Examination of Organic Acid Tolerance in Non-Conventional Yeasts

James Andrew Bagley

A Thesis

In the Department of

Biology

Presented in Partial Fulfilment of the Requirements

for the Degree of Master of Science (Biology) at

Concordia University

Montreal, Quebec, Canada

December 2021

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: James Andrew Bagley

Entitled: Examination of Organic Acid Tolerance in Non-Conventional Yeasts.

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

Master of Science (Biology)

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

Signed by the final Examining Committee:

_____ Chair
Dr. Malcolm Whiteway

_____ Examiner
Dr. Aashiq Kachroo

_____ Examiner
Dr. Malcolm Whiteway

_____ Examiner
Dr. Paul Joyce

_____ Supervisor
Dr. Vincent Martin

Approved by _____

Dr. Robert Weladji, Graduate Program Director

Date _____ 2022 _____

Dr. Pascale Sicotte, Dean of Arts and Science

ABSTRACT

Examination of Organic Acid Tolerance in Non-Conventional Yeasts

James Andrew Bagley

Metabolic engineering of yeasts has proven an effective strategy for producing compounds ranging from commodity chemicals to biologics. However, in the case of certain organic acids, there is a toxicity barrier, which prevents commercial production from being viable. To address this problem, we developed a strategy to characterize non-conventional yeasts and used it to search fungal repositories for desirable phenotypes, in our case tolerance to adipic acid, a nylon 6,6 precursor. From publicly accessible yeast collections we selected and screened a collection of 122 strains of yeasts. After finding strains that were tolerant to high concentrations of adipic acid at an industrially relevant pH, suitable antibiotic markers were found and whole genome sequencing and annotation was performed to enable future metabolic engineering efforts in these strains. Annotated strains were examined for evidence of gene expansion in families commonly associated with organic acid tolerance and candidate genes were identified for further research.

Acknowledgements

I would like to begin by thanking my supervisor Professor Vincent Martin for the opportunities and guidance he has afforded me throughout my time in his lab. I would also like to give special thanks to Kaspar Kevvai for contributing his counsel, his humour, and his broad knowledge of Saccharomycotina yeasts.

I would also like to thank Malcolm Whiteway for his involvement in the project and advice regarding the molecular biology of non-Saccharomyces yeasts. The whole of the Martin lab during my tenure also deserves thanks, so Lauren, Mike, Bjorn, Mo, Will, Meghan, Dan, Mindy, Nick, Smita, Audrey, Iain, and Mathieu thank you all for your help.

Finally, I would like to thank my parents and my brother for their love and support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ATCC – American type culture collection

BLAST – Basic Local Alignment Search Tool

CBS – CBS-KNAW fungal collection

CDS – Coding sequence

HMM – Hidden Markov model

logP – n-octanol water partition coefficient

MUSCLE – Multiple sequence comparison by log expectation

NCYC – National collection of yeast cultures

NHEJ – Non-homologous end joining

NRRL – Northern Regional Research Lab Agricultural Resource Service culture collection

OD600 – Optical density measured at a wavelength of 600 nm

SMRT – Single molecule real time sequencing

YNB – Yeast nitrogen base medium

YPD – Yeast peptone dextrose medium

1. Thesis Goal and Objectives

To date, most metabolic engineering efforts have taken place in a select few microbial species. However, to produce compounds such as organic acids at the high concentrations required to be commercially viable, other robust hosts are necessary due to the inherent cytotoxicity of organic acids. The goal of this thesis was to demonstrate a strategy to rapidly find and genetically characterize organic acid-tolerant non-conventional yeasts. The strategy involved a) initial screening of a collection of yeast species, b) in-depth characterization of relevant phenotypes, c) SMRT whole genome sequencing of candidate strains, d) draft annotation of candidate genomes.

As a demonstration for this strategy, we acquired a collection of non-conventional yeast strains and screened them for tolerance to adipic acid, a commodity chemical used in the synthesis of nylon 6,6 and our metabolic engineering target. To aid both reproducibility and reusability of the generated data we preferentially used type strains and drew on strains from public repositories rather than environmental isolates. After screening against adipic acid, select strains were evaluated for tolerance to a suite of 8 different organic acids, to identify whether stains were only tolerant to adipic acid or if there was cross tolerance to other organic acids. At this stage additional strains that were phylogenetically close to those identified in the initial screen were purchased and again tested for tolerance to organic acids. Seventeen strains were ultimately sequenced and had their genomes assembled, 12 of the 17 were functionally annotated, and key protein families implicated in organic acid tolerance were phylogenetically reconstructed.

2. Literature Review

2.1. Stressor tolerance in Saccharomycotina

As of March 2020, the order Saccharomycotina had 1850 published whole genome assemblies from 460 different species available on GenBank. Being home to *Saccharomyces cerevisiae*, arguably the best studied eukaryotic organism, the order has benefitted from a deep understanding of most of the fundamental metabolic and signalling pathways. However, as the majority of this understanding has been derived from one species, and arguably just one strain of that species, details of how the specifics of metabolism vary from species to species is yet undetermined. Recognising this issue, research has intensified on several other well characterized species in the order. For example, *Komagataella* (*Pichia*) *pastoris*, *Kluyveromyces lactis*, *Candida albicans* and *Yarrowia lipolytica*, have been not only sequenced but also characterized to a high degree, though not to the same extent as *S. cerevisiae* (Correia & Mahadevan, 2020; Shen et al., 2018).

However, these organisms still only represent five species of over 1000 identified in the order, and there may still be many species with great scientific and economic potential that are yet to be discovered. In particular, species with a tolerance for the extreme conditions that can arise during metabolic engineering applications hold great potential to improve bioprocess productivity. Table 1 shows a collection of highly tolerant yeasts and their associated stressors. What is particularly useful about yeasts compared to other taxa is that they remain culturable in standard lab conditions. While extremophilic bacteria may only be culturable in anoxic, or hyperthermic conditions, the overwhelming majority of yeast species can be cultivated and developed in the lab at room temperature with standard yeast media (Kurtzman et al., 2011). While the extremophilic bacteria are more tolerant to the most extreme conditions than are yeasts, working with difficult to cultivate microorganisms

increases the time to market for developing a bioprocess. Given the advantages of short development times, adopting an easier-to-work-with organism that still provides significant stress tolerance may be a trade-off worth making in many cases.

Table 1. Examples of highly tolerant yeasts and relevant stressors.

Species	**Family**	**Stressor**
Yamadazyma farinosa	Debaryomyces	70% sorbitol (Louis et al., 2012)
Yamadazyma farinosa	Debaryomyces	4M NaCl (Louis et al., 2012)
Pichia kudriavzevii	Pichiaceae	pH 2.0 (Park et al., 2018)
Pichia kudriavzevii	Pichiaceae	50°C (Park et al., 2018)
Zygosaccharomyces rouxii	Saccharomycetaceae	70% glucose (Solieri et al., 2014)
Pichia membranifaciens	Pichiaceae	400 mM acetate (Konishi et al., 2017)

2.2. Organic acid tolerance in yeasts

Previous work on the tolerance of microorganisms to adipic acid and the conjugate base, adipate, has found that the environmental pH has a huge impact (Karlsson et al., 2017). While *E. coli* can grow in the presence of adipate at pH 7 with very little disturbance to growth rate, even at concentrations as high as 650 mM, dropping the pH to 6 results in a complete arrest at just 192 mM (Karlsson et al., 2017). *S. cerevisiae* by contrast was more tolerant and able to grow at pH 5, even with a concentration of 650 mM, though with a significant reduction in growth rate (Karlsson et al., 2017). However, both pH 6 and pH 5 are higher than the pKa of adipic acid, meaning a majority of the added adipic acid was in the dissociated form, which is much less toxic (Karlsson et al., 2017). In the case of sorbate exposure in *S. cerevisiae*, a decrease in internal pH and an increase in energy expenditure due to expulsion of weak acid anions are the two best documented effects (Piper et al., 2001; Ullah et al., 2013), in *E. coli* an accumulation of anions in the cell has also been observed to disrupt anion balance (Roe et al., 1998).

Zygosaccharomyces bailii has emerged as a frequently studied model for organic acid tolerance in yeast, and a key tolerance mechanism that has been established is a reduction of organic acid diffusion from the extracellular milieu into the cell. This reduction is mediated by an increased proportion of sphingolipids in the plasma membrane (Lindahl et al., 2016) and reduced intracellular pH (Dang et al., 2012). As is the case in *S. cerevisiae*, *Z. bailii* also uses ATPases and anion pumps to reduce the toxic effects of internal accumulation of anions and protons, however, previously mentioned mechanisms reduce the flux of organic acids into the cell and therefore enable the growth of *Z. bailii* in more acidic conditions (Lindahl et al., 2016; Palma et al., 2018), including growth in high concentrations of acetic acid below its pKa of 4.8 (Stratford et al., 2013). Despite this general increase in tolerance to organic acids, and low pH, *Z. bailii* grew similarly to *S. cerevisiae* in the presence of adipic acid (Karlsson et al., 2017). *P. kudriavzevii* is a less characterized species than *Z. bailii* but was shown to be tolerant to high concentrations of succinic acid at low pH (Xiao et al., 2014); however, comparatively little research has been done on the mechanisms of its acid tolerance. The research that exists suggests the use of arginine catabolism to maintain plasma membrane electric potential, and more robust cell wall proteins are involved (Ji et al., 2020; Matsushika et al., 2016; Park et al., 2018).

The distinction between tolerance for low pH conditions, and tolerance for organic acids has been well documented (Fletcher et al., 2017). A yeast species able to grow in low environmental pH will not necessarily tolerate the presence of organic acids (Fletcher et al., 2017; Stratford et al., 2013). Additionally, the cell's response and mechanisms of tolerance to organic acids can vary both between organic acids and inorganic acids as well as between different organic acids (Fletcher et al., 2017; Kawahata et al., 2006; Pereira et al., 2019). Given this heterogeneity and the broad range of organic acids that are interesting biotechnological targets (Abbott et al., 2009; Lin et al., 2014; Park et al., 2018; Pyne et al.,

2018; J.-L. Yu et al., 2018), investigating strategies to robustly assess tolerance of many yeast species to a specific organic acid could prove highly fruitful.

2.3. Genomics of non-Saccharomyces Saccharomycotina

One aspect that makes broad scale screens of yeasts practical is the small genome sizes of yeasts (typically 15 Mbp or less for a haploid), which makes them relatively easy to sequence, requiring approximately 1/3 of a PacBio cell to sequence one genome with 90X coverage. This has resulted in a plethora of available genomes, 1850 assemblies are available for the species that make up the Saccharomycotina clade, and assembly and annotation tools have been developed to aid in the analysis of new genomes from this clade (LRSDAY (Yue & Liti, 2018), AyBraH (Correia et al., 2019), AyBraHAM (Correia & Mahadevan, 2020), YGOB (Byrne, 2005), MGOB (Douglass et al., 2019), CGOB (Maguire et al., 2013)). These tools have enabled annotation of new species and the dissection of gene function to be easier than ever. The level of sequence availability has also enabled researchers to study the evolutionary patterns of the order as a whole and can provide insights into both current yeasts and the evolutionary conditions that produced them (Krassowski et al., 2019; Riley et al., 2016; Shen et al., 2018).

To facilitate the development of Saccharomycotina species for industrial application, new methods have been developed including multiplexed CRISPR-Cas9 knockouts, synthetic and synthetic-hybrid promoters, standardized cloning methods with interchangeable parts, landing pads and synthetic chromosomes for stable multiplexed integrations (Z. Liu et al., 2017). While some of these methods are specific to *S. cerevisiae*, many can be applied more broadly. Even the projects specific to *S. cerevisiae*, such as creation of curated parts toolkits can be replicated and implemented in novel species, as has been demonstrated with both the

Y. lipolytica and *K. marxianus* golden gate toolkits (Larroude et al., 2019; Rajkumar et al., 2019)

2.4. Metabolic engineering of non-Saccharomyces Saccharomycotina

Due to its status as a model organism, *S. cerevisiae* has the benefit of being one of the best understood potential hosts for metabolic engineering applications (Lian et al., 2018). The native central metabolic pathways are well understood, the phenotypic effects of individual gene knockouts are characterized and there is an abundance of transcriptomics data available to aid in troubleshooting pathway engineering (R. Yu & Nielsen, 2019). This has enabled the metabolic engineering of *S. cerevisiae* to flourish, and the number of compounds that have been made in high titres ranges from complex plant metabolites such as artemisinic acid (Ro et al., 2006) and tetrahydroisoquinoline alkaloids (Pyne et al., 2020) to commodity chemicals like succinic acid (Otero et al., 2013), and lactic acid (Porro et al., 1995). However, *S. cerevisiae* is imperfect for the bioproduction of many products due both to its natural tendency to accumulate ethanol, which diverts carbon away from product formation and to the wide swings in gene expression between fermentative and respiratory growth phases (Ho et al., 2018).

One of the first Saccharomycotina strains seriously investigated for use as a metabolic engineering host, aside from *S. cerevisiae*, was *Y. lipolytica* (Beopoulos et al., 2009). The species' high acetyl-CoA flux is diametrically opposed to the strong glycolysis preference of *S. cerevisiae* making it ideal for producing lipids, TCA intermediates and other acetyl-CoA derived compounds (Beopoulos et al., 2009; H. Liu et al., 2019). The species naturally accumulates lipids and is strictly aerobic and cannot ferment, unlike *S. cerevisiae*, which ferments under aerobic conditions in the presence of high glucose concentrations (known as the Crabtree effect). The result of this is *Y. lipolytica* is well suited for production of anything

derived from acetyl-CoA such as triacyl glycerides, succinate, and carotenoids (Cui et al., 2017; H. Liu et al., 2019; Yuzbashev et al., 2010).

K. marxianus is a thermotolerant, acid tolerant, rapid growing yeast, which has fostered interest in it for metabolic engineering applications (Castro & Roberto, 2014; Marcišauskas et al., 2019; Morrissey et al., 2015; Rajkumar et al., 2019). This has manifested in the development of a *K. marxianus* toolkit, inspired by the yeast toolkit (YTK), which is based on the same MoClo standard as the YTK allowing for interchange of CDSs (Lee et al., 2015). The toolkit has so far been used to produce *K. marxianus* strains that have increased flux to phenylalanine and 2-phenylethanol by adapting an established strategy for *S. cerevisiae*, and to create NHEJ deficient mutants to decrease the risk of random integrations. A manually curated metabolic model of *K. marxianus* strain DMKU3–1042 has also been developed to facilitate metabolic engineering projects (Marcišauskas et al., 2019).

Part of the credit for successes in metabolic engineering of non-conventional strains must be given to the transfer of knowledge from metabolic engineering of model species to non-conventional species without having to replicate the years of systems biology research that has been performed on *S. cerevisiae*. In an extreme case, one group successfully produced lactic acid in four different species of yeast, using a single set of promoters and terminators, and integrating into conserved rDNA, allowing a single genetic construct to be used across species (Li et al., 2017).

While the development of these strains has been successful in their respective goals, the strain selection process for metabolic engineering of non-conventional yeasts has typically been either picked semi-arbitrarily based on general characteristics reported in the literature, e.g., acid tolerance, lactose metabolism, or by environmental isolation. We demonstrate the feasibility of a third approach, building a library of non-conventional strains

and then screening the collection with specific criteria relevant to our end goal, tolerance to adipic acid. This approach is more conducive to replication by other groups than environmental isolation as all the strains were sourced from publicly accessible depositories. As the strains are tested for the specific condition we are interested in as opposed to relying on what is in the pre-existing literature, we also improve our odds of finding the ideal strain for our goal when compared to picking a strain semi-arbitrarily. As the tools to engineer new organisms continue to advance, this strategy will become more and more advantageous

3. Materials and Methods

3.1. Yeast strains selection

A collection of 152 yeast strains were selected based on a mix of existing literature indicating tolerance, while prioritizing phylogenetic diversity. Strains were ordered from various public repositories in North America and Europe (Table 2).

Table 2. List of strains used in this study and deposit IDs. Depositor ID reflects which source was used; alternate IDs are listed for ease of reference.

Code	Species name	Depositor ID	NRRL	CBS	NCYC	ATCC	Phaff
001	*Barnettozyma californica*	8860 CBS		8860			
002	*Barnettozyma californica*	8866 CBS		8866			
003	*Barnettozyma pratensis*	9055 CBS		9055			
004	*Buckleyzyma aurantiaca*	68-251 Phaff	Y-1581	317		32770	68-251
006	*Candida apicola*	1887 CBS		1887			
007	*Candida apicola*	1888 CBS		1888			
008	*Candida argentea*	3753 NCYC	Y-63798		3753		
009	*Candida argentea*	3784 NCYC			3784		
00A	*Candida asparagi*	9770 CBS		9770			
00E	*Candida boidinii*	3092 CBS		3092			
00F	*Candida boidinii*	8251 CBS		8251			
00G	*Candida boleticola*	7847 CBS		7847			
00I	*Candida californica*	06-229 Phaff					06-229
00J	*Candida californica*	04-1050 Phaff					04-1050
00M	*Candida davenportii*	3013 NCYC		9069	3013		
00N	*Candida digboiensis*	9800 CBS		9800			

00O	*Candida ecuadoriensis*	3782 NCYC	Y-63799	12653	3782		
00Q	*Candida etchellsii*	51-33 Phaff	Y-17084	1750		60119	51-33
00R	*Candida etchellsii*	60-8 Phaff		2907			60-8
00S	*Candida friedrichii*	4114 CBS		4114		22970	
00T	*Candida galis*	8842 CBS		8842			
00W	*Candida hawaiiana*	9146 CBS		9146			
00X	*Candida khmerensis*	9785 CBS		9785			
010	*Candida norvegica*	9474 CBS		9474			
013	*Candida qinlingensis*	9768 CBS		9768			
014	*Candida robusta*	Y-144 NRRL	Y-144			2373	
015	*Candida robusta*	365 NCYC			365	16664	
018	*Candida shehatae*	2389 NCYC	Y-12856	4705	2389	58779	12-114
01C	*Candida sorboxylosa*	Y-17669 NRRL	Y-17669	6378	2606	24120	12-150
01D	*Candida succiphila*	1403 NCYC	Y-11998	8003	1403	46049	12-108
01E	*Candida succiphila*	Y-17658 NRRL	Y-17658	7297			
01F	*Candida succiphila*	2625 NCYC	Y-17856	7920	2625		
01G	*Candida tenuis*	Y-17105 NRRL	Y-17105	4113		58781	
01H	*Candida tenuis*	57-18 Phaff	Y-1498	615		10573	57-18
01I	*Candida vanderwaltii*	8270 CBS		8270			
01J	*Candida zemplinina*	06-225 Phaff					06-225
01K	*Cyberlindnera jadinii*	Y-1542 NRRL	Y-1542	1600		18201	
01L	*Cyberlindnera saturnus*	8880 CBS		8880			
01O	*Debaryomyces fabryi*	Y-17914 NRRL	Y-17914	789		20278	
01P	*Debaryomyces hansenii*	74-86 Phaff	Y-7426	767		36239	74-86
01Q	*Debaryomyces hansenii*	475 NCYC	Y-1454	811	475		75-11
01R	*Debaryomyces hansenii*	2530 NCYC		8109	2530		
01S	*Hannaella luteola*	10491 CBS		10491			
01T	*Hanseniaspora uvarum*	04-162 Phaff					04-162
01U	*Hanseniaspora valbyensis*	17 NCYC	Y-1626	479	17	10631	
01V	*Issatchenkia orientalis*	55 NCYC			55		
01W	*Issatchenkia orientalis*	4001 NCYC			4001		
01X	*Kazachstania exigua*	1478 NCYC			1478		
01Y	*Kazachstania exigua*	Y-12640 NRRL	Y-12640	379		10599	55-83
01Z	*Kazachstania exigua*	80-20 Phaff					80-20
020	*Kazachstania unispora*	971 NCYC	Y-1556	398	971	10612	01-160
021	*Kazachstania unispora*	Y-1565 NRRL	Y-1565	399			
022	*Kluyveromyces aestuarii*	7776 CBS		7776			
023	*Komagataella pastoris*	9180 CBS		9180			
024	*Kuraishia hungarica*	9254 CBS		9254			
026	*Leucosporidium scottii*	Y-7185 NRRL	Y-7185			22182	
027	*Meyerozyma caribbica*	5289 CBS		5289			
028	*Meyerozyma guilliermondii*	5483 CBS		5483			

029	*Nadsonia starkeyi-henrici*	12-1050 Phaff	YB-3963	2159		24615	12-1050
02A	*Naumovia castellii*	3006 CBS		3006			
02E	*Pichia etchellsii*	740 NCYC	Y-7121	2011	740	20126	66-23
04N	*Pichia exigua*	Y-10920 NRRL	Y-10920				
02F	*Pichia fermentans*	850 NCYC	Y-1619	187	850	10651	
02G	*Pichia fermentans*	562 NCYC	Y-1879	603	562	9330	
02H	*Pichia fermentans*	Y-11508 NRRL	Y-11508	4807		28526	
02I	*Pichia fermentans*	Y-7181 NRRL	Y-7181	1876		24750	
L105	*Pichia fermentans*	L105					
02J	*Pichia heedii*	1489 NCYC	Y-10967	6930	1489	34936	76-356
02K	*Pichia heedii*	76-503 Phaff	Y-10970	6933		34939	76-503
02L	*Pichia kluyveri*	05-608 Phaff					05-608
02M	*Pichia kudriavzevii*	2658 NCYC	Y-7551	5147	2658	22692	66-21
02N	*Pichia kudriavzevii*	872 NCYC			872		
02O	*Pichia kudriavzevii*	2064 CBS		2064			
L394	*Pichia kudriavzevii*	L394					
L395	*Pichia kudriavzevii*	L395					
02P	*Pichia manshurica*	3374 NCYC			3374		
02Q	*Pichia manshurica*	Y-27978 NRRL	Y-27978	209			
02R	*Pichia membranifaciens*	Y-2026 NRRL	Y-2026	107		26288	57-22
02S	*Pichia membranifaciens*	714 NCYC	Y-6775	5567	714	58071	
02T	*Pichia membranifaciens*	Y-1575 NRRL	Y-1575	191			
02U	*Pichia membranifaciens*	7313 CBS		7313			
02V	*Pichia membranifaciens*	7314 CBS		7314			
04O	*Pichia nakasei*	Y-7686 NRRL	Y-7686				
04P	*Pichia norvegensis*	Y-7687 NRRL	Y-7687				
02W	*Pichia occidentalis*	75-63 Phaff	Y-7552	5459			75-63
02X	*Pichia occidentalis*	Y-7767 NRRL	Y-7767	1910		22686	75-57
04Q	*Pichia occidentalis*	YB-3389 NRRL	YB-3389				
04R	*Pichia occidentalis*	Y-6545 NRRL	Y-6545				
L396	*Pichia occidentalis*	L396					
02Y	*Rhodotorula glutinis*	68-262 Phaff		2203			68-262
02Z	*Rhodotorula mucilaginosa*	63 NCYC	Y-2510	316	63		68-312
030	*Rhodotorula toruloides*	5991 CBS		5991			
031	*Rhodotorula toruloides*	67-55 Phaff	Y-6987	6016			67-55
032	*Rhodotorula toruloides*	68-269 Phaff		2370			68-269
Cen	*S. cerevisae* CEN.PK 113-7D						
033	*Saccharomyces cerevisiae*	3040 NCYC			3040		
034	*Saccharomyces cerevisiae*	1592 CBS		1592			
035	*Saccharomyces cerevisiae*	8859 CBS		8859			
036	*Saccharomycodes ludwigii*	732 NCYC		1169	732		
037	*Saccharomycodes ludwigii*	734 NCYC			734		
038	*Saccharomycodes ludwigii*	731 NCYC	Y-12793	821	731	11313	

039	*Saccharomycodes ludwigii*	7780 CBS		7780			
03A	*Saccharomycopsis schoenii*	9156 CBS		9156			72-139
03B	*Scheffersomyces amazonensis*	12363 CBS		12363			
03C	*Scheffersomyces shehatae*	Y-12858 NRRL	Y-12858	5813		34887	
03D	*Scheffersomyces shehatae*	Y-17102 NRRL	Y-17102			22984	
03E	*Scheffersomyces stipitis*	79-261 Phaff	Y-7124	5773		58376	79-261
03F	*Scheffersomyces stipitis*	7507 CBS		7507			
03H	*Schizoblastosporion starkeyi-henricii*	7647 CBS		7647			
03I	*Schizosaccharomyces pombe*	380 NCYC		10392	380		
03J	*Schizosaccharomyces pombe*	936 NCYC		10394	936		
03K	*Schizosaccharomyces pombe*	683 NCYC		10393	683		
03L	*Schizosaccharomyces pombe*	04-213 Phaff	Y-12796	356			04-213
03M	*Schizosaccharomyces pombe*	3422 NCYC		5680	3422		
03N	*Schwanniomyces etchellsii*	Y-7546 NRRL	Y-7546	2012			
03O	*Spathaspora passalidarum*	Y-27907 NRRL	Y-27907	10155		MYA-4345	
03P	*Tetrapisispora arboricola*	Y-27308 NRRL	Y-27308	8765			
03Q	*Tetrapisispora blattae*	77-7 Phaff	Y-10934	6284		34711	77-7
03R	*Tetrapisispora fleetii*	Y-27350 NRRL	Y-27350	8957			
03S	*Torulaspora delbrueckii*	69-34 Phaff	Y-866	1146		10662	69-34
03T	*Torulaspora delbrueckii*	492 NCYC			492		
03U	*Ustilentyloma graminis*	502 NCYC	Y-2474	2826	502	32768	
019	*Wickerhamiella sorbophila*	Y-7921 NRRL	Y-7921	6739		60130	80-75
01A	*Wickerhamiella sorbophila*	130 NCYC	Y-27074	2280	130		
01B	*Wickerhamiella sorbophila*	82-593 Phaff					82-593
03V	*Wickerhamomyces anomalus*	432 NCYC	Y-366	5759	432	8168	76-71
03W	*Wickerhamomyces anomalus*	375 NCYC			375		
03Y	*Zygoascus meyerae*	Y-6591 NRRL	Y-6591	4099		15542	
03Z	*Zygoascus meyerae*	2544 NCYC		7115	2544		
040	*Zygoascus tannicolus*	75-58 Phaff	Y-7499	6067		22263	75-58
041	*Zygosaccharomyces bailii*	1416 NCYC	Y-2227	680	1416	58445	
042	*Zygosaccharomyces bailii*	1766 NCYC			1766		
043	*Zygosaccharomyces bailii*	1427 NCYC			1427		
044	*Zygosaccharomyces bailii*	417 NCYC			417		
045	*Zygosaccharomyces bailii*	464 NCYC			464		
046	*Zygosaccharomyces bailii*	1085 CBS		1085			
047	*Zygosaccharomyces bailii*	Y-1404 NRRL	Y-1404			11486	
048	*Zygosaccharomyces bailii*	2446 CBS		2446			
049	*Zygosaccharomyces bailii*	2902 CBS		2902			

04A	*Zygosaccharomyces bailii*	563 NCYC	Y-12949	685	563		
04B	*Zygosaccharomyces bailii*	7315 CBS		7315			
04C	*Zygosaccharomyces bailii*	7316 CBS		7316			
04D	*Zygosaccharomyces bisporus*	1515 NCYC			1515	38993	
04E	*Zygosaccharomyces bisporus*	1496 NCYC	Y-7684	1083	1496	52407	
04F	*Zygosaccharomyces bisporus*	1495 NCYC	Y-7558	702	1495	52405	66-24
04G	*Zygosaccharomyces bisporus*	1498 NCYC			1498		
04H	*Zygosaccharomyces bisporus*	1555 NCYC	Y-7253		1555		
04I	*Zygosaccharomyces bisporus*	2449 CBS		2449			
04J	*Zygosaccharomyces lentus*	1601 NCYC		8516	1601		
04K	*Zygosaccharomyces lentus*	Y-27276 NRRL	Y-27276	8574			
04L	*Zygosaccharomyces pseudobailii*	2856 CBS		2856		56074	
04M	*Zygosaccharomyces rouxii*	40-319 Phaff					40-319

3.2. Strains revival and storage

Strains were received from the repositories either as freeze-dried pellets or colonies on agar slants, in both cases they were inoculated into liquid yeast malt medium, containing 5 g/L of peptone, 3 g/L of yeast extract, 3 g/L of malt extract and 10 g/L of dextrose. After inoculation, strains were grown at 30°C until fully saturated, allowing for up to a week at which point they were frozen in 15% glycerol stocks in triplicate and stored at -80°C.

3.3. Screening strains for tolerance to low pH and organic acids

To screen for tolerance to low pH and organic acids, frozen strains were revived in batches of 30 in YPD medium (20 g/L of dextrose, 20 g/L of peptone and 10 g/L of yeast extract), in test tubes at 30°C. Once all strains were visibly turbid, they were back diluted (1:20) into 180 µL of YPD in a shallow 96-well plate and left to grow for 8 hours, or until all strains reached an OD600 of 1.8. After the second incubation the OD600 was read on a plate reader and a custom Python script was used to create a CSV file with the absolute volumes of

diluent and cell culture necessary to dilute each culture to an OD600 of 1.8. Dilutions were then carried out using the Span-8 of a Biomek FXP liquid handling system to create an intermediate plate with 30 strains in triplicate plus *S. cerevisiae* CEN.PK 113-7D as a control, all at consistent OD600 of 1.8, plus media blanks. Both CEN.PK 113-7D and the media blanks were also in triplicate. From the intermediate plates, a 10 µL inoculum was drawn and dispensed into a prepared experimental plate containing 170 µL of the relevant growth medium. The final volume of each well was 180 µL and the starting OD600 of all testing strains + CEN.PK 113-7D was 0.1.

Initial tests for organic acid tolerance were conducted of YPD based media, YPDcit was standard YPD with 0.1 21.0 g/L, of citric acid added to decrease the pH to 3.0, and YPDAA was standard YPD but with 20 g/L of adipic acid added, resulting in a pH of 3.7.

Organic acid tests conducted in minimal media used a common base of 20 g/L of glucose, 5.1 g/L of YNB without amino acids and ammonium sulphate, 5.0 g/L of ammonium sulphate and 150 mM of the organic acid being tested, with HCl added dropwise to bring the media's pH to 2.8.

After inoculation, plates were sealed with parafilm and incubated at 30°C in Tecan Sunrise plate readers for 48 hours. Plates were orbitally shaken for 15 minutes, allowed to rest for 5 minutes before OD600 readings were taken and shaking was restarted.

3.4. Growth rate determination

For the initial set of strains, a simple area under the curve measurement was used to assess the rate of growth, after blanking, the sum of the first 24 hours of OD readings was taken and divided by the total number of reads per hour (3), providing a simple means to compare yeast growth. As many strains were unable to grow, strategies for assessing growth

rate which are dependent on growth curves following a diauxic growth model are unsuitable; a simple area under the curve is more robust.

To compare the relative growth rates of strains in the presence of organic acid inhibitors, growth curve data was exported from the plate readers and processed with a custom Python script to cycle through the OD readings and determine the rate of OD increase across each one-hour period using the formula below, where OD_1 and T_1 refer to the OD reading and read time of the later reading and OD_0, T_0 refer to the OD reading and read time of the earlier reading.

$$\frac{\log_2(OD_1/OD_0)}{T_1-T_0}$$

While the formula can calculate growth between any two arbitrary timepoints, only timepoints 1 hour apart were compared to minimize the impact of artifacts. Likewise, the first 2 hours were not used as the OD readings were often highly volatile immediately post-inoculation, and most strains would be in their lag phase in this period. The result of this analysis was a table of growth rates at each timepoint; to simplify data presentation, the maximum growth rate was used for comparison between strains and growth conditions.

3.5. Genome sequencing

A collection of 17 acid-tolerant strains were selected for sequencing using PacBio SMRT 2.0 technology. Genomic DNA was extracted using Qiagen Genomic DNA buffer set and the Qiagen 20/G Genomic tips. After extraction DNA quality was assessed using both a Tecan M200 infinite plate reader, equipped with a nanoquant adaptor and using the Quant-iT Picogreen dsDNA quantification kit. The Picogreen assay was carried out in accordance with the standard protocol provided by Thermofisher. DNA extracts which registered a DNA

concentration of 200 ng/μL or higher when measured with both techniques and had a 260:280 ratio of 1.8-2.0 were delivered to Génome Québec for PacBio SMRT 2.0 sequencing and DNA sequences were assembled using the PacBio SMRTpipe by Génome Québec. For additional quality control purposes Génome Québec repeated the concentration and purity assessments.

3.6. Genome annotation

To meaningfully evaluate commonalities and differences with respect to species and correlate them to organic acid tolerance, the genomes require assembly and annotation, consisting of gene finding and gene function prediction. The assemblies were provided by Génome Québec as produced by the SMRTpipe assembly pipeline recommended by Pacific Biosciences. Augustus (Stanke et al., 2004) was used to find putative genes in select assemblies using the *S. cerevisiae* gene model packaged with Omicsbox, without hinting and after masking repeated sequences with a hidden Markov model trained off the Dfam database (Hubley et al., 2016). After finding putative ORFs, four different methods were used to create functional annotations, InterProScan 5.52-86.0 (Jones et al., 2014), Blast2GO (Conesa et al., 2005) using non-redundant protein sequences database limited to the Saccharomycotina clade, EggNOG mapper (Huerta-Cepas et al., 2017) and searching the custom AyBraH database for hits using TBLASTN.

3.7. Gene clustering & phylogenetic analysis

Genes were assigned into clusters using the AyBraH database categories, homologous ortholog groups (HOGS) and the more specific fungal ortholog groups (FOGS) (Correia et al., 2019). After assigning genes to HOGS and FOGS, the genes were translated to amino acid sequences and a multiple sequence alignment was generated using Mega-X (Kumar et al., 2018). In addition to the novel genomes sequenced, other Pichiacea species represented in

the AyBraH database along with *Y. lipolytica* were included in the alignments. From the multiple sequence alignments, phylogenetic tree generated with IQ-TREE2 (Minh et al., 2020). Select phylogenetic trees were uploaded to iTOL for visualization and analysis (Letunic & Bork, 2021).

4. Results

4.1. Strain selection

Strains were selected from a range of depositors on the basis of species diversity, reported acid tolerance, and with a preference for type strains to increase repeatability. In total 177 strains were considered from 105 species, of which 153 strains from 83 species were finally ordered. From the 153 strains ordered 122 were successfully resuscitated in YM medium at 30°C and frozen into cryogenic stocks. The final 122 strains represented 68 species from 29 different genera. A complete list of the 122 strains is available in Table 2.

4.2. Initial screening

The first screening experiment conducted in both YPD and YPD with 0.1 M citric acid, the resultant pH in the acidic medium was 3.0. YPD was selected over a defined medium as defined media are typically created to fulfil the nutritional requirements of a specific species and we did not want to bias the results towards strains or species with differing nutritional requirements. After screening 93 strains, the citric acid screen was stopped, as while there were some strains that were unable to grow in YPD_{cit} most strains were able to grow completely uninhibited (Figure 1). This result was unexpected, and likely a result of two features: A) citric acid being most toxic at a neutral pH when it is in its fully protonated form where it can chelate calcium ions in the medium (Nielsen & Arneborg,

2007), and B) strains that had previous literature evidence of acid tolerance were selected, and the pH 3.0 condition was not sufficiently inhibitory.

After those initial results, strains which were unable to grow effectively in the comparatively moderate pH 3.0 conditions were dropped and the selection was switched to 20 g/L of adipic acid. The remaining 29 strains that had not been tested in citric acid were only tested in YPD or YPD + 20 g/L of adipic acid (YPD_{AA}). The addition of adipic acid to the medium dropped the pH from near neutral to approximately 3.7. As the pKa values of adipic acid are 4.43 and 5.41, the drop in pH was sufficient to ensure that adipic acid instead of the conjugate base, adipate, would be present in the media. Despite having a higher pH than the initial citric acid screen, the correlation between growth in YPD and YPD_{AA} media was lower, $R^2 = 0.685$, vs $R^2 = 0.905$ for YPD_{cit} (Figure 2), indicating that adipic acid did influence the growth rate of several strains. This result enabled further testing to be done on a smaller subsample of strains.

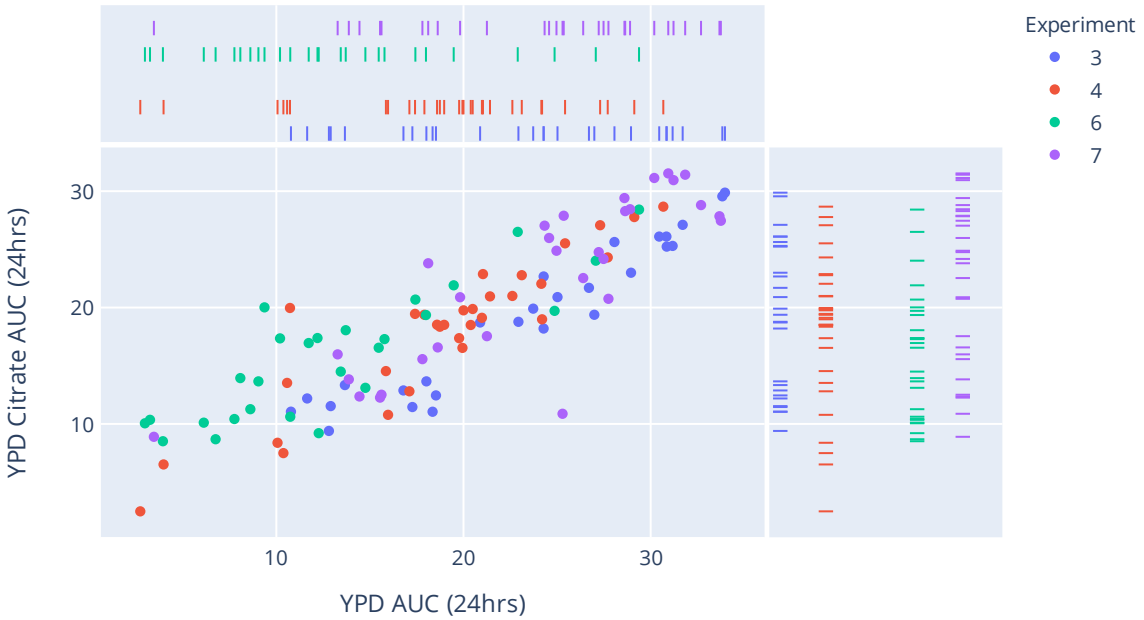

Figure 1. Results of the initial organic acid tolerance screen using YPD and YPD with citric acid. Growth is measured using the area under the curve (AUC) in each medium. Colours are used to group strains run simultaneously. Rug plots on the X and Y axes show the distribution of each variable independently. Each data point represents the average of a biological triplicate, strains on average had a standard deviation of 2.2, indicating that the measures were consistent.

Despite being grown in a rich media some strains were still not able to grow to an OD600 greater than 1.0 even without the presence of inhibitors; the exact reasons are unknown and likely vary by strain but susceptibility to the environmental factors present in microtiter plates such as a low dissolved oxygen, lack of an essential nutrient, rapid loss of viability after exponential growth or specific temperature requirements may have contributed, in addition some strains may have slower growth than expected and would have reached an OD600 greater than one given sufficient time.

To select strains from the adipic acid screen to continue working with, two primary criteria were used: tolerance to adipic acid (measured as relative area under the curve in YPD_{AA} when compared to YPD), and rapid growth in YPD (measured as relative area under the curve of each strain compared to the control strain CEN.PK 113-7D).

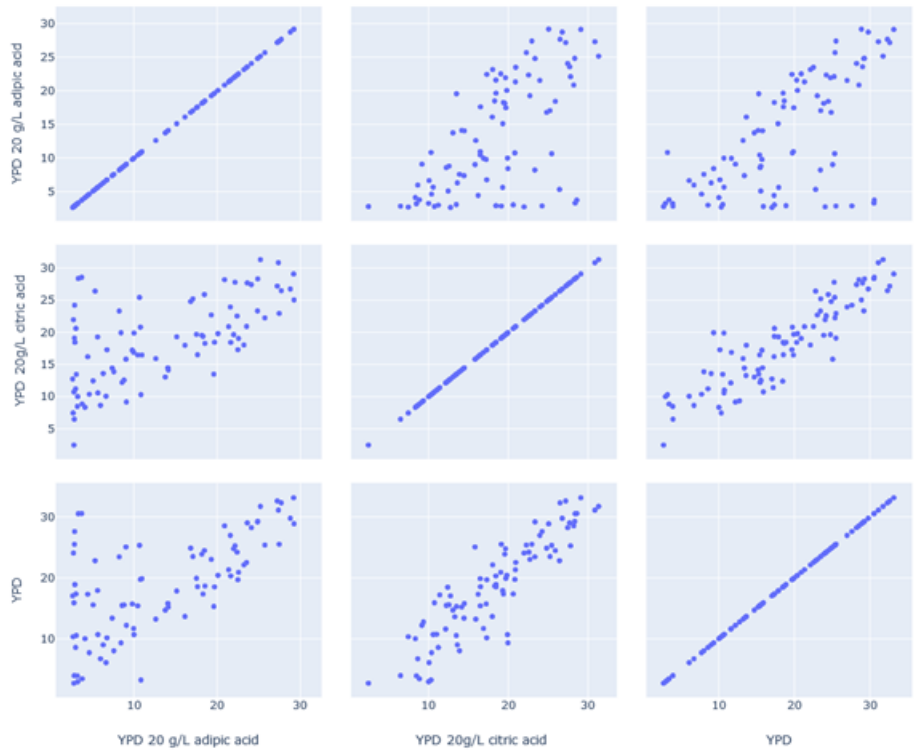

Figure 2. Scatter plots of yeast growth in different media relative to one another, the mean of 3 biological replicates is displayed for each condition, with the growth being assessed by area under the curve. CEN.PK113-7D had an AUC of 23.6 in YPD, 25.3 in YPD_{cit} and 17.1 in YPD_{AA}, averaged across 3 replicates.

The threshold for each criterion was set to 80%, therefore, to reach the next round of screening each strain can grow no more than 20% slower in YPD_{AA} than in YPD, and no more than 20% slower in YPD than CEN.PK 113-7D. Together the criteria ensure selected strains proliferate rapidly, making them easy to work with in the lab, and are able to thrive in the acidic conditions that are conducive to economic organic acid production.

Applying these criteria produces a list of 15 strains (Table 3) down from the original 122. When examining the strains that passed both criteria, there is a clear over-representation of *Pichia* strains relative to the initial collection. Of the fifteen, seven are in the *Pichia* genus, making up 46.7% of the total versus only 10% of the whole strain collection. Not all *Pichia*

species in the initial dataset passed both criteria however, including all 3 *P. membranifaciens* strains (the type species of the *Pichia* genus), and 3 of 4 *Pichia fermentans* strains.

Table 3. List of non-conventional strains determined to be robust potential hosts.

Strain ID	Depositor ID	Species	YPD AUC	YPD_{AA} AUC
00E	3092 CBS	*Candida boidinii*	21.6	19.7
019	Y-7921 NRRL	*Candida sorbophila*	28.5	24.7
01L	8880 CBS	*Cyberlindnera saturnus*	18.7	18.7
01R	2530 NCYC	*Debaryomyces hansenii*	19.9	20.9
01V	55 NCYC	*Pichia kudriavzevii*	19.5	20.5
01W	4001 NCYC	*Pichia kudriavzevii*	30.5	26.5
01X	1478 NCYC	*Kazachstania exigua*	25.6	21.4
020	971 NCYC	*Kazachstania unispora*	17.8	17.4
02G	562 NCYC	*Pichia fermentans*	20.4	16.6
02M	2658 NCYC	*Pichia kudriavzevii*	26.6	22.2
02N	872 NCYC	*Pichia kudriavzevii*	18.2	19.9
02Q	Y-27978 NRRL	*Pichia manshurica*	17.3	14.8
02W	75-63 Phaff	*Pichia occidentalis*	29.7	25.0
03P	Y-27308 NRRL	*Tetrapisispora arboricola*	22.6	19.5
03V	432 NCYC	*Wickerhamomyces anomalus*	27.1	26.1

This overrepresentation of *Pichia* strains presents a useful opportunity to dig deeper into the background and metabolism of these species, to potentially understand the mechanisms underlying tolerance but also to find strains that are better suited to growing in cheap media. Additionally, further probing of this genus may reveal even more tolerant species.

4.3. The *Pichia* clade

The *Pichia* clade was a large and polyphyletic genus prior to 2011 where a taxonomic reorganization reduced the size of the genus down to a monophyletic group of 30 species within the order Pichiaceae. *Pichia* strains have been isolated from a diverse range of habitats including sugarcane (Dhaliwal et al., 2011), sea mud (Qu et al., 2012), spoiled wines (Saez et al., 2011), soil (Labbani et al., 2015) and fish guts (Mandal & Ghosh, 2013). As the *Pichia*

strains already in our collection appeared to have a higher frequency of acid tolerance, we expanded our collection to include 8 more strains, three of those strains were from species previous not represented in the initial screen: *Candida sorboxylosa*, *Pichia norvegensis* and *Pichia exigua*. *C. sorboxylosa* despite its name has been placed in the *Pichia* genus by whole genome phylogenetic analysis (Shen et al., 2018).

Unfortunately, despite a significant research effort there is very little sharing of strains between research groups, often groups isolate their own strain and have a specific phenotype they test for. As a result, it is unclear which phenotypes are specific to strains and which are common to all strains within a species or common to all species within the genus. Despite that, in the literature, many of the isolates appear to grow in harsh acidic environments (Ji et al., 2020; Park et al., 2018).

To characterize our strains systematically, an experiment was conducted to evaluate the ability of each *Pichia* strain to grow in 8 different organic acids of varying chain lengths and with either 1 or 2 carboxyl groups, at a consistent pH and a consistent concentration. The acids tested were acetic acid, propionic acid, butyric acid, valeric acid, succinic acid, glutaric acid, and adipic acid, each acid was tested at a concentration of 150 mM, and a pH of 2.8.

In addition to using a common pH and concentration across acids we sought to evaluate the ability of strains to grow on defined media. An initial test revealed that strains had significant variance in their ability to grow in YNB with standard composition, some strains were unable to reach an OD600 of 1 in YNB, compared to 1.75 in YPD medium. Existing literature indicated that myo-inositol was a limiting growth factor in some strains, but supplementation had no effect in our case [data not shown]. An increase in the amount of YNB from 1.7 g/L (YNB_{reg}) to 5.1 g/L (YNB_{del}) did appear to alleviate the limitation and allowed the affected strains to grow up to a maximum OD of 1.4 (Figure 3). Which

component or components in the YNB were limiting is unknown, YNB is composed of a mixture of 20 minerals and salts, any could have been limiting, glucose and ammonium sulphate were added separately at 20 g/L and 5 g/L respectively. All future experiments in minimal media utilized "YNBdel" as a growth medium to ensure results are reflective of each strains potential. The individual component or components causing the growth malus could have been determined through modifying the concentrations of each YNB component one at a time however for the purposes of assessing organic acid tolerance, however establishing nutritional requirements wasn't a priority in this work.

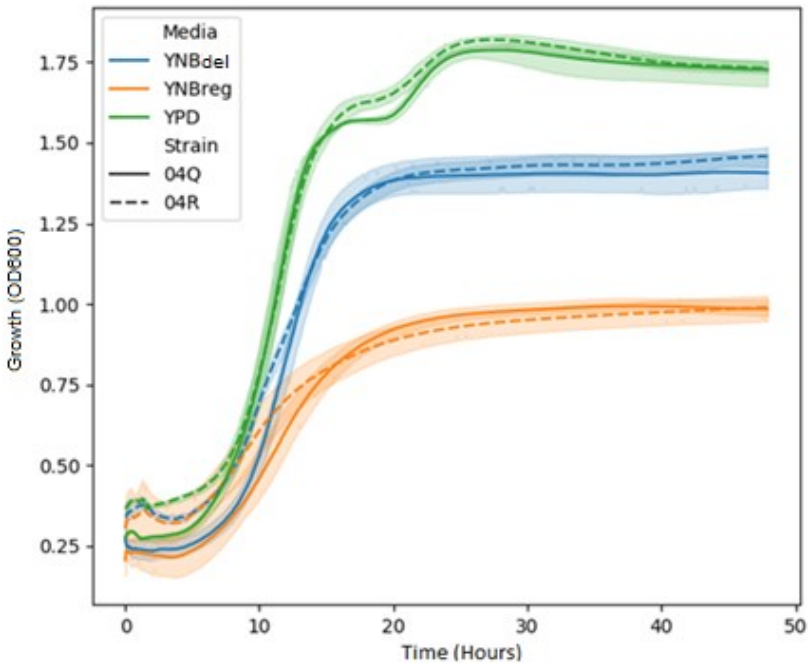

Figure 3. Comparison of growth in standard YNB medium vs growth in YNB_{del}. Highlighted area shows the standard deviation of 3 replicates. Composition of YNB_{reg}: 20 g/L glucose, 1.7 g/L YNB; Composition of YNB_{del}: 20g/L glucose, 5.1 g/L YNB.

Among dicarboxylic acids, the reference strain, *S. cerevisiae* CEN.PK 113-7D saw a steady decrease in growth rate as the chain length of the organic acid increased, this trend was also present in most of the *Pichia* strains, however in *P. occidentalis*, the growth rate trended up with chain length, and in *P. manshurica* there was no consistent trend (Figure 4). The trend for organic acids to become more toxic with increased chain lengths is consistent

with the standard theory of organic acid tolerance, in which acids with longer chain lengths are more lipophilic. In effect this means they can diffuse across the plasma membrane more easily and thus, per molar unit will both acidify the cytoplasm more rapidly and need to be expelled from the cell by transporters at an increased rate (Warth, 1988).

Similar to in the initial screen, several of the *Pichia* strains including *P. heedi*, *P. exigua*, and *P. nakasei* were unable to consistently grow in the control condition (YPD or YNB) (Figure 4). These strains were documented in the literature as being able to grow in YNB, and at 30°C (Kurtzman et al., 2011) however in those experiments the strains were allowed to grow for a week or longer, whereas in our experiment they were given only one day to grow, possibly explaining the disparity in results. *Candida sorboxylosa* appeared to grow more rapidly in acid conditions than in regular YNB (Figure 4), however physical examination of the cultures revealed an extensive pellicle in acidic conditions which interferes with OD600 values and makes the derived growth rates impossible to compare between conditions. Surprisingly, none of the tested strains were able to grow in acetic, propionic, butyric or valeric acids, indicating that at equimolar concentrations they are significantly more toxic than dicarboxylic acids. As acetic acid has a similar logP in octanol/water to adipic acid it was expected to have a similar level of toxicity; it is possible that acetic acid has a lower activation energy for crossing the lipid bilayer due to its lower molecular mass, and this results in the increased toxicity, a molecular dynamics simulation could inform further why this was the case.

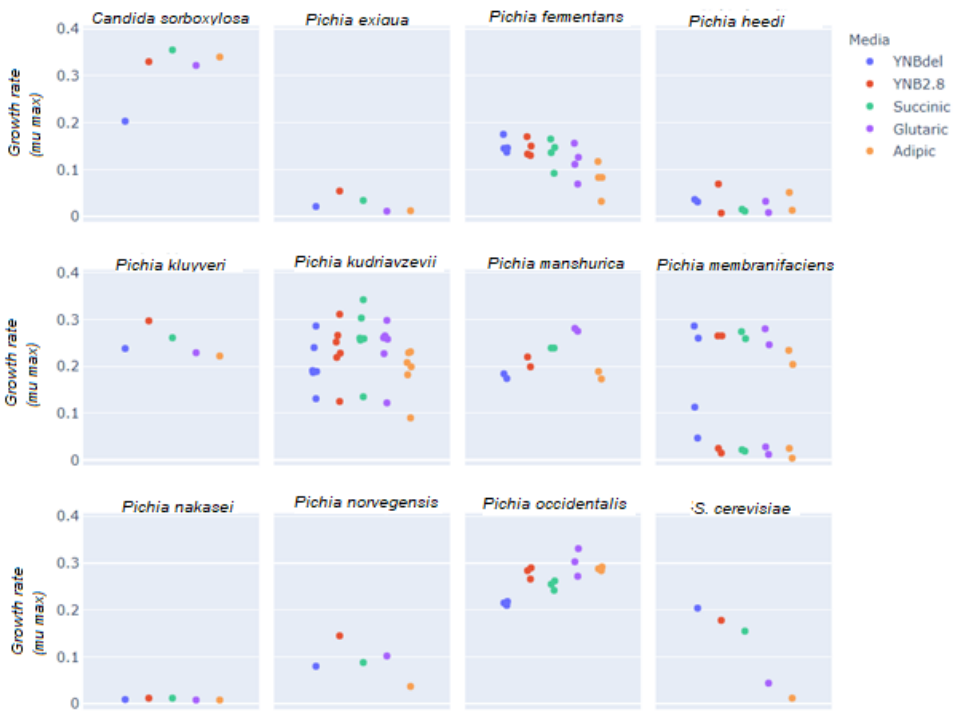

Figure 4. Comparison of *Pichia spp.* growth rates in different organic acids. Each point represents the growth rate of one strain in a specific condition reported as the mean of 3 replicates. All organic acid containing media were pH adjusted to 2.8 after the addition of 150 mM of the respective organic acid, YNB 2.8 is standard YNB based media but with the pH adjusted to 2.8 by addition of HCl.

4.4. Genome sequence and annotation

Given the results of the initial screen and follow up on the *Pichia* clade a hybrid approach was taken for selecting strains to sequence. The top performing strains from the initial screen, except for those which displayed extreme pellicle formation or flocculation, were all sequenced, and additional strains from the *Pichia* clade were sequenced to provide a more complete picture of strain and gene phylogeny.

The strains were sequenced using SMRT 2.0 to create assemblies that are as close to chromosome level as possible (Table 4), the strategy produced good results with the assemblies for most strains having under 20 contigs. Two of the strains, *K. exigua* – 01X and *P. membranifaciens* – 01V appear to be heterozygous diploids, resulting in assemblies with approximately double the expected base pairs and contigs. *Pichia kudriavzevii* – 02N also had

a larger than expected genome size but while the ancestry of 01X and 01V were easily resolved, the assembly for *P. kudriavzevii* – 02N had regions with high similarity to both *P. kudriavzevii* but also some smaller contigs with high similarity to *Debaryomyces hansenii*, examining the read coverage of the contigs showed 90% of contigs had a read coverage between 85X and 95X, of the remaining 10%, 9.3% were between 62.9 and 80.8%. The range of read coverages indicates it is possible that the sample had been contaminated, however not all contigs with high *D. hansenii* similarity had a low read coverage.

Table 4. Summary of strains and corresponding assemblies. Genomic DNA was extracted using Qiagen genomic DNA 20/G tips and sequenced on a Pacbio SMRT 2.0 sequencing platform by Génome Québec. Sequence data was assembled by Génome Québec using SMRTpipe.

Species	**ID**	**Genome size (Mbp)**	**Polished assembly Contigs**	**Max contig size (Mbp)**
Pichia occidentalis	02W	12.6	27	4.3
Pichia occidentalis	04R	11.9	12	2.8
Pichia occidentalis	04Q	11.7	12	2.3
Pichia occidentalis	L396	14.8	20	2.5
Pichia kudriavzevii	02N	22.1	218	2.4
Pichia kudriavzevii	02M	11	10	2.7
Pichia fermentans	02G	13.7	62	1.7
Pichia kluyveri	02L	11.3	20	3
Pichia membranifaciens	01V	19.4	73	1.9
Pichia manshurica	02Q	12	6	3.4
Wickerhamiella sorbophila	019	8.3	10	2.5
S. cerevisiae	033	12	18	1.5
S. cerevisiae	014	11.8	18	1.5
Kazachstania exigua	01X	26	39	2
Scheffersomyces amazonensis	03B	13.9	27	3.4
Wickerhamomyces anomalus	03V	15.1	30	1.8
Candida boleticola	00G	16.1	25	3

Once the genomes were assembled, putative ORFs were annotated using Augustus and repeat masking, as described in the methods. As none of the species had pre-trained gene models, two gene models were used for each species, and if the two models were in agreement (difference in total ORF count of <5%), the gene model from the best documented species was used. If the gene models were not in agreement, then the strain was not annotated further. For CUG-ser yeasts, the two gene models used were from *C. albicans* and *S. stipitis* were used, for non CUG-ser yeasts, *S. cerevisiae* and *K. phaffii* gene models were used.

After predicting ORFs, putative gene ontologies were assigned using Blast2GO, EggNOG and InterProScan (Conesa et al., 2005; Huerta-Cepas et al., 2017; Jones et al., 2014) (Table 5). Predicted ORFs were also aligned to the pan-yeast proteome using a local BLAST database comprised of annotated genes from 33 fungal species (Correia et al., 2019), genes were assigned into the ortholog groups of their closest match to enable manual annotation via examination of phylogenetic trees of orthologous groups.

Table 5. Results of annotating yeast species. GO mapping was accomplished using a combination of Blast2GO, EggNOG and InterProScan, annotations from each program were compiled to predict putative function based on homology.

Species	ID	Predicted ORFs	GO Mapped ORFs
Pichia occidentalis	02W	5016	4105
Pichia occidentalis	04R	5357	4793
Pichia occidentalis	04Q	5299	4743
Pichia kudriavzevii	02N	8894	7942
Pichia kudriavzevii	02M	5227	4716
Pichia fermentans	02G	6323	5288
Pichia kluyveri	02L	5219	4119
Pichia *membranifaciens*	01V	8826	7965
Pichia manshurica	02Q	5351	4758
Wickerhamiella sorbophila	019	4479	4063
S. cerevisiae	033	5643	5547
S. cerevisiae	014	5473	5391
Kazachstania exigua	01X	10298	9894

To evaluate relatedness of the sequenced strains, a set of genes present in all strains were concatenated and then aligned. From this alignment a phylogenetic tree was generated (Figure 5). An interesting feature of the phylogenetic tree is that the placement of species does not seem to correlate strongly to observed acid tolerance, with *P. fermentans*, one of the least acid tolerant strains of the collection being clustered with *P. occidentalis* and *P. kluyveri* strains, which are among the most acid tolerant strains. This suggests that the acid tolerance phenotype may be older than the *Pichia* genus, rather than a recent evolution. The closest related genus to the *Pichia* is the Saturnispora, which have not subject of much published research, however strains of the second closest related genus, Brettanomyces, have been observed growing at pH values as low as 2.0 (Conterno et al., 2006). *Candida boidinii* strains, also in the Pichiaceae clade but more distantly related, have also been recorded as able to grow in harsh acidic conditions (Osawa et al., 2009). As research on the Pichiaceae clade at large is still in its infancy it's difficult to assess how prolific tolerance to organic acids is in this clade.

To date, the only gene from *P. kudriavzevii* or the *Pichia* clade proven to positively influence acid tolerance when expressed in *S. cerevisiae* is the cell wall maintenance protein GAS1, and the mechanism by which it does so is unknown (Matsushika et al., 2016). To identify putative causes of acid tolerance among these species, phylogenetic trees of genes families that fall into a few categories that may be involved in acid tolerance were examined. Gene families were determined based on TBlastN results of each gene against the AyBraH database (Correia et al., 2019). The categories of interest were based off known mechanisms of acid tolerance in other species and include: ABC transporters (Balzi & Goffeau, 1995; Piper et al., 1998), cell wall maintenance proteins (Matsushika et al., 2016), decarboxylases (Cibrario et al., 2020), MFS transporters (Pereira et al., 2019), fatty acid desaturases (Besada-Lombana et al., 2017) and proton pumps.

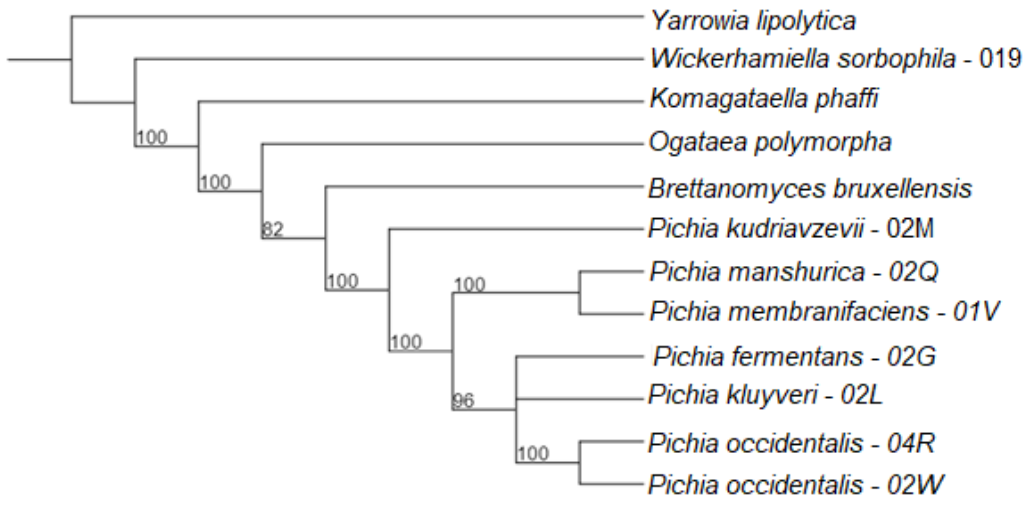

Figure 5. Phylogenetic tree showing relatedness of different *Pichia spp.* select species from AyBraH as a reference. Phylogenetic tree based on concatenated proteins sequences of: NUO.02 (FOG00685), NUO.06 (FOG00690), NUO.20 (FOG00709), CYS2 (FOG01372) and NUO assembly protein (FOG04021). The alignment and phylogenetic tree were generated using MEGA X.

In addition to the sequenced *Pichia* spp., *Komagataella pastoris*, *Brettanomyces bruxellensis*, *Ogataea polymorpha* and *Y. lipolytica* from the AyBraH database were included in the analysis to provide reference. Protein families were aligned with MUSCLE (Edgar, 2004) using the neighbour joining algorithm, and maximum likelihood trees were built using IQ-tree 2 (Minh et al., 2020) with 1000 ultrafast bootstrap replicates and modelfinder (Kalyaanamoorthy et al., 2017) to determine an appropriate substitution model. Predicted genes with internal stop codons were dropped, as were genes that differed from the mean amino acid length by more than 25%.

Pdr12p in *S. cerevisiae* is the principal weak acid pump and has been shown to be induced in *S. cerevisiae* in the presence of adipic acid, and a knockout has increased susceptibility to adipic acid toxicity (Fletcher et al., 2021). The gene is induced by War1p in the presence of organic acids, especially longer chain organic acids (Kren et al., 2003). Examining Pdr12 protein family phylogenetic tree 2provides an interesting and unusual

result: while the family has homologs present in *K. pastoris* and *Y. lipolytica*, there are no hits in AyBraH for Pdr12 in *O. polymorpha* or *B. bruxellensis*. Additionally, there are several orthologs in *Pichia* glade species (Figure 6). *P. occidentalis*, *P. manshurica* and *P. kluyveri* each have 3 copies, *P. fermentans*, *P. membranifaciens* each have 2 (after considering 01V's putative heterozygous diploidy) and *P. kudriavzevii* has one. The Pdr12 family in the *Pichia* clade is also subdivided into two clades, one clade is present only in *P. occidentalis*, *P. manshurica* and *P. kluyveri*, whereas the other, usually in two adjacent copies on the chromosome, is present in most *Pichia* strains. In each strain with a duplicate copy, the duplicate copy is directly adjacent to the original, and the closest relative is the alternate copy, suggesting that post-duplication there have been intramolecular recombination events occurring in these strains, as it is unlikely that each strain experienced a recent gene duplication event independently. As the upper, smaller clade of Pdr12 family proteins is present only in *P. kluyveri*, *P. manshurica* and *P. occidentalis* and those three strains are the most tolerant to the longer organic acids to which Pdr12 is specific, there is a strong possibility that this subclade of transporters is involved as additional copies of *Pdr12* in *S. cerevisiae* have shown to increase tolerance to adipic acid (Pereira et al., 2019).

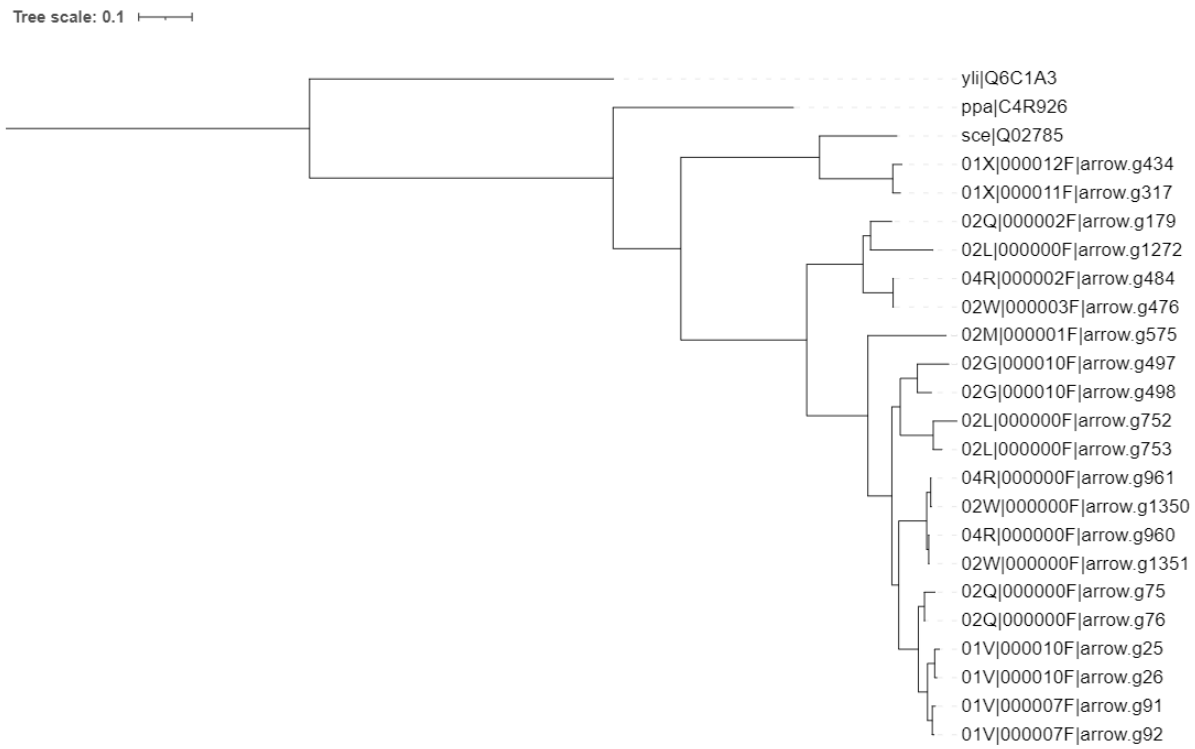

Figure 6. PDR12p phylogenetic tree. Alignment generated with MEGA X, tree generated with IQ-TREE 2, visualization generated with EMBL iTOL. 01X refers to *K. exigua* 1478 NCYC, 01V refers to *P. membranifaciens* 55 NCYC, 02Q refers to *P. manshurica* Y-27978 USDA, 02M refers to *P. kudriavzevii* 2658 NCYC, 04R refers to *P. occidentalis* Y-6545 USDA, 02W refers to *P. occidentalis* 75-63 Phaff, 02G refers to *P. fermentans* 562 NCYC, 02L refers to *P. kluyveri* 05-608 Phaff. yli, ppa, sce, opm and dbx refer to *Y. lipolytica*, *K. phaffi*, *S. cerevisiae*, *O. polymorpha* and *B. bruxellensis* sequences from the Aybrah database.

The Agp2 protein family has also undergone expansion in the *Pichia* clade, while *S. cerevisiae* has only one copy, there are three in most of the *Pichia* strains (Figure 7). The biological function of Agp2 in *S. cerevisiae* is to act as a signal transducer to control expression of other amino acid and polyamine transporters (Aouida et al., 2013). It's possible that this function is conserved in the *Pichia* strains, however given the large evolutionary distance between *S. cerevisiae* and the *Pichia* clade it is difficult to say. What is interesting however is the degree to which the copy numbers are stably preserved in the lower cluster of Agp2 family proteins; given the stable inheritance it's not unlikely that they have diverged, and each have a distinct function.

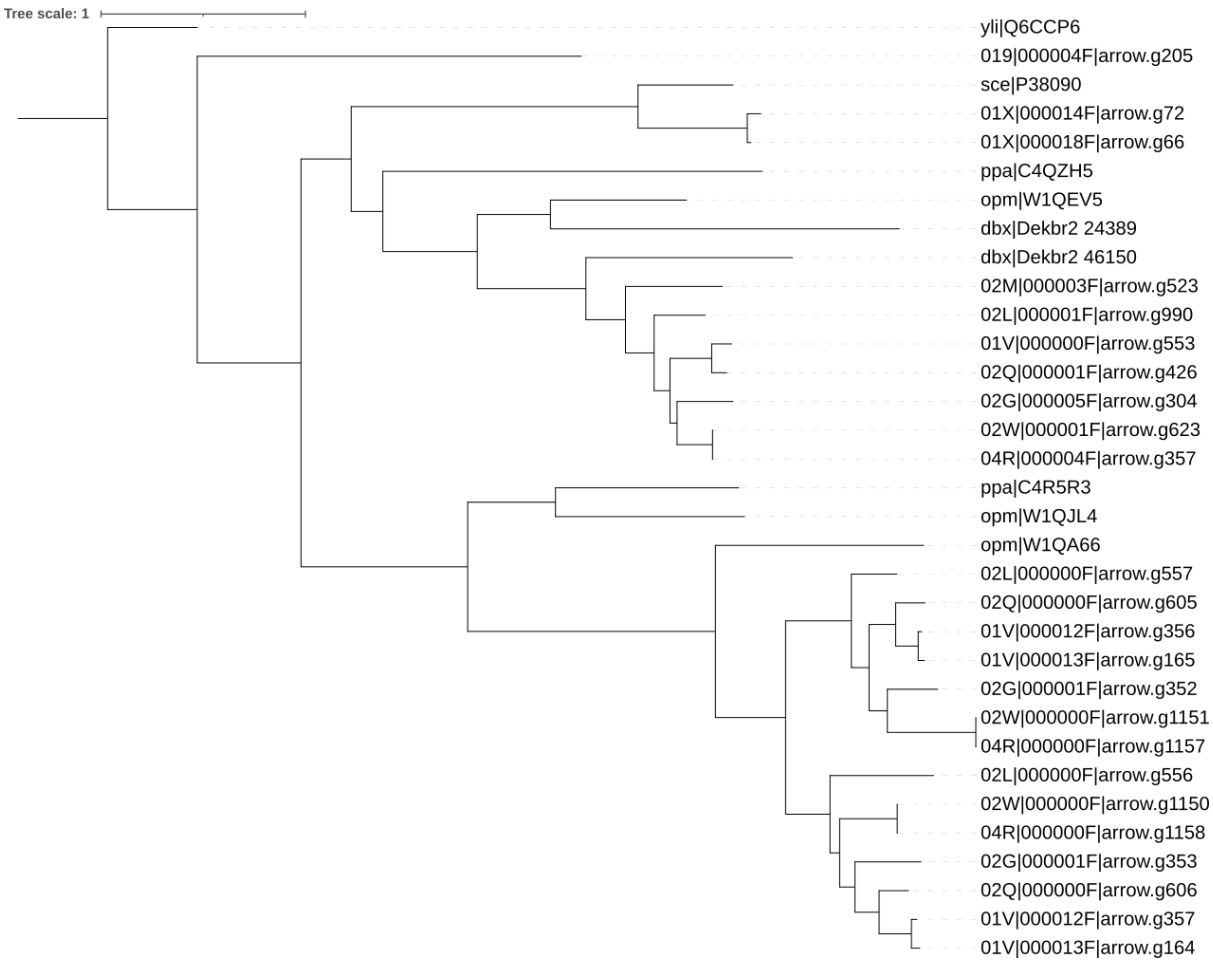

Figure 7. AGP2p phylogenetic tree. Alignment generated with MEGA X, tree generated with IQ-TREE 2, and visualization generated with EMBL iTOL. 01X refers to *K. exigua* 1478 NCYC, 01V refers to *P. membranifaciens* 55 NCYC, 02Q refers to *P. manshurica* Y-27978 USDA, 02M refers to *P. kudriavzevii* 2658 NCYC, 04R refers to *P. occidentalis* Y-6545 USDA, 02W refers to *P. occidentalis* 75-63 Phaff, 02G refers to *P. fermentans* 562 NCYC, 02L refers to *P. kluyveri* 05-608 Phaff. yli, ppa, sce, opm and dbx refer to *Y. lipolytica*, *K. phaffi*, *S. cerevisiae*, *O. polymorpha* and *B. bruxellensis* sequences from the Aybrah database.

PMA1 has had a duplication event in *P. occidentalis*, with one of the copies undergoing rapid divergence, as illustrated by branch lengths, there has also been a similar event in *K. exigua*, with the strain having 4 copies, two of which being significantly diverged (Figure 8). *P. manshurica* has also seen a duplication but there are no ancestral or genus wide events. Lacking evidence to the contrary, a common mechanism for the whole genus is

expected, it seems unlikely that this *PMA1* duplication in *P. manshurica* and *P. occidentalis* is a primary driver of their acid tolerance, though it may boost the tolerance of these strains in particular.

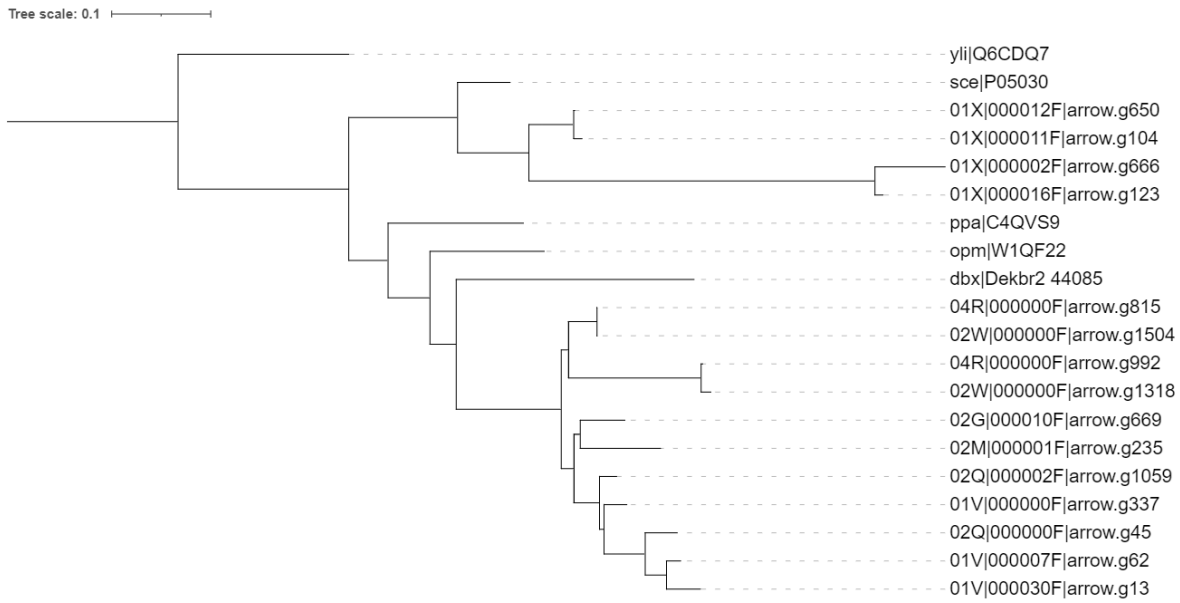

Figure 8. PMA1p phylogenetic tree. Alignment generated with MEGA X, tree generated with IQ-TREE 2, and visualization generated with EMBL iTOL. 01X refers to *K. exigua* 1478 NCYC, 01V refers to *P. membranifaciens* 55 NCYC, 02Q refers to *P. manshurica* Y-27978 USDA, 02M refers to *P. kudriavzevii* 2658 NCYC, 04R refers to *P. occidentalis* Y-6545 USDA, 02W refers to *P. occidentalis* 75-63 Phaff, 02G refers to *P. fermentans* 562 NCYC, 02L refers to *P. kluyveri* 05-608 Phaff. yli, ppa, sce, opm and dbx refer to *Y. lipolytica*, *K. phaffi*, *S. cerevisiae*, *O. polymorpha* and *B. bruxellensis* sequences from the Aybrah database.

Although Qdr3 is a transporter that has previously been shown to increase weak acid tolerance in *S. cerevisiae* (Pereira et al., 2019), there were no members of the family present in the *K. exigua* strain, and no duplications present in the *Pichia* clade strains (Figure 9). Qdr3 is unlikely to be a major mediator of acid tolerance in the *Pichia* clade given that, and the fact that it is not induced by the presence of organic acids in *S. cerevisiae*.

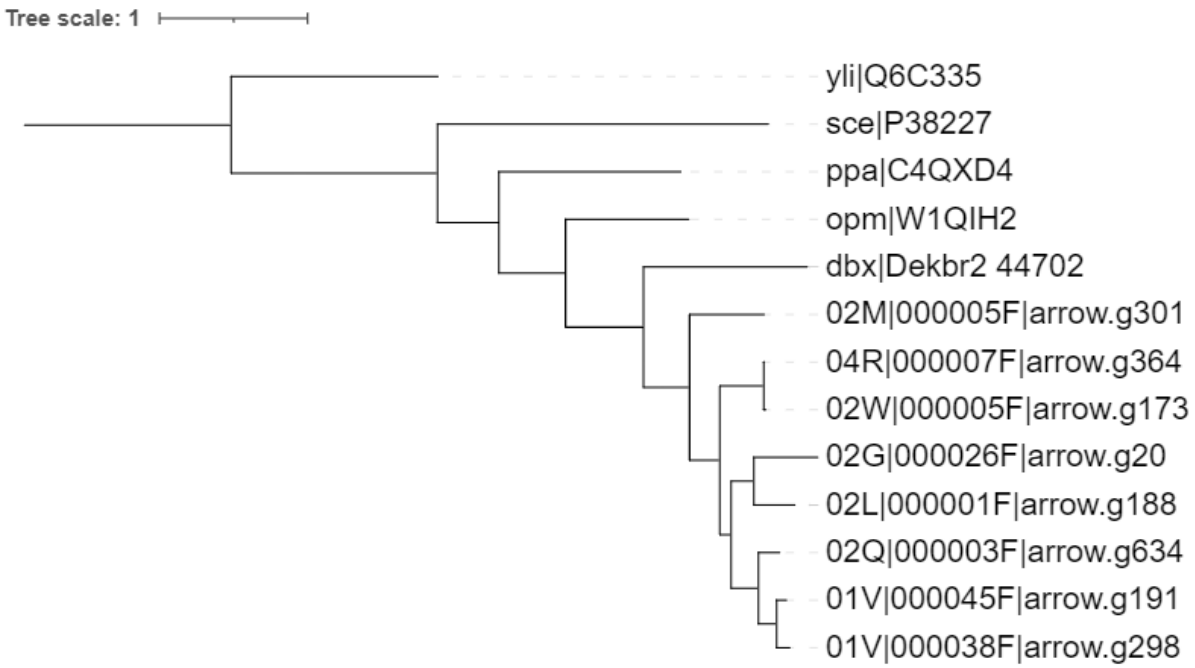

Figure 9. QDR3p Phylogenetic tree. Alignment generated with MEGA X, tree generated with IQ-TREE 2, visualization generated with EMBL iTOL. 01X refers to *K. exigua* 1478 NCYC, 01V refers to *P. membranifaciens* 55 NCYC, 02Q refers to *P. manshurica* Y-27978 USDA, 02M refers to *P. kudriavzevii* 2658 NCYC, 04R refers to *P. occidentalis* Y-6545 USDA, 02W refers to *P. occidentalis* 75-63 Phaff, 02G refers to *P. fermentans* 562 NCYC, 02L refers to *P. kluyveri* 05-608 Phaff. yli, ppa, sce, opm and dbx refer to *Y. lipolytica*, *K. phaffi*, *S. cerevisiae*, *O. polymorpha* and *B. bruxellensis* sequences from the Aybrah database.

The Snq2 family of ABC transporters has undergone many duplications in the *Pichia* strains, and likely numerous homologous recombination events given the consistent high copy number, and the fact that the proteins cluster with species (Figure 10). There is also a cluster of proteins present only in *P. kudriavzevii*, *P. manshurica*, *P. occidentalis* and *P. membranifaciens* (01V), mostly in two copies. Interestingly, the two copies present in *P. membranifaciens* cluster separately with the *P. manshurica* homolog, this is unusual as in most cases where *P. membranifaciens* has two copies due to its presumed heterozygous diploidy they cluster most closely to each other (Figure 10).

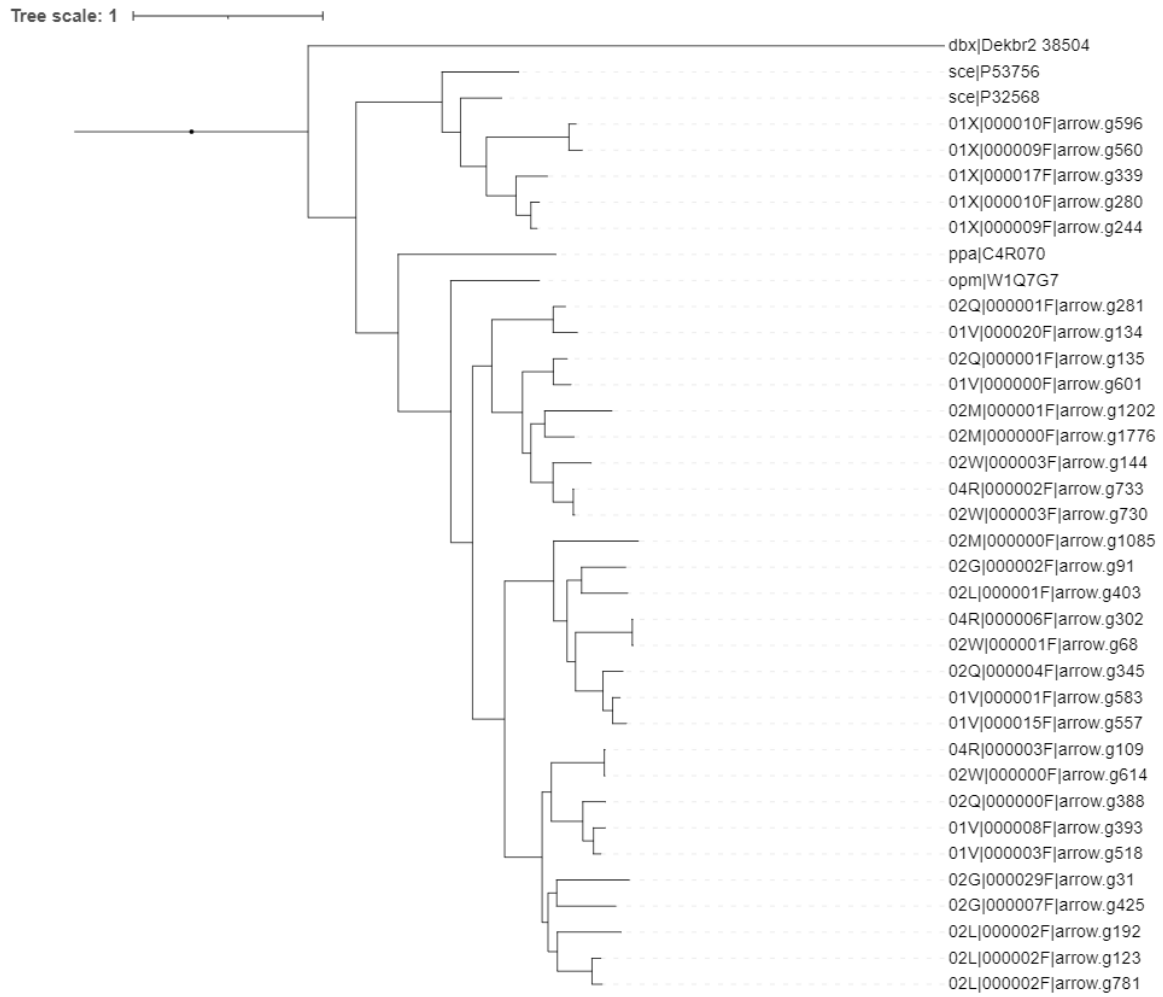

Figure 10. SNQ2p Phylogenetic tree. Alignment generated with MEGA X, tree generated with IQ-TREE 2, and visualization generated with EMBL iTOL. 01X refers to *K. exigua* 1478 NCYC, 01V refers to *P. membranifaciens* 55 NCYC, 02Q refers to *P. manshurica* Y-27978 USDA, 02M refers to *P. kudriavzevii* 2658 NCYC, 04R refers to *P. occidentalis* Y-6545 USDA, 02W refers to *P. occidentalis* 75-63 Phaff, 02G refers to *P. fermentans* 562 NCYC, 02L refers to *P. kluyveri* 05-608 Phaff. yli, ppa, sce, opm and dbx refer to *Y. lipolytica*, *K. phaffi*, *S. cerevisiae*, *O. polymorpha* and *B. bruxellensis* sequences from the Aybrah database.

5. Discussion

The developed workflow and strategy use many strains and few conditions, with the aim to sample as broad a genetic diversity as possible and find the most suitable strain possible. The approach shown enabled us to test a diverse range of organisms that previously have had very little published research, including *P. occidentalis*, which is ultimately the

tested species which appears to be most tolerant to adipic acid (Figure 4). *P. occidentalis* has no prior research as a metabolic engineering host and the research on its acid tolerance has primarily been from the perspective of food contamination (Arroyo López et al., 2007; Arroyo-López et al., 2006), and thus would be an unlikely strain to include if we had reduced screening capacity, though it was also isolated from grape skin and evaluated for potential as a lactic acid producing strain (Park et al., 2018). Additionally, the high throughput yeast cultivation techniques developed here could be paired to any analytic with sufficient throughput, potentially aiding in media optimization or diagnosing issues with engineered strains.

The identification of acid tolerance as a feature common in species from the *Pichia* clade raises further questions about which specific organic acids they are capable of tolerating, and whether the phenotype is also present in the adjacent Saturnispora and Brettanomyces genera. The testing showed that most *Pichia* strains were able to tolerate short to medium chain dicarboxylic acids in concentrations of 150 mM at a low pH with only a minimal hit to the growth rate (Figure 4), and that lowering the pH of YNB with HCl resulted in only a small increase in growth rate. Other groups have independently found certain strains of *Pichia* were tolerant to high concentrations of both succinic acid and lactic acid (Park et al., 2018; Xiao et al., 2014), but the numbers of strains tested are usually small, and they are often environmental isolates, rather than type strains. We think this strategy of utilizing common strains is essential for advancing research in non-conventional yeasts as it improves inter-lab reproducibility and ensures that concepts inferred from one experiment can be applied by other groups.

Looking at the phylogenetic trees revealed that neither *QDR3* nor *PDR12* showed consistently increased copy numbers across all strains, though most had at least one

duplication of *PDR12*. *P. kudriavzevii*, which is among the most acid tolerant strains in the collection showed none. The *PDR12* duplications may contribute to organic acid tolerance in those strains that have them, but they clearly are not the primary driver in all strains, the Pdr12 transporter may still mediate organic acid tolerance in these strains, however.

The *SNQ2* family of transporters saw a comparatively high number of duplications, present in all *Pichia* strains but due to intramolecular recombination, there are potentially significant differences between the transporters in different species. Snq2 has been previously found to be involved in tolerance to oxalic, malonic, formic, acetic, and propionic acids in *S. cerevisiae* (Cheng et al., 2007), given the acid tolerance observed and the *SNQ2* family expansion, some involvement is expected in our strains.

AGP2 codes for an environmental sensor in *S. cerevisiae*, and though the precise substrate it recognizes is unknown it regulates the uptake of L-carnitine and polyamines (Aouida et al., 2013). The *AGP2*-like family has evidence of both recent and ancestral expansion (Figure 7), given the structural similarity between L-carnitine and dicarboxylic acids, which *Pichia* species can import and metabolize (Kurtzman et al., 2011; Xi et al., 2021), it is possible that one of the homologues may act as a regulator of organic acid import, similar to the function of Snf3 with respect to sugars (Özcan & Johnston, 1999). What is clear is that there was an ancestral tandem gene duplication, as indicated by the consecutive gene numbers across all sequenced strains; unlike the SNQ2 gene family there is also a clear taxonomy, and the genes have also diverged since the duplications as indicated by branch lengths (Figure 7).

Further investigation in the form of gene knockouts and overexpression experiments would be necessary to confirm what role if any these genes play in organic acid tolerance, and whether their upregulation could result in even more tolerance to organic acids. The

recent development of a genetic toolkit for *Pichia kudriavzevii* provides a model for how this work could proceed (Cao et al., 2020). Metabolic engineering of *Pichia kudriavzevii* has already displayed that the acid tolerance phenotype results in measurable increases in productivity of succinic acid at low pH (Park et al., 2018), which can potentially reduce the operating costs of an organic acid bioprocess. Given the range of organic acids with industrial relevance, the further development of these strains can be of significant economic interest (J. Liu et al., 2017), and identifying causative proteins and systems in *Pichia* species for organic acid tolerance could allow for their reconstruction in other yeasts, or enhancement in the researched *Pichia* strains.

6. Conclusion

In this work we built on existing literature on non-Saccharomyces yeasts, assessing the tolerance of 122 publicly accessible yeasts to organic acid stress and then sequencing and annotating the genomes of the most tolerant strains. Based on those annotated genomes and literature from other yeast species, we developed several hypotheses about which genes are responsible for the organic acid tolerance phenotype. Acknowledging that while the organic acid tolerance is likely polygenic, prior work showed that individual transporters can significantly impact the ability of yeast to grow in the presence of specific organic acids (Kren et al., 2003), and that even among transporters active on the relevant substrates the degree to which tolerance is provided varies (Pereira et al., 2019). Given that, the identification of new transporters with potentially superior transport activities could prove valuable towards enabling organic acid cell factories.

Additionally, the characterization and sequencing of additional strains in this clade has the potential to enable metabolic engineering of these strains for production of organic acids or other products.

Future work in this area would be to develop genetic tools in theses strains to test gene knock-outs/knock-ins and empirically identify drivers of organic acid tolerance, the same tools could also be used to produce strains that accumulate and excrete organic acids. Key to achieving these goals would be the development of a CRISPR/Cas9 system, characterization of promoters and terminators, an optimized transformation method, and identification of functional centromeres. To improve the screening and characterization process used in this work further, an even larger selection of strains could be used, the 122 strains present in this study still don't represent the full diversity of the Saccharomycotina clade. The incorporation of additional factors of interest into the screen could similarly improve the results, screening for important metabolic characteristics such as accumulation of side products or tolerance for a specific medium could identify strains that would be even better suited for a specific bioprocess

7. References

Abbott, D. A., Zelle, R. M., Pronk, J. T., & van Maris, A. J. A. (2009). Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: Current status and challenges. *FEMS Yeast Research*, *9*(8), 1123–1136. https://doi.org/10.1111/j.1567-1364.2009.00537.x

Aouida, M., Texeira, M. R., Thevelein, J. M., Poulin, R., & Ramotar, D. (2013). Agp2, a Member of the Yeast Amino Acid Permease Family, Positively Regulates Polyamine

Transport at the Transcriptional Level. *PLOS ONE*, *8*(6), e65717. https://doi.org/10.1371/journal.pone.0065717

Arroyo López, F. N., Quintana, M. C. D., & Fernández, A. G. (2007). Modelling of the growth-no growth interface of Issatchenkia occidentalis, an olive spoiling yeast, as a function of the culture media, NaCl, citric and sorbic acid concentrations: Study of its inactivation in the no growth region. *International Journal of Food Microbiology*, *117*(2), 150–159. https://doi.org/10.1016/j.ijfoodmicro.2007.03.005

Arroyo-López, F. N., Durán-Quintana, M., & Garrido-Fernández, A. (2006). Microbial Evolution during Storage of Seasoned Olives Prepared with Organic Acids with Potassium Sorbate, Sodium Benzoate, and Ozone Used as Preservatives. *Journal of Food Protection*, *69*(6), 1354–1364. https://doi.org/10.4315/0362-028X-69.6.1354

Balzi, E., & Goffeau, A. (1995). Yeast multidrug resistance: The PDR network. *Journal of Bioenergetics and Biomembranes*, *27*(1), 71–76. https://doi.org/10.1007/BF02110333

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.-L., Molina-Jouve, C., & Nicaud, J.-M. (2009). Yarrowia lipolytica as a model for bio-oil production. *Progress in Lipid Research*, *48*(6), 375–387. https://doi.org/10.1016/j.plipres.2009.08.005

Besada-Lombana, P. B., Fernandez-Moya, R., Fenster, J., & Da Silva, N. A. (2017). Engineering Saccharomyces cerevisiae fatty acid composition for increased tolerance to octanoic acid. *Biotechnology and Bioengineering*, *114*(7), 1531–1538. https://doi.org/10.1002/bit.26288

Byrne, K. P. (2005). The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, *15*(10), 1456–1461. https://doi.org/10.1101/gr.3672305

Cao, M., Fatma, Z., Song, X., Hsieh, P.-H., Tran, V. G., Lyon, W. L., Sayadi, M., Shao, Z., Yoshikuni, Y., & Zhao, H. (2020). A genetic toolbox for metabolic engineering of Issatchenkia orientalis. *Metabolic Engineering*, *59*, 87–97. https://doi.org/10.1016/j.ymben.2020.01.005

Castro, R. C. A., & Roberto, I. C. (2014). Selection of a Thermotolerant Kluyveromyces marxianus Strain with Potential Application for Cellulosic Ethanol Production by Simultaneous Saccharification and Fermentation. *Applied Biochemistry and Biotechnology*, *172*(3), 1553–1564. https://doi.org/10.1007/s12010-013-0612-5

Cibrario, A., Miot-Sertier, C., Paulin, M., Bullier, B., Riquier, L., Perello, M.-C., de Revel, G., Albertin, W., Masneuf-Pomarède, I., Ballestra, P., & Dols-Lafargue, M. (2020). Brettanomyces bruxellensis phenotypic diversity, tolerance to wine stress and wine spoilage ability. *Food Microbiology*, *87*, 103379. https://doi.org/10.1016/j.fm.2019.103379

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Conterno, L., Joseph, C. M. L., Arvik, T. J., Henick-Kling, T., & Bisson, L. F. (2006). Genetic and Physiological Characterization of Brettanomyces bruxellensis Strains Isolated from Wines. *American Journal of Enology and Viticulture*, *57*(2), 139–147.

Correia, K., & Mahadevan, R. (2020). Pan-Genome-Scale Network Reconstruction: Harnessing Phylogenomics Increases the Quantity and Quality of Metabolic Models. *Biotechnology Journal*, *15*(10), 1900519. https://doi.org/10.1002/biot.201900519

Correia, K., Yu, S. M., & Mahadevan, R. (2019). AYbRAH: a curated ortholog database for yeasts and fungi spanning 600 million years of evolution. *Database*, *2019*. https://doi.org/10.1093/database/baz022

Cui, Z., Gao, C., Li, J., Hou, J., Lin, C. S. K., & Qi, Q. (2017). Engineering of unconventional yeast Yarrowia lipolytica for efficient succinic acid production from glycerol at low pH. *Metabolic Engineering*, *42*, 126–133. https://doi.org/10.1016/j.ymben.2017.06.007

Dang, T. D. T., De Maeseneire, S. L., Zhang, B. Y., De Vos, W. H., Rajkovic, A., Vermeulen, A., Van Impe, J. F., & Devlieghere, F. (2012). Monitoring the intracellular pH of Zygosaccharomyces bailii by green fluorescent protein. *International Journal of Food Microbiology*, *156*(3), 290–295. https://doi.org/10.1016/j.ijfoodmicro.2012.03.028

Dhaliwal, S. S., Oberoi, H. S., Sandhu, S. K., Nanda, D., Kumar, D., & Uppal, S. K. (2011). Enhanced ethanol production from sugarcane juice by galactose adaptation of a newly isolated thermotolerant strain of Pichia kudriavzevii. *Bioresource Technology*, *102*(10), 5968–5975. https://doi.org/10.1016/j.biortech.2011.02.015

Douglass, A. P., Byrne, K. P., & Wolfe, K. H. (2019). The Methylotroph Gene Order Browser (MGOB) reveals conserved synteny and ancestral centromere locations in the yeast family Pichiaceae. *FEMS Yeast Research*, *19*(6). https://doi.org/10.1093/femsyr/foz058

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Fletcher, E., Feizi, A., Bisschops, M. M. M., Hallström, B. M., Khoomrung, S., Siewers, V., & Nielsen, J. (2017). Evolutionary engineering reveals divergent paths when yeast is adapted to different acidic environments. *Metabolic Engineering*, *39*, 19–28. https://doi.org/10.1016/j.ymben.2016.10.010

Fletcher, E., Mercurio, K., Walden, E. A., & Baetz, K. (2021). A yeast chemogenomic screen identifies pathways that modulate adipic acid toxicity. *IScience*, *24*(4), 102327. https://doi.org/10.1016/j.isci.2021.102327

Ho, B., Baryshnikova, A., & Brown, G. W. (2018). Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. *Cell Systems*, *6*(2), 192-205.e3. https://doi.org/10.1016/j.cels.2017.12.004

Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. A., & Wheeler, T. J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Research*, *44*(D1), D81–D89. https://doi.org/10.1093/nar/gkv1272

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Ji, H., Xu, K., Dong, X., Sun, D., Peng, R., Lin, S., Zhang, K., & Jin, L. (2020). Transcriptional profiling reveals molecular basis and the role of arginine in response to low-pH stress in Pichia kudriavzevii. *Journal of Bioscience and Bioengineering*, *130*(6), 588–595. https://doi.org/10.1016/j.jbiosc.2020.07.014

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A.,

Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Karlsson, E., Mapelli, V., & Olsson, L. (2017). Adipic acid tolerance screening for potential adipic acid production hosts. *Microbial Cell Factories*, *16*(1), 20. https://doi.org/10.1186/s12934-017-0636-6

Kawahata, M., Masaki, K., Fujii, T., & Iefuji, H. (2006). 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 Research*, *6*(6), 924–936. https://doi.org/10.1111/j.1567-1364.2006.00089.x

Krassowski, T., Kominek, J., Shen, X.-X., Opulente, D. A., Zhou, X., Rokas, A., Hittinger, C. T., & Wolfe, K. H. (2019). Multiple Reinventions of Mating-type Switching during Budding Yeast Evolution. *Current Biology : CB*, *29*(15), 2555-2562.e8. https://doi.org/10.1016/j.cub.2019.06.056

Kren, A., Mamnun, Y. M., Bauer, B. E., Schuller, C., Wolfger, H., Hatzixanthis, K., Mollapour, M., Gregori, C., Piper, P., & Kuchler, K. (2003). War1p, a Novel Transcription Factor Controlling Weak Acid Stress Response in Yeast. *Molecular and Cellular Biology*, *23*(5), 1775–1785. https://doi.org/10.1128/MCB.23.5.1775-1785.2003

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kurtzman, C. P., Fell, J. W., & Boekhout, T. (2011). *The yeasts: A taxonomic study*. Elsevier Science. https://www.sciencedirect.com/book/9780444521491/the-yeasts

Labbani, F.-Z. K., Turchetti, B., Bennamoun, L., Dakhmouche, S., Roberti, R., Corazzi, L., Meraihi, Z., & Buzzini, P. (2015). A novel killer protein from Pichia kluyveri isolated from an Algerian soil: Purification and characterization of its in vitro activity against food and beverage spoilage yeasts. *Antonie van Leeuwenhoek*, *107*(4), 961–970. https://doi.org/10.1007/s10482-015-0388-4

Larroude, M., Park, Y.-K., Soudier, P., Kubiak, M., Nicaud, J.-M., & Rossignol, T. (2019). A modular Golden Gate toolkit for Yarrowia lipolytica synthetic biology. *Microbial Biotechnology*, *12*(6), 1249–1259. https://doi.org/10.1111/1751-7915.13427

Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, *4*(9), 975–986. https://doi.org/10.1021/sb500366v

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, H., Fan, H., Li, Y., Shi, G. Y., Ding, Z. Y., Gu, Z. H., & Zhang, L. (2017). Construction and application of multi-host integrative vector system for xylose-fermenting yeast. *FEMS Yeast Research*, *17*(6), 1–10. https://doi.org/10.1093/femsyr/fox055

Lian, J., Mishra, S., & Zhao, H. (2018). Recent advances in metabolic engineering of Saccharomyces cerevisiae: New tools and their applications. *Metabolic Engineering*, *50*, 85–108. https://doi.org/10.1016/j.ymben.2018.04.011

Lin, Y., Sun, X., Yuan, Q., & Yan, Y. (2014). Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in Escherichia coli. *Metabolic Engineering*, *23*(2014), 62–69. https://doi.org/10.1016/j.ymben.2014.02.009

Lindahl, L., Genheden, S., Eriksson, L. A., Olsson, L., & Bettiga, M. (2016). Sphingolipids contribute to acetic acid resistance in Zygosaccharomyces bailii. *Biotechnology and Bioengineering*, *113*(4), 744–753. https://doi.org/10.1002/bit.25845

Liu, H., Marsafari, M., Wang, F., Deng, L., & Xu, P. (2019). Engineering acetyl-CoA metabolic shortcut for eco-friendly production of polyketides triacetic acid lactone in Yarrowia lipolytica. *Metabolic Engineering*, *56*, 60–68. https://doi.org/10.1016/j.ymben.2019.08.017

Liu, J., Li, J., Shin, H., Liu, L., Du, G., & Chen, J. (2017). Protein and metabolic engineering for the production of organic acids. *Bioresource Technology*, *239*, 412–421. https://doi.org/10.1016/j.biortech.2017.04.052

Liu, Z., Liang, Y., Ang, E. L., & Zhao, H. (2017). A New Era of Genome Integration—Simply Cut and Paste! *ACS Synthetic Biology*, *6*(4), 601–609. https://doi.org/10.1021/acssynbio.6b00331

Maguire, S. L., ÓhÉigeartaigh, S. S., Byrne, K. P., Schröder, M. S., O'Gaora, P., Wolfe, K. H., & Butler, G. (2013). Comparative Genome Analysis and Gene Finding in Candida

Species Using CGOB. *Molecular Biology and Evolution*, *30*(6), 1281–1291. https://doi.org/10.1093/molbev/mst042

Mandal, S., & Ghosh, K. (2013). Optimization of Tannase Production and Improvement of Nutritional Quality of Two Potential Low-Priced Plant Feedstuffs under Solid State Fermentation by Pichia kudriavzevii Isolated from Fish Gut. *Food Biotechnology*, *27*(1), 86–103. https://doi.org/10.1080/08905436.2012.755929

Marcišauskas, S., Ji, B., & Nielsen, J. (2019). Reconstruction and analysis of a Kluyveromyces marxianus genome-scale metabolic model. *BMC Bioinformatics*, *20*(1), 551. https://doi.org/10.1186/s12859-019-3134-5

Matsushika, A., Negi, K., Suzuki, T., Goshima, T., & Hoshino, T. (2016). Identification and Characterization of a Novel Issatchenkia orientalis GPI-Anchored Protein, IoGas1, Required for Resistance to Low pH and Salt Stress. *PLOS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0161888

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Morrissey, J. P., Etschmann, M. M. W., Schrader, J., & de Billerbeck, G. M. (2015). Cell factory applications of the yeast Kluyveromyces marxianus for the biotechnological production of natural flavour and fragrance molecules. *Yeast*, *32*(1), 3–16. https://doi.org/10.1002/yea.3054

Nielsen, M. K., & Arneborg, N. (2007). The effect of citric acid and pH on growth and metabolism of anaerobic Saccharomyces cerevisiae and Zygosaccharomyces bailii

cultures. *Food Microbiology*, *24*(1), 101–105. https://doi.org/10.1016/j.fm.2006.03.005

Osawa, F., Fujii, T., Nishida, T., Tada, N., Ohnishi, T., Kobayashi, O., Komeda, T., & Yoshida, S. (2009). Efficient production of L-lactic acid by Crabtree-negative yeast Candida boidinii. *Yeast*, *26*(9), 485–496. https://doi.org/10.1002/yea.1702

Otero, J. M., Cimini, D., Patil, K. R., Poulsen, S. G., Olsson, L., & Nielsen, J. (2013). Industrial Systems Biology of Saccharomyces cerevisiae Enables Novel Succinic Acid Cell Factory. *PLOS ONE*, *8*(1), e54144. https://doi.org/10.1371/journal.pone.0054144

Özcan, S., & Johnston, M. (1999). Function and Regulation of Yeast Hexose Transporters. *Microbiology and Molecular Biology Reviews*, *63*(3), 554–569. https://doi.org/10.1128/MMBR.63.3.554-569.1999

Palma, M., Guerreiro, J. F., & Sá-Correia, I. (2018). Adaptive Response and Tolerance to Acetic Acid in Saccharomyces cerevisiae and Zygosaccharomyces bailii: A Physiological Genomics Perspective. *Frontiers in Microbiology*, *9*, 274. https://doi.org/10.3389/fmicb.2018.00274

Park, H. J., Bae, J., Ko, H., Lee, S., Sung, B. H., Han, J., & Sohn, J. (2018). Low-pH production of d-lactic acid using newly isolated acid tolerant yeast Pichia kudriavzevii NG7. *Biotechnology and Bioengineering*, *115*(9), 2232–2242. https://doi.org/10.1002/bit.26745

Pereira, R., Wei, Y., Mohamed, E., Radi, M., Malina, C., Herrgård, M. J., Feist, A. M., Nielsen, J., & Chen, Y. (2019). Adaptive laboratory evolution of tolerance to

dicarboxylic acids in Saccharomyces cerevisiae. *Metabolic Engineering*, *56*, 130–141. https://doi.org/10.1016/j.ymben.2019.09.008

Piper, P., Calderon, C. O., Hatzixanthis, K., & Mollapour, M. 2001. (2001). Weak acid adaptation: The stress response that confers yeasts with resistance to organic acid food preservatives. *Microbiology*, *147*(10), 2635–2642. https://doi.org/10.1099/00221287-147-10-2635

Piper, P., Mahé, Y., Thompson, S., Pandjaitan, R., Holyoak, C., Egner, R., Mühlbauer, M., Coote, P., & Kuchler, K. (1998). The Pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast. *EMBO Journal*, *17*(15), 4257–4265. https://doi.org/10.1093/emboj/17.15.4257

Porro, D., Brambilla, L., Ranzi, B. M., Martegani, E., & Alberghina, L. (1995). Development of metabolically engineered Saccharomyces cerevisiae cells for the production of lactic acid. *Biotechnology Progress*, *11*(3), 294–298. https://doi.org/10.1021/bp00033a009

Pyne, M. E., Kevvai, K., Grewal, P. S., Narcross, L., Choi, B., Bourgeois, L., Dueber, J. E., & Martin, V. J. J. (2020). A yeast platform for high-level synthesis of tetrahydroisoquinoline alkaloids. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17172-x

Pyne, M. E., Narcross, L., Melgar, M., Kevvai, K., Mookerjee, S., Leite, G. B., & Martin, V. J. J. (2018). An Engineered Aro1 Protein Degradation Approach for Increased cis,cis -Muconic Acid Biosynthesis in Saccharomyces cerevisiae. *Applied and Environmental Microbiology*, *84*(17), AEM.01095-18. https://doi.org/10.1128/AEM.01095-18

Qu, Y., Cao, X., Ma, Q., Shi, S., Tan, L., Li, X., Zhou, H., Zhang, X., & Zhou, J. (2012). Aerobic decolorization and degradation of Acid Red B by a newly isolated Pichia sp. TCL. *Journal of Hazardous Materials*, *223–224*, 31–38. https://doi.org/10.1016/j.jhazmat.2012.04.034

Rajkumar, A. S., Varela, J. A., Juergens, H., Daran, J.-M. G., & Morrissey, J. P. (2019). Biological Parts for Kluyveromyces marxianus Synthetic Biology. *Frontiers in Bioengineering and Biotechnology*, *7*, 97. https://doi.org/10.3389/fbioe.2019.00097

Riley, R., Haridas, S., Wolfe, K. H., Lopes, M. R., Hittinger, C. T., Göker, M., Salamov, A. A., Wisecaver, J. H., Long, T. M., Calvey, C. H., Aerts, A. L., Barry, K. W., Choi, C., Clum, A., Coughlan, A. Y., Deshpande, S., Douglass, A. P., Hanson, S. J., Klenk, H.-P., … Jeffries, T. W. (2016). Comparative genomics of biotechnologically important yeasts. *Proceedings of the National Academy of Sciences*, *113*(35), 9882–9887. https://doi.org/10.1073/pnas.1603941113

Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., & Keasling, J. D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, *440*(7086), 940–943. https://doi.org/10.1038/nature04640

Roe, A. J., McLaggan, D., Davidson, I., O'Byrne, C., & Booth, I. R. (1998). Perturbation of Anion Balance during Inhibition of Growth of Escherichia coli by Weak Acids. *Journal of Bacteriology*, *180*(4), 767–772. https://doi.org/10.1128/JB.180.4.767-772.1998

Saez, J. S., Lopes, C. A., Kirs, V. E., & Sangorrín, M. (2011). Production of volatile phenols by Pichia manshurica and Pichia membranifaciens isolated from spoiled wines and cellar environment in Patagonia. *Food Microbiology*, *28*(3), 503–509. https://doi.org/10.1016/j.fm.2010.10.019

Shen, X. X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R. ichiroh, Čadež, N., Libkind, D., … Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, *175*(6), 1533-1545.e20. https://doi.org/10.1016/j.cell.2018.10.023

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*, *32*(suppl_2), W309–W312. https://doi.org/10.1093/nar/gkh379

Stratford, M., Steels, H., Nebe-von-Caron, G., Novodvorska, M., Hayer, K., & Archer, D. B. (2013). Extreme resistance to weak-acid preservatives in the spoilage yeast Zygosaccharomyces bailii. *International Journal of Food Microbiology*, *166*(1), 126–134. https://doi.org/10.1016/j.ijfoodmicro.2013.06.025

Ullah, A., Chandrasekaran, G., Brul, S., & Smits, G. J. (2013). Yeast adaptation to weak acids prevents futile energy expenditure. *Frontiers in Microbiology*, *4*(JUN), 142. https://doi.org/10.3389/fmicb.2013.00142

Warth, A. D. (1988). Effect of Benzoic Acid on Growth Yield of Yeasts Differing in Their Resistance to Preservatives. *Applied and Environmental Microbiology*, *54*(8), 2091–2095. https://doi.org/10.1128/aem.54.8.2091-2095.1988

Xi, Y., Zhan, T., Xu, H., Chen, J., Bi, C., Fan, F., & Zhang, X. (2021). Characterization of JEN family carboxylate transporters from the acid-tolerant yeast Pichia kudriavzevii and their applications in succinic acid production. *Microbial Biotechnology*, *14*(3), 1130–1147. https://doi.org/10.1111/1751-7915.13781

Xiao, H., Shao, Z., Jiang, Y., Dole, S., & Zhao, H. (2014). Exploiting Issatchenkia orientalis SD108 for succinic acid production. *Microbial Cell Factories*, *13*(1), 121. https://doi.org/10.1186/s12934-014-0121-4

Yu, J.-L., Qian, Z.-G., & Zhong, J.-J. (2018). Advances in bio-based production of dicarboxylic acids longer than C4. *Engineering in Life Sciences*, *18*(9), 668–681. https://doi.org/10.1002/elsc.201800023

Yu, R., & Nielsen, J. (2019). Big data in yeast systems biology. *FEMS Yeast Research*, *19*(7). https://doi.org/10.1093/femsyr/foz070

Yue, J.-X., & Liti, G. (2018). Long-read sequencing data analysis for yeasts. *Nature Protocols*, *13*(6), 1213–1231. https://doi.org/10.1038/nprot.2018.025

Yuzbashev, T. V., Yuzbasheva, E. Y., Sobolevskaya, T. I., Laptev, I. A., Vybornaya, T. V., Larina, A. S., Matsui, K., Fukui, K., & Sineoky, S. P. (2010). Production of succinic acid at low pH by a recombinant strain of the aerobic yeast Yarrowia lipolytica. *Biotechnology and Bioengineering*, *107*(4), 673–682. https://doi.org/10.1002/bit.22859