assembly	8.0398E-06	238	cytoplasmic part	0.000360066	2068			

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		modification-dependent macromolecule catabolic process	1.05669E-05	191	cellular bud neck	0.000539118	142			
		vesicle organization	1.10487E-05	170	vacuole	0.000885136	346			
		response to stress	1.23637E-05	555	cell periphery	0.002229077	505			
		establishment of protein localization to organelle	1.58537E-05	245	lytic vacuole membrane	0.002236091	156			
		protein targeting	1.64261E-05	241	ubiquitin ligase complex	0.002315021	68			
		membrane organization	2.28337E-05	230	mitochondrial protein complex	0.002744016	88			
		cellular protein complex assembly	2.52332E-05	206	fungal-type vacuole membrane	0.002873636	155			
		protein complex subunit organization	4.58772E-05	269	chromatin	0.004029974	115			
		protein modification by small protein conjugation or removal	8.3794E-05	163	transferase complex, transferring phosphorus-containing groups	0.004029974	115			
		single-organism membrane organization	8.64446E-05	215	intrinsic component of mitochondrial membrane	0.00489256	55			
		regulation of molecular function	0.000110762	300	site of polarized growth	0.006974481	188			
		protein-DNA complex subunit organization	0.00021114	143	organelle subcompartment	0.010582067	70			
		nucleobase-containing compound biosynthetic process	0.000273119	681	cell projection	0.013563189	98			
		negative regulation of biological process	0.000432874	419	mating projection	0.014202785	96			
		positive regulation of biological process	0.000499303	462	endoplasmic reticulum part	0.014421842	268			
		single-organism process	0.000652449	2095	integral component of mitochondrial membrane	0.015774256	52			
		macromolecular complex assembly	0.000689075	413	nuclear outer membrane-endoplasmic reticulum membrane network	0.018343731	259			
		transcription from RNA polymerase II promoter	0.000967847	365	endoplasmic reticulum membrane	0.019378798	251			

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		positive regulation of cellular process	0.001040195	451	cell projection part	0.019857679	93			
		positive regulation of metabolic process	0.001105115	366	cell cortex part	0.021579039	100			
		regulation of catalytic activity	0.001266934	283	storage vacuole	0.023993393	297			
		negative regulation of cellular process	0.001318419	411	fungal-type vacuole	0.023993393	297			
		positive regulation of cellular metabolic process	0.00167775	362	lytic vacuole	0.026090501	298			
		proteasomal protein catabolic process	0.00170302	121	chromosomal region	0.029719719	142			
		membrane fusion	0.001851922	101	nuclear lumen	0.048701845	558			
		autophagy	0.002048321	133						
		cellular response to DNA damage stimulus	0.002369438	253						
		cellular macromolecular complex assembly	0.002383028	389						
		single-organism cellular process	0.002835684	1811						
		regulation of biological quality	0.003227204	372						
		proteasome-mediated ubiquitin-dependent protein catabolic process	0.004100867	118						
		protein ubiquitination	0.004700014	109						
		Golgi vesicle transport	0.004971877	150						
		cell cycle process	0.005421556	436						
		single-organism localization	0.006132154	509						
		protein modification by small protein conjugation	0.006303914	131						
		endocytosis	0.006971141	99						
		positive regulation of biosynthetic process	0.006978518	292						

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		positive regulation of cellular biosynthetic process	0.006978518	292						
		positive regulation of nitrogen compound metabolic process	0.009604701	299						
		organic substance catabolic process	0.009798001	488						
		regulation of cellular component organization	0.009958845	304						
		positive regulation of molecular function	0.011294593	187						
		cellular catabolic process	0.012089516	426						
		negative regulation of metabolic process	0.012343872	308						
		catabolic process	0.012952822	491						
		membrane invagination	0.01589636	83						
		positive regulation of macromolecule metabolic process	0.018260806	345						
		DNA repair	0.018516671	216						
		negative regulation of biosynthetic process	0.018729451	257						
		negative regulation of cellular biosynthetic process	0.018729451	257						
		chromatin remodeling	0.020486444	73						
		protein import	0.02102021	106						
		negative regulation of cellular metabolic process	0.021108208	305						
		histone modification	0.027946732	105						
		regulation of cell cycle	0.028750689	203						
		single-organism transport	0.031965164	467						
		mitochondrial respiratory chain complex IV biogenesis	0.031966295	27						

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
		nucleosome organization	0.033334733	62						
		positive regulation of catalytic activity	0.036661091	178						
		organelle fusion	0.037824632	106						
		vacuole organization	0.041364763	80						
		nucleotide-excision repair	0.045880089	51						
Null-Up	626	cell cycle	0.000366784	117	microtubule cytoskeleton	6.63027E-06	34	double-stranded DNA binding	0.011273841	39
		reproduction	0.003320998	80	cell periphery	0.000142952	112	sequence-specific DNA binding	0.014241085	46
		biological regulation	0.003552248	237	cytoskeleton	0.000397397	46	sequence-specific double-stranded DNA binding	0.032891815	32
		positive regulation of biological process	0.005477186	98	microtubule	0.000424642	20			
		positive regulation of cellular process	0.006730697	98	chromosome	0.001503817	69			
		reproductive process	0.007348359	77	chromosomal part	0.001988098	65			
		cellular response to stimulus	0.010289438	140	spindle	0.00327333	19			
		cell cycle process	0.01248731	93	cytoskeletal part	0.003421556	42			
		regulation of cellular process	0.013077037	195	supramolecular complex	0.006393128	20			
		regulation of biological process	0.017898981	200	supramolecular polymer	0.006393128	20			
		microtubule polymerization or depolymerization	0.029852703	13	supramolecular fiber	0.006393128	20			
		response to stimulus	0.034634646	154	polymeric cytoskeletal fiber	0.006393128	20			
					microtubule organizing center	0.006635697	19			
					spindle pole body	0.006635697	19			
					chromosomal region	0.017158814	37			
Up-Down	7	none			none			none		
Up-Null	233	single-organism process	0.003501	136	peroxisome	0.007776083	12	none		
					microbody	0.007776083	12			

Expression Trend	Number of Genes	Biological Process			Cellular Localization			Molecular Function		
		Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes	Ontology	*p*-value	# of genes
					peroxisomal matrix	0.0438193	5			
					microbody lumen	0.0438193	5			
Up-Up	98	none			none			none		