Toward the development of an adipic acid-tolerant production host in *Saccharomyces cerevisiae*

Iain Summerby-Murray

A thesis
in
The Department
of
Biology

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Science (Biology)
at Concordia University
Montréal, Québec, Canada

April 2022

Concordia University
School of Graduate Studies

This is to certify that the thesis prepared

By: **Iain Summerby-Murray**
Entitled: **Toward the development of an adipic acid-tolerant production host in *Saccharomyces cerevisiae***

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

Master of Science (Biology)

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

Signed by the final examining commitee:

_____ Chair
Dr. Malcolm Whiteway

_____ Examiner
Dr. Laurent Potvin-Trottier

_____ Examiner
Dr. Paul Joyce

_____ Supervisor
Dr. Vincent Martin

Approved by _____
Dr. Robert Weladji, Graduate Program Director

Date _____ _____
Dr. Pascale Sicotte, Dean of Arts and Science

Abstract

Toward the development of an adipic acid-tolerant production host in *Saccharomyces cerevisiae*

Iain Summerby-Murray

Adipic acid is used in the production of a range of important products, notably nylon, but its production using traditional means relies on petroleum precursors and is environmentally damaging. The production of adipic acid from renewable sources using microbial cell factories has generated significant interest yet fails to compete economically with its petro-based alternative, partly due to the large expense required for post-fermentation processing. Using a low pH fermentation strategy, downstream acidification costs can be decreased, which will help with raising the economic viability of bio-based adipic acid. We screened a large collection of tentatively acid tolerant strains of *Saccharomyces cerevisiae* to assess adipic acid tolerance and later characterised top adipic acid tolerant strains for suitability as adipic acid production hosts. Top strains were mostly diploid and had growth yields in adipic acid roughly 5x higher than the non-tolerant strain CEN.PK113-7D. Furthermore, we show that adipic acid tolerance is an inducible phenotype – pre-adaptation in adipic acid reduces latency upon subsequent exposure to adipic acid. Next, we investigated transcriptional changes by way of RNAseq. Gene content was largely the same across 4 different strains of *S. cerevisiae*. Exposure to adipic acid caused cell-wide transcriptional changes in both tolerant and non-tolerant strains. Differences in gene expression, especially in stress response, provided an explanation for the adipic acid tolerant phenotype but, for a more complete picture, more work is still required. Finally, as a proof of concept for the validity of acid tolerant production strains, we expressed *aroZ*, necessary for the first step in adipic acid production from the shikimate pathway, in an adipic acid tolerant strain and showed increased PCA, an adipic acid precursor, accumulation when grown at low pH and under adipic acid stress compared to a similarly engineered adipic acid-sensitive CEN.PK113-7D. Our results emphasize the utility of adipic acid tolerant strains for the production of adipic acid. Other carboxylic acid may also benefit from similarly acid tolerant production strains, which could help improve bio-based alternatives in the chemicals industry and reduce our reliance on petroleum-based products.

Acknowledgments

I would like to thank my supervisor Dr. Vincent Martin, for his mentorship and advice throughout my degree, and past and current members of the Martin Lab including Audrey Morin, Olivier Rouseau, Dr. James Dhaliwal, Meghan Davies, Dr. Bjorn Bean, Dr. Jaya Joshi, Mackenzie Thornbury, Dr. Lauren Narcross, Dr. Michael Pyne, Mohamed Nasr, Mathieu Husser, Christien Dykstra, James Bagley, Dr. Kealan Exley, Dr. Kaspar Kevvai, and Daniel Tsyplenkov.

Thanks also to my parents and family, and especially to my partner Krista Jager, for their support over the last several years.

Author contributions

Experimental design by Iain Summerby-Murray and Dr. Vincent Martin. The *AroZ* expression cassette described in section 2.3.4 was made by Mohamed Nasr. Code used for the high-throughput screening via the BioMek liquid handler and that used for ploidy determination is based on work by James Bagley. Analysis and writing by Iain Summerby-Murray.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Petroleum-based chemical production and the shift toward bio-manufacturing

The worldwide use of fossil fuels is known to be detrimental to the environment, largely due to the production of greenhouse gasses, a major driver of climate change[1]. The chemicals industry represents a major energy consumer and contributes roughly 7% of man-made greenhouse gas emissions worldwide[2]. Due to its reliance on fossil fuels and its greenhouse gas emissions, there is growing concern for the detrimental environmental effects of the chemicals sector[3].

One solution to deal with excessive petroleum usage is the development and use of carbon-neutral bio-manufacturing processes whereby materials are converted into useful products by exploiting microorganisms, either through endogenous pathways or with the introduction of engineered metabolic pathways[4]. The yeast *Saccharomyces cerevisiae*, filamentous fungus *Aspergillus niger*, and bacterium *Escherichia coli* are among the most commonly used and best studied microbial cell factories although non-traditional platform species are becoming more widely used[5]. The production of a wide range of useful chemical products has been achieved in a variety of different microorganisms including biofuels[6], pharmaceuticals[7], and food additives[8]. Some chemicals are specific targets of bio-manufacturing, yet others are used as building blocks — so called platform chemicals – for later conversion to valuable end products[9]. Of these, carboxylic acids are an important kind of chemical building block: their characteristic carboxyl group[10] makes them versatile

and able to undergo diverse chemical conversions to different products[5;9]. In 2004 the US Department of Energy identified many carboxylic acids as having high potential to be produced from biomass[11]; since then a number of these potential high value products have been produced in microbial cell factories such as lactic acid[12], succinic acid[13], and 3-hydroxypropionic acid[14].

1.2 Adipic acid

Adipic acid is a dicarboxylic acid[15] primarily used in the production of nylon, as well as in plasticizers, lubricants, and as a food additive[16;17]. Adipic acid ranks as one of the top bulk chemical products, with annual production exceeding 3 million tons[17;18]. Despite its importance as a bulk chemical product, there are major concerns over the detrimental environmental effects associated with adipic acid production – in 2016 adipic acid accounted for roughly 20% of total petrochemical-related CO_2 release in the USA[19;20]. Current petroleum-based production of adipic acid relies on the conversion of benzene to a mixture of cyclohexanone and cyclohexanol via the oxidation of cyclohexane. The cyclohexanone:cyclohexanol mixture is then converted to adipic acid by catalytic conversion in the presence of nitric acid[21] (Figure 1.1). This process releases 1 mole nitrous oxide, a potent greenhouse gas, per mole adipic acid[22].

Due to its ubiquity as a chemical product, the noxious environmental damage caused by its production using traditional means, and fluctuating cost of petroleum precursors, there is substantial need to develop bio-based adipic acid. In addition, adipic acid is well suited for bio-manufacturing due to its potential as a platform chemical[23] and is uniquely situated as a dicarboxylic acid with no known natural production[18;24].

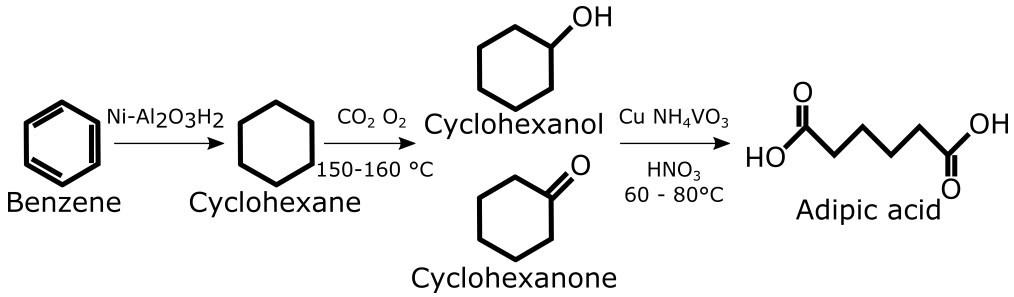

Figure 1.1: Adipic acid is traditionally produced from petrochemical precursors. Benzene, a petroleum derivative, is reduced to cyclohexane via nickel catalysis at high pressure. Cyclohexane is oxidized to a cyclohexanone:cyclohexanol mixture which is converted to adipic acid in the presence of nitric acid. Adapted from Van de Vyver et al.[21]

1.2.1 Bio-manufacturing adipic acid and its precursor *cis,cis*-muconic acid

A number of different strategies in the production of adipic acid have been employed with varying degrees of success. Production of adipic acid and its precursor *cis,cis*-muconic acid (ccM) from a variety of pathways and organisms is shown in Table 1.1. Over the last two decades end product yields have improved steadily both in traditional microbial factories as well as in emerging microbial cell factories such as in *Pichia* species[5;25].

Table 1.1: Strategies used in the production of adipic acid and its precursor *cis,cis*-muconic acid. Unless otherwise noted, production refers to adipic acid.

Date	**Species***	**Description**	Production	**Reference**
2002	*E. coli*	Degradation of cyclohexanol and cyclohexanone by orthologous expression of genes from *Acinetobacter*, *Arthrobacter*, and *Rhodococcus* species	NA	26;27
2002	*E. coli*	ccM production via shikimate intermediates using orthologous Kp**AroZp**, Kp**AroYp**, and Ac**CatAp** expression followed by chemo-catalytic hydrogenation to adipic acid	36.8 g/L ccM	28
2009	NA	Degradation of lysine to adipic acid	NA	29
2012	*S. cerevisiae*	ccM production by way of shikimate intermediates	1.56 mg/L ccM	30
2015	*Thermobifida fusca*	Reverse adipate pathway. First known natural production	2.23 g/L	24
2015	NA	Conversion of glutaric acid or furan dicarboxylic acids to adipic acid by oxidation-hydrodeoxygenation	NA	31;32
2016	*Pseudomonas putita*	Improved ccM production through improved PCA decarboxylase activity by co-expression of flavin-derived cofactors	4.92 g/L ccM	33
2018	*S. cerevisiae*	ccM production by way of shikimate intermediates along with feedback inhibition of aromatic amino acid metabolism	5.1 g/L ccM	34
2018	*E. coli*	Reconstructed reverse adipate pathway from *T. fusca*	68 g/L	35
2018	*Candida tropicalis*	Degredation of fatty acids to adipic acid by disrupting C6 acyl chain degradation	50 g/L	36
2019	*E. coli*	Codon optimized conversion of DHS to ccM in a DHS-overproducing strain	64.5 g/L ccM‡	37
2020	*S. cerevisiae*	Biosensor-aided mutageneis to identify mutants with improved DHS dehydrogenase and catechol1,2-dioxygenase activity	20.8 g/L ccM‡	38

* NA denotes that the process is a proposed method that hasn't necessarily been shown in practice

‡ current highest known production

Initial attempts showed some success by producing adipic acid from the degradation of cyclohexane and cyclohexanol using genes from *Acinetobacter*, *Arthrobacter*, and *Rhodococcus* species[27]. However, since this approach required the use of cyclohexanol, a petroleum derived product, as a feedstock, this approach is not preferred. Other approaches make use of the β-ketoadipate pathway found in *Pseudomonas*[39] and *Acinetobacter*[40] species. Here, catechol, an intermediate in the conversion of benzoate to acetyl-coA and succinyl-coA, is converted to ccM. Again, this approach is not preferred since benzene is derived from petroleum and is toxic to the production host. Alternatively, adipic acid from sugar is often preferred due to the high yield – theoretically 0.9 moles adipic acid per mole glucose – as

well as lower estimated operating costs[17].

Of particular interest to this thesis is the heterologous ccM production from the shikimate pathway, first shown in *S. cerevisiae* by Weber et al.[30], in which carbon flux is diverted from the native metabolism to the orthologous pathway at 3-dehydroshikimate (DHS) (Figure 1.2). DHS is oxidized by the DHS dehydrogenase AroZp to produce protocatechuic acid (PCA) then PCA undergoes a conversion by the action of the PCA decarboxylase AroYp to catechol (CAT). In the first iterations of this pathway CAT was converted to ccM by CatAp, however more recent iterations make use of the catechol dioxygenase Hqd2p instead[34]. Recent studies have greatly improved metabolite yields through this pathway – from 1.56 mg/L ccM[30] to 20.8 g/L in *S. cerevisiae*[38] and 64.5 g/L in *E. coli*[37] – yet the final conversion of ccM to adipic acid remains poor. The conversion of ccM to adipic acid can be carried out using two groups of enzymes: the Old Yellow Enzymes (OYE) and enoate reductases (ER). OYEs have low activity and are generally overshadowed by ERs. Recent discoveries and improvements in ERs have shown significant promise[41;42]. However, until this strategy becomes more widely implemented, many studies continue to report ccM yield as a proxy for adipic acid.

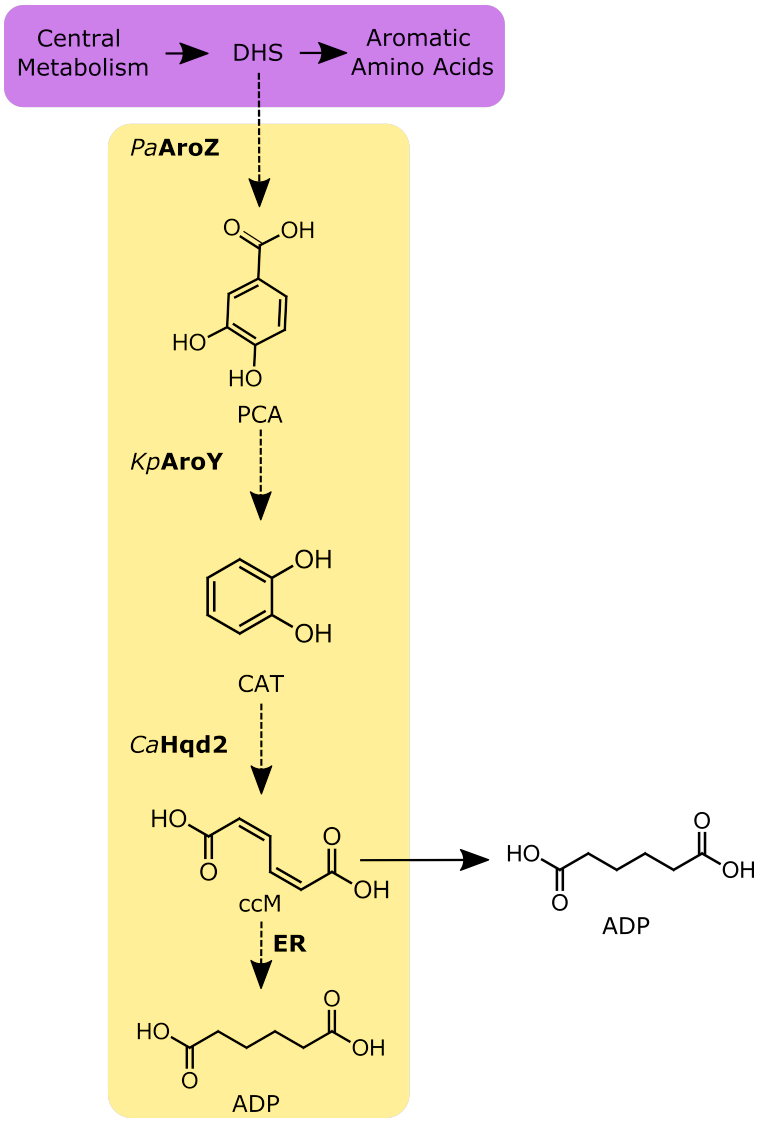

Figure 1.2: The introduction of heterologous genes allows for diversion of 3-dehydroshikimate (DHS) toward adipic acid (ADP). DHS is converted to protocatechuic acid (PCA) by the DHS dehydrogenase AroZp. The conversion of PCA to catechol (CAT) is carried out by the PCA decarboxylase AroY followed by conversion of CAT to *cis,cis*-muconic acid (ccM) by the catechol dioxygenase Hqd2. Initial attempts at producing ADP relied on the chemical conversion of ccM to ADP by metal catalysts whereas recent advances in enoate reductases (ER) improved the biological conversion[41]. Adapted from Pyne et al.[34] and Joo et al.[41].

Despite recent improvements in bio-manufacturing, many products are not economically competitive with their petroleum-based counterparts[17;43]. Economic viability is dependent on four main factors: yield, titer, price of feedstock, and downstream processing

costs. Recent analysis suggests that improvements in these main areas will allow bio-based production of 26 chemicals to become economically viable with 17 of those being as or more cost-effective than their petro-based counterparts[43]. According to this model, current bio-adipic acid has an internal rate of return (IRR) of 3% compared to an IRR of roughly 30% for petro-adipic acid. By increasing titers to 100 g/L, decreasing downstream processing cost and feedstock costs, and improving yield, the IRR for bio-adipic acid could rise as high as 55%. While the complete implementation of this model is still a long way off, improvements in yields and titers are steadily improving.

One area which is not improving at the same rate is the cost of downstream processing, which can account for as much as 50% of the total production cost[44]. Due to the acidic nature of the product, traditional bio-production of adipic acid uses base addition to counteract the buildup of H+ ions (Figure 1.3), which would otherwise impose significant toxicity upon the microbial cell factories. However, in conditions that are optimal for production in current *E. coli* and *S. cerevisiae* production strains, adipic acid exists largely in its dissociated and semi-dissociated state. Acidification is required to purify the acid form for later use; this adds significant cost since acidifying 1 mole of adipic acid requires 1 mole of sulfuric acid[44].

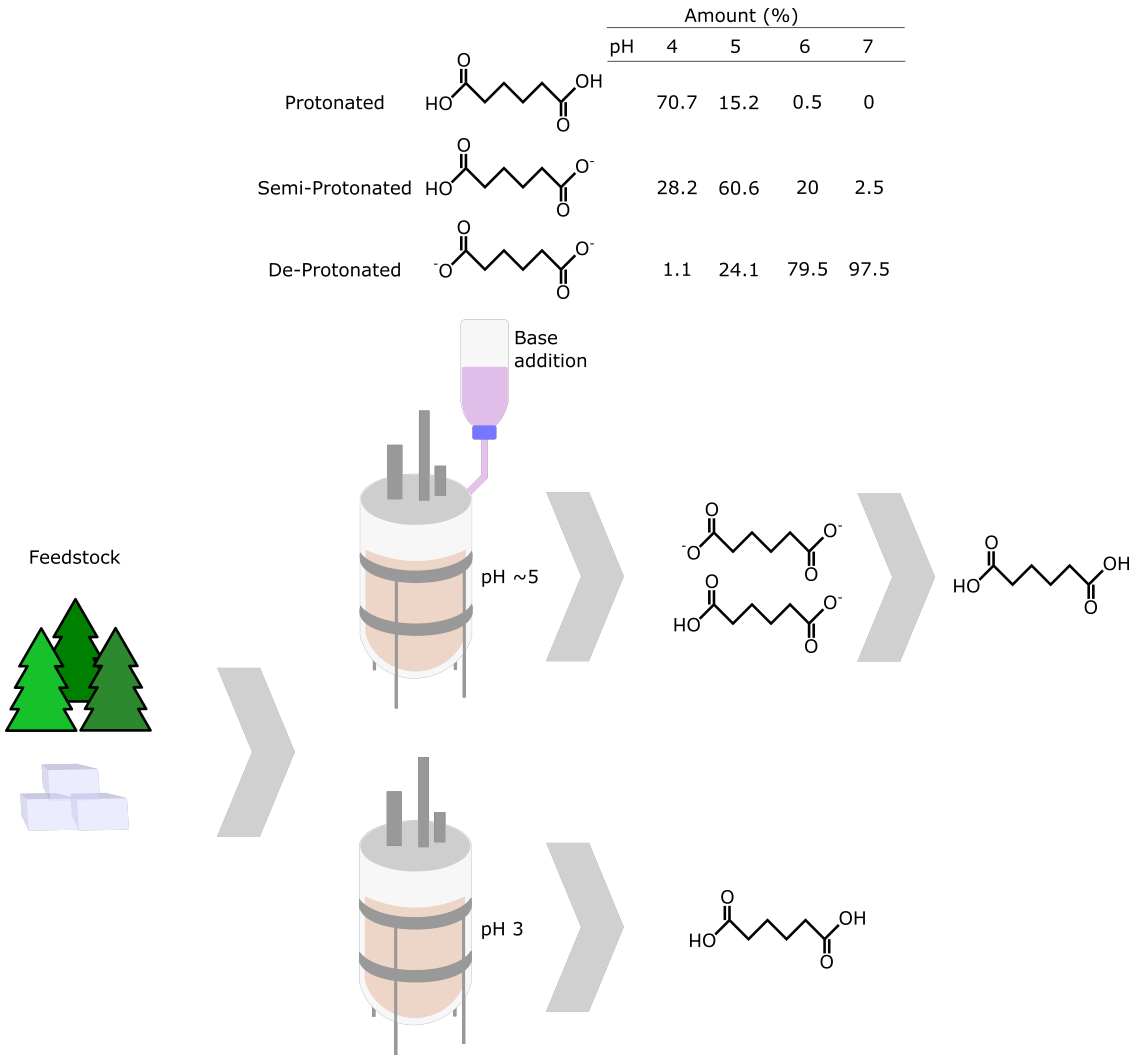

Figure 1.3: Adipic acid losses its protons sequentially. At neutral pH, adipic acid exist primarily in its dissociated form but as the pH decreases, more and more adipic acid exists in its un-dissociated form. Producing adipic acid at pH 5 results in only 15% production of the protonated form, requiring later acidification of the semi-protonated and de-protonated forms. Most yeasts grow readily at pH 5. The media pH drops as adipic acid is produced, which requires the addition of base to maintain a set pH. Additionally, at pH 5, adipic acid is produced primarily in its dissociated and semi dissociated state, which necessitates downstream treatment with acid to revert to the protonated form. This acidification adds substantially to the cost of biomanufacturing. The use of acid tolerant strains at low pH avoids the use of base addition and results in the protonated acid form being the dominant product. Adapted from Karlsson et al.[45]

An alternative approach is to use acid tolerant production strains, either by

engineering current production strains to be more tolerant or by adapting naturally tolerant strains to produce the compound of interest. These approaches generally focus on yeasts, which tend to perform better at low pH than bacteria[46]. By using a strain that can tolerate low pH and high levels of acid, base addition is no longer required, and the product is predominantly in its undissociated form, which avoids the need for downstream acidification. The combination of removing base addition and acidification contribute to greatly reducing the cost of production. Previous attempts at using acid tolerant strains have shown success. For example, in 2018 an acid tolerant strain of *Pichia kudriavzevii* was used to produce 135 g/L lactic acid at pH 3.6[47].

1.3 Exploring causes of acid tolerant phenotypes

1.3.1 Carboxylic acid tolerance in yeast

Due to its role as a model organism and the importance it has in a range of industries, *S. cerevisiae* has been studied extensively to elucidate its response to carboxylic acids and low pH. Due to the ubiquity of carboxylic acids and the potentially harmful effects to intracellular homeostasis, natural processes exist to adapt to and diminish the effects of these compounds on cell fitness. Much is already known about the effects of various organic acids, genes related to acid tolerance, and the transcriptional changes that these acids invoke (reviewed in Mira et al.[48]).

Transcriptomic studies in yeast have shown that carboxylic acid induced stress is complex and tolerance to this stress can be dealt with using different mechanisms. Exposure to carboxylic acids triggers multifaceted gene expression remodelling and changes to biochemical processes, including expression of proton-removing transporters, cell wall reorganization to restrict acid diffusion, changes to metal metabolism[49], and the activation of anti-reactive oxygen species pathways[50]. This response is mediated by several transcription factors that affect a cell-wide network of genes. For example, the transcription factor Haa1 regulates roughly 85 different genes in *S. cerevisiae*[51]. Protective mechanisms can work against a broad range of acid stresses, such as by the recruitment of multi drug resistance transporters and H+ ATPases[52] or the activation of heat shock proteins[53], or be specific to the characteristic of the acid involved, as seen with the expression of *HAA1*, which decreases with increased acid lipophilicity[54].

1.3.2 Non-traditional yeast strains represent a poorly understood source of tolerant phenotypes

Despite the presence of acid tolerant gene networks in *S. cerevisiae*, most common laboratory strains are not acid tolerant. As pH drops below 5 most *S. cerevisiae* experiences reduced growth; below pH 2.5 growth is severely restricted and cells quickly die[53;55]. Conversely, many non-*Saccharomyces* yeast can withstand low pH environments as well as high levels of carboxylic acids. For example, *Zygosaccharomyces bailii* can withstand up to 800 mM acetic acid whereas *S. cerevisiae* dies after exposure to 20-120 mM acetic acid[56]. Indeed, the number of extremophile phenotypes in non-*Saccharomyces* yeast make them candidates of significant interest for several applications yet challenges with genetic manipulation of these strains often make *S. cerevisiae* a more desirable host due to how well studied it is.

Between these two extremes are an increasingly well characterized group of non-conventional, i.e. non-lab, *S. cerevisiae* strains that offer a compromise between tolerant but poorly studied yeast species and well characterized but acid susceptible strains of *S. cerevisiae*. Recent efforts to better characterize these strains have shown that a huge variation of phenotypes exists within this single species[57]. For example, the strain *S. cerevisiae* var. *boulardii*, despite having a near identical genome to the reference strain S288c[58], has many different metabolic and physiological characteristics: it is tolerant to low pH, extreme temperature, and organic acids[59], probably due to a number of gene duplications[60].

1.4 RNAseq - a tool for exploring gene expression and more

Over the last two decades, as sequencing technologies have improved, RNA sequencing (RNAseq) experiments have taken over from microarray-based gene expression experiments. RNAseq improves upon many of the challenges faced by microarray experiments – such as nonspecific probe-oligomer binding, the necessity of prior knowledge of the sequences being tested, and lack of comparability between samples[61] – while also offering single nucleotide resolution[62]. RNAseq experiments generally follow a standard approach: RNA is extracted from cells, often exposed to different conditions then reverse transcribed to cDNA. Next, cDNA libraries are prepared and sequenced using Next-Generation Sequencing

(NGS) technologies. Bioinformatic tools are used to map reads to a reference genome or transcriptome; reads per feature are then counted and analysed[63]. Using RNAseq not only provides insight into gene expression but can also be used to explore genetic variations, gene fusions[64;65], and novel transcript expression[66]. The RNAseq process requires several steps and tools to take RNA from different samples and convert it to gene expression differences. This process can vary greatly depending on different factors including species, depth of coverage, sample conditions, and sequencing parameters. Despite the complexity, an ever-expanding toolkit of software[67] makes RNAseq experiments increasingly accessible and powerful.

Many studies use RNAseq to identify phenotype-genotype interactions, which can be especially useful when looking for certain phenotypes. In 2017 Dong et al.[55] took an RNAseq approach to investigate the effects of acetic acid on gene expression. They reported changes in expression for roughly 800 different genes, including those associated with amino acid and carbohydrate metabolism, ribosome and protein synthesis, transporters and permeases, and cell wall organization. Since many of the effects of acid stress are shared – for example *QDR3*[68] and *PMA1*[68;69] – across different acids[48], transcriptomics studies for any similar carboxylic acid provide insight into the effects of adipic acid on gene expression. While other studies have investigated the causes of improved adipic acid tolerance[68], to our knowledge, a comprehensive investigation of the genes associated with adipic acid stress has not yet been conducted using RNAseq.

1.5 Aims and rationale

Given the need to lower our environmental footprint from the production of adipic acid, our reliance on adipic acid-based products, and the fluctuating cost of petroleum precursors, it is essential to produce adipic acid from renewable sources. Current bio-based adipic acid fails to compete with its petro-based alternative yet it could eventually out-compete petro-based if the cost associated with its production could be lowered. One key area in which cost could be lowered is in downstream processing, of which a major component is the cost associated with post-production acidification to convert acid salts into their acid form. By using acid tolerant production strains, we can keep the pH during production relatively low to keep the acid in its protonated form, thereby lowering production costs.

Here, we screen a large collection of *S. cerevisiae* for adipic acid tolerant phenotypes

and characterize these adipic acid tolerant strains for later use. Next, we investigate the causes of adipic acid tolerance by comparing transcript levels between adipic acid tolerant strains and the non-tolerant CEN.PK113-7D during adipic acid stress by way of RNAseq. Finally, we demonstrate the utility of adipic acid tolerant strains for the production of adipic acid and its precursors by showing improved accumulation of the adipic acid precursor PCA in an engineered adipic acid tolerant strain compared to CEN.PK113-7D.

Chapter 2

Materials & Methods

2.1 Yeast strain screening

2.1.1 Strains and media

A collection of yeast strains was received from Lallemand Inc., denoted in this thesis as L001-L501. A selection of strains is described in Table 2.1 while the complete list of strains is found in the appendix. Yeast cultures were grown in YPD (10 g/L yeast extract (YE), 20 g/L tryptone, 20 g/L dextrose), YPS (10 g/L YE, 20 g/L tryptone, 20 g/L sucrose), or YNBG (6.8 g/L Yeast Nitrogen Base supplemented with 20 g/L glucose) supplemented as necessary with adipic acid. Where noted, media were supplemented with potassium citrate and HCl or NaOH to maintain a set pH.

Table 2.1: Selection of screened strains used in this thesis

Strain	Description *	Source
CEN.PK113-7D	MAT_{α} ura3-52 MAL2-8C SUC2	Euroscarf
L433	Wine	Lallemand
L248	Wine	Lallemand
L401	Wine	Lallemand
L451	Wine	Lallemand
L411	Wine	Lallemand
L434	Wine	Lallemand
L465	Wine	Lallemand
L489	Ingredients	Lallemand
L295	Beer	Lallemand
O33	Fruit juice drink	NCYC accession 3040

* Descriptions of Lallemand strains provided by Lallemand

2.1.2 Library screening

An automated screening method was developed to ensure accurate and repeatable results across the large number of strains screened in this thesis. Strains arrayed on YPD-agar were used to inoculate liquid media cultures, which were grown overnight. Cultures were diluted into an intermediate culture plate such that a chosen volume of inoculum resulted in a set starting OD_{600} using a custom Python script (available on github at ConcordiaMartinLab) and a Biomek FXp automated liquid handler (Beckman Coulter, USA). Strains were inoculated to an OD_{600} of 0.1 in 100 μL YNBG with (pH 2.8) and without (pH 4.8) the addition of 20 g/L adipic acid and cultured with 200 rpm shaking for 3 days at 30 °C . OD_{600} readings were taken every 24 hours using a plate reader (TECAN, CH). Using the screening method described above, select acid tolerant strains were re-screened in a larger volume to ensure reproducibility at scale. Strains were cultured in 1 mL YNBG with and without adipic acid. OD_{600} readings were taken every 24 hours using Sunrise (TECAN, CH) plate readers.

2.1.3 Ploidy determination

Strain ploidy was determined for select acid tolerant strains by flow cytometry following a previously described method[70]. In short, exponential cultures (OD_{600} 0.6) were fixed in 1 mL 70% (v/v) ethanol overnight before being washed and suspended in 50 mM sodium citrate. Cultures were treated with 0.5 mg/mL RNase A then stained with

25 ug/mL propidium iodide in 50 mM sodium citrate overnight. For each sample 10,000 events were captured using an Accuri C9 flow cytometer (BD Biosciences, USA). Data were analysed using Accuri flow software and a custom Python script (available on github at ConcordiaMartinLab) to compare fluorescent peaks and estimate genome size compared to the known haploid strain CEN.PK113-7D.

2.1.4 Acid tolerance inducibility

To determine if tolerance to adipic acid is an inducible phenotype, select strains were pre-cultured in YNBG with 0, 10, or 20 g/L adipic acid for 4 hours before being each inoculated into YNBG with 0 or 20 g/L adipic acid. Cultures were kept at 30 °C with 200 rpm shaking for 48 hours. OD_{600} was measured every 5 minutes using Sunrise plate readers (TECAN, CH). Growth characteristics were analysed using a custom R script (available on github at ConcordiaMartinLab).

2.2 Transcriptomics analysis

2.2.1 RNA extraction and sequencing

In order to assess causes of adipic acid tolerance in the identified adipic acid tolerant strains, gene expression was compared between non-tolerant and tolerant strains after exposure to adipic acid using RNAseq. Strains CEN.PK113-7D, L433, L248, L295, and O33 were cultured in potassium citrate-buffered YNBG (pH 3.2) with and without the addition of 5 g/L adipic acid and grown to an OD_{600} of 0.8. RNA was extracted using the Ambion RiboPure – Yeast kit (Thermo Fisher Scientific, USA). In short, cell membranes were disrupted with Zirconia beads and shaking in phenol:chloroform:isoamyl alcohol. RNA was separated from polysaccharides, proteins, and cell debris by removing the aqueous layer then purified using glass fiber filters. Residual DNA was removed by treating samples with RNAse free DNaseI. RNA sample quality control was carried out using a Tapestation 4150 (Agilent, USA). Next, cDNA libraries for sequencing were generated by the Centre d'expertise et de services Génome Québec from RNA samples using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB). These cDNA libraries were sequenced using the Illumina NovaSeq6000 platform at Génome Québec using the following

parameters: paired end 100 bp reads and 25 million reads per sample. Sequencing data were accessed from the Génome Québec Nanuq server for analysis.

2.2.2 Bioinformatic workflow

In order to assess differential gene expression between the various strains and conditions, RNAseq sample data were processed following the well described STAR-HTseq-EdgeR pipeline (reviewed by Corchete et al[71]) or, when reference genomes were not available, using the Trinity pipeline[72] in combination with EdgeR[73]. The analysis was carried out in Omicsbox version 2.0.36 except where otherwise noted. Figure 2.1 shows a schematic of the analysis pipeline including an additional workflow required for strains without a reference genome.

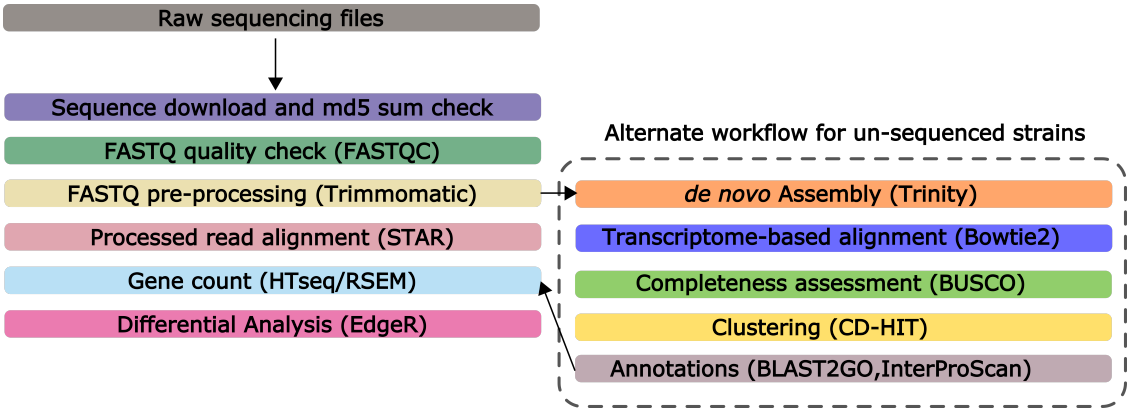

Figure 2.1: RNAseq analysis workflow used in this thesis. Samples from all strains and conditions were pre-processed using FastQC and Trimmomatic. Strains with reference genomes were analysed using the STAR-HTseq-EdgeR pipeline. Strains without reference genomes had reference transcriptomes built using Trinity. The *de novo* transcriptome assembly was verified using BUSCO and Bowtie2, clustered using CD-HIT, with functional annotations provided by Blast2Go and InterProScan. After completing the Trinity alternate pipeline, samples without reference genomes returned to the standard pipeline by way of RSEM gene counts. Differential gene expression analysis was done using EdgeR. All of the above was carried out using the Omicsbox interface.

2.2.3 Quality control and pre-processing

Sequence data were received from Génome Québec in FASTQ format. Raw sequences underwent quality control using FastQC[74], which assesses sequences according to 9 different metrics, including quality per sequence and per base, GC content, and length distribution across a given library. Sequence files were pre-processed to trim low quality bases, filter short reads, and remove adapter and contaminating sequences using Trimmomatic[75] with the following parameters: paired-end reads, custom adapter sequences (NEBnext Dual), Adaptive Quality trimming, with other parameters set to default.

2.2.4 Read alignment

Two different strategies were used for read alignment depending on whether reference sequences were available. Reads from strain CEN.PK113-7D were aligned directly to a previously sequenced reference genome[76] using the Spliced Transcripts Alignment to a Reference (STAR) algorithm[77] version 2.7.9a with default parameters. Strains L433, L248, L295, and O33 were aligned to *de novo* transcriptomes, described below, using RNASeq by Expectation Maximization (RSEM)[78] version 1.3.3 following *de novo* transcriptome assembly with Trinity[72].

2.2.5 *de novo* transcriptome assembly

For reads to be counted and gene expression analysis performed, reads must be aligned to a reference. Since reference genome sequences are not available for strains L433, O33, L248, and L295, it was necessary to build a reference transcriptome using a workflow that makes use of the *de novo* sequence assembler Trinity[72;79] version 2.12.0 and its associate tools. Following the reconstruction of strain transcriptomes, reads were aligned using Bowtie2[80] version 2.4.1 to confirm the integrity of the assemblies.

In order to quantify the quality of the *de novo* assembly, expected gene content was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO)[81] from OrthoDB[82]. Each strain was compared to the *Saccharomycetes* class using a Basic Local Alignment Search Tool (BLAST) e-value cutoff of 10^{-3}. The Trinity process results in multiple sequences per "Trinity gene", i.e. for a 1000 bp gene Trinity sees a read from 1-800 bp as separate from a

read from 1-1000 bp even though both reads are from the same biological source. In order to reduce the redundancy created during *de novo* assembly, similar sequences were pared down to the fewest distinct sequences using the clustering algorithm CD-HIT[83]. Following clustering, "Trinity genes" were annotated with Blast2Go[84] and InterProScan[85] to provide functional annotations.

2.2.6 Differential expression analysis

Gene count tables were constructed using HTseq[86] and RSEM. Where necessary, sequences were queried against the *Saccharomycotina* clade with BLAST. Successful hits were used to link biological genes to the "Trinity genes". Gene names were then appended to *de novo* read counts using a custom R script. Differential Expression (DE) analysis was performed with EdgeR[73;87] with data normalized using the Trimmed Mean of M (TMM) method[88]. For each strain, pairwise DE analysis was carried out between samples derived from cultures in which adipic acid had been supplemented and their corresponding adipic acid-free cultures. In Omicsbox, the "Robust" option was enabled to better handle outliers. For each strain MDS plots were generated in Omicsbox to visualize differences in samples due to culture conditions. DE genes (DEGs) were determined by considering Fold Change (FC) and False Discovery Rate (FDR). For the purposes of this study, genes with a log_2(FC) of 1 or greater were considered to have a difference in gene expression and those with an FDR less than 0.05 were considered to be significant. Functional enrichment analysis was visualized using gProfiler[89] and STRING[90] using the list of DEGs identified using EdgeR.

2.3 Yeast strain engineering

2.3.1 Plasmids

Plasmids, described in Table 2.2, were maintained in *E. coli* grown in Lysogeny Broth (LB) supplemented with antibiotics where appropriate. Plasmids were purified from *E. coli* stocks using the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific). DNA used as parts for plasmid construction was amplified using Phusion High Fidelity DNA polymerase.

Table 2.2: Plasmids used in this thesis

Plasmid	Progenitor	Description	Reference
pBB94	NA	ColE1^{ori} 2μ^{ori} P$_{TEF}$-**KanMX**^{R}-T$_{TEF}$ P$_{RNR2}$-**Cas9**-T$_{CYC1}$ [2x BsaI protospacer]	[91]
pBB95	NA	ColE1^{ori} 2μ^{ori} P$_{TEF}$-**Hyg**^{R}-T$_{TEF}$ P$_{RNR2}$-**Cas9**-T$_{CYC1}$ [2x BsaI protospacer]	[91]

2.3.2 pCas and gRNA design

In order to carry out genetic manipulations, CRISPR-Cas9 directed modifications of the yeast genome were carried out using a 20 nucleotide guide sequence with homology to a specific locus to direct Cas9 endonuclease activity and cause a double stranded break at the target site[92]. The inclusion of a donor DNA fragment, with homology regions flanking the double stranded break, allows for homology directed repair (HDR) which results in the integration of the donor DNA sequence at the targeted site[93;94]. Guide RNA sequences were designed using Benchling with off target scores calculated as previously described[95]. Guide RNA expression cassettes were created using overlapping oligonucleotides containing the guide sequence as well as overhanging ends compatible with a pre-built Cas9 plasmid system[91]; the unique guide sequence was then assembled into the Cas9 plasmid *in vivo* by homologous recombination.

2.3.3 Yeast transformation

Yeast was transformed using a lithium acetate (LiAc)/polyethylene glycol (PEG) protocol adapted from Gietz and Schiestl[96]. Strains were cultured overnight before being diluted to an OD_{600} of 0.1 and grown to a final OD_{600} of 0.8. Pellets from the cultures were then washed in water followed by 100 μM LiAc before being suspended in transformation mix – 100 μL 50% (w/v) PEG 3350, 5.6 μL 3M LiAc, and 4.4 μL 10 mg/mL salmon sperm DNA per reaction – and added to 40 μL of DNA in water. Samples were then incubated at 30 °C for 30 minutes followed by 42 °C for 30 minutes before being recovered in 600 μL YPD and plated on selective media.

2.3.4 Colony PCR

Following genetic manipulation via CRISPR-Cas9 and selective plating, the resulting colonies were analyzed by colony PCR to confirm the desired genetic modification had occurred. Cells from individual colonies were suspended in 40 μL water; half the suspended cells were kept in reserve while the rest were boiled for 2 minutes. PCR was carried out using 2.5 μL of the boiled yeast-water mix along with 3 μL of 2x Phire Plant Direct PCR Master Mix (Thermo) and 600 nM of each forward and reverse primer per reaction. PCR products were visualized by gel electrophoresis and band size compared to the theoretical size given a correct manipulation.

2.3.5 *aroZ* integration

The first step in the production of adipic acid from yeast using the shikimate pathway makes use of the DHS dehydrogenase AroZp. An expression cassette for *AROZ* built using *Podospora anserina AROZ* (GenBank accession CAD60599) under the control of the strong promoter P_{pgk1}[97] with flanking regions specific for Flagdeldt site 18[98] had previously been built and integrated into CEN.PK113-7D by Mohamed Nasr in our laboratory (unpublished). This *AROZ* cassette was amplified off genomic DNA from the previous integration using primers ISMFgF18F and ISMFgF18R (Table 2.3) and sequence verified. Using this as a template for homology directed repair, the resulting PCR product was transformed into strains CEN.PK113-7D and L433 along with a CRISPR-Cas9 expression plasmid[91] containing a geneticin or hygromycin resistance cassette and gRNA specific for Flagfeldt site 18, resulting in strains S17 and S20 (Table 2.3). Following selective plating for the CRISPR-Cas9 plasmid, resulting colonies were verified for the correct integration of the *AROZ* cassette by colony PCR using primers ISMaroZF and ISMaroZR and sequence verified.

Table 2.3: Strains and primers used for *AROZ* integration

Strains

Name	Progenitor	Description	Reference
CEN.PK113-7D	-	MAT_{α} ura3-52 MAL2-8C SUC2	Euroscarf
L433	-	-	Lallemand
S17	CEN.PK113-7D	P_{pgk1}-**PaAROZ**-T_{adh1}	This study
S20	L433	P_{pgk1}-**PaAROZ**-T_{adh1}	This study

Primers

Name	Sequence (5'-3')	Description
ISMFgF18F	TGTGCACAAAGGCCATAATA	Amplify *AROZ* cassette out of Flagfeldt site 18
ISMFgF18R	AAAGCTGGCTCCCCTTAGACAA	Amplify *AROZ* cassette out of Flagfeldt site 18
ISMaroZF	CAAGGCAGCGGATAAAGACAAC	Confirm integration of *AROZ*
ISMaroZR	CAAGGCAGCGGATAAAGACAAC	Confirm integration of *AROZ*

2.3.6 Metabolite analysis

To demonstrate the viability of adipic acid tolerant production strains we first sought to show that these strains are better at producing the adipic acid precursor PCA at low pH and under adipic acid stress than strains from the the non-tolerant CEN.PK113-7D background. Metabolite production was analyzed using High Performance Liquid Chromatography (HPLC), as previously described[34]. Strains S17 and S20 were cultured in 15 mL of potassium citrate-buffered YPS, supplemented with 20 g/L adipic acid where needed, and incubated at 30 °C with 200 rpm shaking. Media were supplemented at 72 hours and again at 144 hours with 1/10th volume 40% (w/v) sucrose to account for evaporation. After 8 days, 500 μL of each culture was centrifuged at 15,000 x g for 5 minutes. The supernatants were passed through a 0.2 μM filter and diluted 10x in 50 mM phosphate buffered saline (PBS, pH 7.6) so that sample concentrations fell within the linear range during quantification. Culture supernatants were analyzed using a Thermo Finnigan Surveyor HPLC equipped with a FAST-ACID column (Bio-Rad) coupled with a Waters 3100 refractive index (RI) detector. Samples were resolved isocratically in 10 mM sulfuric acid at 65 °C with a flow rate of 0.8 mL/min. Protocatechuic acid was detected using a UV/vis detector at 210 nm. The determination of metabolite concentration from chromatogram data was done by comparing sample peak area to a linear standard curve produced using reference standards (Sigma-Aldrich) from 0 to 6.25 mM. The limit of detection (LOD) was calculated using $LOD = 3\sigma/m$ where σ is the standard deviation of known concentrations and m is the slope from the standard curve; LOD was determined to be 14 ng/L. Samples

were run in quadruplicate for each strain-condition combination and data is shown as mean ± standard error.

Chapter 3

Results

3.1 Screening and characterisation of adipic acid tolerant strains

A collection of 500 yeast strains were screened to select for adipic acid tolerance first using a high-throughput method then again using a low-throughput but higher volume method to confirm the results of the high-throughput method for strains exhibiting tolerance to adipic acid. Top adipic acid tolerant strains were selected based on a comparison of growth yields – i.e. how much biomass compared to starting biomass, in our case measured by optical density/turbidity to determine number of cells– in adipic acid and without the addition of adipic acid. Ploidy of the top acid tolerant strains was determined by propidium iodide staining and showed ploidy distribution similar to previously reported levels[57]. Finally, the adipic acid tolerant phenotype was shown to be inducible: pre-treatment in adipic acid improved growth and reduced latency in strains exposed to adipic acid stress.

3.1.1 Our collection of strains displays a wide range of adipic acid tolerant phenotypes

A collection of 500 tentatively acid tolerant strains were obtained from Lallemand Inc. In order to assess tolerance to adipic acid, these strains were screened using a high-throughput method to compare growth yields in adipic acid to growth without adipic

acid. To do so, strains were inoculated in 100 μL YNB supplemented with 20 g/L glucose with and without 20 g/L adipic acid and grown for 48 hours. During this time, growth was measured by OD_{600} in a TECAN plate reader. Upon completion, the strains showed a wide range of adipic acid tolerant phenotypes (Figure 3.1): some strains grew well in both conditions while others were unable to grow at all in the acid condition.

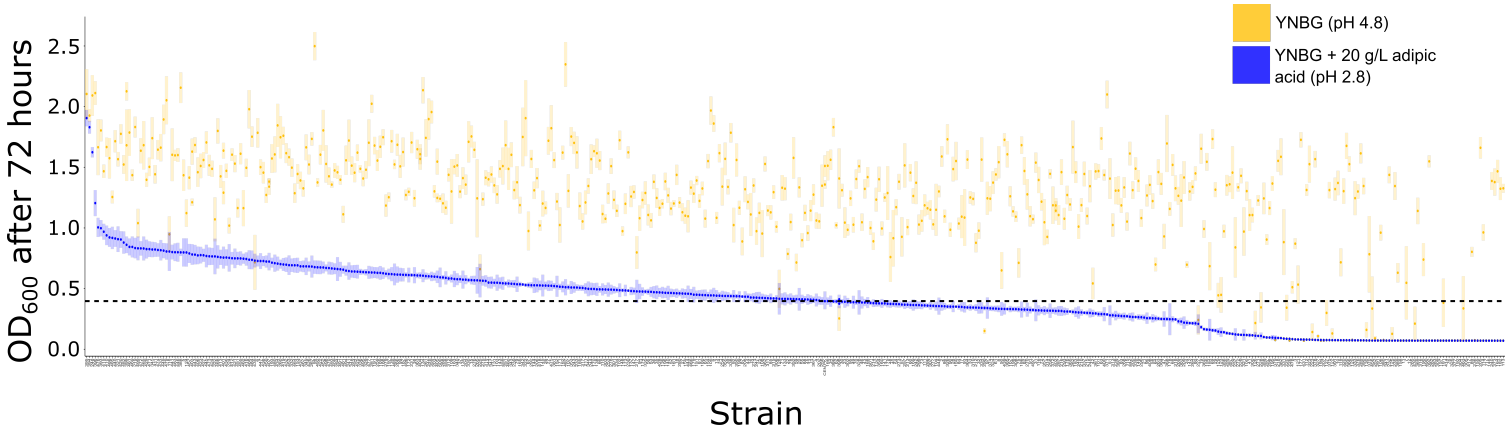

Figure 3.1: Comparison of growth of our strain collection in adipic acid shows a wide range of adipic acid tolerant phenotypes. Strain growth was determined by measuring OD_{600} for strains grown with (blue) and without (yellow) adipic acid. Strain CEN.PK113-7D, a well known lab strain, was used as a comparison; the dotted grey line represents the OD_{600} reached by CEN.PK113-7D under adipic acid stess. Data is ordered by decreasing OD_{600} in adipic acid and shown as mean $\pm$ standard error of 4 or 6 biological replicates.

3.1.2 Top adipic acid strains grow well with and without adipic acid supplementation

Since the ultimate goal is to produce adipic acid, we wanted the top strains to be able to grow well in adipic acid to reduce product toxicity. A subset of the most suitable strains was selected from the strain collection by comparing growth yields for each strain in adipic acid to its growth without adipic acid (Figure 3.2). Strains with a growth ratio of 1, i.e. strains that can grow as well in adipic acid as in non-acid media, are desirable. However, growth ratio alone is not enough to select the top strains since some strains have a growth ratio near 1 but poor growth yields in both conditions. Consequently, the list of top candidate strains was determined by selecting strains with an OD_{600} after 48 hours greater than 1.5 in YNBG and a growth ratio close to 1. Note that, while non-*S. cerevisiae* strains were present in the 500 strain collection, only the 473 *S. cerevisiae* strains were considered for work following the screening phase.

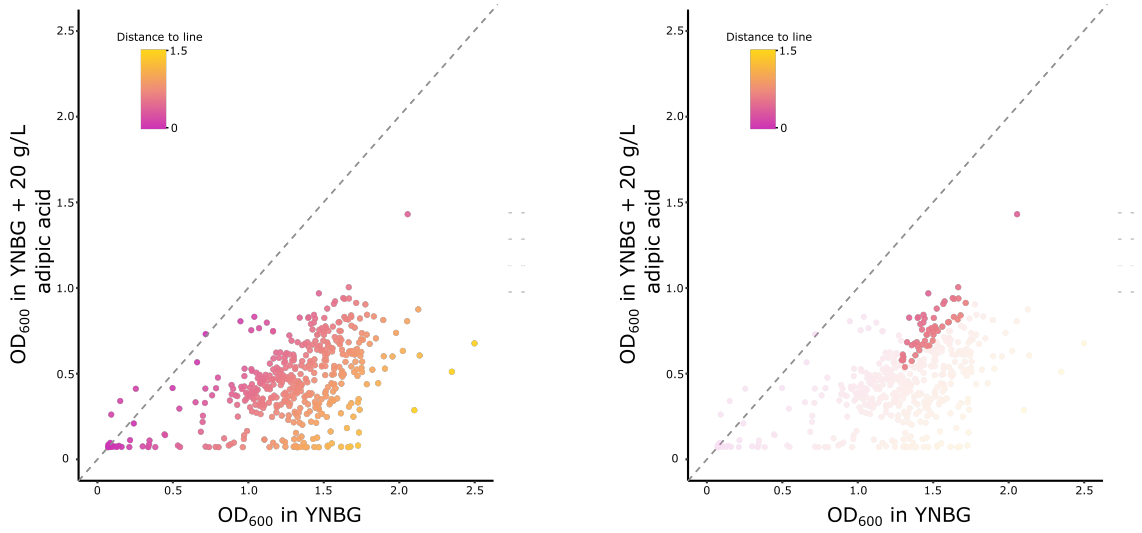

Figure 3.2: Top adipic acid strains were identified by comparing growth yields after 48 hours with and without adipic acid. Strains with a ratio close to 1 and with an OD_{600} greater than 1.5 in YNBG were selected for further characterisation (right).

3.1.3 Top strains outperform lab strain CEN.PK113-7D in adipic acid

Next, strain growth yield was assessed in larger volumes to confirm the high-throughput results and to assess scalability . Top adipic acid tolerant strains were inoculated in 1 mL YNB with 20 g/L glucose with and without 20 g/L adipic acid. OD_{600} was measured after 48 hours using a TECAN plate reader. The top adipic acid tolerant strains showed growth yields 2-5 times greater in adipic acid compared to the lab strain CEN.PK113-7D (Figure 3.3). Most strains, including CEN.PK113-7D, grew equally well in YNBG yet, in adipic acid, even the adipic acid tolerant strains showed variable growth – e.g. L433 has higher OD_{600} than L410.

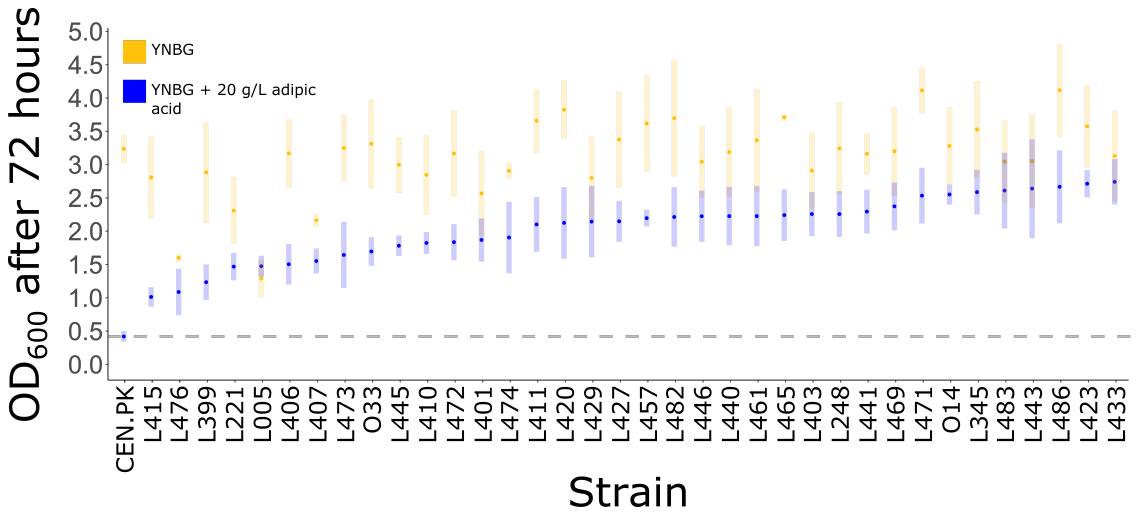

Figure 3.3: In larger volumes, top adipic acid tolerant strains consistently showed more desirable growth characteristics than the lab strain CEN.PK113-7D. Data is ordered by decreasing OD_{600} in adipic acid and shown as mean ± standard error of 3 replicates.

3.1.4 Top adipic acid tolerant strains are mostly diploid

Changes to ploidy can have significant effects on fitness in yeast[99]. Many commercial and tolerant yeasts are found with >2n genomes and genome-wide copy number changes[100], which are associated with adaptation to stresses and improved fitness both in *S. cerevisiae*[101] and in other fungi[102]. Ploidy also influences later genetic manipulations, with haploids being generally preferred to manipulate while diploids and aneuploids are more tolerant to different conditions. To assess ploidy in our top adipic acid tolerant strains, each

strain was stained with propidium iodide, a DNA intercalating agent. Using flow cytometry, propidium iodide induced fluorescence – an approximation of DNA content within each cell – was compared to the lab strain CEN.PK113-7D, a known haploid (Table 3.1). Most of the acid tolerant strains had an estimated genome size roughly twice as large as the haploid CEN.PK113-7D and were consequently considered diploid. Others, notably L221, L443, and L394, had significantly larger estimated genomes and were closer to tetraploidy. Of the strains shown here only four, L441, L393, L473, and L420, had estimated genomes only slightly larger than CEN.PK113-7D.

Table 3.1: Most of the top adipic acid tolerant strains are diploid. Ploidy was approximated by measuring propidium iodide stained DNA and comparing levels with the known haploid strain CEN.PK113-7D. Ploidy is presented as a ratio of the haploid: a ploidy of 2 represents twice as much DNA as CEN.PK113-7D. Genome size estimates are based on comparison to CEN.PK113-7D.

Strain	Estimated Genome Size (Mbp)	Ploidy	Strain	Estimated Genome Size (Mbp)	Ploidy
CEN.PK113-7D	11.65	1	L445	28.20	2.42
L005	28.20	2.42	L446	22.68	1.94
L221	45.37	3.89	L433	23.91	2.05
L345	25.13	2.15	L457	21.46	1.84
L403	25.13	2.15	L465	24.52	2.10
L394	38.13	3.27	L461	23.91	2.05
L395	24.52	2.10	L469	25.75	2.21
L396	20.23	1.73	L471	22.07	1.89
L406	23.91	2.05	L473	18.39	1.57
L410	25.13	2.15	L472	22.68	1.94
L407	25.75	2.21	L420	15.94	1.36
L474	24.52	2.10	L423	26.36	2.26
L476	22.68	1.94	L427	24.52	2.10
L401	23.91	2.05	L429	25.13	2.15
L399	29.43	2.52	L411	23.91	2.05
L486	24.52	2.10	L415	25.13	2.15
L219	25.75	2.21	L482	26.36	2.26
L441	17.78	1.52	L483	22.07	1.89
L440	23.91	2.05	L393	16.55	1.42
L443	33.72	2.89	L485	24.52	2.10

3.1.5 Tolerance to adipic acid is an inducible phenotype in the top adipic acid tolerant strains

Previous reports have shown that exposure to acid stress has inhibitory effects on growth[48;69;103]. After exposure, only a fraction of cells are able to continue growth. This results in a latency period where growth is not observed until the tolerant cells within the population can repopulate the culture. The latency period caused by acid stress is dose- and strain-dependent[104] and gives the cells time to adapt their cellular processes to deal with the acid stress[48]. Latency can be reduced by pre-treatment – cells that are pre-adapted to a stress show decreased latency upon subsequent exposure to that acid stress[69;103].

In order to assess if tolerance to adipic acid is similarly inducible, the adipic acid tolerant strain L433 and non-tolerant strain CEN.PK113-7D were pre-cultured in YNBG supplemented with 0, 10, or 20 g/L adipic acid for 4 hours. Each pre-culture was inoculated to a set starting OD_{600} of 0.1 into YNBG supplemented with 0, 10, or 20 g/L adipic acid and OD_{600} was measure over 48 hours (Figure 3.4). For cultures grown in YNBG alone, growth was largely independent of pre-culture. For CEN.PK113-7D, growth in 20 g/L adipic acid was largely inhibited regardless of the pre-culturing regime. However, for L433, cultures grown in YNBG supplemented with 20 g/L adipic acid performed better after being pre-cultured in adipic acid compared to pre-cultured in YNBG alone. Pre-culture in 20 g/L adipic acid made L433 better able to grow in high levels of adipic acid compared to the intermediate pre-culture of 10 g/L adipic acid which, in turn, provided better growth than pre-culturing without adipic acid.

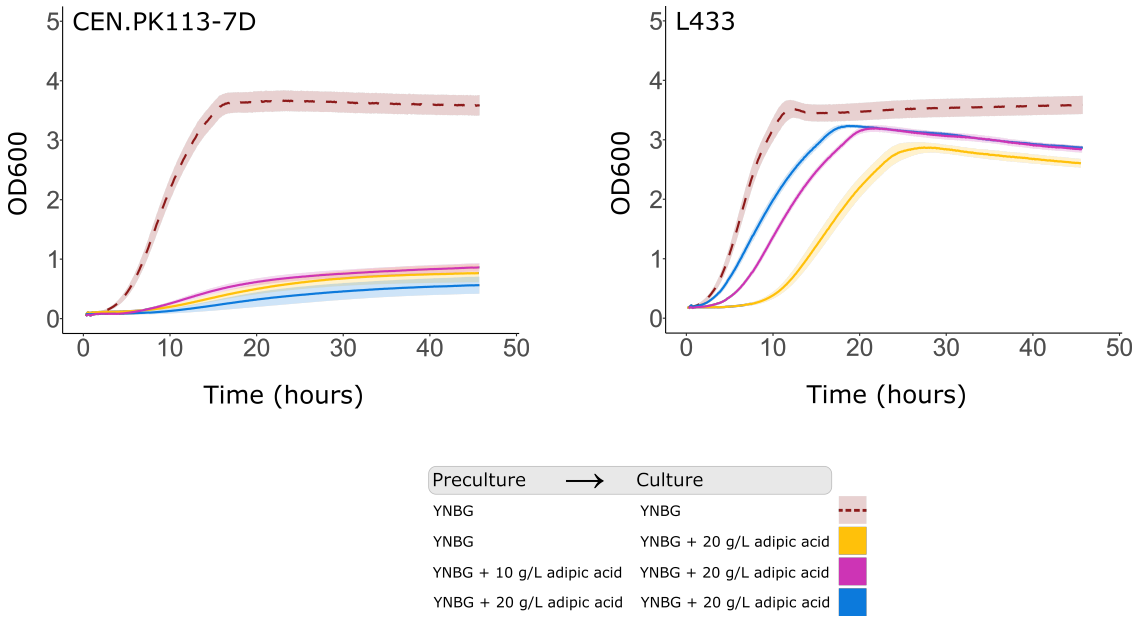

Figure 3.4: Tolerance to adipic acid is inducible. Strains were pre-cultured in YNBG supplemented with 0, 10, or 20 g/L adipic acid before being inoculated in YNBG supplemented with 0 or 20 g/L adipic acid. OD_{600} measurements were taken every 5 minutes for 48 hours using a TECAN Sunrise platereader. Data are shown as mean of 4 replicates ± standard error

3.2 Comparative gene expression

Comparative gene expression analysis from RNAseq experiments provides insight into the causes behind a phenotype. Previous work, such as Dong et al[55], has shown the wide range of genes affected by acid stress. Here, we show that gene expression is altered in both adipic acid tolerant and non-tolerant strains that undergo adipic acid stress. Furthermore, we provide comparisons between the adipic acid stress response of adipic acid tolerant strains to the adipic acid non-tolerant strain CEN.PK113-7D. Five strains, the adipic acid tolerant L433, L248, and O33 and non-tolerant strains CEN.PK113-7D and L295, were cultured with and without adipic acid stress. RNA from each strain and condition was extracted, converted to cDNA, and sequenced for gene expression analysis.

3.2.1 Sample separability determines identification of Differentially Expressed Genes

The identification of Differentially Expressed Genes (DEGs) is predicated on expression differences between different conditions, in our case presence or absence of adipic acid in the culture medium. It is therefore necessary to assess RNAseq samples both for consistency within treatment – i.e. that samples in the same treatment group have a similar response – and for separability between conditions. To do so, multidimensional scaling (MDS) analysis was carried out using EdgeR[73]. Briefly, each sample is represented by multiple dimensions with each dimensional axis representing the expression data of a single gene. Dimensionality reduction algorithms collapse samples data to compare overall dimensional similarities between samples. Samples with similar gene expression profiles across all axes are separated by less distance than samples with less similar total gene expression profiles. For each strain, sample similarities were compared to determine how well we can separate gene expression values from adipic acid treated cells from untreated samples (Figure 3.5). Percent variability explained by the first and second dimension are 49.8 and 30.8 for CEN.PK113-7D, 58.4 and 28.7 for L433, 41.7 and 34.7 for O33, 53.4 and 23.9 for L248, and 47.9 and 33.6 for L295. In the cases of CEN.PK113-7D, L433, O33, and L248, adipic acid treated samples cluster distinctly from the untreated samples. However, within these strains there exists differences in gene expression profile distinguishability, which influences downstream identification of DEGs. For example, L248 shows separation between treated samples and untreated samples only on the x-axis but, when considering the y-axis alone, sample conditions are indistinguishable. The same is true to a lesser extent in O33. For O33 and L248 the number of DEGs is noticeably lower than CEN.PK113-7D and L433, both of which had much more convincing separability (see Figure 3.6b below). For L295, four of the six samples, two from the untreated group and two from the adipic acid treated group, are indistinguishable using this method. Consequently, since sample-condition distinguishability is essential for the identification of DEGs, samples from L295 were removed from further analysis.

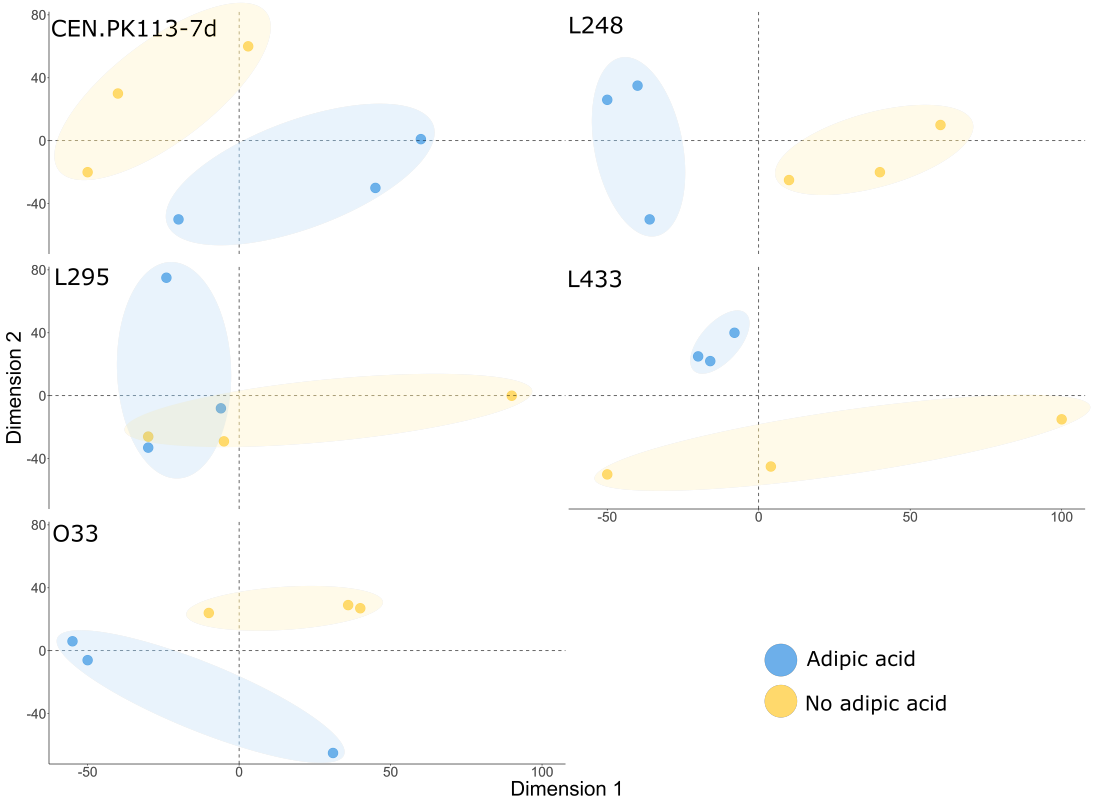

Figure 3.5: MDS plots reveal good separation of conditions per strain except in L295. MDS was used to visualise differences in samples from different conditions for each strain. X-axis represents dimension 1. Strains CEN.PK113-7D, L433, L248, and O33 show distinguishable grouping of the acid-treated samples compared to the un-treated samples. For strain L295 sample variability is harder to attribute to condition; consequently DE is obscured.

After confirming that there was sufficient sample variability between conditions, DEGs were determined from gene expression data by pairwise differential expression analysis using EdgeR[73]. Expression data were normalized using the Trimmed Mean of M-values method[88;105]. Genes were considered DE by selecting those with a fold change (FC) between the treated samples to untreated samples greater than 2 and with a False Discovery Rate (FDR) less than 0.05. DEGs for each strain are represented in Figure 3.6a, with significantly up-regulated genes shown as blue points and significantly down-regulated genes shown in yellow. Genes that were below the FDR cutoff are shown in grey while genes that are significant but who do not show sufficient FC are shown in black. The number of DEGs varied greatly between strains (Figure 3.6b). For CEN.PK113-7D, 356 genes were up-regulated and 210 down-regulated under adipic acid stress. Likewise, 225 up-regulated and 240 down-regulated genes were identified in L433; 177 up-regulated and 93 down-regulated in

L248; 51 up-regulated and 6 down-regulated in O33. The full list of DEGs identified per strain is available in the appendix.

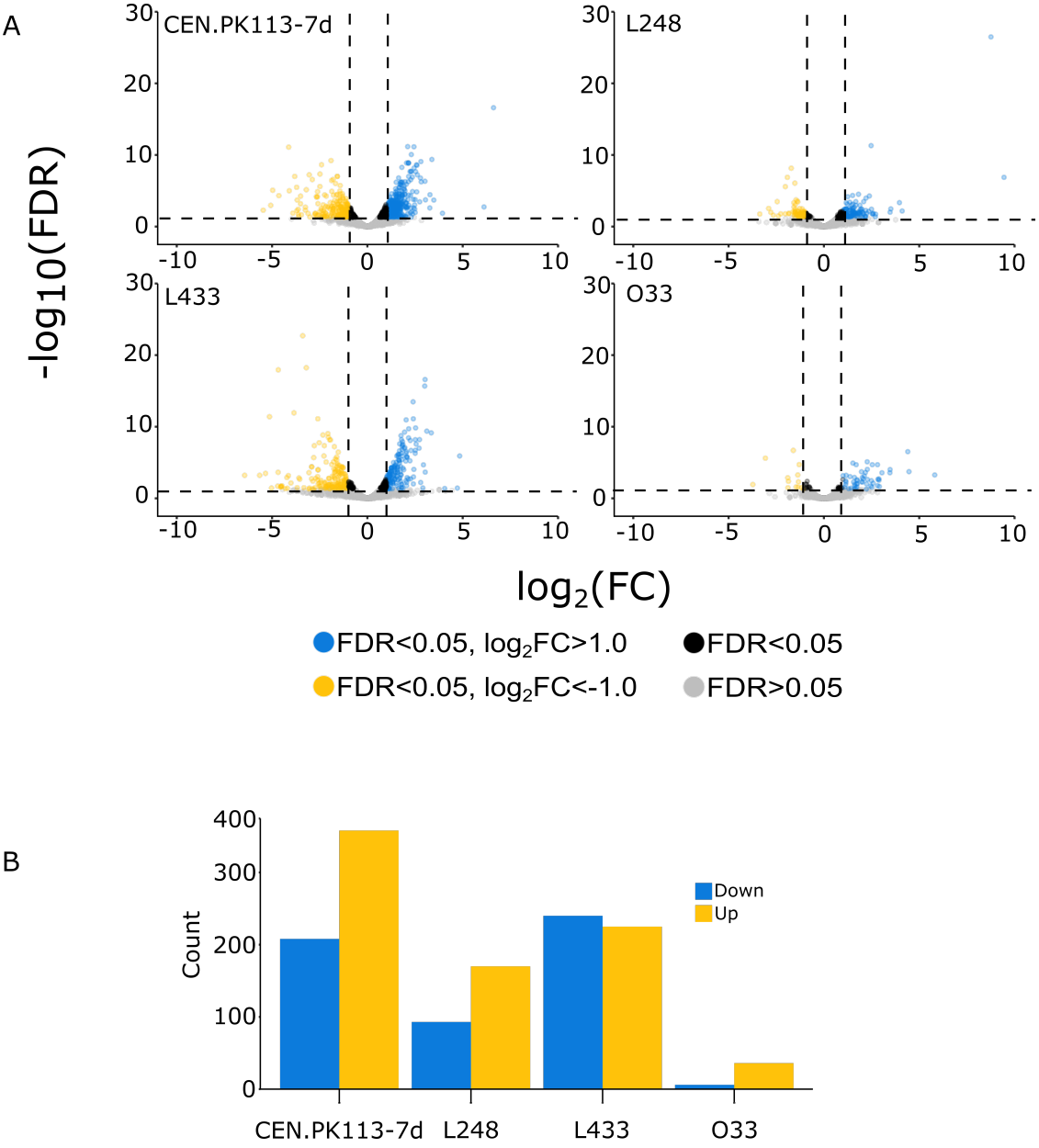

Figure 3.6: **A** DEGs identified in strains CEN.PK113-7D, L433, L248, and O33 after exposure to adipic acid. Each DEG is represented by a coloured dot: significantly up-regulated and significantly down-regulated in blue and yellow. Genes are not considered DE if they have a FC between -2 and 2 or if the FDR is greater 0.05. **B** Total number of DEGs identified in CEN.PK113-7D, L433, L248, and O33.

3.2.2 Adipic acid stress results in cell-wide transcriptional changes

Exposure to adipic acid results in the differential expression of various genes. Within the four strains a total of 1366 genes were differentially expressed. The top differentially expressed genes, ranked by FDR, for each strain are shown in Figure 3.7, Figure 3.8, Figure 3.9, and Figure 3.10. Relative gene expression is shown by normalised counts with genes highly expressed in the adipic acid-treated samples (A1, A2, A3) in blue and genes lowly expressed in the treated samples in yellow, compared to the un-treated samples (Y1, Y2, Y3). For each strain, a maximum of 75 genes, selected by lowest FDR, is shown; the full list of DEGs per strain is available in the appendix.

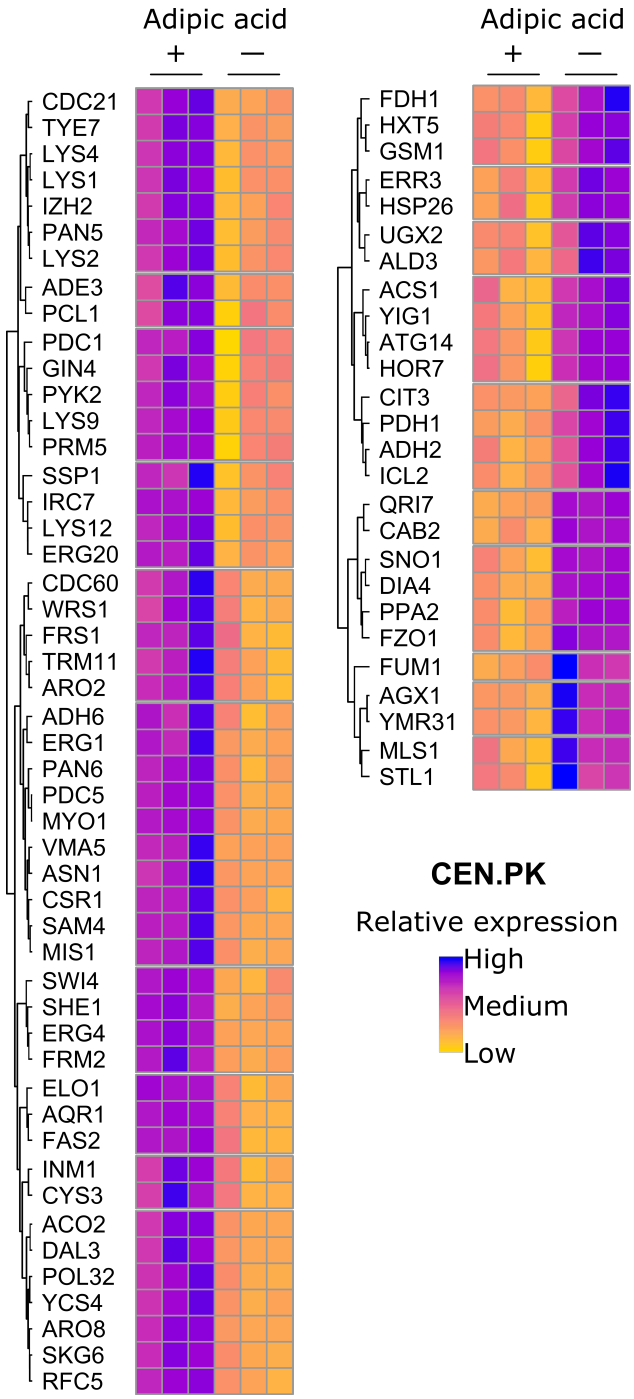

Figure 3.7: Top differentially expressed genes identified in CEN.PK113-7D. DEGs were identified using EdgeR[73] with only the top 75, ranked by lowest FDR, shown here. Relative gene expression, as shown by normalised gene count, is shown in blue and yellow for genes with high and low expression, respectively, from the adipic acid-treated samples compared to un-treated samples.

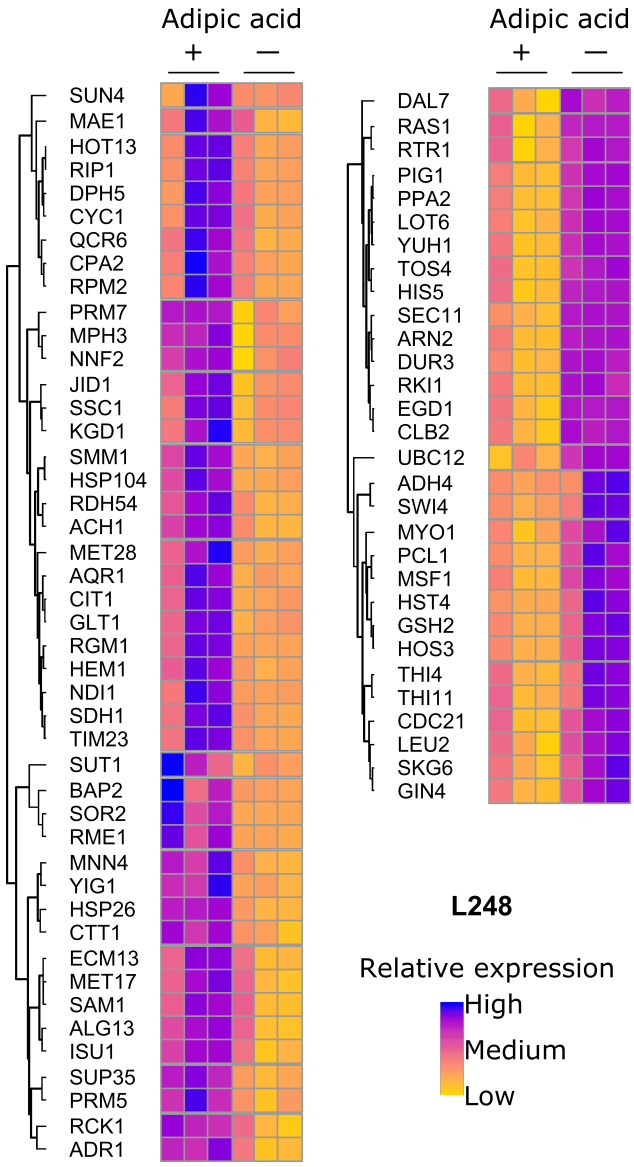

Figure 3.8: Top differentially expressed genes identified in L248. DEGs were identified using EdgeR[73] with only the top 75, ranked by lowest FDR, shown here. Relative gene expression, as shown by normalised gene count, is shown in blue and yellow for genes with high and low expression, respectively, from the adipic acid-treated samples compared to un-treated samples.

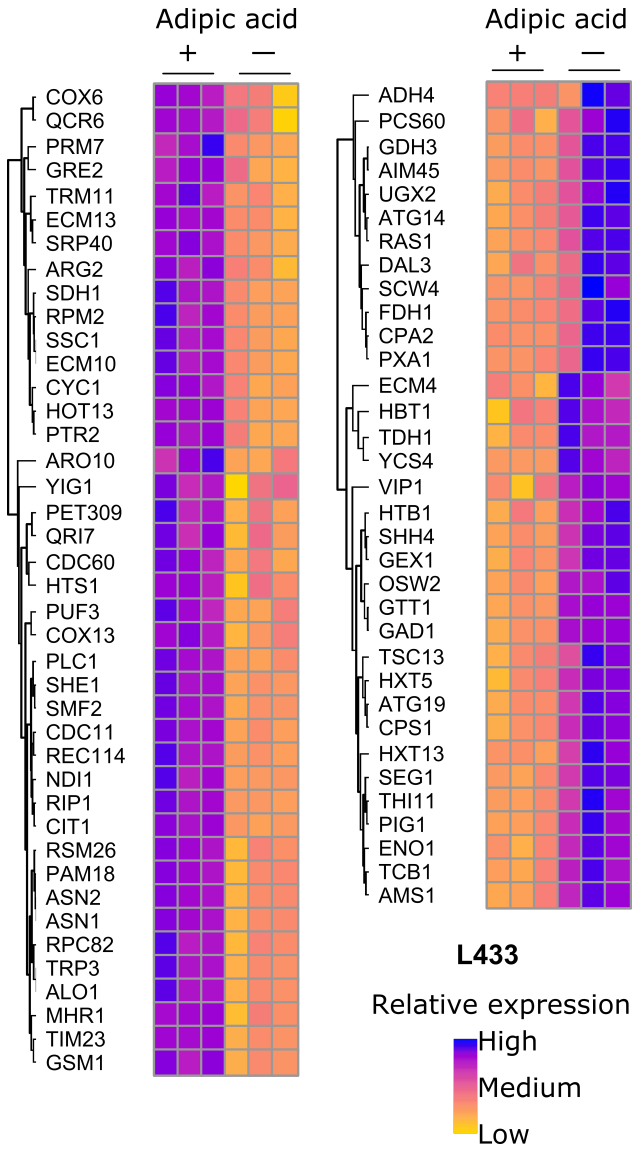

Figure 3.9: Top differentially expressed genes identified in L433. DEGs were identified using EdgeR[73] with only the top 75, ranked by lowest FDR, shown here. Relative gene expression, as shown by normalised gene count, is shown in blue and yellow for genes with high and low expression, respectively, from the adipic acid-treated samples compared to un-treated samples.

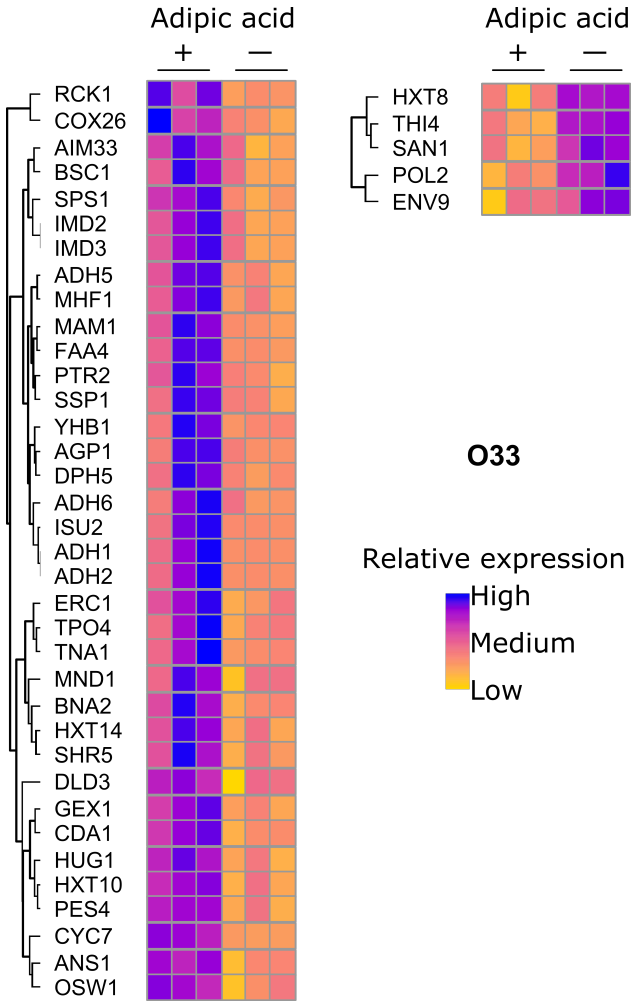

Figure 3.10: Top differentially expressed genes identified in O33. DEGs were identified using EdgeR[73]. Relative gene expression, as shown by normalised gene count, is shown in blue and yellow for genes with high and low expression, respectively, from the adipic acid-treated samples compared to un-treated samples.

The large number of DEGs among different strains makes analysis complex and necessitates a bioinformatic workflow to derive biological significance from the genes identified. Indeed, direct gene-by-gene comparisons are cumbersome and do not provide much functional value[106]. Comparisons of Gene Ontology (GO) offer a simpler and more complete way of assessing large gene lists for functional enrichment patterns. In this method, each gene is described by molecular function, pathways in which it is involved, and cellular localization; by grouping genes by function or process we can gain better insight as to the biological effects of adipic acid stress.

Identified DEGs were annotated with GO terms using gProfiler[89] and STRING[90].

94.8% of the identified DEGs were linked to GO terms. Overall, 259 unique GO terms were assigned to DEGs, representing a wide range of cellular responses including translation and ribosomal protein production, carboxylic acid metabolism, aerobic respiration, and oxidoreductase activity. Within these terms there is significant overlap. Not only are genes multiply assigned with different terms but many of the terms, due to the hierarchical format of GO terms, are redundant – e.g. *ALD2* is annotated with GO:0006082, GO:0019752, GO:0032787, and GO:0043436, among others, which translates to organic acid metabolism, carboxylic acid metabolism, monocarboxylic acid metabolism, oxoacid metabolism. The term-within-term approach of GO terms provides depth to biological function yet can also provide a false sense of the range of affected processes. While enriched terms such as cellular process, metabolic pathways, and cytoplasm are very broad and found in both the up-regulated and down-regulated gene lists, they do provide evidence of the cell-wide transcriptional changes caused by adipic acid stress.

3.2.3 Gene clustering identifies up- and down-regulated subsets of differentially expressed genes in adipic acid tolerant and non-tolerant strains

Gene clustering identifies up- and down-regulated subsets of differentially expressed genes in adipic acid tolerant and non-tolerant strains. Up- and down-regulated gene lists were grouped using a previously described Markov cluster algorithm (MCL) for proteins[107] in STRING[90]. Briefly, gene-based protein similarities are represented as a network of connected nodes. Pairwise node interactions are iteratively weighted and compared until a stable gene network is achieved. MCL clustering of up-regulated genes reveals 109 different clusters (Figure 3.11a), of which 57 contained at least 3 genes. The top clusters by gene count were related to translational and ribosomal genes (90), RNA processing and biogenesis (65), DNA replication, DNA repair, and cycle related genes (31), respiration, redox, and proton transmembrane transport (16), fatty acid biosynthesis and cell membrane regulation (12), aromatic amino acid biosynthesis(11), sterol biosynthesis (11), protein folding and localization (10), and cell wall organization (10). Down-regulated genes were clustered into 158 groups (Figure 3.11b), with 33 containing at least 3 genes. Among the largest gene clusters were those related to respiration and proton transmembrane transport (12), fatty acid and carboxylic acid metabolism (10), autophagy (10), carbohydrate metabolism (9), and TCA cycle (7).

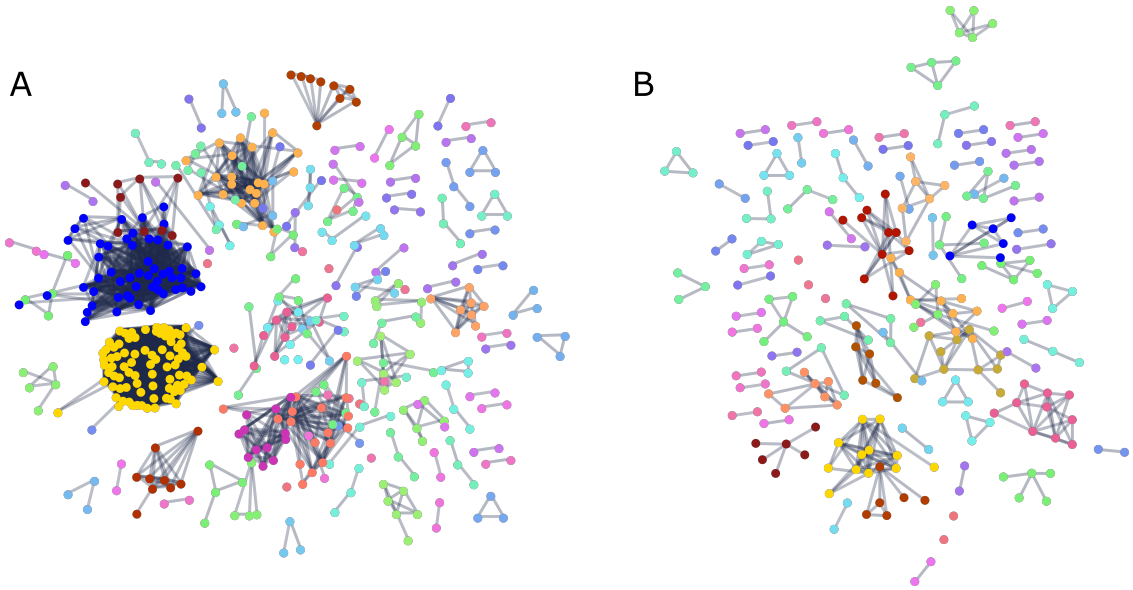

Figure 3.11: MCL clustering of total up-regulated (A) and down-regulated (B) genes identified under adipic acid stress in strains CEN.PK113-7D, L433, L248, and O33 show distinct grouping. Each node represents a gene product with edges showing interaction. Up-regulated gene clusters showed enrichment of genes related to ribosomes and translation, RNA processing, DNA repair, respiration and redox, cell membrane regulation, and protein folding. Down-regulated genes clustered in fewer number but represent proton transport, fatty acid and carboxylic acid metabolism, and carbohydrate metabolism.

3.2.4 Strain-by-strain responses to adipic acid stess reveal similar overarching pathway enrichment

DEG analysis on a strain-by-strain basis reveals differential gene expression profiles. Clusters of genes were primarily associated with gene expression, respiration and the mitochondria, oxidoreductase activity, and transporter activity. Below, prominent gene clusters are described for up- and down-regulated processes per strain.

CEN.PK113-7D

CEN.PK113-7D had a large number of up-regulated genes including those related to translation (80 genes such as *RPP*-, *RPL*-, and *RPS*-family genes), DNA replication (*MCM5*, *MCM6*, *MCM2*, *POL1*, *POL2*, *POL32*, *POL30*) and repair (*RCF5*, *MSH2*, *RFC3*), ribosome

biogenesis (*NOP58*, *NOP13*, *HAS1*, *RRP3*, *RPA43*) and RNA modification (*TRM11*, *PUS4*, *EMG1*), tRNA metabolism (*HTS1*, *ALA1*, *MES1*, *FRS1*, *TYS1*, *CDC80*), sterol biosynthesis (*ERG1*, *ERG2*, *ERG3*, *ERG4*, *ERG11*, *ERG26*, *ERG20*, *NCP1*, *MVD1*, *ELO1*, *ELO3*, *OLE1*) and fatty acid biosynthesis (*IZH2*), lysine biosynthesis (*LYS1*, *LYS2*, *LYS4*, *LYS9*, *LYS12*, *LYS21*, *ACO2*) and related 2-oxocarboxylic acid metabolism (*LYS4*, *LYS21*, *LYS12*, *ACO2*), cysteine and methionine biosynthesis (*MET3*, *MET10*, *MET14*, *MET16*, *MET22*, *SAM4*, *IRC7*, *CYS3*), cell wall metabolism (*GAS5*, *UTR2*, *YPS3*, *CRH1*), aromatic amino acid biosynthesis (*ARO8*, *ARO4*, *TYR1*, *TRP2*), protein folding (*SSZ1*, *ZUO1*, *ZPR1*, *YAR1*), and MAPK signalling (*PXL1*, *RHO2*, *MYO1*). Of the down-regulated genes in CEN.PK113-7D, the largest clusters included fatty acid catabolism (*CIT3*, *PDH1*, *ICL2*) and oxidation (*POT1*, *MDH3*, *FOX2*, *DCI1*, *ECI1*), amino acid catabolism (*EDH3*, *KGD2*, *MDH3*, *FOX2*, *DCI1*, *POT1*, *FDH1*, *CIT3*, *ICL2*, *PDH1*, *PDC5*, *ADH2*, *ARO10*, *ARO9*, *UGA1*, *CTT1*), transmembrane transporter activity (*PMA2*, *ARP4*, *ATP6*, *ATP15*, *COX3*, *COX5B*, *COX6*, *COX7*, *COX8*, *COX13*, *COX17*, *CTR1*, *CTR3*, *HXT5*, *ARO2*, *POR2*, *CRC1*, *DUR3*, *SSU1*), various oxidoreducatases (e.g. *FDH1*, *ADH2*, *COX*-family genes, *ALD5*), and glutothione transferases (*GTO1*, *GRX1*, *ECM4*, *GTT1*).

L433

Similarly, up-regulated genes in strain L433 included those related to gene expression (e.g. *PPM2*, *PDR16*, *MOT1*, *TRM11*) especially of genes related to the mitochondrion (*HTS1*, *MRPS34*, *RMS26*, *MRPL22*, *MRPL35*, *MRP7*) as well as RNA processing and maturation (*DEG1*, *RPM2*, *PRP16*, *PAP1*, *PPM2*, *TRM11*, *NOP4*, *NOP9*, *RRP9*). Furthermore, genes related to mitochondrial translation were enriched (*AEP2*, *MRPL22*, *MRPL35*, *MRPL40*, *MHR1*, *MRPS35*, *HTS1*, *NAM9*, *TUF1*) as were processes related to targeting of proteins to the mitochondrion (*TIM12*, *TIM23*, *TOM71*, *ECM10*, *HOT13*, *HSP60*, *SSC1*). Respiration (*COX6*, *COX13*, *PUF3*, *RIP1*, *QCR2*, *QCR6*, *CYC1*, *ALG2*, *CIT1*) and transmembrane transport were also up-regulated (*FCY2*, *HOT13*, *JEN1*, *POR1*, *ATP16*, *QCR2*, *QCR6*, *TIM23*, *TIM11*, *PEX25*, *SSC1*, *PAM18*, *COX6*, *COX13*, *PTR2*). Finally, DNA repair processes (*ECO1*, *PSF2*, *MCM4*, *POL32*, *TOP1*) and protein re-folding (*SSC1*, *HSP60*, *ECM10*) were also up-regulated under adipic acid stress. Meanwhile, down regulated genes included those related to carbohydrate metabolism (*ALD3*, *HBT1*, *IMA4*) and import (*HXT5*, *HXT7*, *HXT11*, *HXT13*), MAPK signalling (*MID2*, *KDX1*, *MKK2*), glutathione transferases (*ECM4*, *GTT1*), thiamine synthesis (*SNO2*, *THI20*, *THI11*, *THI22*), reduction in expression of H+ ATPases (*VMA5*, *VMA6*), and

decreased transporter expression. Interestingly, the up-regulated transporters tended to be related to the mitochondrial membrane and respiration (*COX6*, *COX13*, *RIP1*, *SSC1*, *POR1*, *SPF1*, *TIM23*, *QCR6*, *QCR21*, *ATP16*, *HOT13*, *TIM11*, *TIM71*) whereas the down-regulated transporters were related to sugar uptake or associated with the plasma membrane (*VAM5*, *VMA6*, *MTM1*, *HXT5*, *HXT6*, *HXT7*, *HXT11*, *HXT13*, *SAM3*, *UGA4*, *ATY7*, *PXA1*, *FAT1*).

L248

L248 showed increased expression of genes related to stress responses (*HSP104*, *HSP26*, *HSP78*, *CTT1*, *TPS1*, *TSL1*, *NTH1*, *HSP42*, *IGD1*, *PIC2*, *SUP35*), increased respiration (*QCR6*, *QCR9*, *CYC1*, *COR1*, *SDH1*, *COA2*, *COA4*, *COX15*) and related proton transmembrane transport (*QCR6*, *QCR9*, *COR1*, *RIP1*) as well as up-regulation of NADH oxidation (*NDI1*, *NDE1*). Additionally, amino acid export was increased (*AQR1*, *PRM7*) along with increased carboxylic acid metabolism (*CIT1*, *ACH1*, *MAE1*, *KGD1*). Fewer and less clear clusters were identified of the down-regulated genes but those identified included decreased expression of genes related to cell division (*GIN4*, *MYO1*, *ACM11*, *CLB2*), autophagy (*PEP3*, *ERS1*, *RAS1*, *YPT7*), and mitochondrial membrane proteins (*MDM10*, *SAM50*).

O33

While relatively few DEGs were identified in strain O33, clustering analysis still managed to show enriched pathways including the up-regulation of ethanol and related amino acid catabolism to alcohols (*ADH1*, *ADH2*, *ADH5*), increased oxidoreductase activity (*DLD3*, *ADH1*, *ADH2*, *ADH5*, *ADH6*), and increased GTP and GMP biosynthesis (*IMD2*, *IMD3*). Due to the low number of down-regulated DEGs no clustering was possible. Notably, the hexose transporter *HXT8* was down-regulated.

3.2.5 Gene content, but not gene expression, is consistent between adipic acid tolerant and non-tolerant strains

One of the ways that different phenotypes can arise is through different gene content. For example, in section 3.3 below we show that the integration of a single gene changes non PCA producing strains into PCA producing strains. Analysis of the *S. cerevisiae* pan-genome, i.e. the collection of all genes contained within the species, estimates that it comprises between 7,000 and 8,000 genes[57;108;109]. Of these, roughly 75% are considered core genes that are found across 95% of all species of *S. cerevisiae*[109]. While non-core genes may offer avenues for different phenotypes, often gene content alone is not enough to explain intra-species variation[109], which can rely instead on factors complementary to gene content such as chromosomal rearrangement[110] and differential gene expression (review by Sumner and Avery[111]).

Together, the genomes of L433, L248, and O33 contain 1941 unique – i.e. not repeated between the three strains – genome features not found in CEN.PK113-7D. Features were further clarified by removing putatively annotated and broadly defined – e.g. 'reductase activity' rather than a gene identifier – features, which left 515 genes not shared between CEN.PK113-7D and the adipic acid tolerant strains. On its face, this suggests that gene content may be the driver of different levels of adipic acid tolerance. However, many of these non-shared genes were not DE (453/515) and, of those that were DE, many of the DEGs took part in similar processes or had the same function as in CEN.PK113-7D. For example, *SDH3* was DE in CEN.PK113-7D but not in any of the adipic acid tolerant strains while *SHH3*, an *SDH3* paralog, was DE in the adipic acid tolerant strains. Similarly, *COX17* was found in adipic acid tolerant strains but not CEN.PK113-7D whereas the *CXO17* homolog *COX23* was. Of the genes not common between the adipic acid tolerant strains and CEN.PK113-7D 166 were determined to be paralogs. Additionally, 27 DEGs were identified as being non-common between CEN.PK113-7D and the adipic acid tolerant group after accounting for paralogous genes and non-specific annotations. These genes include *ARN2*, *THI11*, *COA2*, *AAD16* (a relative of CEN.PK113-7D:*AAD14*), *SOR2*, *HSP42*, *MPH3*, *IMA5* (similar to CEN.PK113-7D:*IMA1* and CEN.PK113-7D:*IMA2*), *PRM7*, *MCH2* (similar to CEN.PK113-7D:*MCH1*, CEN.PK113-7D:*MCH4*, and CEN.PK113-7D:*MCH5*) *COS7*, *SAM3*, *SNO3*, *HXT11*, *RRT15*, *SGE1* aka *NOR1*, *JEN1*, *FSP2* (similar to CEN.PK113-7D:*IMA1*), *ENV9*, *HXT8*, *SHR5*, *IMD2* (similar to CEN.PK113-7D:*IMD3*). Within this subset of DEGs, only some stand out as having potentially beneficial function for improved adipic acid tolerance. For example, the multidrug transporter *SGE1*, the

carboxylate/proton symporter *JEN1*, and the monocarboxylic acid transporter *MCH2*.

While gene content may be nearly identical between CEN.PK113-7D and the adipic acid tolerant strains, gene expression and consequently phenotype varies. Differential gene expression analysis shows that a variety of genes are oppositely expressed in the adipic acid tolerant group compared to CEN.PK113-7D, highlighted in Table 3.3.

Table 3.3: Certain genes are oppositely expressed between the adipic acid tolerant strains and CEN.PK113-7D

Gene	Acid tolerant			Non-tolerant
	L433	O33	L248	CEN
ADH2		up		down
ALG13			up	down
ARO10	up			down
COX13	up			down
COX6	up			down
CTT1			up	down
DCI1			up	down
FZO1	up			down
GSM1	up			down
HSP26			up	down
PTR2	up	up		down
QRI7	up			down
REC114	up			down
RIP1	up		up	down
SDH1	up		up	down
STL1	up			down
TPN1	up			down
YIG1	up		up	down
CDC21			down	up
DAL3	down			up
DAL7			down	up
ERG3			down	up
GEX1	down			up
GIN4			down	up
HXT6	down			up
LYS2	down			up
MTQ1	down			up
MYO1			down	up
NUP188			down	up
PCL1			down	up
POL2		down		up
RFC5	down			up
RPS1b			down	up
SCW4	down			up
SKG6			down	up
SWI1			down	up
TOS4			down	up
VMA5	down			up

Continued on next page

Gene	Acid tolerant			Non-tolerant
	L433	O33	L248	CEN
YCS4	down			up

3.2.6 Gene expression analysis reveals different pathway effects in adipic acid tolerant strains compared to CEN.PK113-7D

DEG comparison between CEN.PK113-7D and the adipic acid tolerant group shows that some pathways are similarly affected during adipic acid stress while others are differentially expressed. Both the adipic acid tolerant strains and CEN.PK113-7D undergo increased gene expression, although the exact genes involved vary as shown by the up-regulation of *RPS*- and *RPL*-family genes in CEN.PK113-7D and by up-regulation of *DRS1*, *RPA4*, *MAK21*, *BUD22*, *NOP7*, and *UTP*-family genes in the adipic acid tolerant group. Pathways with different expression during adipic acid stress include genes related to stress responses, particularly protein folding. In the adipic acid tolerant strains genes such as *HSP104*, *HSP60*, *HSP26*, *HSP42*, *HSP78*, *ECM10*, *MDJ1*, and *SSC1* were up-regulated whereas only one similar gene, *HSP105*, was enriched in CEN.PK113-7D. Furthermore, the adipic acid tolerant strains showed increased oxidoreductase activity through genes related to respiration and central metabolism such as *IMD2*, *IMD3*, *ADH5*, *ADH1*, *ADH2*, *ADH6*, *CTT1*, *KGD1*, *SOR1* and related NADH dehydrogenase activity (*COX6*, *COX15*, *COX26*, *CYC7*, *CYC1*, *QCR6*, *QCR9*). In CEN.PK113-7D oxidoreductase activity is also increased but is largely related to different processes from the adipic acid tolerant strains such as ergosterol biosynthesis (*ERG1*, *ERG4*, *ERG3*, *ERG11*, *ERG20*, *NCP1*), aromatic amino acid metabolism (*ARO2*, *ARO4*, *PDC1*, *PDC5*, *ALD5*, *TYR1*), and lysine biosynthesis from aminoadipate (*LYS12*, *LYS1*, *LYS2*, *LYS9*, *ACO2*). Both groups show some variability when it comes to overall transporter activity. In the adipic acid tolerant strains genes with mitochondrial transporter activity are up-regulated (*TIM23*, *NDI1*, *COX6*, *COX15*, *QCR6*, *QCR9*, *COR1*, *HOT13*, *PAM18*) while similar genes are down-regulated in CEN.PK113-7D (*COX3*, *COX7*, *COX13*, *COX5B*, *COX4*, *COX6*, *COX8*, *ATP15*, *ATP4*, *SHA2*, *SDH3*, *SDH4*, *MDH1*), consistent with previous work[55].

In terms of carbohydrate metabolism, the two groups had varying expression. Some genes were up-regulated in both adipic acid tolerant strains (*ADH5*, *ADH6*, *ADH1*, *ADH2*, *YIG1*) and CEN.PK113-7D (*ADL5*, *ADH6*, *PDC5*, *PDC1*) while others were down-regulated

in adipic acid tolerant strains (*GTT1*, *ECM4*, *UGX2*, *ADH4*, *ALD3*, *ENO1*, *MIG3*) and CEN.PK113-7D (*UGX2*, *ADH2*, *ALD2*, *ALD3*). Notably, a number of hexose transporters (*HXT6*, *HXT5*, *HXT7*, *HXT11*, *HXT8*, *HXT13*) and related genes such as *IMA4*, which encodes an α-glucosidase, were down-regulated in the adipic acid tolerant strains but not in CEN.PK113-7D. Thiamine related genes (*THI5*, *THI11*, *THI22*, *THI20*, *THI73*) were down-regulated only in adipic acid tolerant strains as was amino acid biosynthesis (*HIS5*, *LEU2*, *LYS2*) and fatty acid oxidation (*PCS60*, *PAA2*, *PAT1*, *PXA1*).

3.3 Production of adipic acid precursors in acid tolerant yeast

The identification of differential gene expression between adipic acid tolerant strains and the non-tolerant CEN.PK113-7D shows that, while gene content remains similar, genomic regulation influences adipic acid tolerance. In the context of adipic acid production, an ideal production strain would make use of this differential gene expression to better convert biomass to adipic acid at low pH and with minimal effects of product toxicity. Here we confirm that an adipic acid tolerant strain is better able to produce protocatechuic acid, an adipic acid precursor, at low pH and under adipic acid stress.

3.3.1 The adipic acid tolerant L433 has improved production of the adipic acid precursor protocatechuic acid during low pH and adipic acid stress

The previously described pathway to produce adipic acid and its precursor ccM requires carbon flux out of the shikimate pathway. The first heterologous step is the conversion of DHS to PCA by the DHS-dehydratase AroZp. In order to evaluate the viability of ccM or adipic acid production from an adipic acid tolerant host, a copy of *AROZ* was integrated at Flagfeldt site 18 under control of the strong promoter P_{pgk1} in the acid tolerant strain L433 and in the non-tolerant strain CEN.PK113-7D, resulting in strains S20 and S17 respectively. Each of these four strains was cultured in potassium citrate buffered-YPS set to pH 3 or pH 5 or supplemented with 20 g/L adipic acid set to pH 3. Cultures were maintained at 30 °C for 8 days and fed with 1/10th volumes of 40% (w/v) sucrose on days

3 and 6 to account for evaporation and to add sugar. Following this culture regime, sample supernatants were analyzed by HPLC for PCA production (Figure 3.12). In the parent strains CEN.PK113-7D and L433 no PCA was detected in any of the growth conditions. In terms of absolute PCA production, S17 and S20 accumulated similar amounts of PCA when grown at pH 5: 0.27 ± 0.03 g/L and 0.26 ± 0.04 g/L, respectively. At pH 3 S17 outperformed S20 and produced 0.6 ± 0.02 g/L compared to 0.28 ± 0.03. When grown in adipic acid S20 produced 0.6 ± 0.08 g/L compared to S17 production of 0.25 ± 0.05 g/L. When normalized to cell density, PCA production per OD600 in S17 was 0.02 ± 0.003, 0.03 ± 0.001, and 0.031 ± 0.006 g/L/OD_{600} for pH 5, pH 3, and 20 g/L adipic acid respectively. For S20, PCA production per OD was 0.02 ± 0.003, 0.02 ± 0.002, 0.07 ± 0.003 g/L/OD_{600}.

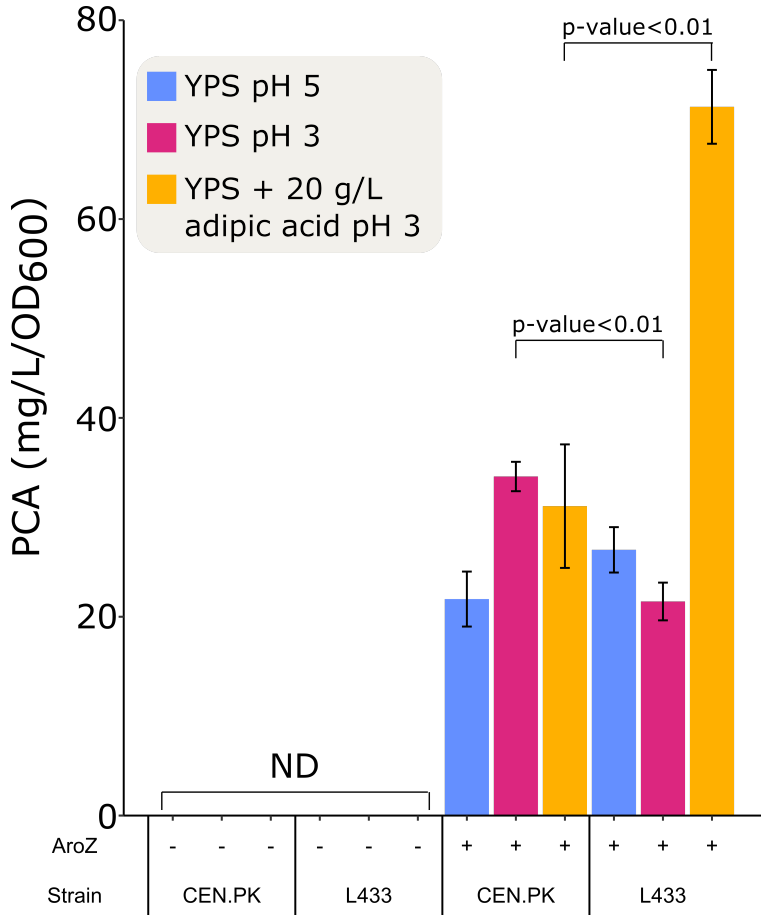

Figure 3.12: The adipic acid tolerant strain S20 accumulates more PCA than the non-tolerant S17 at low pH and under adipic acid stress. PCA accumulation was measured by HPLC in strains CEN.PK113-7D, L433, S17, and S20 after 8 days of culturing in YPS at pH 3, pH5, or with 20 g/L adipic acid (pH 3). PCA was not detected in the parent strains CEN.PK113-7D and L433 but was detected in S17 and S20. PCA accumulation in S20 supplemented with 20 g/L adipic acid was significantly greater than all other strain-condition pairs (p-value $<$ 0.01) compared using a two sample t-test. PCA accumulation was determined by comparing HPLC spectra peak height with that of a linear standard curve and normalised to OD_{600}. Data are shown as mean of 4 replicates $\pm$ standard error.

Chapter 4

Discussion

Bio-based production of adipic acid offers a less environmentally damaging and potentially more economically viable alternative to traditional petroleum based methods. To make this switch not only do microbial cell factories have to withstand and produce high levels of adipic acid they also would, ideally, do so at low pH so that downstream processing, particularly the major cost of acidification of adipate to adipic acid, can be minimised. Here we emphasize the utility of acid tolerant strains of *S. cerevisiae* as production hosts and sources of important phenotypes for the industrial production of organic acids.

4.1 Screening

Our initial high-throughput screen for adipic acid tolerance showed that our 500 strain collection had a wide range of phenotypes after exposure to adipic acid. Roughly half of the strains had a higher growth yield in adipic acid than CEN.PK113-7D, with the top strains having a growth yield roughly twice as high as that of CEN.PK113-7D (Figure 3.1). The use of micro-titer plates for initial screening, especially when coupled with large collections, is widely used due to compatibility with automation and high-throughput methods as well as lower costs compared to shake flasks[112]. However, these high-throughput methods do not always scale well and can necessitate secondary, high volume but low-throughput screening[113]. To confirm results of the high-throughput screening we implemented a secondary screening using larger volumes. In this second screen, the difference between our top adipic acid strains and CEN.PK113-7D was even greater – a 5 fold increase in media

volume led to the top strains reaching a growth yield nearly 5 times greater than that of CEN.PK113-7D (Figure 3.3).

One of the important factors in selecting a production strain is the amount of resources required to maintain the strain during production. Consequently, prototrophs are often preferred over auxotrophs, which would otherwise require supplementation, e.g. with vitamins or amino acids, resulting in increased costs. Indeed, the prevalence of prototropic industrial strains can sometimes hinder strain development due to the lack of auxotrophic markers used for selection during genetic manipulation[114] and can necessitate the use of alternative selection such as through the use of antibiotics. Out of the 500 strains screened for adipic acid tolerance, growth in YNBG was consistent across most strains (Figure 3.1), suggesting that most strains were prototropic. Of the strains that grew poorly in YNBG most also grew poorly in adipic acid. It is unclear if this is due to auxotrophic strains being limited by metabolic deficiencies or if mis-inoculation occurred during the screening process. Due to the large number of samples to be screened – 500 strains in 2 conditions and 4 replicates each – and due to the high-throughput approach necessary for this screening, mis-inoculation was not immediately identifiable. It is therefore possible that some strains identified as having low adipic acid tolerance are actually highly tolerant but were mis-inoculated and consequently mis-identified. However, the high number of strains with high adipic acid tolerance identified in the screening make it unlikely that we are missing much when it comes to a theoretical strain masked by mis-inoculation. Interestingly, strains that grew well in adipic acid tended come from L400-L500 series strains. Further clarification from Lallemand showed that these strains were largely isolates from winemaking. Wine strains have been shown to have high tolerance for ethanol but also for organic acids such as citric acid[115]. Tolerance to adipic acid is perhaps not surprising since there is well documented overlap between mechanisms against ethanol stress and organic acid stress, such as the expression of H+ ATPases[52;115] and changes to plasma membrane lipid composition[48;51;115].

Ploidy determination revealed that most of the top adipic acid tolerant strains identified in this screening were diploid or close to diploid based on comparison with the genome of CEN.PK113-7D (Table 3.1). However, when using the propidium iodide staining method there is some uncertainty over absolute classification of strain ploidy. For example, strain L473 was identified as having a genome 1.57 times greater in size than the haploid CEN.PK113-7D – should this be classified as a haploid, perhaps with an increased number of non-core genes absent in CEN.PK113-7D, or as a diploid with a smaller total gene content, or as some kind of aneuploid? Additional methods, such as karyotyping via

electron miscroscopy[116] or in combination with quantitative PCR[117], may be necessary to fully elucidate ploidy in this strain collection. Regardless, the prevalence of non-haploids in our subset of adipic acid tolerant strains is expected. In a large scale characterisation of yeast strains, Peter et al. showed that 87% (694/794) of strains from a wide range of sources were diploid and further demonstrated that diploid strains had significantly increased overall fitness compared to those of any other ploidy level[57]. In this same study, aneuploids were shown to have overall less fitness than euploids as a whole, although the overwhelming majority of euploids were diploid whereas triploid and tetraploid strains alone actually had lower fitness than either aneuploids or diploids. On the surface, the lowered fitness of aneuploid strains may seem detrimental for adaptation in environmental stress. However, reports suggests that aneupoidy provides an increased fitness advantage in the form of "bet hedging": under normal conditions aneuploids have decreased fitness and increased proteomic stress[101;118] but are sometimes better able to withstand extreme stress due to allelic variation[119]. With this in mind, it is perhaps not surprising that the haploid CEN.PK113-7D is less tolerant than our diploid adipic acid tolerant strains. However, ploidy itself was not enough to increase adipic acid tolerance: growth characterisation of the diploid CEN.PK122 under adipic acid stress had both a final growth yield and maximum growth rate similar to CEN.PK113-7D – i.e. much lower than the adipic acid tolerant strains (not shown).

4.2 Comparative gene expression

Transcript analysis of strains under adipic acid stress identified DEGs, which can provide insight into causes behind adipic acid tolerant phenotypes. However, between the four strains in which DEGs were identified, there were noticeable differences in the number of DEGs per strain – CEN.PK113-7D (358 up, 208 down), L433 (198 up, 204 down), L248 (145 up, 73 down), O33 (36 up, 6 down), L295 (0 up, 0 down). This difference is most likely due to differences in sample variability – i.e. unwanted variation –, which influences how well variation between conditions can be measured, and not actual differences between strains. When sample variability was low between samples of the same treatment group but high between the two treatment conditions, we were able to detect DEGs in similar quantities to previous work. Dong et al. reported similar numbers of DEGs – 300-500 both up- and down-regulated – at three different times points after exposure to acetic acid[55]. But when sample variability was poor – i.e. when dimensionality analysis showed overlap between samples from different conditions – our ability to detect DEGs was greatly reduced.

In the most extreme case of poor sample variation between conditions, that of L295, no DEGs were detected. The direct link between sample variation and DEG determination make it unlikely that L295 truly had no differential gene expression under adipic acid stress. Rather, the lack of identifiable DEGs may be remedied by improved sample preparation and experimental design. In a remarkable RNAseq study using 48 replicates per condition, Gierlinski et al. show that so-called "bad" replicates can significantly affect gene count and recommend the use of at least 3 and sometimes up to 6 replicates[120]. Even in strains in which we identified DEGs, the use of increased replicate number may provide an avenue to improve distinguishability between conditions when paired with the removal of outlier replicates. For example, sample A2 from L248 had, in the case of several DEGs, relative expression closer to that of the 3 untreated samples (Figure 3.8). Conesa et al. emphasise that there are no set standards for assessing variations between samples but support the use of principal component (PC) analysis, analogous to the MDS analysis performed above, to ensure that samples cluster together per condition[121]. Recently, Chen et al. reported the use of robust PC analysis methods that could detect outliers not detected using classical PC analysis, and later validated that the removal of these outliers resulted in gene expression that more closely matched biological reality[122]. Overall, a more robust approach is to increase the number of replicates so that poor samples, as determined by PC analysis, can be removed while still allowing significant findings to be derived from the remaining replicates.

Differential gene expression analysis revealed cell-wide transcriptional changes upon adipic acid stress both in the non-tolerant CEN.PK113-7D and in the adipic acid tolerant strains. Using MCL to group related genes we show that common transcriptional changes exist between tolerant and non-tolerant strains (Figure 3.11). Many of the pathways highlighted during clustering – such as pH homeostasis, amino acid metabolism, protein folding stabilization – are broadly similar with previous works[48;55;123]. Previously the carbonic anhydrase-expressing *NCE103* was reported to be down-regulated under acetic acid stress[55]. We report that *NCE103* expression was similarly down-regulated in CEN.PK113-7D under adipic acid stress but that *NCE103* expression was actually up-regulated in L433. Nce103p, previously linked with improved tolerance to oxidative stress[124], catalyses the hydration of CO_2 to HCO_3 reversibly and may play a role in improved intracellular de-acidification of L433 over CEN.PK113-7D. The H+ ATPase Pma1p has previously been linked to weak acid tolerance[48;125]. Dong et al. report down-regulation of *PMA1* and its relative *PMA2* under acetic acid stress; we show that *PMA2* is heavily down-regulated – a nearly 25-fold decrease – in CEN.PK113-7D while, strangely, *PMA1* is actually up-regulated 3 fold. Interestingly, neither *PMA1* nor *PMA2* is DE in any of the

acid tolerant strains. Previous work has put forward the use of Pma1p overexpression as a mechanism for generating industrial strains with improved acid tolerance[125]; perhaps the lack of *PMA1/PMA2* down-regulation in our adipic acid tolerant strains is a natural analog for this engineered approach. Genes related to protein folding and stress response play an important role in maintaining proper cellular function[126]. Under adipic acid stress, we show that genes related to the stress response and protein folding were DE such as *HSP60*, *ECM10*, *HSP104*, *HSP26*, and *SUP35*. Of the 13 common stress-related genes found under acetic acid stress and adipic acid stress, 11 were DE similarly to those affected by acetic acid[55]. Protein synthesis is crucial to produce the proteins necessary to carry out cellular activity. Here, we report that the expression of genes related to protein synthesis is greatly increased after exposure to adipic acid stress: 90 genes were up-regulated across our 4 strains. However, other reports show that gene expression related to protein synthesis is generally decreased upon cellular stress[55;127;128]. Out of 31 DEGs common between our adipic acid exposure and the acetic acid exposure of Dong et al[55], all but 2 are oppositely expressed – down-regulated in acetic acid and up-regulated in adipic acid. While there are differences in differential gene expression specific to different acids depending on lipophilicity, length, and functional groups[48], the disparity in protein synthesis-related gene expression may be better explained not as a difference in acid-specific transcriptional responses but rather by the time of measurement after response. Dong et al collected RNA samples 45, 120, and 240 minutes post-exposure whereas we collected RNA only 240 minutes after adipic acid exposure. While Dong et al report wide spread down-regulation of genes related to protein synthesis, these are found mostly in samples collected 45 minute post-exposure. They also show that the expression of genes related to protein synthesis rebound from the initial down-regulation after 2.5 hours. The temporal component is especially noticeable when comparing genes associated with transcription networks. Previous works have identified several genes whose expression influence the acid stress response, notably *WAR1*, *HAA1*, *MSN2*, *MSN4*, *RIM101*, and *PDR1*[48;54]. With the exception of *MSN4* and *HAA1* in L433, each of these transcription factor-expressing genes were identified during transcriptional analysis in all 4 strains. However, none of the genes were identified as DE. Due to the broad specificity of these transcription factor genes, it is unlikely that adipic acid does not affect their expression. By collecting RNA samples sooner after adipic acid exposure we might expect to see expression changes in these genes similar to those identified in similar organic acids[48;55].

Gene content was compared and showed very little difference between the adipic acid tolerant strains and CEN.PK113-7D. Previous works have identified core genes within

S. cerevisiae and suggest that, even in the most extreme cases, nearly all genes are similar between strains[108;109]. We show that the number of genes was comparable between our strains – 4320 for CEN.PK113-7D, 4472 for L248, 4102 for O33, and 4900 for L433 – and that many of the genes that are different between strains were not DE under adipic acid stress. Of the DEGs found in the adipic acid tolerant groups that were not identical to CEN.PK113-7D, most were either paralogous or homologous with genes from CEN.PK113-7D. Furthermore, even un-shared DEGs that were not homologs or paralogs were often later identified as genes with different names but similar functions to genes found in CEN.PK113-7D. Despite this, there were some genes that were identified as being present in the adipic acid tolerant strains but not in CEN.PK113-7D including *AAD16*, *ARN2*, *THI11*, *RRT15*, *HXT11*, *COA2*, and *JEN1*. *JEN1* is notable among these genes since previous work has linked *JEN*-family genes to carboxylic acid transport and tolerance in *S. cerevisiae*[129], *P. kudriavzevii*[130], and *K. marxianus*[131]. While *JEN1* was only DE in a single adipic acid tolerant strain, L433, its expression in the other two adipic acid tolerant strains was just below the cutoff for being considered a DEG – L248 exhibited a 1.9 FC increase while O33 showed a 2.4 FC increase but an FDR of 0.1. Other than *JEN1* there seems to be no immediately identifiable single mechanism by which the genes unique to the adipic acid tolerant strains could influence adipic acid tolerance, although perhaps more complex synergistic effects are at play. Strangely, some of the genes that we identified as unique to the adipic acid tolerant strains have previously been shown to exist in the CEN.PK background, including *JEN1* and *SGE1*. During testing of the *de novo* transcriptome assembly workflow we used CEN.PK113-7D to compare the reference-based and *de novo*-based methods and corroborated the high comparability between Trinity and sequence-first methods shown by Grabherr et al[72]. While *JEN1*, *SGE1*, and some other genes were not identified using a reference-based approach, *de novo* assembly based differential gene expression analysis (not shown) suggests that *JEN1* and *SGE1* are part of the CEN.PK113-7D genome. The reason for this disparity is not immediately clear but further supports that gene expression, not gene content, plays a larger role in adipic acid tolerance. Overall, we show that the transcriptional response to adipic acid stress is complex and multi-faceted. While variation in the expression of various genes exists between the adipic acid tolerant strains and the non-tolerant CEN.PK113-7D, further work is still required to confirm the main causative genes behind adipic acid tolerance. We have identified genes not found or not DE in CEN.PK113-7D (see appendix) and genes that are DE but in opposite ways between the tolerant and non-tolerant strains (Table 3.3). Future work to elaborate on these genes might benefit from the use of combinatorial gene overpression/repression to identify which genes and combinations of genes provide improved adipic acid tolerance with the smallest cost.

4.3 Utility of adipic acid tolerant strains in producing adipic acid precursors

The oxidation of DHS by AroZp is the first step in the heterologous pathway that diverts flux from the shikimate pathway to adipic acid[30;34]. We were unable to complete this full pathway at this time but instead used PCA, a product of AroZp's action on DHS, accumulation as a proof of concept. We show that PCA accumulation at pH 5 was the same in the S17 and S20, both in terms of absolute PCA and PCA production normalised to OD_{600}. At low pH without adipic acid, S17 outperformed S20 in total PCA accumulation but not after normalisation. At low pH and under adipic acid stress S20 accumulated roughly twice as much PCA than S17 both in absolute PCA and after normalisation. Normalised PCA accumulation was significantly higher in S20 at pH 3 with adipic acid than any other condition, which supports the use of adipic acid tolerant strains as adipic acid production hosts. Pyne et al showed nearly 2 g/L PCA in a CEN.PK113-7D background strain[34] yet, under similar conditions, our CEN.PK113-7D-derived S17 only produced 0.27 g/L PCA. Importantly, the strain used by Pyne et al was missing *ARO3* and *ARO4* but expressed a feedback resistant copy of *ARO4* to prevent feedback inhibition by aromatic amino acids (Figure 1.2). Assuming that the this 7 fold increase in PCA can be achieved through *ARO3* and *ARO4* deregulation alone, we might expect to see substantial increases in PCA production in the adipic acid tolerant strains by using a similar approach, especially when comparing PCA production at low pH and under adipic acid stress where PCA production was already nearly twice as high in the adipic acid strain. Future work is still required to achieve ccM or adipic acid production in our adipic acid tolerant strains but, when coupled with methods such as deregulation of *aro1* [34] and gene overexpression[38], this seems promising in the search to produce adipic acid and its precursors in high quantities at low pH.

4.4 Conclusion

This study showed that a wide range of adipic acid tolerant phenotypes exists across a large collection of suspected acid tolerant strains. Further characterisation of top adipic acid strains revealed that they are largely diploid and that tolerance to adipic acid can be induced by pre-adaptation. Gene expression comparisons suggest that gene expression, but not gene content, is the driver of adipic acid tolerance within our top adipic acid tolerant

strains. The introduction of the orthologous DHS dehydratase AroZp into an adipic acid tolerant strain resulted in increased accumulation of PCA, an adipic acid precursor, at low pH and under adipic acid stress compared to the non-tolerant CEN.PK113-7D. Future work is necessary to implement the complete adipic acid pathway in acid tolerant yeast but our preliminary results support the utility of adipic acid tolerant strains of *S. cerevisiae* as adipic acid production hosts. By improving the low pH fermentation of adipic acid we can lower overall costs of renewable based adipic acid and reduce our dependency on, and the detrimental effects of, petroleum-based adipic acid.

Chapter 5

Appendix

5.0.1 Tentatively acid tolerant strains

Table 5.1: Strains obtained from Lallemand used in our screening for adipic acid. Strains are, here known, annotated with information regarding their isolation location.

List of tentatively acid tolerant strains

Identifier	Description	Identifier	Description	Identifier	Description	Identifier	Description
L001	Wine	L444	Wine	L501	Cheese	L102	Lab
L002	Beer	L456	Wine	L343	Wine	L203	Sake
L003	Cheese	L434	Wine	L432	Beer	L301	Wine
L004	Wine	L444	Wine	L501	Cheese	L102	Lab
L005	Beer	L456	Wine	L343	Wine	L203	Sake
L006	Cheese	L434	Wine	L432	Beer	L301	Wine
L007	Wine	L444	Wine	L501	Cheese	L102	Lab
L008	Beer	L456	Wine	L343	Wine	L203	Sake
L009	Cheese	L434	Wine	L432	Beer	L301	Wine
L010	Wine	L444	Wine	L501	Cheese	L102	Lab
L011	Beer	L456	Wine	L343	Wine	L203	Sake
L012	Cheese	L434	Wine	L432	Beer	L301	Wine
L013	Wine	L444	Wine	L501	Cheese	L102	Lab
L014	Beer	L456	Wine	L343	Wine	L203	Sake
L015	Cheese	L434	Wine	L432	Beer	L301	Wine
L016	Wine	L444	Wine	L501	Cheese	L102	Lab
L017	Beer	L456	Wine	L343	Wine	L203	Sake
L018	Cheese	L434	Wine	L432	Beer	L301	Wine
L019	Wine	L444	Wine	L501	Cheese	L102	Lab
L020	Beer	L456	Wine	L343	Wine	L203	Sake
L021	Cheese	L434	Wine	L432	Beer	L301	Wine
L022	Wine	L444	Wine	L501	Cheese	L102	Lab
L023	Beer	L456	Wine	L343	Wine	L203	Sake
L024	Cheese	L434	Wine	L432	Beer	L301	Wine
L025	Wine	L444	Wine	L501	Cheese	L102	Lab
L026	Beer	L456	Wine	L343	Wine	L203	Sake
L027	Cheese	L434	Wine	L432	Beer	L301	Wine
L028	Wine	L444	Wine	L501	Cheese	L102	Lab
L029	Beer	L456	Wine	L343	Wine	L203	Sake
L030	Cheese	L434	Wine	L432	Beer	L301	Wine
L031	Wine	L444	Wine	L501	Cheese	L102	Lab
L032	Beer	L456	Wine	L343	Wine	L203	Sake
L033	Cheese	L434	Wine	L432	Beer	L301	Wine

Continued on next page

List of tentatively acid tolerant strains

Identifier	Description	Identifier	Description	Identifier	Description	Identifier	Description
L034	Wine	L444	Wine	L501	Cheese	L102	Lab
L035	Beer	L456	Wine	L343	Wine	L203	Sake
L036	Cheese	L434	Wine	L432	Beer	L301	Wine
L037	Wine	L444	Wine	L501	Cheese	L102	Lab
L038	Beer	L456	Wine	L343	Wine	L203	Sake
L039	Cheese	L434	Wine	L432	Beer	L301	Wine
L040	Wine	L444	Wine	L501	Cheese	L102	Lab
L041	Beer	L456	Wine	L343	Wine	L203	Sake
L042	Cheese	L434	Wine	L432	Beer	L301	Wine
L043	Wine	L444	Wine	L501	Cheese	L102	Lab
L044	Beer	L456	Wine	L343	Wine	L203	Sake
L045	Cheese	L434	Wine	L432	Beer	L301	Wine
L046	Wine	L444	Wine	L501	Cheese	L102	Lab
L047	Beer	L456	Wine	L343	Wine	L203	Sake
L048	Cheese	L434	Wine	L432	Beer	L301	Wine
L049	Wine	L444	Wine	L501	Cheese	L102	Lab
L049	Wine	L444	Wine	L501	Cheese	L102	Lab
L049	Wine	L444	Wine	L501	Cheese	L102	Lab

5.0.2 DEGs identified by RNAseq in strains CEN.PK113-7D, L433, L248, and O33

Table 5.3: List of identified differentially expressed genes in strains CEN.PK113-7D, L433, L248, and O33

CEN.PK113-7D		L433		L248		O33	
up	down	up	down	up	down	up	down
RRP3	FMP16	ISC10	VPS35	MSA1	PBY1	BSC1	THI4
ELO1	FMP10	RRP9	MTC1	STP3	MSF1	AGP1	ENV9
ELO3	AYR1	RPA190	GEX1	MSH4	TOS4	HUG1	HXT8
TPI1	FMP23	PPR1	GAP1	PTC2	APN1	PTR2	POL2
LYS2	FMP46	SMF2	ADH4	RDH54	GIN4	COX26	YJU2
LYS4	FMP45	GAS1	YMR018	HLR1	YDL121C	TPO4	SAN1
LYS1	FMP40	CBC2	VEL1	RIM8	PCL1	PES4	
LYS9	MCH1	SQT1	RAS1	IGD1	NUP192	RCK1	
RRT2	TPK2	DEG1	GAT3	YBR184W	NUP188	CYC7	
GAS5	AI1	TPN1	SIP18	NNF2	SUV3	YHB1	
DIN7	GRX1	PUF3	SCW10	RIP1	RAS1	OSW1	
GRX3	MCH4	AIM32	MCD1	SDH1	THI4	BNA2	
MCH5	GRX5	RFC4	LYS2	HOT13	MDM10	MND1	
GRX8	TPN1	PRP16	MTF2	COX15	RPS1A	AIM33	
HEM13	FOX2	MCM4	CSE4	UBC8	RPS1B	FAA4	
PDC5	PDC6	YPL272C	SMK1	IXR1	DAL7	DLD3	
PDC1	ZIP2	FKS3	YNL010W	SUN4	GSH2	ISU2	
RAX1	VPS73	NOC4	MCH2	SUT1	EDS1	SPS1	
RAX2	MCP1	FCF2	HEM14	SUR2	ZRT1	SHR5	
OAC1	PDH1	FLO5	YPL041C	SUR1	BUD8	GEX1	
MCM5	HED1	YNL035C	HIF1	VPS13	AAD3	ADH5	
MCM6	MDM36	NOG2	MTM1	PLM2	YLL058W	ADH6	
MCM2	RSB1	LIA1	CBF1	UBR2	RKI1	ADH1	
TYE7	COG5	DAK2	RFA3	PUF3	ARN2	ADH2	
RPS15	PMA2	YHR097C	AUS1	COA4	SWI4	MHF1	
RPS11B	PUT1	TUF1	MCK1	COA2	SWI5	IMD2	
RPS12	AQY1	RBG1	CSH1	SMM1	PEP3	IMD3	
RPS13	COQ4	SMY2	RFC5	TPS1	UBC12	ERC1	
RPS31	GSM1	ARO10	AIM45	DRE2	YPT7	HXT14	
AQR1	AIM2	TRM11	PHM7	YKR075C	RTR1	HXT10	
RPS1B	ARA2	NOP4	ICY1	AQR1	IRR1	ANS1	
RPS10B	MLS1	NOP2	UGA4	DAK2	HOS3	TNA1	

Continued on next page

CEN.PK113-7D		L433		L248		O33	
up	down	up	down	up	down	up	down
RPS20	COX3	NOP7	TSC13	MLF3	THI11	CDA1	
RPS0A	COX4	NOP9	YOR059C	MLH2	YHR192W	SSP1	
DAL3	COX6	GSM1	MDM30	YFR057W	ENV11	MAM1	
MTW1	COX7	HEM1	HEF3	HEM1	EGD1	DPH5	
DAL7	COX8	DRS1	YSP3	HEM3	YIM1		
HSP150	MDH1	QCR6	YLR012C	COR1	ACM1		
DAK2	MDH3	QCR2	DAL3	YOR1	GEA1		
THI6	SNO1	YPR027C	MTW1	MLP1	CDC73		
RPS16A	SFA1	DNM1	RSF2	TIM23	CDC21		
RPS14B	ARO9	SER2	ORM2	LAC1	ECM10		
PMA1	HFD1	TYW1	DNF2	UTR2	GEF1		
RPS8A	SNZ1	COT1	SEG1	MDJ1	PPA2		
RPS19B	RCF1	HRK1	YKL161C	MUM3	DUR3		
PUS4	COX5B	TIM12	YNL115C	CPA2	GNA1		
PUS1	KGD2	RBS1	BPT1	DSE1	ADH4		
RPS7A	KGD1	TIM11	AQY1	DSE2	MXR2		
RPS6A	GTO1	FCY2	RSC30	YML122C	KRS1		
TRM11	COX13	TIM23	RBK1	ARP7	SAM50		
TRM10	COX17	COX6	RFS1	YCR006C	TUB2		
RPS17A	MEC3	NCE103	YLR149C	SWH1	YNR066C		
TYR1	SOD2	YNL017C	FAR11	IRA2	HIS5		
EMG1	GTT1	ARG2	AMS1	KGD1	YML037C		
RPS9A	GCV1	MHR1	IMO32	AAD16	MYO1		
RPS18B	MEF2	UTP4	TDH1	RCK1	PIG1		
THS1	MEI5	ERG20	COS7	CYC1	LEU2		
TYS1	OM45	AEP2	ROG3	YPS6	ERG3		
THR1	ZTA1	FLO5	BUD3	YPS1	CLB2		
RPS4A	GCY1	PMT4	YIL089W	YIR042C	SEC27		
RSR1	MEK1	PMT2	YPL216W	RTS3	SEC11		
SET6	HOP2	YEL025C	MDE1	CHL4	PRI1		
SET2	HOR7	CTK1	TIM17	HOM3	SEC39		
SES1	COB	YCG1	SAN1	CYT2	PRM1		
ARD1	HSP26	FHL1	SNF6	SOR2	ERS1		
UTR2	HSP12	ARO2	SAM3	MVP1	CDC9		
PEA2	FZO1	MAK21	PIG1	HOS1	HST4		
AAH1	DCI1	CDC60	YKL069W	HSP26	SKG6		
PMU1	HSP31	SFB3	UGX2	DCI1	YUH1		
CTF18	RSM10	REC114	ULA1	HSP42	TEX1		
DSF2	GUD1	IMP3	YDL218W	GLT1	TFC8		
CGR1	FAA2	ECM13	SRT1	HSP78	LOT6		
ARO8	FIT3	ECM16	BUL1	FRE1	FET4		
ARO4	QRI7	UPS3	MDL1	GLY1			
ARO2	FYV10	ECM10	CPA2	MNN4			
SWE1	HXT5	PET309	YOR012W	GUP2			
MAK21	GMC2	GCD7	GOR1	CAF120			
DED81	SPO22	GCD6	AVT1	YIG1			
IZH2	SPG1	CDC14	SNO2	TSL1			
RPS23A	RMD5	CLB4	ROT2	RME1			
RPS24B	SPG5	CDC11	HNT1	ACH1			
RPS21A	SPG4	CCS1	EIS1	ISU1			
SWI4	GUT1	SSC1	IZH4	IST2			
ART5	YIG1	YDL180W	FDH1	ISW2			
RPS22A	ATG3	DBP2	ERG1	CIT1			
SEC63	PXA1	SFK1	HTB1	SUP35			
WRS1	ATG9	NPP2	PRE5	PXR1			
SFG1	ATG8	YER137C	RTC4	CDC20			
RPS27A	POR2	MRPS35	YLR040C	ECM11			
SFH5	ACH1	YPP1	HXT13	CDC34			
RPS28A	POT1	PRM7	HXT11	ECM13			
RPS25A	CRC1	MID1	SSA3	NRP1			
ARX1	ISU1	RPA43	YPI1	EGT2			
RPS26B	CAB2	CYC1	GPA1	ZDS1			
PER1	SPR3	COX13	SEC17	ARP10			
IRC7	ATO2	FUS1	YCS4	RPP1B			
TIP1	MAM33	UTP25	DBP1	MET17			
RPA43	ATP6	GPI1	HFM1	CBF2			
RPS29A	ATP4	UTP13	MID2	MET28			
TIR4	CIT2	MMR1	CPS1	RNR2			
CGI121	CIT3	PAP1	MIG3	PHO4			
TAE1	ACO1	BMS1	CPR2	ADR1			

Continued on next page

CEN.PK113-7D		L433		L248		O33	
up	down	up	down	up	down	up	down
NPT1	SHE4	AFI1	COX10	MPH3			
YPS3	RMR1	YNL058C	COX18	ADY2			
MVD1	ALD3	RPM2	RRT15	CSR2			
RPL2A	ALD2	TRM1	PRR1	CBP2			
RPL38	ACS1	TRP3	UTP11	ALG13			
RPL39	PEX22	SRP40	YDR374C	QCR6			
RPL30	REG2	RLI1	OSW2	QCR9			
RPL32	PPA2	MRD1	RTR1	SRM1			
ASC1	TKL1	PNT1	GCR2	IMA5			
NPY1	DUR3	NDI1	SPC97	PIC2			
RPL1A	DMC1	ASN2	ENO1	SAM1			
RPL25	IKS1	ASN1	YIL060W	HSP104			
RPL28	HYR1	FLO1	GPM2	HAP4			
RPL29	BOP2	MPA43	GTT1	NTH1			
HOG1	NAS2	FZO1	ATG17	TEC1			
RLF2	CAT2	RSM27	ATG19	RGM1			
RPL4A	CAT8	RSM26	YOR019W	CTS1			
TRM9	YMR31	PSF2	ATG14	SSC1			
EFB1	GND1	YPL014W	ASH1	CTT1			
TRP2	FBP1	GLR1	LSG1	PRK1			
RPL10	BXI1	PSE1	SOL4	GCD10			
YHM2	ICL1	ESF2	HBT1	PRM7			
DCD1	ICL2	PWP1	THI73	PRM5			
CYS3	VAB2	AST1	HOR7	GPI2			
HO	ADH2	HSP60	THI20	RPM2			
RPL9A	EHD3	MRP7	THI22	RAD16			
CHS3	AIM36	FAF1	LSP1	NDE1			
CHS7	AIM46	SXM1	HOS3	MAE1			
ASN1	MRPL6	STE4	SGE1	NDI1			
RPL6A	AIM41	SKT5	THI11	VID24			
RPL8B	AIM19	MSS51	GDB1	COS111			
MET3	UGA1	GPI12	TRX1	NUM1			
RPL7B	UGA4	QRI7	YNL018C	JID1			
SCP160	MPM1	SGN1	HTB1	RAD59			
MES1	ARO80	MRT4	FAA2	MAM3			
BNI5	CBP4	PDR16	NUT1	DPH5			
BNI4	MPS2	PAM18	STB4	BAP2			
PFK27	MPS3	BFA1	GYP7	MRS3			
EFT1	ARO10	STL1	DTR1	PSK2			
ISC1	SAF1	URB1	YKL151C	SKS1			
THI80	AMS1	URB2	GDH3	YOR228C			
OLE1	SAE3	JEM1	BAT2	IRC18			
HXT6	LEE1	FUN30	YCR101C	YDR089W			
FRM2	XYL2	SPF1	MAS1				
DCV1	NCE102	YIG1	TSA2				
POL1	IME4	YTA12	HPA2				
POL2	IME1	JEN1	FAL1				
POL5	UGX2	MBA1	JIP3				
LCB1	SRT1	MSC6	IFM1				
YIH1	RGI1	HTS1	HXT6				
ATF2	FDH1	POR1	HXT7				
FAS1	REC114	YBR184W	HXT5				
FAS2	CTR1	YEH2	INP54				
FRS1	CTR3	MSI1	PCS60				
FAR1	PAI3	TOM1	SPO13				
EXG2	CTT1	MRPL35	TFS1				
HPM1	TMT1	MRPL40	YEA4				
YIP3	FUM1	MRPL22	PXA1				
ATM1	IBA57	CIT1	FAT1				
CIS3	PRM2	SHE2	MSE1				
UNG1	ERR3	SHE1	APC2				
ACM1	ATG38	YGR015C	APC5				
PXL1	ATG31	SPT8	CAB5				
NUP188	ATG36	LCP5	CAB2				
ALA1	CDA2	TOP1	NRD1				
NOP13	SSQ1	POL32	VMA5				
ACO2	RAD28	PEX25	SDS24				
SPS1	ATG16	RPC17	VMA6				
SHE3	ATG19	ALG7	SCW4				
SHE1	ATG14	HDA1	PTA1				

Continued on next page

CEN.PK113-7D		L433		L248		O33	
up	down	up	down	up	down	up	down
CRH1	ATG11	NAM9	CAD1				
ALD5	PRX1	FCY21	PGK1				
POL30	INH1	RIP1	ALD3				
POL32	HBT1	GRE2	ADE12				
NOP58	SSU1	ATP16	REG2				
RPC11	RAD34	YNG1	PEX27				
ZUO1	OXR1	YGR054W	RIM8				
FSF1	CMC2	ECO1	SHH4				
ALG5	MRP8	SLX5	YET3				
ALG3	ALG13	SDH1	YET2				
NRK1	STL1	PTR2	XBP1				
ATX2	EST2	HOT13	YND1				
FSH1	OPT2	ALO1	YRM1				
RPC25	PDR10	TGL1	YBR016W				
DUT1	DHH1	MOT1	ECM3				
ALK1	AGX1	TOM71	ECM4				
DUS3	CEM1	YJR054W	ECM5				
YAR1	ECI1	PLC1	PET130				
SEN34	PCH2	EPL1	YJL160C				
CCW14	SUE1	YKR005C	GAD1				
CAX4	PCI8	BUD22	TCB1				
RPC40	RIP1	PPM2	USA1				
KIP2	ECM4	YPR022C	MKK2				
ADE3	GRE1	TPA1	YKL050C				
RPL16B	ATP15	RPC40	FSP2				
RPL17B	SDH6	RPC82	CNN1				
ADH6	SDH5		RIX1				
RPL14B	SDH4		KIN3				
RPL15A	SDH3		YDR249C				
MET16	SDH2		DIC1				
MET22	SDH1		RAM1				
MET14	GAD1		VIP1				
MET10	PTR2						
RTT109	ETR1						
PRP43	RRI2						
PYK2	DIA4						
ADO1							
CBF5							
CSI2							
RNR1							
LDS2							
PHO3							
RFC5							
RFC3							
TMA19							
KAR1							
PMI40							
VAS1							
TUB4							
TIF11							
ILS1							
CSR1							
ILV1							
SRB2							
ILV3							
HIS7							
NMA111							
SIS1							
SAC1							
FTR1							
RPL12B							
RPL13A							
YSY6							
RPL11A							
SRM1							
HAM1							
SRL1							
HRP1							
MYO1							
RPL26B							

Continued on next page

CEN.PK113-7D		L433		L248		O33	
up	down	up	down	up	down	up	down
RPL27A							
SAM4							
HAS1							
ERG11							
ERG20							
ERG26							
LYS21							
LYS12							
KSS1							
PIS1							
CDC45							
CDC60							
ECM22							
ERG1							
ERG3							
ERG2							
ERG4							
CDC21							
CTP1							
CTS2							
SSA2							
NCP1							
SSB1							
RPB8							
YCS4							
TEF4							
ERO1							
RPL20A							
PRM5							
RPL21A							
MIG2							
UTP22							
PAN6							
PAN5							
UTP15							
HSL1							
RPL24B							
RPL22A							
CLN2							
RPL23A							
RPL5							
RPL3							
DOT1							
SNU13							
WWM1							
PRY3							
PRY1							
KCH1							
SSP1							
MRC1							
SAP155							
RPP0							
RPS2							
MIS1							
SKG6							
RPS5							
INM1							
SSZ1							
YDJ1							
PSA1							
LOH1							
DPB2							
RHO2							
HCH1							
PDR17							
RPL33A							
URA2							
URA6							
STM1							
RPL36B							

Continued on next page

CEN.PK113-7D		L433		L248		O33	
up	down	up	down	up	down	up	down
RPL37B							
OYE2							
RPL34B							
RPL35B							
ZPR1							
CMR1							
BSC1							
TOF1							
HTS1							
YRF1-5							
GIC1							
YEH1							
YVH1							
RIF1							
VMA5							
VMA7							
SCW4							
MSH2							
PKR1							
DHR2							
DYS1							
GIM5							
TOS6							
TOS4							
HUA2							
GIN4							
BBP1							
JJJ3							
SLT2							
GRC3							
RPL40A							
PCL1							
RPP1A							
RPP1B							
RTT10							
RPL42A							
USB1							
RPP2A							
RPP2B							
RPL43B							
SUN4							
SHB17							
UBC4							
BUD31							
RRM3							
BUD17							

Bibliography

[1] S A Montzka, E J Dlugokencky, and J H Butler. Non-CO 2 greenhouse gases and climate change. *Nature*, 2011. doi: 10.1038/nature10322.

[2] Peter G. Levi and Jonathan M. Cullen. Mapping Global Flows of Chemicals: From Fossil Fuel Feedstocks to Chemical Products. *Environ. Sci. Technol.*, 52(4):1725–1734, feb 2018. doi: 10.1021/ACS.EST.7B04573.

[3] Eugene A. Rosa and Thomas Dietz. Human drivers of national greenhouse-gas emissions. *Nat. Clim. Chang.*, 2(8):581–586, 2012. doi: 10.1038/nclimate1506.

[4] Tino Polen, Markus Spelberg, and Michael Bott. Toward biotechnological production of adipic acid and precursors from biorenewables. *J. Biotechnol.*, 167(2):75–84, 2013. doi: 10.1016/j.jbiotec.2012.07.008.

[5] Jian Li, Lanxin Rong, Yu Zhao, Shenglong Li, Cuiying Zhang, Dongguang Xiao, Jee Loon Foo, and Aiqun Yu. Next-generation metabolic engineering of non-conventional microbial cell factories for carboxylic acid platform chemicals. *Biotechnol. Adv.*, 43(1), nov 2020. doi: 10.1016/J.BIOTECHADV.2020.107605.

[6] Sung Kuk Lee, Howard Chou, Timothy S. Ham, Taek Soon Lee, and Jay D. Keasling. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.*, 19(6):556–563, 2008. doi: 10.1016/J.COPBIO.2008.10.014.

[7] Michael E. Pyne, Kaspar Kevvai, Parbir S. Grewal, Lauren Narcross, Brian Choi, Leanne Bourgeois, John E. Dueber, and Vincent J.J. Martin. A yeast platform for high-level synthesis of tetrahydroisoquinoline alkaloids. *Nat. Commun.*, 11(1):1–10, 2020. doi: 10.1038/s41467-020-17172-x.

[8] Shota Isogai, Tomonori Matsushita, Hiroyuki Imanishi, Jirasin Koonthongkaew, Yoichi Toyokawa, Akira Nishimura, Xiao Yi, Romas Kazlauskas, and Hiroshi Takagi. High-level production of lysine in the yeast Saccharomyces cerevisiae by rational design of homocitrate synthase. *Appl. Environ. Microbiol.*, 87(15):1–12, 2021. doi: 10.1128/AEM.00600-21/ASSET/4E6CCD2E-1EDF-44D2-824A-88024FB7D8FB.

[9] Derek A Abbott, Rintze M Zelle, Jack T Pronk, and Antonius J A Van Maris. Metabolic engineering of Saccharomyces cerevisiae for production of carboxylic acids: current status and challenges. *FEMS Yeast Res.*, 9(8):1123–1136, 2009. doi: 10.1111/j.1567-1364.2009.00537.x.

[10] G. P. Moss, P. A.S. Smith, and D. Tavernier. Glossary of class names of organic compounds and reactive intermediates based on structure. *Pure Appl. Chem.*, 67(8-9): 1307–1375, jan 1995. doi: 10.1351/PAC199567081307.

[11] T Werpy and G Petersen. Top Value Added Chemicals from Biomass Volume I-Results of Screening for Potential Candidates from Sugars and Synthesis Gas. Technical report, 2004.

[12] Christopher W. Johnson and Gregg T. Beckham. Aromatic catabolic pathway selection for optimal production of pyruvate and lactate from lignin. *Metab. Eng.*, 28:240–247, 2015. doi: 10.1016/J.YMBEN.2015.01.005.

[13] Jiang feng Ma, Ming ke Wu, Chang qing Zhang, Ai yong He, Xiang ping Kong, Gan lu Li, Ce Wei, and Min Jiang. Coupled ARTP and ALE strategy to improve anaerobic cell growth and succinic acid production by Escherichia coli. *J. Chem. Technol. Biotechnol.*, 91(3):711–717, 2016. doi: 10.1002/JCTB.4633.

[14] Hyun Gyu Lim, Myung Hyun Noh, Jun Hong Jeong, Sunghoon Park, and Gyoo Yeol Jung. Optimum Rebalancing of the 3-Hydroxypropionic Acid Production Pathway from Glycerol in Escherichia coli. *ACS Synth. Biol.*, 5(11):1247–1255, 2016. doi: 10.1021/ACSSYNBIO.5B00303.

[15] John J McKetta, editor. *Chemical Processing Handbook*. CRC Press, 1993.

[16] Michael Tuttle Musser. Adipic acid. In *Ullmann's Encycl. Ind. Chem.* Wiley, 7th edition, 2000. doi: 10.1002/14356007.

[17] Jeovanna Rios, Juliana Lebeau, Tian Yang, Shuai Li, and Michael D Lynch. A critical review on the progress and challenges to a more sustainable, cost competitive synthesis of adipic acid. *Green Chem.*, 23:3172, 2021. doi: 10.1039/d1gc00638j.

[18] Emma Skoog, Jae Ho Shin, Veronica Saez-Jimenez, Valeria Mapelli, and Lisbeth Olsson. Biobased adipic acid – The challenge of developing the production host. *Biotechnol. Adv.*, 36(8):2248–2263, dec 2018. doi: 10.1016/J.BIOTECHADV.2018.10.012.

[19] Energy Information Administration. Annual Energy Outlook with Projections: 2015 with Projections To 2040. Technical report, 2015.

[20] Environmental Protection Agency. Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2019. Technical report, 2021.

[21] Stijn Van De Vyver and Yuriy Román-Leshkov. Emerging catalytic processes for the production of adipic acid. *Catal. Sci. Technol.*, 3:1465, 2013. doi: 10.1039/c3cy20728e.

[22] R. A. Reimer, C. S. Slaten, M. Seapan, M. W. Lower, and P. E. Tomlinson. Abatement of N2O emissions produced in the adipic acid industry. *Environ. Prog.*, 13(2):134–137, may 1994. doi: 10.1002/EP.670130217.

[23] Ed de Jong, Adrian Higson, Patrick Walsh, and Maria Wellisch. Bio-Based Chemicals Value Added Products from Biorefineries. Technical report, International Energy Agency, 2012.

[24] Y Deng and Y Mao. Production of adipic acid by the native-occurring pathway in Thermobifida fusca B6. *J. Appl. Microbiol.*, 119(4):1057–1063, 2015. doi: 10.1111/jam.12905.

[25] Taicheng Zhu, Hongbing Sun, Meiyu Wang, and Yin Li. Pichia pastoris as a Versatile Cell Factory for the Production of Industrial Enzymes and Chemicals: Current Status and Future Perspectives. *Biotechnol. J.*, 14(6), 2019. doi: 10.1002/BIOT.201800694.

[26] P C Brzostowicz, M S Blasko, and P E Rouvière. Identification of two gene clusters involved in cyclohexanone oxidation in Brevibacterium epidermidis strain HCU. *Appl. Microbiol. Biotechnol.*, 58(6):781–789, 2002. doi: 10.1007/s00253-002-0968-x.

[27] Q Cheng, S Thomas, and P E Rouvière. Biological conversion of cyclic alkanes and cyclic alcohols into dicarboxylic acids: biochemical and molecular basis. *Appl. Microbiol. Biotechnol.*, 58(6):704–711, 2002. doi: 10.1007/s00253-002-0958-z.

[28] Wei Niu, K M Draths, and J W Frost. Benzene-Free Synthesis of Adipic Acid. *Biotechnol. Prog.*, 18(2):201–211, 2002. doi: 10.1021/bp010179x.

[29] Anthony P Burgard, Priti Pharkya, and Robin E Osterhout. Microorganisms for the production of adipic acid and other compounds, mar 2009.

[30] Christian Weber, Christine Brückner, Sheila Weinreb, Claudia Lehr, Christine Essl, and Eckhard Boles. Biosynthesis of cis,cis-Muconic Acid and Its Aromatic Precursors, Catechol and Protocatechuic Acid, from Renewable Feedstocks by Saccharomyces cerevisiae. *Appl. Environ. Microbiol.*, 78(23):8421–8430, 2012. doi: 10.1128/AEM.01983-12.

[31] Jan C.J. Bart and Stefano Cavallaro. Transiting from Adipic Acid to Bioadipic Acid. Part II. Biosynthetic Pathways. *Ind. Eng. Chem. Res.*, 54(2):567–576, jan 2015. doi: 10.1021/IE502074D.

[32] Xin Jin, Mengyuan Liu, Guangyu Zhang, Jinyao Wang, Qi Xia, Yu Sun, Ziqi Zhou, Wenxiang Zhang, Shu Wang, Chun Ho Lam, Jian Shen, Chaohe Yang, and Raghunath Vitthal Chaudhari. Chemical Synthesis of Adipic Acid from Glucose and Derivatives: Challenges for Nanocatalyst Design. *ACS Sustain. Chem. Eng.*, 8(51): 18732–18754, 2020. doi: 10.1021/ACSSUSCHEMENG.0C04411.

[33] Christopher W. Johnson, Davinia Salvachúa, Payal Khanna, Holly Smith, Darren J. Peterson, and Gregg T. Beckham. Enhancing muconic acid production from glucose and lignin-derived aromatic compounds via increased protocatechuate decarboxylase activity. *Metab. Eng. Commun.*, 3:111–119, 2016. doi: 10.1016/J.METENO.2016.04.002.

[34] Michael E. Pyne, Lauren Narcross, Mindy Melgar, Kaspar Kevvai, Shoham Mookerjee, Gustavo B. Leite, and Vincent J.J. Martin. An Engineered Aro1 Protein Degradation Approach for Increased cis,cis-Muconic Acid Biosynthesis in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.*, 84(17), sep 2018. doi: 10.1128/AEM.01095-18.

[35] Mei Zhao, Dixuan Huang, Xiaojuan Zhang, Mattheos A G Koffas, Jingwen Zhou, and Yu Deng. Metabolic engineering of Escherichia coli for producing adipic acid through the reverse adipate-degradation pathway. *Metab. Eng.*, 47:254–262, 2018. doi: 10.1016/j.ymben.2018.04.002.

[36] Thomas Beardslee and Stephen Picataggio. Bio-based adipic acid from renewable oils. *Lipid Technol.*, 24(10):223–225, 2012. doi: 10.1002/lite.201200230.

[37] Si Sun Choi, Seung Yeul Seo, Sun Ok Park, Han Na Lee, Ji Soo Song, Ji Yeon Kim, Ji Hoon Park, Sangyong Kim, Sang Joung Lee, Gie Taek Chun, and Eung Soo Kim. Cell

Factory Design and Culture Process Optimization for Dehydroshikimate Biosynthesis in Escherichia coli. *Front. Bioeng. Biotechnol.*, 7:241, oct 2019. doi: 10.3389/FBIOE.2019.00241/BIBTEX.

[38] Guokun Wang, Leyman Øzmerih, Rio Guerreiro, Ana C Meireles, Ana Carolas, Nicholas Milne, Michael K Jensen, Bruno S Ferreira, and Irina Borodina. Improvement of cis,cis-Muconic Acid Production in Saccharomyces cerevisiae through Biosensor-Aided Genome Engineering. *ACS Synth. Biol.*, 9:634–646, 2020. doi: 10.1021/acssynbio.9b00477.

[39] Bin Cao, Anli Geng, and Kai Chee Loh. Induction of ortho- and meta-cleavage pathways in Pseudomonas in biodegradation of high benzoate concentration: MS identification of catabolic enzymes. *Appl. Microbiol. Biotechnol.*, 81(1):99–107, nov 2008. doi: 10.1007/S00253-008-1728-3/TABLES/2.

[40] Lauren S. Collier, George L. Gaines, and Ellen L. Neidle. Regulation of Benzoate Degradation in Acinetobacter sp. Strain ADP1 by BenM, a LysR-Type Transcriptional Activator. *J. Bacteriol.*, 180(9), 1998. doi: 10.1128/jb.180.9.2493-2501.1998.

[41] Jeong Chan Joo, Anna N. Khusnutdinova, Robert Flick, Taeho Kim, Uwe T. Bornscheuer, Alexander F. Yakunin, and Radhakrishnan Mahadevan. Alkene hydrogenation activity of enoate reductases for an environmentally benign biosynthesis of adipic acid. *Chem. Sci.*, 8(2):1406–1413, jan 2017. doi: 10.1039/C6SC02842J.

[42] Kaushik Raj, Siavash Partow, Kevin Correia, Anna N. Khusnutdinova, Alexander F. Yakunin, and Radhakrishnan Mahadevan. Biocatalytic production of adipic acid from glucose using engineered Saccharomyces cerevisiae. *Metab. Eng. Commun.*, 6:28–32, jun 2018. doi: 10.1016/J.METENO.2018.02.001.

[43] Jennifer N Hennigan, Philip S Wagner, Chris Burk, John P Efromson, Zhixia Ye, Matthew L Lipscomb, and Michael D Lynch. A Technoeconomic Evaluation of the Potential of Industrial Biotechnology for the Competitive Production of Commodity and Bulk Chemicals. *ChemRxiv*, nov 2020. doi: 10.26434/CHEMRXIV.13238996.V1.

[44] Jufang Wang, Meng Lin, Mengmeng Xu, and Shang-Tian Yang. *Anaerobic Fermentation for Production of Carboxylic Acids as Bulk Chemicals from Renewable Biomass in "Anaerobes in Biotechnology"*. Springer International Publishing, 2016. doi: 10.1007/10_2015_5009.

[45] Emma Karlsson, Valeria Mapelli, and Lisbeth Olsson. Adipic acid tolerance screening for potential adipic acid production hosts. *Microb. Cell Fact.*, 16(1):1–17, 2017. ISSN 14752859. doi: 10.1186/s12934-017-0636-6.

[46] Amita Gupta, Michael A. Hicks, Shawn P. Manchester, and Kristala L.J. Prather. Porting the synthetic D-glucaric acid pathway from Escherichia coli to Saccharomyces cerevisiae. *Biotechnol. J.*, 11(9):1201–1208, 2016. doi: 10.1002/BIOT.201500563.

[47] Hyun Joo Park, Jung Hoon Bae, Hyeok Jin Ko, Sun Hee Lee, Bong Hyun Sung, Jong In Han, and Jung Hoon Sohn. Low-pH production of d-lactic acid using newly isolated acid tolerant yeast Pichia kudriavzevii NG7. *Biotechnol. Bioeng.*, 115(9):2232–2242, sep 2018. doi: 10.1002/BIT.26745.

[48] Nuno P. Mira, Miguel Cacho Teixeira, and Isabel Sá-Correia. Adaptive response and tolerance to weak acids in Saccharomyces cerevisiae: A genome-wide view. *Omi. A J. Integr. Biol.*, 14(5):525–540, 2010. doi: 10.1089/omi.2010.0072.

[49] Miho Kawahata, Kazuo Masaki, Tsutomu Fujii, and Haruyuki Iefuji. Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in Saccharomyces cerevisiae cultures induce expression of intracellular metal metabolism genes regulated by Aft1p. *FEMS Yeast Res.*, 6(6):924–936, 2006. doi: 10.1111/J.1567-1364.2006.00089.X.

[50] V. Yu Sekova, N. N. Gessler, E. P. Isakova, A. N. Antipov, D. I. Dergacheva, Y. I. Deryabina, and E. V. Trubnikova. Redox status of extremophilic yeast Yarrowia lipolytica during adaptation to pH-stress. *Appl. Biochem. Microbiol.*, 51(6):649–654, 2015. doi: 10.1134/S0003683815060137.

[51] Nuno P. Mira, Jorg D. Becker, and Isabel Sá-Correia. Genomic Expression Program Involving the Haa1p-Regulon in Saccharomyces cerevisiae Response to Acetic Acid. *OMICS*, 14(5):587, 2010. doi: 10.1089/OMI.2010.0048.

[52] Yvonne Nygård, Dominik Mojzita, Mervi Toivari, Merja Penttilä, Marilyn G. Wiebe, and Laura Ruohonen. The diverse role of Pdr12 in resistance to weak organic acids. *Yeast*, 31(6):219–232, 2014. doi: 10.1002/YEA.3011.

[53] H. F. De Melo, B. M. Bonini, J. Thevelein, D. A. Simões, and M. A. Morais. Physiological and molecular analysis of the stress response of Saccharomyces cerevisiae imposed by strong inorganic acid with implication to industrial fermentations. *J. Appl. Microbiol.*, 109(1):116–127, 2010. doi: 10.1111/J.1365-2672.2009.04633.X.

[54] A. R. Fernandes, N. P. Mira, R. C. Vargas, I. Canelhas, and I. Sá-Correia. Saccharomyces cerevisiae adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. *Biochem. Biophys. Res. Commun.*, 337(1):95–103, nov 2005. doi: 10.1016/J.BBRC.2005.09.010.

[55] Yachen Dong, Jingjin Hu, Linlin Fan, and Qihe Chen. RNA-Seq-based transcriptomic and metabolomic analysis reveal stress responses and programmed cell death induced by acetic acid in Saccharomyces cerevisiae. *Sci. Rep.*, 7(1):1–16, 2017. doi: 10.1038/srep42659.

[56] Margarida Palma, Joana F. Guerreiro, and Isabel Sá-Correia. Adaptive response and tolerance to acetic acid in Saccharomyces cerevisiae and Zygosaccharomyces bailii: A physiological genomics perspective. *Front. Microbiol.*, 9:274, 2018. doi: 10.3389/FMICB.2018.00274/BIBTEX.

[57] Jackson Peter, Matteo De Chiara, Anne Friedrich, Jia Xing Yue, David Pflieger, Anders Bergström, Anastasie Sigwalt, Benjamin Barre, Kelle Freel, Agnès Llored, Corinne Cruaud, Karine Labadie, Jean Marc Aury, Benjamin Istace, Kevin Lebrigand, Pascal Barbry, Stefan Engelen, Arnaud Lemainque, Patrick Wincker, Gianni Liti, and Joseph Schacherer. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature*, 556(7701):339–344, 2018. doi: 10.1038/s41586-018-0030-5.

[58] Juliana L.R. Fietto, Raquel S. Araújo, Frederico N. Valadão, Luciano G. Fietto, Rogelio L. Brandão, Maria J. Neves, Fátima C.O. Gomes, Jacques R. Nicoli, and Ieso M. Castro. Molecular and physiological comparisons between Saccharomyces cerevisiae and Saccharomyces boulardii. *Can. J. Microbiol.*, 50(8):615–621, aug 2004. doi: 10.1139/W04-050.

[59] Indu Khatri, Rajul Tomar, K. Ganesan, G. S. Prasad, and Srikrishna Subramanian. Complete genome sequence and comparative genomics of the probiotic yeast Saccharomyces boulardii. *Sci. Rep.*, 7(1):1–12, 2017. doi: 10.1038/s41598-017-00414-2.

[60] Laura Edwards-Ingram, Paul Gitsham, Nicola Burton, Geoff Warhurst, Ian Clarke, David Hoyle, Stephen G. Oliver, and Lubomira Stateva. Genotypic and physiological characterization of Saccharomyces boulardii, the probiotic strain of Saccharomyces cerevisiae. *Appl. Environ. Microbiol.*, 73(8):2458–2467, 2007. doi: 10.1128/AEM.02201-06.

[61] John C. Marioni, Christopher E. Mason, Shrikant M. Mane, Matthew Stephens, and Yoav Gilad. RNA-seq: An assessment of technical reproducibility and comparison

with gene expression arrays. *Genome Res.*, 18(9):1509–1517, 2008. doi: 10.1101/GR.079558.108.

[62] Zhong Wang, Mark Gerstein, and Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, 10(1):57–63, 2009. doi: 10.1038/NRG2484.

[63] Xiaobei Zhou, Helen Lindsay, and Mark D. Robinson. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Res.*, 42 (11):e91, jun 2014. doi: 10.1093/NAR/GKU310.

[64] Henrik Edgren, Astrid Murumagi, Sara Kangaspeska, Daniel Nicorici, Vesa Hongisto, Kristine Kleivi, Inga H. Rye, Sandra Nyberg, Maija Wolf, Anne Lise Borresen-Dale, and Olli Kallioniemi. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol.*, 12(1):1–13, 2011. doi: 10.1186/GB-2011-12-1-R6/FIGURES/5.

[65] Christopher A Maher, Chandan Kumar-Sinha, Xuhong Cao, Shanker Kalyana-Sundaram, Bo Han, Xiaojun Jing, Lee Sam, Terrence Barrette, Nallasivam Palanisamy, and Arul M Chinnaiyan. Transcriptome sequencing to detect gene fusions in cancer. *Nature*, 458:97–101, 2009. doi: 10.1038/nature07638.

[66] Mitchell Guttman, Manuel Garber, Joshua Z Levin, Julie Donaghey, James Robinson, Xian Adiconis, Lin Fan, Magdalena J Koziol, Andreas Gnirke, Chad Nusbaum, John L Rinn, Eric S Lander, and Aviv Regev. Ab initio reconstruction of cell type–specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.*, 28(5):503–510, 2010. doi: 10.1038/nbt.1633.

[67] Charlotte Soneson and Mauro Delorenzi. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14(1):1–18, mar 2013. doi: 10.1186/1471-2105-14-91/TABLES/3.

[68] Rui Pereira, Yongjun Wei, Elsayed Mohamed, Mohammad Radi, Carl Malina, Markus J. Herrgård, Adam M. Feist, Jens Nielsen, and Yun Chen. Adaptive laboratory evolution of tolerance to dicarboxylic acids in Saccharomyces cerevisiae. *Metab. Eng.*, 56:130–141, 2019. doi: 10.1016/J.YMBEN.2019.09.008.

[69] Daniel González-Ramos, Arthur R. Gorter De Vries, Sietske S. Grijseels, Margo C. Van Berkum, Steve Swinnen, Marcel Van Den Broek, Elke Nevoigt, Jean Marc G. Daran, Jack T. Pronk, and Antonius J.A. Van Maris. A new laboratory evolution approach to select for constitutive acetic acid tolerance in Saccharomyces cerevisiae

and identification of causal mutations. *Biotechnol. Biofuels*, 9(1):1–18, 2016. doi: 10.1186/S13068-016-0583-1/FIGURES/6.

[70] Robert Todd, Ann Braverman, and Anna Selmecki. Flow Cytometry Analysis of Fungal ploidy. *Curr. Protoc. Microbiol.*, 50(1):139–148, 2018. doi: 10.1002/cpmc.58.Flow.

[71] Luis A. Corchete, Elizabeta A. Rojas, Diego Alonso-López, Javier De Las Rivas, Norma C. Gutiérrez, and Francisco J. Burguillo. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Sci. Rep.*, 10(1):1–15, nov 2020. doi: 10.1038/s41598-020-76881-x.

[72] Manfred G. Grabherr, Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica Di Palma, Bruce W. Birren, Chad Nusbaum, Kerstin Lindblad-Toh, Nir Friedman, and Aviv Regev. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.*, 29(7):644, jul 2011. doi: 10.1038/NBT.1883.

[73] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, jan 2010. doi: 10.1093/BIOINFORMATICS/BTP616.

[74] Simon Andrews, Felix Krueger, Anne Segonds-Pichon, Laura Biggins, Christel Krueger, and Steven Wingett. FastQC, 2012.

[75] Anthony M Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014. doi: 10.1093/bioinformatics/btu170.

[76] Alex N. Salazar, Arthur R.Gorter de Vries, Marcel van den Broek, Melanie Wijsman, Pilar de la Torre Cortés, Anja Brickwedde, Nick Brouwers, Jean Marc G. Daran, and Thomas Abeel. Nanopore sequencing enables near-complete de novo assembly of Saccharomyces cerevisiae reference strain CEN.PK113-7D. *FEMS Yeast Res.*, 17(7):1–11, 2017. doi: 10.1093/femsyr/fox074.

[77] Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR:

ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, jan 2013. doi: 10.1093/BIOINFORMATICS/BTS635.

[78] Bo Li and Colin N. Dewey. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1):1–16, aug 2011. doi: 10.1186/1471-2105-12-323.

[79] Brian J. Haas, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood, Joshua Bowden, Matthew Brian Couger, David Eccles, Bo Li, Matthias Lieber, Matthew D. Macmanes, Michael Ott, Joshua Orvis, Nathalie Pochet, Francesco Strozzi, Nathan Weeks, Rick Westerman, Thomas William, Colin N. Dewey, Robert Henschel, Richard D. Leduc, Nir Friedman, and Aviv Regev. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat. Protoc.*, 8(8):1494–1512, aug 2013. doi: 10.1038/NPROT.2013.084.

[80] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, 9(4):357, apr 2012. doi: 10.1038/NMETH.1923.

[81] Mathieu Seppey, Mosè Manni, and Evgeny M. Zdobnov. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol.*, 1962:227–245, 2019. doi: 10.1007/978-1-4939-9173-0_14.

[82] Evgenia V. Kriventseva, Dmitry Kuznetsov, Fredrik Tegenfeldt, Mosè Manni, Renata Dias, Felipe A. Simão, and Evgeny M. Zdobnov. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.*, 47(1):807–811, jan 2019. doi: 10.1093/NAR/GKY1053.

[83] Weizhong Li and Adam Godzik. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659, jul 2006. doi: 10.1093/BIOINFORMATICS/BTL158.

[84] Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Tim D. Williams, Shivashankar H. Nagaraj, María José Nueda, Montserrat Robles, Manuel Talón, Joaquín Dopazo, and Ana Conesa. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.*, 36(10):3420–3435, jun 2008. doi: 10.1093/NAR/GKN176.

[85] Matthias Blum, Hsin Yu Chang, Sara Chuguransky, Tiago Grego, Swaathi Kandasaamy, Alex Mitchell, Gift Nuka, Typhaine Paysan-Lafosse, Matloob Qureshi,

Shriya Raj, Lorna Richardson, Gustavo A. Salazar, Lowri Williams, Peer Bork, Alan Bridge, Julian Gough, Daniel H. Haft, Ivica Letunic, Aron Marchler-Bauer, Huaiyu Mi, Darren A. Natale, Marco Necci, Christine A. Orengo, Arun P. Pandurangan, Catherine Rivoire, Christian J.A. Sigrist, Ian Sillitoe, Narmada Thanki, Paul D. Thomas, Silvio C.E. Tosatto, Cathy H. Wu, Alex Bateman, and Robert D. Finn. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.*, 49(1):344–354, jan 2021. doi: 10.1093/NAR/GKAA977.

[86] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, jan 2015. doi: 10.1093/BIOINFORMATICS/BTU638.

[87] Davis J. McCarthy, Yunshun Chen, and Gordon K. Smyth. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.*, 40(10):4288–4297, may 2012. doi: 10.1093/NAR/GKS042.

[88] Yunshun Chen, Aaron T L Lun, and Gordon K Smyth. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, 5(1438):1–51, 2016. doi: 10.12688/f1000research.8987.1.

[89] Uku Raudvere, Liis Kolberg, Ivan Kuzmin, Tambet Arak, Priit Adler, Hedi Peterson, and Jaak Vilo. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists. *Nucleic Acids Res.*, 47(1):191–198, jul 2019. doi: 10.1093/NAR/GKZ369.

[90] Damian Szklarczyk, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda T. Doncheva, John H. Morris, Peer Bork, Lars J. Jensen, and Christian Von Mering. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, 47(1):607–613, 2019. doi: 10.1093/NAR/GKY1131.

[91] Björn D.M. Bean, Malcolm Whiteway, and Vincent J.J. Martin. The MyLo CRISPR-Cas9 Toolkit: A Markerless Yeast Localization and Overexpression CRISPR-Cas9 Toolkit. *bioRxiv*, page 2021.12.15.472800, 2021. doi: 10.1101/2021.12.15.472800.

[92] Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. A programmable dual-RNA-guided DNA endonuclease

in adaptive bacterial immunity. *Science (80-.).*, 337(6096):816–821, aug 2012. doi: 10.1126/SCIENCE.1225829.

[93] Kevin R. Oldenburg, Kham T. Vo, Susan Michaelis, and Chris Paddon. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.*, 25(2):451, 1997. doi: 10.1093/NAR/25.2.451.

[94] T. L. Orr-Weaver, J. W. Szostak, and R. J. Rothstein. Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. U. S. A.*, 78(10):6354, 1981. doi: 10.1073/PNAS.78.10.6354.

[95] Patrick D. Hsu, David A. Scott, Joshua A. Weinstein, F. Ann Ran, Silvana Konermann, Vineeta Agarwala, Yinqing Li, Eli J. Fine, Xuebing Wu, Ophir Shalem, Thomas J. Cradick, Luciano A. Marraffini, Gang Bao, and Feng Zhang. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol. 2013 319*, 31(9):827–832, jul 2013. doi: 10.1038/nbt.2647.

[96] R Daniel Gietz and Robert H Schiestl. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, 2(1):31–35, 2007. doi: 10.1007/978-1-4939-1363-3_1.

[97] Michael E. Lee, William C. DeLoache, Bernardo Cervantes, and John E. Dueber. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.*, 4(9):975–986, 2015. doi: 10.1021/sb500366v.

[98] Dongmei Bai Flagfeldt, Verena Siewers, Le Huang, and Jens Nielsen. Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae. *Yeast*, 26:545–551, 2009. doi: 10.1002/yea.1705.

[99] Gareth A. Cromie and Aimée M. Dudley. Aneuploidy: Tolerating Tolerance. *Curr. Biol.*, 25(17):771–773, aug 2015. doi: 10.1016/J.CUB.2015.06.056.

[100] W. Albertin, P. Marullo, M. Aigle, A. Bourgais, M. Bely, C. Dillmann, D. De Vienne, and D. Sicard. Evidence for autotetraploidy associated with reproductive isolation in Saccharomyces cerevisiae: towards a new domesticated species. *J. Evol. Biol.*, 22(11): 2157–2170, nov 2009. doi: 10.1111/J.1420-9101.2009.01828.X.

[101] Norman Pavelka, Giulia Rancati, Jin Zhu, William D. Bradford, Anita Saraf, Laurence Florens, Brian W. Sanderson, Gaye L. Hattem, and Rong Li. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature*, 468(7321):321–325, oct 2010. doi: 10.1038/nature09529.

[102] Anna M. Selmecki, Keely Dulmage, Leah E. Cowen, James B. Anderson, and Judith Berman. Acquisition of Aneuploidy Provides Increased Fitness during the Evolution of Antifungal Drug Resistance. *PLOS Genet.*, 5(10):e1000705, 2009. doi: 10.1371/JOURNAL.PGEN.1000705.

[103] Cristina A. Viegas, Paulo F. Almeida, Miguel Cavaco, and Isabel Sá-Correia. The H+-ATPase in the plasma membrane of Saccharomyces cerevisiae is activated during growth latency in octanoic acid-supplemented medium accompanying the decrease in intracellular pH and cell viability. *Appl. Environ. Microbiol.*, 64(2):779–783, 1998. doi: 10.1128/AEM.64.2.779-783.1998/ASSET/8F9ACFCB-7DA4-4CC1-8AB6-CD2C0E299A72.

[104] Steve Swinnen, Miguel Fernández-Niño, Daniel González-Ramos, Antonius J.A. van Maris, and Elke Nevoigt. The fraction of cells that resume growth after acetic acid addition is a strain-dependent parameter of acetic acid tolerance in Saccharomyces cerevisiae. *FEMS Yeast Res.*, 14(4):642–653, 2014. doi: 10.1111/1567-1364.12151.

[105] Mark D. Robinson and Alicia Oshlack. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, 11(3):1–9, mar 2010. doi: 10.1186/GB-2010-11-3-R25.

[106] Mikhail Pomaznoy, Brendan Ha, and Bjoern Peters. GOnet: A tool for interactive Gene Ontology analysis. *BMC Bioinformatics*, 19(1):1–8, dec 2018. doi: 10.1186/S12859-018-2533-3.

[107] A. J. Enright, S. Van Dongen, and C. A. Ouzounis. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, 30(7):1575–1584, 2002. doi: 10.1093/NAR/30.7.1575.

[108] Charley G.P. McCarthy and David A. Fitzpatrick. Pan-genome analyses of model fungal species. *Microb. Genomics*, 5(2), 2019. doi: 10.1099/MGEN.0.000243.

[109] Gang Li, Boyang Ji, and Jens Nielsen. The pan-genome of Saccharomyces cerevisiae. *FEMS Yeast Res.*, 19(7), 2019. ISSN 15671364. doi: 10.1093/femsyr/foz064.

[110] Jing Hou, Anne Friedrich, Jacky De Montigny, and Joseph Schacherer. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in Saccharomyces cerevisiae. *Curr. Biol.*, 24(10):1153, may 2014. doi: 10.1016/J.CUB.2014.03.063.

[111] Edward R Sumner and Simon V Avery. Phenotypic heterogeneity : differential stress resistance among individual cells of the yeast Saccharomyces cerevisiae. *Microbiology*, 148:345–351, 2002.

[112] Quan Long, Xiuxia Liu, Yankun Yang, Lu Li, Linda Harvey, Brian McNeil, and Zhonghu Bai. The development and application of high throughput cultivation technology in bioprocess development. *J. Biotechnol.*, 192:323–338, 2014. doi: 10.1016/J.JBIOTEC.2014.03.028.

[113] Sarah Teworte, Koray Malcı, Laura E. Walls, Murni Halim, and Leonardo Rios-Solis. Recent advances in fed-batch microscale bioreactor design. *Biotechnol. Adv.*, 55:107888, 2022. doi: 10.1016/J.BIOTECHADV.2021.107888.

[114] Guo Chang Zhang, In Iok Kong, Heejin Kim, Jing Jing Liu, Jamie H.D. Cate, and Yong Su Jin. Construction of a quadruple auxotrophic mutant of an industrial polyploid Saccharomyces cerevisiae strain by using RNA-guided Cas9 nuclease. *Appl. Environ. Microbiol.*, 80(24):7694–7701, dec 2014. doi: 10.1128/AEM.02310-14.

[115] F. Aguilera, R. A. Peinado, C. Millán, J. M. Ortega, and J. C. Mauricio. Relationship between ethanol tolerance, H+-ATPase activity and the lipid composition of the plasma membrane in different wine yeast strains. *Int. J. Food Microbiol.*, 110(1): 34–42, 2006. doi: 10.1016/J.IJFOODMICRO.2006.02.002.

[116] Breck Byers and Loretta Goetsch. Electron microscopic observations on the meiotic karyotype of diploid and tetraploid Saccharomyces cerevisiae (synaptonemal complex/recombination nodules). Technical Report 12, 1975.

[117] Jin Zhu, Norman Pavelka, William D. Bradford, Giulia Rancati, and Rong Li. Karyotypic Determinants of Chromosome Instability in Aneuploid Budding Yeast. *PLOS Genet.*, 8(5):e1002719, may 2012. doi: 10.1371/JOURNAL.PGEN.1002719.

[118] Ana B Oromendia, Stacie E Dodgson, and Angelika Amon. Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.*, 26(24):2696–2708, 2012. doi: 10.1101/gad.207407.112.

[119] Avihu H Yona, Yair S Manor, Rebecca H Herbst, Gal H Romano, Amir Mitchell, Martin Kupiec, Yitzhak Pilpel, and Orna Dahan. Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U. S. A.*, 109(51):21010–21015, 2012. doi: 10.1073/pnas.1211150109.

[120] Marek Gierliński, Christian Cole, Pietà Schofield, Nicholas J. Schurch, Alexander Sherstnev, Vijender Singh, Nicola Wrobel, Karim Gharbi, Gordon Simpson, Tom Owen-Hughes, Mark Blaxter, and Geoffrey J. Barton. Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. *Bioinformatics*, 31(22): 3625, may 2015. doi: 10.1093/BIOINFORMATICS/BTV425.

[121] Ana Conesa, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michal Wojciech Szcześniak, Daniel J. Gaffney, Laura L. Elo, Xuegong Zhang, and Ali Mortazavi. A survey of best practices for RNA-seq data analysis. *Genome Biol.*, 17(1):1–19, 2016. doi: 10.1186/S13059-016-0881-8.

[122] Xiaoying Chen, Bo Zhang, Ting Wang, Ting Wang, Azad Bonni, and Guoyan Zhao. Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. *BMC Bioinformatics*, 21(1):1–20, 2020. doi: 10.1186/S12859-020-03608-0.

[123] Tiantian Tian, Dianhui Wu, Chan Tat Ng, Hua Yang, Jun Liu, Junyong Sun, and Jian Lu. Uncovering mechanisms of greengage wine fermentation against acidic stress via genomic, transcriptomic, and metabolic analyses of Saccharomyces cerevisiae. *Appl. Microbiol. Biotechnol.*, 104(17):7619–7629, 2020. doi: 10.1007/S00253-020-10772-Z/FIGURES/3.

[124] Daniel Clark, Roger S. Rowlett, John R. Coleman, and Daniel F. Klessig. Complementation of the yeast deletion mutant DeltaNCE103 by members of the beta class of carbonic anhydrases is dependent on carbonic anhydrase activity rather than on antioxidant activity. *Biochem. J.*, 379(3):609, 2004. doi: 10.1042/BJ20031711.

[125] Yeji Lee, Olviyani Nasution, Young Mi Lee, Eunjung Kim, Wonja Choi, and Wankee Kim. Overexpression of PMA1 enhances tolerance to various types of stress and constitutively activates the SAPK pathways in Saccharomyces cerevisiae. *Appl. Microbiol. Biotechnol.*, 101(1):229–239, jan 2017. doi: 10.1007/S00253-016-7898-5.

[126] Priyanka Saini, Arun Beniwal, Anusha Kokkiligadda, and Shilpa Vij. Response and tolerance of yeast to changing environmental stress during ethanol fermentation. *Process Biochem.*, 72:1–12, 2018. doi: 10.1016/J.PROCBIO.2018.07.001.

[127] Audrey P. Gasch, Paul T. Spellman, Camilla M. Kao, Orna Carmel-Harel, Michael B. Eisen, Gisela Storz, David Botstein, and Patrick O. Brown. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, 11(12):4241–4257, oct 2000. doi: 10.1091/MBC.11.12.4241.

[128] Robert A. Crawford and Graham D. Pavitt. Translational regulation in response to stress in Saccharomyces cerevisiae. *Yeast*, 36(1):5–21, 2019. doi: 10.1002/YEA.3349.

[129] Arne Peetermans, María R. Foulquié-Moreno, and Johan M. Thevelein. Mechanisms underlying lactic acid tolerance and its influence on lactic acid production in Saccharomyces cerevisiae. *Microb. Cell*, 8(6):111–130, 2021. doi: 10.15698/MIC2021.06.751.

[130] Yongyan Xi, Tao Zhan, Hongtao Xu, Jing Chen, Changhao Bi, Feiyu Fan, and Xueli Zhang. Characterization of JEN family carboxylate transporters from the acid-tolerant yeast Pichia kudriavzevii and their applications in succinic acid production. *Microb. Biotechnol.*, 14(3):1130–1147, 2021. doi: 10.1111/1751-7915.13781.

[131] Dongmei Wang, Dan Wu, Xiaoxue Yang, and Jiong Hong. Transcriptomic analysis of thermotolerant yeast Kluyveromyces marxianus in multiple inhibitors tolerance. *RSC Adv.*, 8(26):14177–14192, 2018. doi: 10.1039/C8RA00335A.