SCUOLA DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

University of Milano-Bicocca | University College Cork

Department of Biotechnology and Biosciences | School of Microbiology

PhD program in Converging Technologies for Biomolecular Systems (TeCSBI) | PhD program in Microbial Biotechnology

Cycle XXXIII

Application of direct and indirect strain engineering approaches to unlock the potential of the yeasts *Zygosaccharomyces parabailii and Kluyveromyces marxianus* for bio-based processes

Pooja Jayaprakash

Registration number: 836891

Supervisors: Prof. Paola Branduardi (University of Milano-Bicocca)

Prof. John Morrissey (University College Cork)

May 2022

“When the going gets tough, the tough get going”

To all those who believed in me....

List of Abbreviations

ABC: ATP-binding cassette

AFEX: Ammonia fibre expansion

ALE: Adaptive laboratory evolution

BLAST: Basic Local Alignment Search Tool

$CaCl_2.2H_2O$: Calcium chloride dihydrate

CAPEX: Capital expenditure

CRISPR: Clustered regularly interspaced short palindromic repeats

Cas9: CRISPR associated protein 9

cDNA: complementary DNA

CDS: Coding sequence

crRNA: CRISPR RNA

Cryo-EM: Cryogenic electron microscopy

Deep TMHMM: Deep Learning Model for Transmembrane Topology Prediction and Classification

DSB: Double-strand break

$FeSO_4.7H_2O$: Iron sulphate heptahydrate

G418: Geneticin

GMO: Genetically modified organism

GRAS: Generally regarded as safe

gRNA: guide RNA

H_2O_2: Hydrogen peroxide

H_2SO_4: Sulphuric acid

HDR: homology directed repair

HDV: Hepatitis D ribozyme

HH: Hammerhead ribozyme

HMF: Hydroxymethylfurfural

HR: Homologous recombination

KH_2PO_4: Potassium phosphate dibasic

LCB: Lignocellulosic biomass

LCC: Lignin-carbohydrate complex

LiAc: Lithium acetate

LPMO: Lytic polysaccharide monooxygenase

MAFFT: Multiple Alignments using Fast Fourier Transform

$Mg_2SO_4.7H_2O$: Magnesium sulphate heptahydrate

$MnSO_4.H_2O$: Manganese sulphate monohydrate

MUSCLE: Multiple Sequence Comparison by Log-Expectation

Na_2HPO_4: Sodium phosphate dibasic

NaOH: Sodium hydroxide

NBD: Nucleotide-binding domain

NCBI: National Center for Biotechnology Information

NCY: Non-conventional yeasts

NHEJ: Non-homologous end joining

OD: Optical density

ORF: Open reading frame

PAM: Protospacer adjacent motif

PCR: Polymerase chain reaction

PDB: Protein Data bank

PEG: Polyethylene glycol

pH: Potential of hydrogen

pKa: Acid dissociation constant

PyMOL: Molecular Visualization System

QPS: Qualified presumption of safety

RNA: Ribonucleic acid

ROS: Reactive oxygen species

rRNA: Ribosomal RNA

RT-qPCR: Quantitative reverse transcription PCR

SBP: Sugar beet pulp

SGR: Specific growth rate

SHF: Separate hydrolysis and fermentation

SNP: Single nucleotide polymorphism

SPOT-RNA: RNA Secondary Structure Predictor

ssDNA: Single-stranded DNA

SSF: Simultaneous saccharification and fermentation

TMD: Transmembrane domain

tracrRNA: transactivating CRISPR RNA

tRNA: Transfer RNA

v/v: Volume per volume

w/v: Weight per volume

WGD: Whole genome duplication

WIS: Water insoluble solids

WOA: Weak organic acids

WT: Wild type

YNB: Yeast nitrogen base medium

YPD: Yeast extract-peptone-dextrose medium

YPF: Yeast extract-peptone-fructose medium

$ZnSO_4.7H_2O$: Zinc sulphate heptahydrate

Table of Contents

Abstract

The efficient implementation of biorefineries to produce bio-based chemicals and fuels requires a sustainable source of feedstock and robust microbial factories. Among others, lignocellulosic biomass represents cheap and sugar-enriched feedstock. The conversion of lignocellulosic biomass into the desired products using microbial cell factories is a promising option to replace the fossil based petrochemical refinery. Minimum nutritional requirements and robustness have made yeasts a class of microbial hosts widely employed in industrial biotechnology, exploiting their natural abilities as well as genetically acquired pathways for the production of natural and recombinant products, including bulk chemicals such as organic acids. In bio-based industrial processes, microorganisms are subjected to different kinds of stresses associated with process conditions. These stressors are known to inhibit cellular metabolism and compromise the performances of a fermentative process, being an important limitation to an effective marketability of biobased microbial products. Therefore, the exploration of yeast biodiversity to exploit unique native features and the understanding of mechanisms to endure harsh conditions are essential to develop viable and competitive bioprocesses. In the chapter 1 of this thesis we reviewed the link between LCBs composition, choice of enzymatic cocktail and selection of yeast species and strains that need to be considered in an integrated fashion to enable the development of an efficient bio-based process. We discussed the pivotal role of enzymatic cocktail optimization to unlock the potential of non-Conventional yeasts, which, thanks to broader substrate utilization, inhibitor resistance and peculiar metabolism, can widen the array of feedstock and products of biorefineries.

The aim of this PhD work was to expand the industrial potential of two non-conventional yeasts, *Zygosaccharomyces parabailii* and *Kluyveromyces marxianus,* by applying direct and indirect strain engineering approaches. These yeasts possess desirable characteristics. *K. marxianus* has broad specificity for both hexose and pentose sugars as carbon and energy source. Apart from this, its thermotolerance, fast growth and the ability to thrive at pH below 3, make it ideal for industrial use. However, the lack of tolerance of this yeast to inhibitory compounds, particularly weak organic acid produced during LCB pretreatment, hinders its use when this biomass is

used as substrates. Although the use of synthetic biology techniques has started to be employed to understand the robustness of *K. marxianus* and for the production of various chemicals, the mechanisms related to organic acid tolerance are yet to be deciphered. To match this goal, we used Adaptive Laboratory Evolution (ALE), an indirect strain engineering approach, alternative and often complementary to direct engineering. In chapter 2, we aimed to improve the tolerance of *K. marxianus* to sugar beet pulp (SBP) hydrolysate at pH 3.0 at two different temperatures, 30 ^{o}C and 40 ^{o}C. Using the ALE approach, we selected *K. marxianus* evolved isolates with robust phenotype compared to the parental strains, at 30 ^{o}C.

Differently to *K. marxianus*, the hybrid yeast *Z. parabailii* exhibits resistance to weak organic acids (WOA) also at low pH. Understanding the mechanism involved in tolerance to WOA can be used for avoiding the growth of this yeast in food production pipelines as well as for promoting its use as a cell factory for the production of organic acids and other bio-products. In chapter 3 of this study, our aim was to understand the phenotype-genotype correlation involved in the WOA tolerance trait. Using direct engineering method, we constructed and characterised single and double *Z. parabailii pdr12* mutants. This study revealed that Pdr12p is involved in tolerance to acetic and butyric acids and not in tolerance towards sorbic and benzoic acids. Furthermore, analysis of the Pdr12p sequence provided insights in the amino acids differences. The *pdr12* mutants were constructed by the classical tool of exploiting deletion cassettes. Advances in metabolic engineering and synthetic biology have increased the need for creating techniques such as CRISPR-Cas9 for faster and more efficient genome editing. In chapter 4 of this study our aim was to develop a CRISPR-Cas9 system for simultaneous disruption or deletion of two alleles of a gene in *Z. parabailii*. We evaluated the use of four different gRNA expression systems consisting of combinations of tRNAs, tRNA and ribozyme or ribozymes as self-cleaving flanking elements. The functionality of the gRNA systems was tested by analyzing the inactivation of the *ADE2* gene in the wild type strain and the most efficient gRNA system was used to successfully construct a *Z. parabailii dnl4* mutant. This mutant exhibited improved homologous recombination in the deletion of both *ADE2* alleles. Analysis of mutations in the gRNA target regions of both *ADE2* and *DNL4* genes suggested inter-allelic rearrangements between the two gene loci, as well as absence

of large regions of chromosomes.

Overall, this work contributes to the vast array of studies that are shedding light on yeasts biodiversity, both as a way for understanding their natural potential and as an instrument for tailoring novel cell factories.

Riassunto

L'efficiente implementazione di bioraffinerie microbiche per la produzione di sostanze chimiche e combustibili di origine rinnovabile (*bio-based*) richiede una fonte sostenibile di materie prime e robuste *cell factory*. Tra le varie opzioni, le biomasse lignocellulosiche rappresentano una materia prima economica e ricca di zuccheri. La conversione della biomassa lignocellulosica (LCB) nei prodotti desiderati mediante l'uso di *cell factory* microbiche è una alternativa promettente per sostituire le raffinerie petrolchimiche basate su risorse fossili. I requisiti nutrizionali minimi e la robustezza hanno reso i lieviti una classe di microrganismi ampiamente utilizzata nelle biotecnologie industriali, sfruttando le loro capacità naturali e le vie metaboliche geneticamente implementate per l'ottenimento di prodotti naturali e ricombinanti, compresi gli acidi organici. Nei processi industriali *bio-based* i microrganismi sono soggetti a diversi tipi di stress associati alle condizioni di processo.

È noto che questi fattori di stress inibiscono il metabolismo cellulare e ne compromettono le prestazioni di processo, diventando un importante limite ad una possibile commerciabilità dei prodotti microbici *bio-based*. Pertanto, l'esplorazione della biodiversità dei lieviti per sfruttarne le caratteristiche naturali uniche e la comprensione dei meccanismi che permettono loro di sopportare condizioni ambientali avverse sono essenziali per sviluppare bioprocessi competitivi. Nel capitolo 1 di questa tesi abbiamo esaminato l'interdipendenza tra la composizione delle LCB, la scelta del cocktail enzimatico per la degradazione e la selezione di specie e ceppi di lievito, che devono essere considerati in modo integrato per consentire lo sviluppo di un efficiente bioprocesso. Abbiamo discusso il ruolo fondamentale dell'ottimizzazione del cocktail enzimatico per sbloccare il potenziale dei lieviti non convenzionali, che, grazie alla capacità di metabolizzare molti substratie alla resistenza agli inibitori possono ampliare la gamma di materie prime e prodotti delle bioraffinerie.

L'obiettivo di questo lavoro di dottorato è di espandere il potenziale industriale di due lieviti non convenzionali, *Zygosaccharomyces parabailii* e *Kluyveromyces marxianus*, applicando approcci diretti e indiretti di ingegneria metabolica. Questi lieviti possiedono caratteristiche desiderabili. *K. marxianus* ha un'ampia specificità sia per gli zuccheri esosi che per quelli pentosi come fonte di carbonio e di energia. Oltre a

questo, la termotolleranza, la rapida crescita e la capacità di crescere a pH inferiore a 3 lo rendono ideale per l'uso industriale. Tuttavia, la scarsa tolleranza agli acidi organici deboli liberati durante il pretrattamento di LCB, ne ostacola l'uso quando questa biomassa viene utilizzata come substrato.

Sebbene negli ultimi anni lo sviluppo di tecniche di biologia sintetica stai facilitando lo studio e l'impiego di *K. marxianus* per la produzione di varie sostanze chimiche, i meccanismi relativi alla tolleranza agli acidi organici devono ancora essere decifrati. Per raggiungere questo obiettivo, abbiamo utilizzato una *Adaptive Laboratory Evolution* (ALE), un approccio di ingegneria indiretta dei ceppi, alternativo e spesso complementare all'ingegneria diretta. Nel capitolo 2, abbiamo mirato a migliorare la tolleranza di *K. marxianus* all'idrolizzato di polpa di barbabietola da zucchero (SBP) a pH 3,0 a due diverse temperature, 30 °C e 40 °C. Utilizzando l'approccio ALE, abbiamo selezionato isolati di *K. marxianus* evoluti con fenotipo robusto rispetto ai ceppi parentali, a 30 °C.

A differenza di *K. marxianus*, il lievito ibrido *Z. parabailii* mostra resistenza agli acidi organici deboli (WOA) anche a pH bassi. La comprensione dei meccanismi coinvolti nella tolleranza a WOA può permettere di evitare la crescita di questo lievito nelle filiere di produzione alimentare, nonché promuoverne l'uso come *cell factory* per la produzione di acidi organici e altri bioprodotti. Nel capitolo 3 di questo studio, il nostro obiettivo era comprendere la correlazione fenotipo-genotipo coinvolta nella tolleranza a WOA, ed abbiamo concentrato l'attenzione sul trasportatore di membrana Pdr12, creando singoli e doppi mutanti. Questo studio ha rivelato che Pdr12p è coinvolto nella tolleranza agli acidi acetico e butirrico e non nella tolleranza agli acidi sorbico e benzoico. Inoltre, l'analisi della sequenza delle due copie di Pdr12p ha fornito informazioni sulle differenze amminoacidiche. I progressi nell'ingegneria metabolica e nella biologia sintetica hanno spinto verso la necessità di sviluppare tecniche come CRISPR-Cas9 per una modifica del genoma più rapida ed efficiente. Nel capitolo 4 di questo studio il nostro obiettivo era quello di sviluppare un sistema CRISPR-Cas9 per l'inattivazione o la delezione simultanea di due alleli di un gene in *Z. parabailii*. Abbiamo valutato l'uso di quattro diversi sistemi di espressione di gRNA costituiti da combinazioni di tRNA, tRNA e ribozima o ribozimi come sequenze fiancheggianti di processamento. La funzionalità dei sistemi gRNA è stata testata analizzando

l'inattivazione del gene *ADE2* nel ceppo *wild type* ed è stato utilizzato il sistema gRNA più efficiente per costruire con successo un mutante *Z. parabailii* dnl4. Questo mutante ha mostrato ricombinazione omologa nell'eliminazione di entrambi gli alleli *ADE2*. L'analisi delle mutazioni nelle regioni bersaglio del gRNA di entrambi i geni *ADE2* e *DNL4* ha mostrato riarrangiamenti interallelici tra i due loci, nonché la perdita di estese regioni di cromosomi.

Nel complesso, questo lavoro contribuisce alla vasta gamma di studi che stanno facendo luce sulla biodiversità dei lieviti, sia come modo per comprenderne il potenziale naturale, sia come strumento per personalizzare nuove cell factory.

.

Contributions to the studies

Chapter1: SB, PJ, PB and JM contributed to the idea of the review. PJ and SB were involved in writing the different parts of the review. PB and JM corrected the manuscript.

Chapter 2: PJ, FM, PB and JM contributed to the experimental design of the work. PB and JM provided inputs for the workflow. PJ was involved in performing the experiments, analyzing the data and writing of the chapter.

Chapter 3: PJ, PB and JM contributed to the experimental design of the work. PJ performed the experiments, the analysis and writing of the chapter.

Chapter 4: PJ, LB, MV, PB and JM were involved in designing the experimental work plan. PJ performed parts of the experimental work. LB performed parts of the experimental work along with MV. PJ and LB were involved in analyzing data and in writing of the manuscript. MV was involved in making of some figures and writing of the materials and methods. PB and JM provide inputs for the manuscript and were involved in the final editing of the manuscript.

Introduction

The concept of circular bioeconomy and biorefineries

In a world jeopardized by climate change, moving towards a green economy is no longer a choice, but an obligation. Environmental concerns and the global economic landscape are forcing society and businesses to seek out renewable, environmentally benign, and economically viable alternatives. Science and technology are delivering new solutions to sustainably meet the basic requirements of the growing population with limited resources. Replacing petrol as fuel source, with renewable energy options like solar and wind are among these solutions. This also contributes to a bigger objective of creating a more sustainable economy that is less reliant on fossil fuels as a source of energy or chemical precursors (Mahesh Kumar, 2012). The creation of an economy based on sustainable feedstock and zero carbon emissions is a major objective.

In this context, the term ‘bioeconomy’, refers to the possibility of reconciling economic expansion with ecologically responsible activities (Staffas et al., 2013). According to The Global Bioeconomy Summit (German Bioeconomy Council, 2018), Bioeconomy has been defined as a transformative process that contributes to the Sustainable Development Goals. Moreover, a circular bioeconomy is a conceptual framework for transforming and managing our land, food, health, and industrial systems by utilizing renewable natural capital, with the objective of attaining sustainable wellbeing in harmony with nature (Palahi, 2022). The principles of circular economy — reuse, repair and recycle — are a fundamental part of the bioeconomy. Through reuse, repair and recycling, the total amount of waste and its impact is reduced. It also saves energy, minimizes pollution of soil, air and water, and thus prevents damage to the environment, climate and biodiversity (Key to sustainable development - Iberdrola, 2022). Moreover, the objective of European research framework initiatives such as 'HORIZON 2020' (https://ec.europa.eu/research/bioeconomy/) aim to promote innovation in research fields related to the growing bioeconomy (European Commission, 2012).

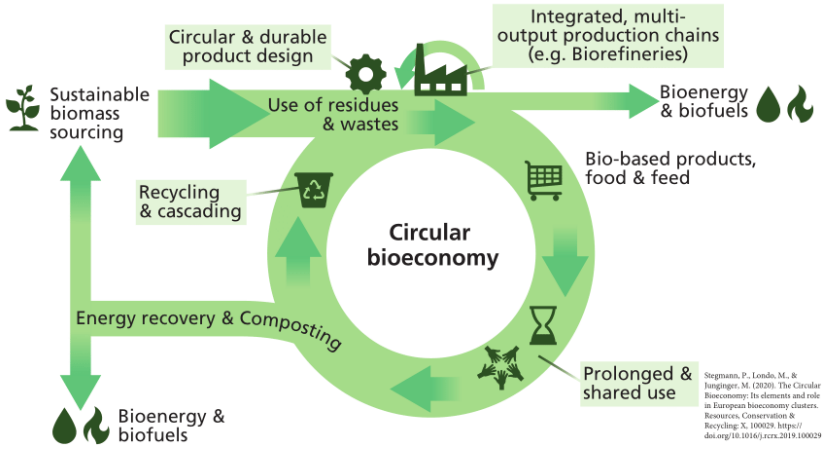


Figure 1: The concept of circular bioeconomy (Stegmann et al., 2020).

Biorefinery serves as a strategic mechanism for the implementation of a circular bioeconomy, given the growing focus on circular economy during the last half-decade, with an emphasis on holistically addressing economic, environmental, and social elements of the industrial sector (Ubando et al., 2020). As summarized in Figure 1, biorefineries are viewed as viable alternatives to petroleum-based refineries as they optimize the sustainable production of food, feed, materials, chemicals, fuels, power and heat from renewable biomass (Clark and Deswarte, 2015). A biorefinery utilizes all kinds of biomass from forestry, agriculture, aquaculture, and residues from industry and household including wood, agricultural crops, organic and forest residues and aquatic biomass to produce a spectrum of marketable products. Among the diverse renewable feed stocks, lignocellulosic biomass (LCB) is the most appealing and exploited owing to its huge abundance and availability. Moreover, LCB relates to second-generation biomass and represents an alternative to first-generation biomass feed stocks that compete with food crops for land (Naik et al., 2010).

The major constituents of LCB are cellulose (30-50%), hemicellulose (20-35%) and lignin (10-20%) and their relative concentrations depends largely on the plant material used (Figure 2a) (Haghighi Mood et al., 2013). Cellulose, the most abundant LCB polymer, is composed of ß-D-glucose units linked by ß-(1,4) glycosidic bonds, with cellobiose as the fundamental repeating dimeric unit. Unlike cellulose, hemicellulose has a random and amorphous structure, which is composed of several heteropolymers including xylan, galactomannan, glucuronoxylan, arabinoxylan, glucomannan and xyloglucan. The heteropolymers of hemicellulose are composed of different 5 and 6

carbon monosaccharide units: pentoses (xylose, arabinose), hexoses (mannose, glucose, and galactose) and acetylated sugars. Hemicelluloses are embedded in the plant cell walls to form a complex network of bonds that provide structural strength by linking cellulose fibers to microfibrils and cross-linking with lignin. Finally, lignin is a three-dimensional polymer of phenylpropanoid units. It functions as the cellular glue which provides compressive strength to the plant tissue and its individual fibers, stiffness to the cell wall and resistance against insects and pathogens. Its high polyphenolic content makes it the ideal natural source of aromatic compounds and (poly-)phenols (Ma et al., 2019).

These three polymers, cellulose, hemicellulose and lignin, are interlinked, resulting in a complex and rigid three-dimensional structure that hinders its industrial utilization (Zoghlami and Paës, 2019). The recalcitrant nature of lignocellulosic materials necessitates a pretreatment process followed by enzymatic hydrolysis to release monomeric sugars. The main goals of pretreatment include: i) producing highly digestible solids that increase sugar yields during enzyme hydrolysis, ii) avoiding the degradation of sugars including those derived from hemicellulose (primarily pentoses), iii) minimizing the formation of inhibitors for subsequent fermentation steps, iv) recovering lignin for conversion into valuable by-products, and v) being cost effective in operation conditions (Jönsson and Martín, 2016). The pretreatment step is one of the most important steps in the entire process as both the efficiency of the conversion and downstream processing of the LCB depend on it (Bertacchi et al., 2022). As illustrated in Figure 2b, the pretreatment process destroys the lignin barrier and partially converts cellulose and hemicelluloses into fermentable sugars such as glucose, xylose, arabinose galactose, and mannose. A subsequent enzymatic hydrolysis process transforms any remaining cellulose and hemicellulose into fermentable sugars. Depending on the type of pretreatment and process parameters, different inhibitory compounds may be released such as: phenolic compounds, furfural, hydroxymethylfurfural (HMF) and different organic acids such as, acetic, levulinic, formic, ferulic, and glucuronic (Chandel et al., 2018; Jönsson and Martín, 2016; Palmqvist and Hahn-Hägerdal, 2000) as outlined in Figure 2. Acetic acid is the most common and abundant weak acid released, and it derives mostly from acetyl groups on hemicellulose polymers (Almeida et al., 2007; Jönsson and Martín, 2016). The

amount and type of inhibitory compounds in the hydrolysates is also dependent on the lignocellulosic source (Ko et al., 2015).

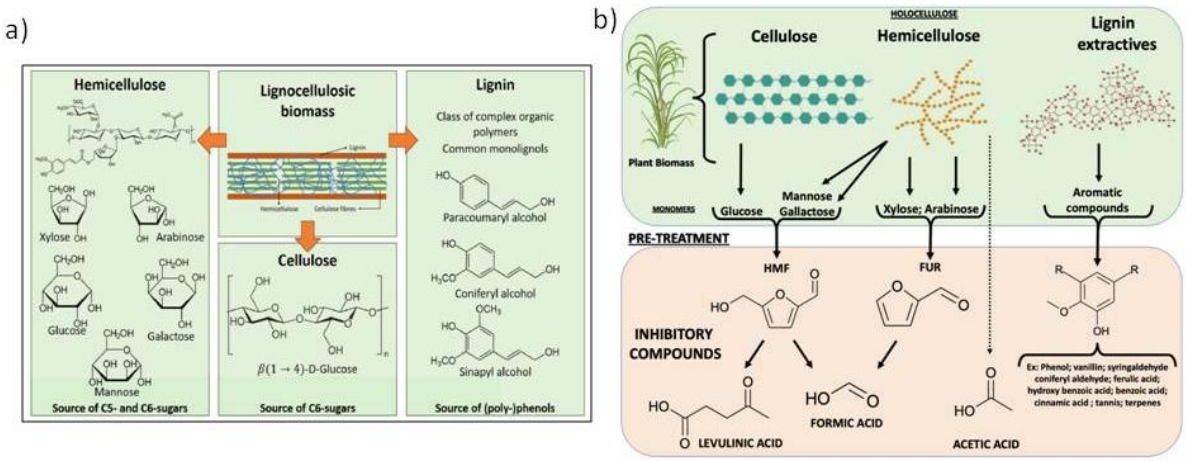


Figure 2: Lignocellulosic biomass composition (a) (Isikgor and Becer, 2015) and structure and types of inhibitory by-products formed after pretreatment (b) (Tramontina et al., 2020).

The biorefinery concept encompasses a wide range of technologies from biomass pretreatment through final delivery. These technologies can be grouped into thermochemical, mechanical, and biochemical processes. One of the fundamental instruments in biochemical processes is to exploit microorganisms, the so-called microbial cell factories, whose role is to convert the biomass provided into desired end products (Dahiya et al., 2018) or into intermediate platforms that can be further converted to products. Fletcher and coworkers (2016) summarized the key characteristics of an ideal microbial cell factory as, 1) easy to engineer, 2) tolerant to industrial stresses, 3) capable of expressing complex heterologous metabolic pathways and 4) efficient producer (high metrics on titer, yield and productivity) (Fletcher et al., 2016). Moreover, an ideal cell factory should be non-pathogenic, genetically stable and must offer desirable cultivation characteristics such as the ease of cultivation and wide range of substrate utilization (Navarrete and L. Martínez, 2020). The microbial cell factory should be able to withstand the inhibitory compounds produced during the pretreatment step and should possess tolerance to compounds such as organic acids or solvents, when these are the desired end products. Organic acids act as a double edge sword in bio-based processes since they can be released from the pretreated biomass, inhibiting the fermentative process but, at the same time, they can be desirable products obtained by microbial fermentation as valuable building block

chemicals with many applications in different industrial sectors (Heer and Sauer, 2008). Consequently, the wider array of microorganisms available makes it possible to have microbial cell factories ready for use in these diverse conditions. The diverse nature of the biomass feedstock, together with the biodiversity of microorganisms, has the potential to lead to the production of a diverse array of molecules of industrial interest (Branduardi, 2021).

Bioprospecting for microbial cell factories

Microorganisms have colonized nearly every niche on this planet. Their survival and replication depend on their ability to produce appropriate responses to the environmental conditions in which they live (Lund et al., 2020). As a result of the physiological diversity of microorganisms, isolated from very diverse environments, they have a remarkable ability to use a wide range of substrates and produce an array of different products (Pham et al., 2019). Bioprospecting involves screening microorganisms from diverse sources for novel and functional characteristics that might be relevant for the production of various bio based products (Adegboye et al., 2021).

Thousands of microorganisms have been catalogued, despite the fact that the vast majority are yet to be identified. Due to this accessible microbial portfolio, extensive bioprospecting may be used to identify and select microbial cell factories for use in specific bioprocesses (Alperstein et al., 2020). In this regard, yeasts and particularly *Saccharomyces cerevisiae* has been the chosen microorganism in many industrial processes, not only because they have a long history of application within the brewing and bread industry, but also because of their natural resilience to low pH and ease of manipulation in haploid and diploid forms. This yeast has been exploited for its natural and engineered features for obtaining products such as bioethanol, organic acids, carotenoids, vitamins and pharmaceuticals (Acevedo-Rocha et al., 2019; Alperstein et al., 2020; Li and Borodina, 2015). Nonetheless, in the emerging bioeconomy, microbial cell factories must efficiently use more sustainable, cheaper, and widely available carbon sources, particularly lignocellulose, and *S. cerevisiae* cannot directly use xylose (the second most abundant sugar in lignocellulosic biomass after glucose) and arabinose (Cai et al., 2012; Zhou et al., 2012). Also, *S. cerevisiae* is an excellent producer of alcohols, esters, and organic acids, but it does not naturally accumulate large amounts of intracellular lipids or consume other residual carbon sources such as short chain fatty acids, limiting its use as a cell factory especially in the case of LCB as feedstock (Zhang et al., 2021).

During bio-based industrial processes, microorganisms are also subjected to different kinds of stresses associated with industrial process conditions, such as non-optimal temperature, inhomogeneous mixing, unfavorable pH, high osmotic pressure, variable

oxygenation levels, presence of inhibitory compounds and final product toxicity (Wehrs et al., 2019). These stress factors are known to inhibit cellular metabolism and compromise the performances of the fermentative process. As a result, either extensive *S. cerevisiae* strain development is required, or the biotech sector must look for other microbes that are inherently more adapted to the tasks at hand. The use of yeasts, other than *S. cerevisiae,* that have natural traits such as utilization of different carbon sources and remarkable stress tolerance are being explored to be used as novel cell factories. These yeast species, generally known as non-conventional yeasts (NCY), are garnering increased attention as new potential hosts for biotechnological applications (Navarrete and L. Martínez, 2020).

Non-conventional yeasts

NCY collectively refers to a group of yeasts, which are progressively gaining more attention as new potential hosts for biotechnological applications compared to *S. cerevisiase.* NCY are far more complex than *S. cerevisiae* alone (Geijer et al., 2022). Many NCY can grow and ferment in a wide range of conditions, and use a wide range of carbon sources present in renewable feedstock in addition to glucose (Ravn et al., 2021; Shen et al., 2018). Some NCY species are particularly resistant to bioprocess-induced stressors such as low pH, high temperatures, and high osmolarities, as well as inhibitory chemicals generated during the pretreatment of biomass (Deparis et al., 2017; Mukherjee et al., 2017; Navarrete and L. Martínez, 2020; Palma et al., 2017). The use of NCY in biotechnological applications has advanced significantly in the recent decade, as well as the development of methods to increase their efficiency (Geijer et al., 2022). Some of the NCY that have been studied in great detail and are employed as cell factories are *K. marxianus* for the production of bioethanol, aroma compounds and biosurfactants (Karim et al., 2020), *Komagataella phaffii* for recombinant proteins and enzymes production (Heistinger et al., 2020) and *Yarrowia lipolytica* for biosurfactants**,** carotenoids, and lipids production (Ledesma-Amaro and Nicaud, 2016). A number of less well-characterized yeasts, such as *Zygosaccharomyces bailii* for organic acids and both *Candida intermedia* and *Candida boidinii* for sugar alcohols (Geijer et al., 2022; Kuanyshev et al., 2021), are emerging as possible cell factories. *Z. parabailii* and *K. marxianus* are two promising non-conventional yeasts among others as mentioned previously, with traits useful for industrial applications.

K. marxianus is a traditional food microbe that is found naturally in many fermented dairy products including cheese and kefir. It is also used in non-dairy processes, such as natural fermentation of Agave-based alcoholic drinks (e.g. tequila or mezcal), is part of the natural flora of Agaves (Lappe-Oliveras et al., 2008), and contributes to chocolate fermentation (Ho et al., 2014). *K. marxianus* belongs to the Saccharomycotina (budding yeast) subphylum and of the family Saccharomycetaceae which also includes *S. cerevisiae* (Blandin et al., 2000). Furthermore, similar to *S. cerevisiae*, numerous species of *K. marxianus* have been granted with “qualified presumption of safety” (QPS) or “generally regarded as safe” (GRAS) status, making it particularly interesting

from an industrial standpoint. Many studies have been conducted to utilize *K. marxianus* for bioethanol production from lignocellulosic biomass (Goshima et al., 2013) or whey due to its respiro-fermentative metabolism and certain unique features (Guimarães et al., 2010). This yeast has numerous significant characteristics that make it ideal for industrial use. *K. marxianus* has broad specificity for both hexose and pentose sugars deriving from plant sources owing to the presence of various sugar transporters (Donzella et al., 2021; Varela et al., 2019). Apart from this, its thermotolerance (up to 52 ^{o}C), fast growth compared to *S. cerevisiae*, the ability to thrive at pH levels below 3 (Lane and Morrissey, 2010) and its robustness to a wide variety of stresses, such as, salt and cell wall stress (Lane et al., 2011) are some of the interesting traits in industrial processes. Since robustness is an important requirement for a production strain, understanding the basis of a strain's response to stress is fundamental to improve it. Moreover, development of molecular tools and synthetic biology approaches for rational engineering has facilitated the use of this yeast in novel applications (Rajkumar et al., 2019; Rajkumar and Morrissey, 2022). Recently, *K. marxianus* was engineered to produce aromatic products, such as 2-phenylethanol (2-PE), through a series of metabolic engineering strategies, including pathway modification and overexpression to be resistant to feedback inhibition, for overproducing phenylalanine de novo from synthetic minimal medium (Rajkumar and Morrissey, 2020). In the genome editing of *K. marxianus*, the CRISPR-Cas9 system has been widely tested. For example, (Löbs et al., 2017a) established a new RNA-polymerase III promoter for single guide RNA (sgRNA) production in *K. marxianus*, allowing for the disruption of genes for the study of ethanol biosynthesis. The mating-type switch in *K. marxianus* has been inactivated using genome editing to accomplish targeted gene disruption (Cernak et al., 2018). Juergens et al. (2018) constructed a Cas9 plasmid using ribozymes for gRNA production, with a gene disruption efficacy of up to 100%. Moreover, in order to increase *K. marxianus* HDR efficiency, *KU80* (Choo et al., 2014), *KU70* and *DNL4* (Rajkumar et al., 2019) which encode integral parts of the NHEJ mediated DSB repair mechanism, were deleted in this yeast, disabling the NHEJ mechanism and increasing the targeted integration efficiency. However, owing to the genetic diversity of *K. marxianus* strains and the limited research on its physiology, biochemistry, and genetics, we are just at the beginning of fully harnessing its

potential as a microbial cell factory (Fonseca et al., 2008; Lane et al., 2011; Rocha et al., 2011). Additionally, the lack of tolerance of this yeast to inhibitory compounds, particularly weak organic acid produced during pretreatment hinders its use when hydrolyzed LCB is used as substrates.

Another NCY *Zygosaccharomyces bailii*, belonging to the Saccharomycotina subphylum also belongs to the Saccharomycotina (budding yeast) subphylum and of the family Saccharomycetaceae. This yeast is well-known for being one of the most aggressive food spoilage yeasts, thriving in acidic environments (Stratford et al., 2013). In addition to remarkable osmotic and ethanol tolerance, *Z. bailii* exhibits an exceptional resistance to weak organic acids (WOA), including acetic, lactic, sorbic, benzoic, and propionic acids (Martorell et al., 2007). Due to these characteristics, including fermentative action in aerobic and anaerobic settings, *Z. bailii* is a potential option for microbial cell factories. Also, this yeast has been exploited for the production of L-ascorbic acid (Branduardi et al., 2004; Sauer et al., 2004), heterologous proteins (Vigentini et al., 2005), lactic acid (Dato et al., 2010) bioethanol (Paixão et al., 2013) and flavor compounds in liquor fermentation (Xu et al., 2017). The phylogenetic re-evaluation of many industrial isolates formerly identified as *Z. bailii* along with substantial changes in rRNA gene sequences led to the discovery of two novel *Zygosaccharomyces* species, *Z. parabailii* and *Z. pseudobailii*, which are closely related to *Z. bailii* (Suh et al., 2013). The genome assembly and annotation of the hybrid yeast *Z. parabailii* achieved in a work by Ortiz and co-workers (2017) suggested that *Z. bailii sensu lato* is highly divergent and has the ability to exploit comparable potential as S. cerevisiae industrial hybrids. (Krogerus et al., 2017; Steensels et al., 2014). The interspecies hybrid *Z. parabailii* is formed by mating between *Z. bailii* and an unidentified *Zygosaccharomyces* species that differs by 7-10% in its nucleotide sequence (Ortiz-Merino et al., 2017). *Z. parabailii* displays loss of heterozygosity at the *MAT* locus, causing them to behave as haploid cells capable of mating-type switching. Only the *MAT* locus originating from the *Z. bailii* parent is active in *Z. parabailii*, and it is thought that the second *MAT* locus was destroyed during futile attempts to switch mating types (Ortiz-Merino et al. 2017).

Additionally, previous work from our group suggested that *Z. parabailii* ATCC60483 is more robust than *Z. bailii* haploid strains. *Z. parabailii* is known to be more resistant to

weak organic acids particularly lactic acid which is one of the important commodity chemicals produced by microbial fermentation (Kuanyshev et al., 2016). One of the main drawbacks in exploiting this hybrid yeast for industrial purposes is the challenging genetic manipulation and scarcity of efficacious molecular tools. Recently in a work by Kuanyshev and co- workers, CRISPR-Cas9 system was developed for construction of a *Z. bailii* strain for lactic acid production. However, this method was developed for the haploid *Z. bailii* but no methods are yet available for the hybrid *Z. parabailii* which carries two copies of most genes. Moreover, in order to further domesticate the robust hybrid yeast *Z. parabailii,* it is also necessary to understand the mechanisms underlying its peculiar traits, particularly tolerance to weak organic acids and to develop molecular tools for efficient gene editing which forms the basis of this work.

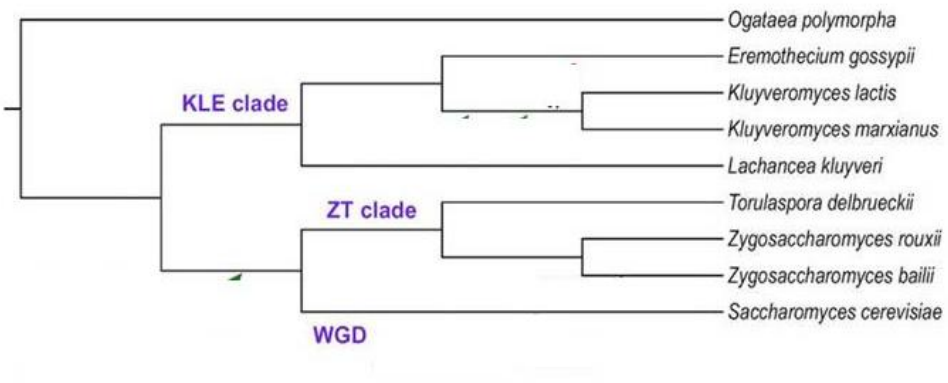


Figure 3: The phylogeny of selected species in the family Saccharomycetaceae. WGD refers to whole genome duplication on how *S. cerevisiae* and its close relatives evolved. ZT clade refers to *Zygosaccharomyces/Torulaspora* sp and KLE clade refers to *Kluyveromyces /Lachancea/Eremothecium* sp (Solieri et al., 2021).

Comparing the whole genomes of multiple closely related species, in the context of well-resolved phylogenetic trees, can reveal genomic events and evolutionary processes that result in functional variations across species (Capra et al., 2010). Recent advancements, based on additional genome sequences, has offered new perspectives that the whole genome duplication (WGD) that occurred ~100 million years ago was a hybridization event between a strain from the *Zygosaccharomyces/Torulaspora* (ZT) clade and a strain from *Kluyveromyces/Lachancea*/*Eremothecium* (KLE) clade (Figure 3) (Marcet-Houben and Gabaldón, 2015; Wolfe, 2015). Gene duplication has been

recognized as an important source of novel genes and genetic variation (Capra et al., 2010). The further away from *S. cerevisiae* in the yeast evolutionary tree, species such as *Zygosaccharomyces bailii* and *Kluyveromyces marxianus* with peculiar traits that can be useful for bio based industrial processes can be identified.

Engineering non-conventional yeasts for bio-based processes

The primary barriers in the development of synthetic biology methods for NCY lies in the lack of understanding their vast diversity of metabolic pathways and production capability, as well as the restricted -omics (genomics, proteomics, transcriptomic, and metabolomics) studies (Geijer et al., 2022). A number of NCY such as *Y. lipolytica* (Kerkhoven et al., 2016), *K. phaffii* (Cankorur-Cetinkaya et al., 2017) *O. polymorpha* (Liebal et al., 2021), *K. lactis* (Dias et al., 2014), *K. marxianus* (Marcišauskas et al., 2019), and *R. toruloides* (Dinh et al., 2019) have curated genome-scale metabolic models that can be integrated with –omics results to guide strain design strategies.

In order to fully exploit the advantages of yeasts in bio-based processes as summarized in figure 4, the following challenges must be taken into consideration: 1) In bio-based industrial production, considerable characterisation and testing of the microorganism(s) of interest is required before novel strains and species are used. 2) When lignocellulosic biomass is the substrate for yeasts, a thorough understanding of the substrate composition and how it interacts with the microorganism of choice is essential. 3) To improve production titers, yields and productivity, process improvement, including optimization of NCY culture parameters and modes, is frequently required. 4) Strain development by targeted and non-targeted engineering methods, is frequently required to enhance the intriguing features (Geijer et al., 2022).

Although the use of synthetic biology techniques has widely been employed to understand the robustness of *K. marxianus* and for the production of various chemicals as mentioned previously, the mechanisms related to organic acid tolerance are yet to be deciphered. To overcome limitations of ‘rational’ metabolic engineering, a ‘bottom-up’ approach called ‘inverse metabolic engineering’ was defined by Bailey and coworkers (Bailey et al., 2002) as an alternative strategy. In particular, continuous evolution procedures based on the application of a selection pressure to obtain a desired phenotype are referred to as ‘evolutionary engineering’ or Adaptive laboratory evolution (ALE).

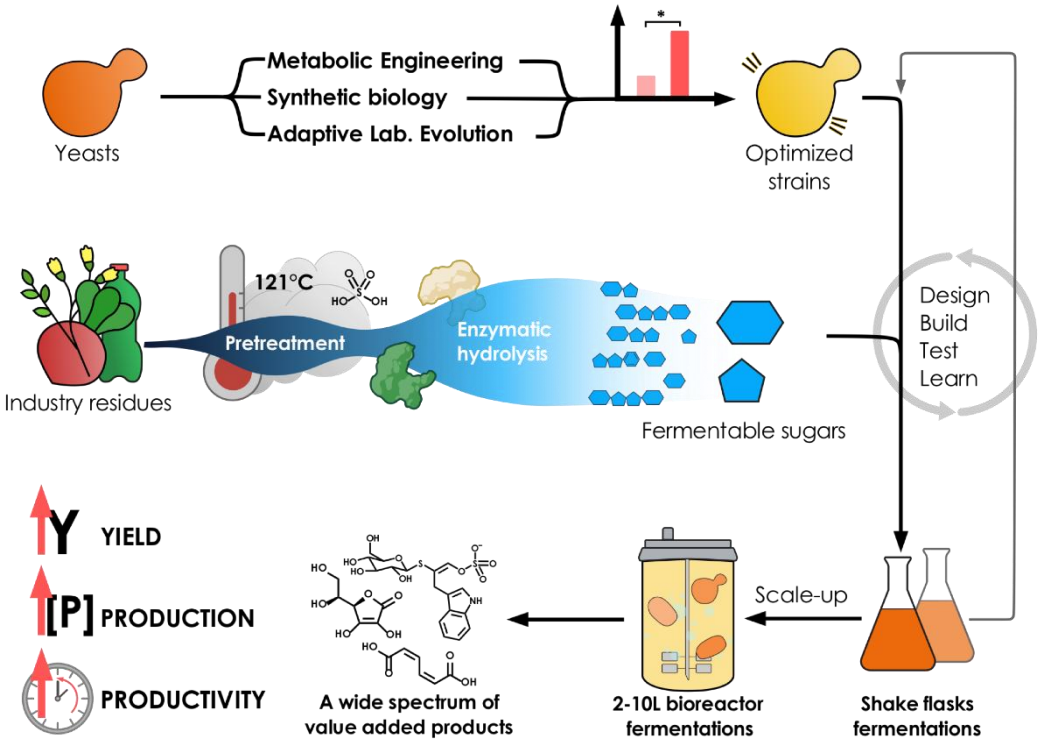


Figure 4: Schematic representation of bio-based microbial processes, involving the tailoring microbial cell factories for the production of value added products by valorizing industrial residues, from https://indbiotechlab.btbs.unimib.it/

ALE, unlike comparative genomics, allows phenotypic changes to be unambiguously linked to a specific growing environment, which leads to trait selection. Furthermore, phenotype-genotype correlations may be easily established by whole genome re-sequencing, thanks to relatively recent technologies such as extensive next-generation DNA sequencing. ALE can be performed in: 1) repeated batch cultivations performed in the presence of a selective pressure or alternatively, 2) prolonged chemostat cultivations performed under selective conditions. Typical ALE experiments are performed for 100 to 2000 generations and usually take a few weeks up to a few months. Earlier studies predicted that a fitness gain of 50–100% may be reached in 100 to 500 generations (equivalent to up to 2 months of selection for a normal *E. coli* or *S. cerevisiae* culture). ALE approach can be applied to traits that are arbitrarily complex, to organisms that are not amenable to genetic engineering or again in the case in which it would be not possible to use GMO for the desired application (i.e food and drink). In ALE, use of chemical mutagens, mutants lacking DNA repair capabilities and transposon libraries can broaden the range of potential selection targets and speed up the selection process. Moreover, selecting for better phenotypes using random

mutagenesis is also a highly common method. Mutations are the basis underlying the genetic change and the selection of improved phenotypes both in nature as well in the laboratory. During ALE, several mutations occur such as SNPs (61%) accounting for the majority of observed genetic changes followed by deletions (29%), insertions (7%) and insertion sequence movements (3%) (Conrad et al., 2011). These genetic changes give rise to new strains with novel phenotypes. During ALE, certain trait values change and are related with higher (Darwinian) fitness, therefore an improved phenotype or feature is frequently equal to greater fitness. Competitive fitness tests typically require batch culture growth and are balanced for all growth stages (lag, exponential, and stationary). Parameters such as μ_{max}, survival rates in toxic concentrations of certain chemical compounds and absolute biomass yield are appropriate fitness criteria. During ALE, several phenotypes occur in the beginning and compete for 'dominance' in the overall population, which can contribute to large population heterogeneity (Dragosits and Mattanovich, 2013; Mans et al., 2018). As a result, it is not possible to obtain a homogeneous population at any moment during a laboratory evolution experiment. It should also be noted that selection for improved fitness in a specialized environment often leads to significant trade-offs in other stressful or selective conditions (Çakar et al., 2005). In this context, the ideal phenotype for biotechnological goals is the one that demonstrates better performance and least trade-offs in different environmental situations, rather than the one with the maximum fitness in a specific state. Sub-culturing the better clones in non-inhibitory media and reassessing the superiority of the phenotypes should be done to ensure the stability of the evolved strains. Following confirmation of the superior phenotype, the next step is to identify the genetic change responsible for the enhanced phenotype at the molecular phenotypic level. To summarize, ALE is an effective method for enhancing industrially significant and genetically complicated features of *S. cerevisiae* and other microorganisms.

Contrary to *K. marxianus*, the hybrid yeast *Z. parabailii* possess tolerance to organic acids at low pH and this characteristic can be exploited for use of this yeast in lignocellulosic biomass for production of value added compounds such as organic acids. However, development of synthetic biology and molecular tools for efficient gene editing in this hybrid yeast is limited. In order to disrupt a gene function in this

hybrid yeast, both the alleles of the gene differing should be simultaneously nonfunctional. The use of the classical method with deletion cassettes containing long homology arms has been successful in obtaining *leu2* mutants in *Z. parabailii* (Dato et al., 2010). This method of deletion was also used in this work for *PDR12* gene disruptions to characterize the role in organic acid tolerance trait. Classical genome editing techniques have found widespread use in functional genomics and biotechnology. Furthermore, advances in metabolic engineering and synthetic biology have increased the need for creating techniques such as CRISPR-Cas9, for faster and more efficient introduction of multiple genome modifications (David and Siewers, 2015). Efficient gene editing techniques is essential to expand the industrial application of *Z. parabailii*. CRISPR-Cas9 has emerged as a powerful and versatile tool for the precise editing of genomes in various species. The CRISPR-Cas9 system has been successfully applied in several NCY such as *Scheffersomyces stipitis, Komagataella phaffi, Ogataea polymorpha, Kluyveromyces lactis, Kluyveromyces marxianus* and *Yarrowia lipolytica* (Cai et al., 2019). The type II CRISPR-Cas9 system comprises of two main parts: short chimeric guide RNA (gRNA), formed by the fusion of CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA), which together forms a secondary structure loop to recruit Cas9 that undergoes conformational changes leading to double stranded break (DSB) in the DNA (Jinek et al., 2012). One of the key steps in successful engineering using CRISPR-Cas9 includes design, expression and delivery of the gRNA. In different yeast systems, this is achieved using either an RNA polymerase III (Pol III) promoter that provides a leader sequence to be cleaved after molecular maturation or a polymerase II (Pol II) promoter flanked with ribozyme cleaving systems, or even tRNAs. The two major pathways for repair of DNA double-strand breaks (DSBs) are homologous recombination (HR) and non-homologous end joining (NHEJ) (Liu et al., 2017). In *S. cerevisiae*, following the DSB by CRISPR-Cas9, HR is the most common DNA repair mechanism harbored. In most other yeasts, however, NHEJ is the preferred DNA repair process, and HR genome engineering is ineffective. As a result of NHEJ, engineering of non-conventional yeasts is frequently accomplished by random integration. The random integration of the altered expression cassette might cause open reading frames or other genomic elements to be disrupted (Löbs et al., 2017b). Hence, establishing an efficient HR in these yeasts is important and this

was also one of the objectives achieved in our work on *Z. parabailii* in chapter 4. The only hybrid yeast successfully engineered using CRISPR-Cas9 is *Saccharomyces pastorianus* (de Vries et al., 2017). In this work we established a CRISPR-Cas9 system in the hybrid *Z. parabailii* for simultaneous *indel* of two alleles of a gene.

Scope and Outline of this work

The aim of this work, as summarized in Figure 5, was to improve the robustness and expand the industrial potential of two non-conventional yeasts, *Z. parabailii* and *K. marxianus*, by applying direct and indirect strain engineering approaches.

In **chapter 1** of this work titled, "Interdependence between lignocellulosic biomasses, enzymatic hydrolysis and yeast cell factories", we reviewed the link between LCBs composition, choice of enzymatic cocktail and selection of yeast species and strains that need to be considered in an integrated fashion to enable the development of an efficient bio-based process. We discussed the pivotal role of enzymatic cocktail optimization to unlock the potential of non-Conventional yeasts, which, thanks to broader substrate utilization, inhibitor resistance and peculiar metabolism, can widen the array of feedstock and products of biorefineries.

In **chapter 2** of this work "Adaptive Laboratory Evolution to enhance *Kluyveromyces marxianus* tolerance to sugar beet pulp hydrolysates", we used an ALE approach to improve the tolerance of *K. marxianus* to SBP hydrolysate at low pH and at two different temperatures, 30^{o}C and 40^{o}C, the second to favour the maintenance of the desired thermotolerance of this yeast. Preliminary growth analysis in different medium such as YPD, Verduyn+2% glucose and sugar beet pulp hydrolysate at 30 °C and 40 °C suggested a lack of growth of this yeast in SBP pH 3.0 in both the temperatures. Therefore, ALE was designed with the objective of improving the lag phase and specific growth rate of the *K. marxianus* parental strain in SBP at pH 3.0 at 30 ^{o}C and 40 ^{o}C. For screening the best performing evolved isolates, growth kinetics were performed in SBP pH 3.0 and lag phase and specific growth rate of isolates from different days of evolution were compared. Taking advantage of the growth of this yeast in SBP pH 4.3, we investigated the positive effect of pre-culturing the cells in SBP at pH 4.3 prior to growth in SBP pH 3.0. However, in industrial productions, a stable robust yeast strain that does not require pre-culturing would be preferable. Further, molecular characterization of pre-cultured parental strain indicated the role of genes involved in oxidative stress tolerance.

Unlike *K. marxianus*, the hybrid yeast *Z. parabailii* is tolerant to weak organic acids at low pH, but the mechanism underlying this unique trait is yet to be unraveled. In

chapter 3 of this work "Contribution of *PDR12* to weak organic acids resistance in the yeast *Zygosaccharomyces parabailii",* the focus was to investigate the role of *PDR12* to if this gene was involved in weak organic acid resistance trait of *Z. parabailii*. For this reason, we investigated the role of *PDR12* gene in the tolerance towards different weak organic acids, such as acetic, butyric, lactic, and sorbic and benzoic acids. In this work, we employed direct engineering method for construction and characterisation of a set of single and double *Z. parabailii pdr12* mutants. Furthermore, we also looked into the possible divergence in the role of the two alleles of *PDR12* towards tolerance to organic acids through mutagenesis and phenotypic characterisation. Understanding the mechanism underlying the tolerance can further be used for developing more industrially relevant strains. Moreover, gaining knowledge over *Z. bailii* tolerance to organic acids can help in developing preventative measures against food spoilage for which *Z. bailii* is responsible.

In **chapter 4** of this work "Establishment of a novel CRISPR-Cas9 system in the hybrid yeast *Zygosaccharomyces parabailii* reveals allele exchange mechanism", we developed a gene editing system for simultaneous disruption or deletion of two alleles of a gene in *Z. parabailii*. We evaluated the use of four different gRNA systems consisting of combinations of tRNAs, tRNA and ribozyme or ribozymes as self-cleaving flanking. The functionality of the gRNA systems was tested by analysing the inactivation of the *ADE2* gene in the wild type *Z. parabailii* and the most efficient gRNA system was used to successfully construct a *Z. parabailii DNL4* mutant. This mutant exhibited improved homologous recombination (HR) in the deletion of both *ADE2* alleles compared to the lack of double deletion by HR observed in the wild type strain. Moreover, analysis of mutations in the gRNA target regions of both *ADE2* and *DNL4* genes suggested inter-allelic rearrangements between the two gene loci, as well as absence of large regions of chromosomes. The allelic exchange mechanism seen in the *Z. parabailii* WT strain as well as in the *dnl4* mutant gives a starting point for assessing inter allelic rearrangements and to understand if one DNA repair fragment can suffice the deletion of two alleles of a gene. Since *Z. parabailii* is resilient to genetic modifications, these advances greatly increase its genetic accessibility thereby facilitating future research into its complex hybrid genome. The findings and methods

described in this manuscript enable avenues of research both in genome engineering and exploration of *Z. parabailii* for industrial uses.

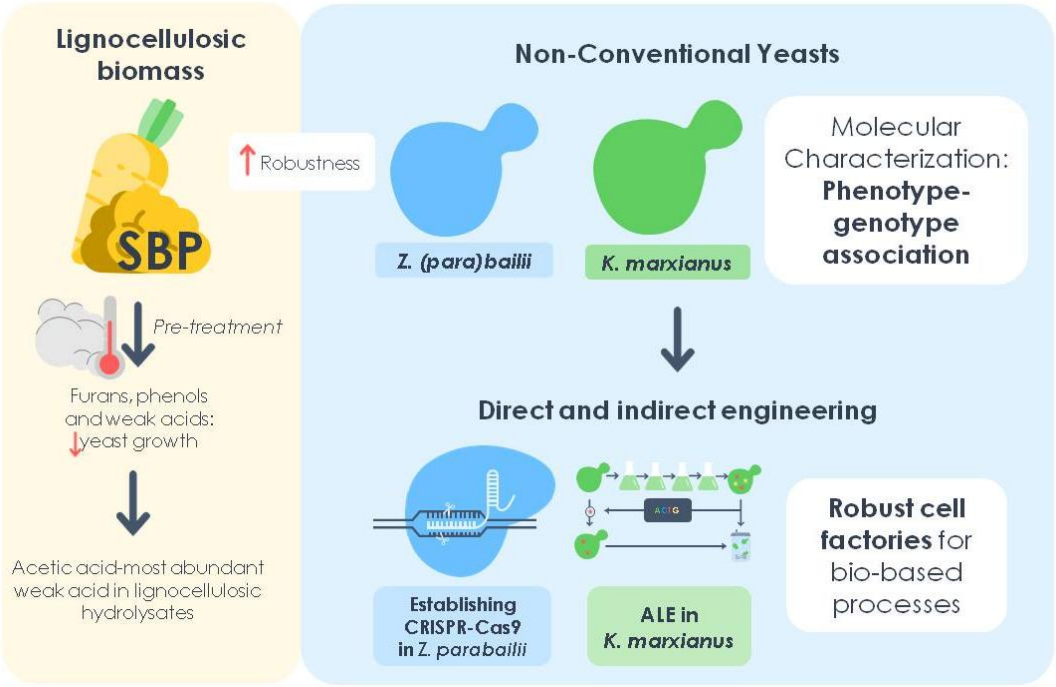


Figure 5: Graphical abstract outlining the aim of this work.

References

Acevedo-Rocha, C.G., Gronenberg, L.S., Mack, M., Commichau, F.M., Genee, H.J., 2019. Microbial cell factories for the sustainable manufacturing of B vitamins. Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2018.07.006

Adegboye, M.F., Ojuederie, O.B., Talia, P.M., Babalola, O.O., 2021. Bioprospecting of microbial strains for biofuel production: metabolic engineering, applications, and challenges. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-020-01853-2

Almeida, J.R.M., Modig, T., Petersson, A., Hähn-Hägerdal, B., Lidén, G., Gorwa-Grauslund, M.F., 2007. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. J. Chem. Technol. Biotechnol. https://doi.org/10.1002/jctb.1676

Alperstein, L., Gardner, J.M., Sundstrom, J.F., Sumby, K.M., Jiranek, V., 2020. Yeast bioprospecting versus synthetic biology—which is better for innovative beverage fermentation? Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-020-10364

Bailey, J.E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W.A., Tsai, P.S., 2002. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. Biotechnol. Bioeng. https://doi.org/10.1002/bit.10441

Baruah, J., Nath, B.K., Sharma, R., Kumar, S., Deka, R.C., Baruah, D.C., Kalita, E., 2018. Recent trends in the pretreatment of lignocellulosic biomass for value-added products. Front. Energy Res. https://doi.org/10.3389/fenrg.2018.00141

Bertacchi, S., Jayaprakash, P., Morrissey, J.P., Branduardi, P., 2022. Interdependence between lignocellulosic biomasses, enzymatic hydrolysis and yeast cell factories in biorefineries. Microb. Biotechnol. https://doi.org/10.1111/1751-7915.13886

Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F., Dujon, B., 2000. Genomic exploration of the Hemiascomycetous yeasts: 13. *Pichia angusta*. FEBS Lett. https://doi.org/10.1016/S0014-5793(00)02284-5

Branduardi, P., 2021. Closing the loop: the power of microbial biotransformations from

traditional bioprocesses to biorefineries, and beyond. Microb. Biotechnol. https://doi.org/10.1111/1751-7915.13713

Branduardi, P., Valli, M., Brambilla, L., Sauer, M., Alberghina, L., Porro, D., 2004. The yeast *Zygosaccharomyces bailii*: A new host for heterologous protein production, secretion and for metabolic engineering applications. FEMS Yeast Res. https://doi.org/10.1016/S1567-1356(03)00200-9

Cai, P., Gao, J., Zhou, Y., 2019. CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications. Microb. Cell Fact. https://doi.org/10.1186/s12934-019-1112-2

Cai, Z., Zhang, B., Li, Y., 2012. Engineering *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: Reflections and perspectives. Biotechnol. J. https://doi.org/10.1002/biot.201100053

Çakar, Z.P., Seker, U.O.S., Tamerler, C., Sonderegger, M., Sauer, U., 2005. Evolutionary engineering of multiple-stress resistant *Saccharomyces cerevisiae*. FEMS Yeast Res. https://doi.org/10.1016/j.femsyr.2004.10.010

Cankorur-Cetinkaya, A., Dikicioglu, D., Oliver, S.G., 2017. Metabolic modeling to identify engineering targets for *Komagataella phaffii*: The effect of biomass composition on gene target identification. Biotechnol. Bioeng. https://doi.org/10.1002/bit.26380

Capra, J.A., Pollard, K.S., Singh, M., 2010. Novel genes exhibit distinct patterns of function acquisition and network integration. Genome Biol. https://doi.org/10.1186/gb-2010-11-12-r127

Cernak, P., Estrela, R., Poddar, S., 2018. Engineering *Kluyveromyces marxianus* as a Robust Synthetic. Am. Soc. Microbiol. https://doi.org/10.1128/mBio.01410-18

Chandel, A.K., Garlapati, V.K., Singh, A.K., Antunes, F.A.F., da Silva, S.S., 2018. The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2018.06.004

Cherubini, F., 2010. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2010.01.015

Choo, J.H., Han, C., Kim, J.Y., Kang, H.A., 2014. Deletion of a KU80 homolog enhances homologous recombination in the thermotolerant yeast *Kluyveromyces marxianus*. Biotechnol. Lett. https://doi.org/10.1007/s10529-014-1576-4

Clark, J., Deswarte, F., 2015. Introduction to Chemicals from Biomass Wiley Series in Renewable Resources.

Conrad, T.M., Lewis, N.E., Palsson, B.O., 2011. Microbial laboratory evolution in the era of genome-scale science. Mol. Syst. Biol. https://doi.org/10.1038/msb.2011.42

Dahiya, S., Kumar, A.N., Shanthi Sravan, J., Chatterjee, S., Sarkar, O., Mohan, S.V., 2018. Food waste biorefinery: Sustainable strategy for circular bioeconomy. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2017.07.176

Dato, L., Branduardi, P., Passolunghi, S., Cattaneo, D., Riboldi, L., Frascotti, G., Valli, M., Porro, D., 2010. Advances in molecular tools for the use of *Zygosaccharomyces bailii* as host for biotechnological productions and construction of the first auxotrophic mutant. FEMS Yeast Res. https://doi.org/10.1111/j.1567-1364.2010.00668.x

David, F., Siewers, V., 2015. Advances in yeast genome engineering. FEMS Yeast Res. https://doi.org/10.1111/1567-1364.12200

de Vries, A.R.G., de Groot, P.A., van den Broek, M., Daran, J.M.G., 2017. CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus*. Microb. Cell Fact. https://doi.org/10.1186/s12934-017-0835-1

Deparis, Q., Claes, A., Foulquié-Moreno, M.R., Thevelein, J.M., 2017. Engineering tolerance to industrially relevant stress factors in yeast cell factories. FEMS Yeast Res. https://doi.org/10.1093/femsyr/fox036

Dias, O., Pereira, R., Gombert, A.K., Ferreira, E.C., Rocha, I., 2014. iOD907, the first genome-scale metabolic model for the milk yeast *Kluyveromyces lactis*.

Biotechnol. J. https://doi.org/10.1002/biot.201300242

Dinh, H. V., Suthers, P.F., Chan, S.H.J., Shen, Y., Xiao, T., Deewan, A., Jagtap, S.S., Zhao, H., Rao, C. V., Rabinowitz, J.D., Maranas, C.D., 2019. A comprehensive genome-scale model for *Rhodosporidium toruloides* IFO0880 accounting for functional genomics and phenotypic data. Metab. Eng. Commun. https://doi.org/10.1016/j.mec.2019.e00101

Donzella, L., Varela, J.A., Sousa, M.J., Morrissey, J.P., 2021. Identification of novel pentose transporters in *Kluyveromyces marxianus* using a new screening platform. FEMS Yeast Res. https://doi.org/10.1093/femsyr/foab026

Dragosits, M., Mattanovich, D., 2013. Adaptive laboratory evolution - principles and applications for biotechnology. Microb. Cell Fact. https://doi.org/10.1186/1475-2859-12-64

European Commission, 2012. Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions - Innovating for Sustainable Growth: A Bioeconomy for Europe, Official Journal of the European Union.

Fletcher, E., Krivoruchko, A., Nielsen, J., 2016. Industrial systems biology and its impact on synthetic biology of yeast cell factories. Biotechnol. Bioeng. https://doi.org/10.1002/bit.25870

Fonseca, G.G., Heinzle, E., Wittmann, C., Gombert, A.K., 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-008-1458-6

Geijer, C., Ledesma-Amaro, R., Tomas-Pejo, E., 2022. Unraveling the potential of non-conventional yeasts in biotechnology. FEMS Yeast Res. https://doi.org/10.1093/femsyr/foab071

Goshima, T., Tsuji, M., Inoue, H., Yano, S., Hoshino, T., Matsushika, A., 2013. Bioethanol production from lignocellulosic biomass by a novel *Kluyveromyces marxianus* strain. Biosci. Biotechnol. Biochem. https://doi.org/10.1271/bbb.130173

Guimarães, P.M.R., Teixeira, J.A., Domingues, L., 2010. Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. Biotechnol. Adv. https://doi.org/10.1016/j.biotechadv.2010.02.002

Haghighi Mood, S., Hossein Golfeshan, A., Tabatabaei, M., Salehi Jouzani, G., Najafi, G.H., Gholami, M., Ardjmand, M., 2013. Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2013.06.033

Heer, D., Sauer, U., 2008. Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. Microb. Biotechnol. https://doi.org/10.1111/j.1751-7915.2008.00050.x

Heistinger, L., Gasser, B., Mattanovich, D., 2020. Microbe profile: *Komagataella phaffii*: A methanol devouring biotech yeast formerly known as Pichia pastoris. Microbiol. https://doi.org/10.1099/mic.0.000958

Ho, V.T.T., Zhao, J., Fleet, G., 2014. Yeasts are essential for cocoa bean fermentation. Int. J. Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2013.12.014

Isikgor, F.H., Becer, C.R., 2015. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. Polym. Chem. https://doi.org/10.1039/c5py00263j

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. https://doi.org/10.1126/science.1225829

Jönsson, L.J., Martín, C., 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2015.10.009

Karim, A., Gerliani, N., Aïder, M., 2020. *Kluyveromyces marxianus*: An emerging yeast cell factory for applications in food and biotechnology. Int. J. Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2020.108818

Kerkhoven, E.J., Pomraning, K.R., Baker, S.E., Nielsen, J., 2016. Regulation of amino-

acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. npj Syst. Biol. Appl. https://doi.org/10.1038/npjsba.2016.5

Ko, J.K., Um, Y., Park, Y.C., Seo, J.H., Kim, K.H., 2015. Compounds inhibiting the bioconversion of hydrothermally pretreated lignocellulose. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-015-6595-0

Krogerus, K., Magalhães, F., Vidgren, V., Gibson, B., 2017. Novel brewing yeast hybrids: creation and application. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-016-8007-5

Kuanyshev, N., Ami, D., Signori, L., Porro, D., Morrissey, J.P., Branduardi, P., 2016. Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation. FEMS Yeast Res. https://doi.org/10.1093/femsyr/fow058

Kuanyshev, N., Rao, C. V., Dien, B., Jin, Y.S., 2021. Domesticating a food spoilage yeast into an organic acid-tolerant metabolic engineering host: Lactic acid production by engineered *Zygosaccharomyces bailii*. Biotechnol. Bioeng. https://doi.org/10.1002/bit.27576

Lane, M.M., Burke, N., Karreman, R., Wolfe, K.H., O'Byrne, C.P., Morrissey, J.P., 2011. Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. https://doi.org/10.1007/s10482-011-9606-x

Lane, M.M., Morrissey, J.P., 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. Fungal Biol. Rev. https://doi.org/10.1016/j.fbr.2010.01.001

Lappe-Oliveras, P., Moreno-Terrazas, R., Arrizón-Gaviño, J., Herrera-Suárez, T., García-Mendoza, A., Gschaedler-Mathis, A., 2008. Yeasts associated with the production of Mexican alcoholic nondistilled and distilled Agave beverages. FEMS Yeast Res. https://doi.org/10.1111/j.1567-1364.2008.00430.x

Ledesma-Amaro, R., Nicaud, J.M., 2016. *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. Prog. Lipid Res. https://doi.org/10.1016/j.plipres.2015.12.001

Li, M., Borodina, I., 2015. Application of synthetic biology for production of chemicals in yeast *Saccharomyces cerevisiae*. FEMS Yeast Res. https://doi.org/10.1111/1567-1364.12213

Liebal, U.W., Fabry, B.A., Ravikrishnan, A., Schedel, C.V., Schmitz, S., Blank, L.M., Ebert, B.E., 2021. Genome-scale model reconstruction of the methylotrophic yeast *Ogataea polymorpha*. BMC Biotechnol. https://doi.org/10.1186/s12896-021-00675-w

Liu, Z., Liang, Y., Ang, E.L., Zhao, H., 2017. A New Era of Genome Integration - Simply Cut and Paste! ACS Synth. Biol. https://doi.org/10.1021/acssynbio.6b00331

Löbs, A.K., Engel, R., Schwartz, C., Flores, A., Wheeldon, I., 2017a. CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-017-0854-5

Löbs, A.K., Schwartz, C., Wheeldon, I., 2017b. Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. Synth. Syst. Biotechnol. https://doi.org/10.1016/j.synbio.2017.08.002

Lund, P.A., De Biase, D., Liran, O., Scheler, O., Mira, N.P., Cetecioglu, Z., Fernández, E.N., Bover-Cid, S., Hall, R., Sauer, M., O'Byrne, C., 2020. Understanding How Microorganisms Respond to Acid pH Is Central to Their Control and Successful Exploitation. Front. Microbiol. https://doi.org/10.3389/fmicb.2020.556140

Ma, J., Shi, S., Jia, X., Xia, F., Ma, H., Gao, J., Xu, J., 2019. Advances in catalytic conversion of lignocellulose to chemicals and liquid fuels. J. Energy Chem. https://doi.org/10.1016/j.jechem.2019.04.026

Mahesh Kumar, 2012. Social, Economic, and Environmental Impacts of Renewable Energy Resources. Intech 13. https://doi.org/10.5772/intechopen.89484

Mans, R., Daran, J.M.G., Pronk, J.T., 2018. Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2017.10.011

Marcet-Houben, M., Gabaldón, T., 2015. Beyond the whole-genome duplication: Phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. PLoS Biol. https://doi.org/10.1371/journal.pbio.1002220

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

Martorell, P., Stratford, M., Steels, H., Fernández-Espinar, M.T., Querol, A., 2007. Physiological characterization of spoilage strains of *Zygosaccharomyces bailii* and *Zygosaccharomyces rouxii* isolated from high sugar environments. Int. J. Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2006.09.014

Mukherjee, V., Radecka, D., Aerts, G., Verstrepen, K.J., Lievens, B., Thevelein, J.M., 2017. Phenotypic landscape of non-conventional yeast species for different stress tolerance traits desirable in bioethanol fermentation. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-017-0899-5

Naik, S.N., Goud, V. V., Rout, P.K., Dalai, A.K., 2010. Production of first and second generation biofuels: A comprehensive review. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2009.10.003

Navarrete, C., L. Martínez, J., 2020. Non-conventional yeasts as superior production platforms for sustainable fermentation based bio-manufacturing processes. AIMS Bioeng. https://doi.org/10.3934/bioeng.2020024

Ortiz-Merino, R.A., Kuanyshev, N., Braun-Galleani, S., Byrne, K.P., Porro, D., Branduardi, P., Wolfe, K.H., 2017. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. PLoS Biol. https://doi.org/10.1371/journal.pbio.2002128

Paixão, S.M., Teixeira, P.D., Silva, T.P., Teixeira, A. V., Alves, L., 2013. Screening of novel yeast inulinases and further application to bioprocesses. N. Biotechnol. https://doi.org/10.1016/j.nbt.2013.02.002

Palma, M., Münsterkötter, M., Peça, J., Güldener, U., Sá-Correia, I., 2017. Genome sequence of the highly weak-acid-tolerant *Zygosaccharomyces bailii* IST302,

amenable to genetic manipulations and physiological studies. FEMS Yeast Res. https://doi.org/10.1093/femsyr/fox025

Palmqvist, E., Hahn-Hägerdal, B., 2000. Fermentation of lignocellulosic hydrolysates. II: Inhibitors and mechanisms of inhibition. Bioresour. Technol. https://doi.org/10.1016/S0960-8524(99)00161-3

Pham, J. V., Yilma, M.A., Feliz, A., Majid, M.T., Maffetone, N., Walker, J.R., Kim, E., Cho, H.J., Reynolds, J.M., Song, M.C., Park, S.R., Yoon, Y.J., 2019. A review of the microbial production of bioactive natural products and biologics. Front. Microbiol. https://doi.org/10.3389/fmicb.2019.01404

Rajkumar, A.S., Morrissey, J.P., 2022. Protocols for marker-free gene knock-out and knock-down in *Kluyveromyces marxianus* using CRISPR/Cas9. FEMS Yeast Res. https://doi.org/10.1093/femsyr/foab067

Rajkumar, A.S., Varela, J.A., Juergens, H., Daran, J.M.G., Morrissey, J.P., 2019. Biological parts for *Kluyveromyces marxianus* synthetic biology. Front. Bioeng. Biotechnol. https://doi.org/10.3389/fbioe.2019.00097

Ravn, J.L., Engqvist, M.K.M., Larsbrink, J., Geijer, C., 2021. CAZyme prediction in Ascomycetous yeast genomes guides discovery of novel xylanolytic species with diverse capacities for hemicellulose hydrolysis. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-021-01995-x

Rocha, S.N., Abrahão-Neto, J., Gombert, A.K., 2011. Physiological diversity within the *Kluyveromyces marxianus* species. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. https://doi.org/10.1007/s10482-011-9617-7

Sauer, M., Branduardi, P., Valli, M., Porro, D., 2004. Production of L-ascorbic acid by metabolically engineered *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii.* Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.70.10.6086-6091.2004

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., Rosa, C.A., DeVirgilio, J., Hulfachor, A.B., Groenewald, M., Kurtzman, C.P., Hittinger, C.T., Rokas, A., 2018. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell. https://doi.org/10.1016/j.cell.2018.10.023

Solieri, L., Cassanelli, S., Huff, F., Barroso, L., Branduardi, P., Louis, E.J., Morrissey, J.P., 2021. Insights on life cycle and cell identity regulatory circuits for unlocking genetic improvement in *Zygosaccharomyces* and *Kluyveromyces* yeasts. FEMS Yeast Res. https://doi.org/10.1093/femsyr/foab058

Staffas, L., Gustavsson, M., McCormick, K., 2013. Strategies and policies for the bioeconomy and bio-based economy: An analysis of official national approaches. Sustain. https://doi.org/10.3390/su5062751

Steensels, J., Snoek, T., Meersman, E., Nicolino, M.P., Voordeckers, K., Verstrepen, K.J., 2014. Improving industrial yeast strains: Exploiting natural and artificial diversity. FEMS Microbiol. Rev. https://doi.org/10.1111/1574-6976.12073

Stegmann, P., Londo, M., Junginger, M., 2020. The circular bioeconomy: Its elements and role in European bioeconomy clusters. Resour. Conserv. Recycl. https://doi.org/10.1016/j.rcrx.2019.100029

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.* Int. J. Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2013.06.025

Suh, S.O., Gujjari, P., Beres, C., Beck, B., Zhou, J., 2013. Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii.* Int. J. Syst. Evol. Microbiol. https://doi.org/10.1099/ijs.0.048058-0

Tramontina, R., Brenelli, L.B., Sodré, V., Franco Cairo, J.P., Travália, B.M., Egawa, V.Y., Goldbeck, R., Squina, F.M., 2020. Enzymatic removal of inhibitory compounds from lignocellulosic hydrolysates for biomass to bioproducts applications. World J. Microbiol. Biotechnol. https://doi.org/10.1007/s11274-020-02942-y

Ubando, A.T., Felix, C.B., Chen, W.H., 2020. Biorefineries in circular bioeconomy: A comprehensive review. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2019.122585

Varela, J.A., Puricelli, M., Montini, N., Morrissey, J.P., 2019. Expansion and diversification of MFS transporters in *Kluyveromyces marxianus*. Front. Microbiol. https://doi.org/10.3389/fmicb.2018.03330

Vigentini, I., Brambilla, L., Branduardi, P., Merico, A., Porro, D., Compagno, C., 2005. Heterologous protein production in *Zygosaccharomyces bailii*: Physiological effects and fermentative strategies. FEMS Yeast Res. https://doi.org/10.1016/j.femsyr.2004.11.006

Wehrs, M., Tanjore, D., Eng, T., Lievense, J., Pray, T.R., Mukhopadhyay, A., 2019. Engineering Robust Production Microbes for Large-Scale Cultivation. Trends Microbiol. https://doi.org/10.1016/j.tim.2019.01.006

Wolfe, K.H., 2015. Origin of the yeast whole-genome duplication. PLoS Biol. https://doi.org/10.1371/journal.pbio.1002221

Xu, Y., Zhi, Y., Wu, Q., Du, R., Xu, Y., 2017. *Zygosaccharomyces bailii* is a potential producer of various flavor compounds in Chinese Maotai-flavor liquor fermentation. Front. Microbiol. https://doi.org/10.3389/fmicb.2017.02609

Zhang, Y., Nielsen, J., Liu, Z., 2021. Yeast based biorefineries for oleochemical production. Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2020.11.009

Zhou, H., Cheng, J. sheng, Wang, B.L., Fink, G.R., Stephanopoulos, G., 2012. Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by *Saccharomyces cerevisiae*. Metab. Eng. https://doi.org/10.1016/j.ymben.2012.07.011

Zoghlami, A., Paës, G., 2019. Lignocellulosic Biomass: Understanding Recalcitrance and Predicting Hydrolysis. Front. Chem. https://doi.org/10.3389/fchem.2019.00874

Chapter 1

Interdependence between lignocellulosic biomasses, enzymatic hydrolysis and yeast cell factories in biorefineries

Stefano Bertacchi[1#], Pooja Jayaprakash[1, 2#], John P. Morrissey[2*] and Paola Branduardi[1*]

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, Milano 20126, Italy.

[2]School of Microbiology, Environmental Research Institute, APC Microbiome Institute, University College Cork, Cork, T12 K8AF, Ireland

#Equal contribution

Published in Microbial Biotechnology, https://pubmed.ncbi.nlm.nih.gov/34289233/

Abstract

Biorefineries have a pivotal role in the bioeconomy scenario for the transition from fossil-based processes towards more sustainable ones relying on renewable resources. Lignocellulose is a prominent feedstock since its abundance and relatively low cost. Microorganisms are often protagonists of biorefineries, as they contribute both to the enzymatic degradation of lignocellulose complex polymers and to the fermentative conversion of the hydrolyzed biomasses into fine and bulk chemicals. Enzymes have therefore become crucial for the development of sustainable biorefineries, being able to provide nutrients to cells from lignocellulose. Enzymatic hydrolysis can be performed by a portfolio of natural enzymes that degrade lignocellulose, often combined into cocktails. As enzymes can be deployed in different operative settings, such as separate hydrolysis and fermentation (SHF) or simultaneous saccharification and fermentation (SSF), their characteristics need to be combined with microbial ones to maximize the process. We therefore reviewed how the optimization of lignocellulose enzymatic hydrolysis can ameliorate bioethanol production when *Saccharomyces cerevisiae* is used as cell factory. Expanding beyond biofuels, enzymatic cocktail optimization can also be pivotal to unlock the potential of non-*Saccharomyces* yeasts, which, thanks to broader substrate utilization, inhibitor resistance and peculiar metabolism, can widen the array of feedstocks and products of biorefineries.

Introduction

Biorefineries can be described as "the sustainable processing of biomass into a spectrum of marketable products (food, feed, materials, chemicals) and energy (fuels, power, heat)" (IEA Bioenergy Task42, 2014). Biorefineries indeed aim to provide a broad portfolio of products alongside classical bio-based molecules such as biofuels or biogas (European Commission, 2018; Rosales-Calderon and Arantes, 2019; Stegmann *et al.*, 2020). Consistently, they are considered as one of the key technologies in the circular bioeconomy scenario, presenting different opportunities and challenges across countries, as they need to be organically integrated in the territories' landscape and infrastructure.

In order to widen possible outcomes of biorefineries and, in some cases, minimize environmental impacts, it is possible to exploit microorganisms, the so-called microbial cell factories, whose role is to convert the provided biomass(es) into the desired product(s) (Dahiya *et al.*, 2018). As a consequence, it is crucial that nutrients released from biomasses can match microbial requirements. In the case of lignocellulosic biomasses (LCBs), constituted by cellulose, hemicellulose and lignin in different ratios, a pretreatment step to open-up the recalcitrant macromolecular structure is followed by enzymatic hydrolysis: This step is preferred to chemical treatment (Galbe and Wallberg, 2019) (*e.g.* acid) as enzymes operate under conditions that are more compatible with microbial growth. Different hydrolyses can generate different mixtures of sugars and other nutrients, both in terms of composition and relative quantities. Notably, enzymes can be applied in two quite distinct processes, namely, separate hydrolysis and fermentation (SHF) and simultaneous saccharification and fermentation (SSF), in order to match microbial cell factories' characteristics (Kawaguchi *et al.*, 2016). Here, we recapitulate recent literature on this subject, and our aim is to underline the tight correlation between biomass, enzymes and yeast cell factories: optimizing the enzymatic cocktail and its operative conditions can unlock the potential of a lignocellulosic biomass and/or of a yeast cell factory (Figure 1). This synergy is crucial for assessing or improving the viability of the overall process, and therefore its economic feasibility (Pellis *et al.*, 2018). This review is a qualitative description on the importance of a link between LCBs composition, choice of enzyme

cocktail and selection of yeast species and strains that need to be considered in an integrated fashion to enable the development of an efficient process.

LCBs as preferred feedstock in biorefineries

LCB as a recalcitrant and uneven feedstock

Annually, about 1.3 billion tons of lignocellulosic biomasses are generated all around the world but only a small fraction is exploited to produce biochemicals, (Baruah *et al.*, 2018). For these reasons, by 2030 the Bio-based consortium aims to replace the 30% of the overall chemical production in the EU with biomolecules derived from biomass (Hassan *et al.*, 2019).

The main examples of LCBs are plant straw, coconut husk, corn stover, sugarcane bagasse, and woody materials in general. The structure of lignocellulose comprises three main biological polymers, lignin (20-30%), hemicellulose (20-40%) and cellulose (40-50%), held together by different types non-covalent bonds and covalent cross-linkages (Hosseini Koupaie *et al.*, 2019). Cellulose, the most abundant LCB polymer, is composed of β-D-glucose units linked by β-(1,4) glycosidic bonds, with cellobiose as the fundamental repeating dimeric unit. Around 500-1400 glucose molecules compose the cellulose chains forming the microfibrils that are embedded in the lignocellulosic matrix, which, in turn, makes it resistant to enzymatic hydrolysis (Zoghlami and Paës, 2019). Hemicelluloses are composed of heterogeneous groups of biopolymers containing various monosaccharide subunits to form xylans, xyloglucan, mannans and glucomannans (McKendry, 2002). They are amorphous with very little physical strength but act as a physical barrier for enzyme accessibility. Lignin, responsible for the hydrophobicity and structural rigidity, binds hemicelluloses, which in turn adhere to cellulose microfibrils in the cell wall (Zoghlami and Paës, 2019). It is a complex amorphous heteropolymer of phenylpropanoid building units (Agbor *et al.*, 2011). Sugar release from LCBs is influenced by factors such as total lignin content, lignin composition and structure (Santos *et al.*, 2012). In addition to blocking access to (hemi)cellulases, the hydrophobic structural features of lignin can also irreversibly adsorb enzymes during hydrolysis (Zeng *et al.*, 2014). Cellulose and hemicelluloses are linked together through hydrogen bonds, while lignin is covalently linked to

hemicelluloses to form lignin-carbohydrate complex (LCC). There are five different types of LCC bonds (phenyl glycosides, benzyl ethers, γ-esters, ferulate/coumarate esters and acetal linkages), that involve the 4-OH and 4-O positions of the lignin moieties (Tarasov *et al.*, 2018). The interaction between lignin and cellulose microfibrils and/or hemicelluloses, driven by LCC linkages, reduces the area of cellulose accessible for enzymes, significantly affecting the enzymatic hydrolysis of LCBs (Du *et al.*, 2014).

Before the enzyme hydrolysis, a pretreatment step is required to destabilize the recalcitrant structure of LCB, thereby enabling access of the enzymes to their substrates. Depending on the type of LCB used and the end product of interest, there are several physical, chemical, thermal and biological pretreatment methods that can also be used individually or in combination (Hosseini Koupaie *et al.*, 2019). Some of the preferred pretreatment methods used in second generation biorefineries are listed in Figure 2, together with examples of enzymes used in processes for specific products, which might change for matching different combinations of biomasses and microbial cell factories.

In any biorefinery, the choice of pretreatment method is determined in part by the type of LCB and the infrastructure available, but it must also consider subsequent steps as these are differentially affected by particular pretreatments (Baruah *et al.*, 2018). Acid or alkaline based methods may lead to the release of carboxylic acids, phenolic compounds, and furans, whereas the main byproducts of the pretreatment by hydrothermal processing are acetic acid and furan aldehydes (*i.e.* furfural and hydroxymethylfurfural) (Jönsson and Martín, 2016; Kim, 2018). Enzyme cocktails can be inhibited by solubilized aromatic compounds, such as phenols, as well as solid components like lignin, and residual hemicellulose (Jönsson and Martín, 2016). Pretreatment methods such as steam explosion lead to minimal release of furanic compounds but releases acetyl groups from lignocellulose that normally are linked to the hemicellulose moiety, exacerbating the impairment of enzymes and microbial activity (Sun *et al.*, 2016). There is a considerable ongoing effort to integrate knowledge of the potential negative effects of pretreatment into process development to ensure that efficient enzyme hydrolysis and subsequent microbial fermentation is possible. This can include the development of enzyme cocktails and microbial strains

that are less sensitive to inhibition by by-products from particular types of pretreatment.

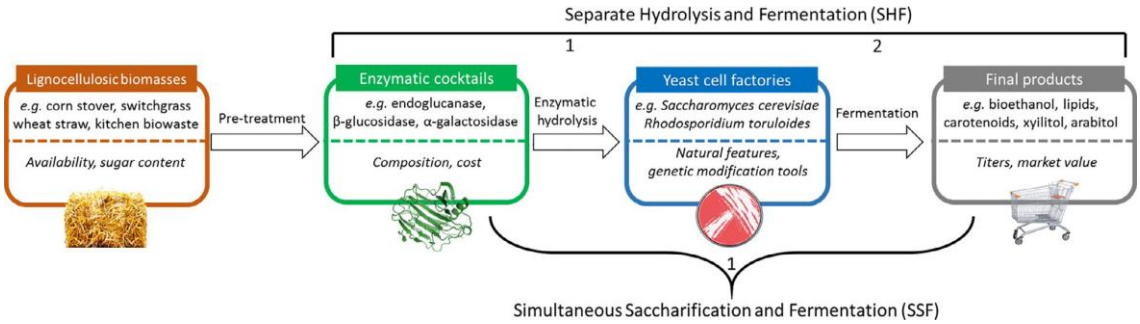


Figure 1. Overview of the processes/factors involved in the conversion of lignocellulosic biomass into final products in second-generation biorefineries. For each step of the overall process (coloured boxes), the main parameters to be considered when establishing a (yeast-based) biorefinery are indicated under the dotted line. Pre-treatment is needed to weaken the intertwined structure of LCBs prior to enzymatic hydrolysis. In SHF, hydrolysis and fermentation are performed as sequential steps, whereas in SSF, they are combined into a single one.

Hydrolysis of LCBs by enzymatic cocktails: unity is strength

According to the International Union of Biochemistry and Molecular Biology (IUBMB), most of the (hemi)cellulases and other polysaccharide degrading enzymes are grouped in the family of O-glycoside hydrolases (GH) and further sub-classified into different families based on primary structure of catalytic domains (Houfani *et al.*, 2020). Given the variation between LCBs from different sources, and the heterogeneity that arises following pretreatment, exploiting the natural biodiversity of GHs is pivotal to address the hydrolysis of the different components of these matrixes. Cellulases and hemicellulases, often combined in cocktails, are employed to hydrolyse cellulose to cellobiose and glucose, and hemicellulose to diverse pentose and hexose sugars (Houfani *et al.*, 2020). Multiple factors such as temperature, pH, rate of mixing, substrate concentration, enzyme loading and addition of surfactants influence dimeric and monomeric sugar yields from LCBs (Sarkar *et al.*, 2012). The complete degradation of cellulose can occur by combined and simultaneous action of three distinct classes of cellulolytic enzymes; namely, endoglucanases, cellobiohydrolases and β-glucosidases. Endoglucanases are involved in the cleavage of internal β-glucosidic bonds thereby providing accessible cellulose chain ends to cellobiohydrolases. These enzymes release cellobiose, which is further hydrolyzed to glucose by β-glucosidase (Yennamalli *et al.*, 2017). In contrast, as the composition of hemicellulose varies depending on the type of LCB, with multiple different monomeric units and bond types, hemicellulolytic enzymes are accordingly more diverse.

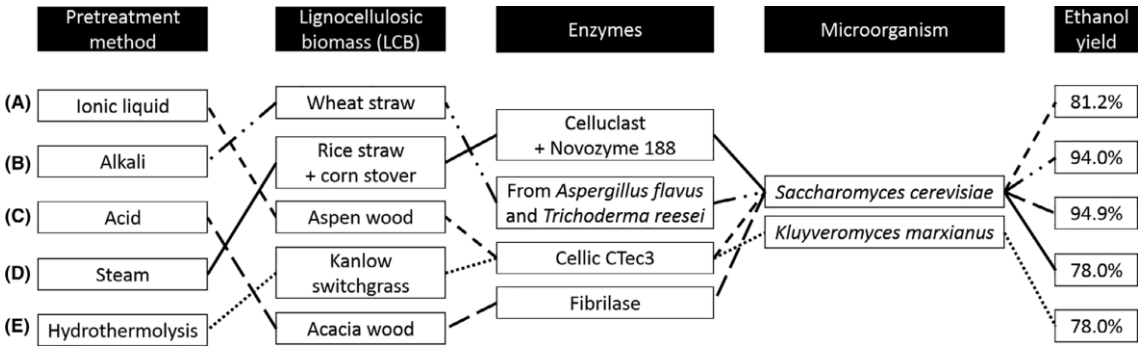


Figure 2: Examples of pre-treatment methods used in combination with enzymatic hydrolysis and microbial fermentation to obtain ethanol. Equa line styles allow reconstructing experimental data reported in literature, and without this guideline, it would be not possible to forecast the correct combinations. Ethanol yields displayed are reported as percentage,

considering 100% the theoretical yield. Ethanol yields are calculated as gram of produced product per gram of total consumed sugars. References: (A) (Goshadrou et al., 2013), (B) (Singh and Bishnoi, 2012), (C) (Lee and Yu, 2020), (D) (Nielsen et al., 2020) and (E) (Suryawati et al., 2008).

For example, endo-β-1,4-xylanase is involved in breaking down internal bonds of xylan, a major polymer found in hemicellulose, leading to the release of xylo-oligosaccharides, the non-reducing ends of which can then be hydrolyzed by β-xylosidase. Accessory enzymes such as α-L-arabinofuranosidase, α-glucuronidase, α-galactosidase, acetylxylan esterase and ferulic acid esterase (Maitan-Alfenas *et al.*, 2015) are often important to increase sugar yields from hemicellulose during the saccharification process (Robl *et al.*, 2013). The correct combination and ratio of hydrolytic and accessory enzymes is very important to effectively liberate the monomeric sugars and to reduce the inhibitory effect of lignin (Van Dyk and Pletschke, 2012). For example, corn stover pretreated by ammonia fiber expansion, hydrolyzed by using six core enzymes (cellobiohydrolase 1, cellobiohydrolase 2, endo-β1,4-glucanase, β-glucosidase, endo-β1,4-xylanase 3 and β-xylosidase), led to a glucose yield of 38.5%; when these enzymes were combined with five accessory enzymes the yield increased to 52.1% (Banerjee *et al.*, 2010).

Furthermore, the use of additives such as non-catalytic proteins and surfactants may reduce lignin adsorption of enzymes and improve the interaction between cellulases and cellulose fibers, thereby enhancing the overall hydrolysis of LCBs (Xu *et al.*, 2019).

Numerous bacterial and fungal species produce (hemi)cellulases, with the filamentous fungal genera *Trichoderma* and *Aspergillus* of particular interest (Maitan-Alfenas *et al.*, 2015). *Trichoderma* ssp. are widely exploited owing to their natural ability to produce two cellobiohydrolases, five endoglucanases and three endoxylanases, but they have lower β-glucosidase activity (Bischof *et al.*, 2016). Efficient β-glucosidase producers are *Aspergillus* ssp., being therefore able to complement the missing activity (Sarkar *et al.*, 2012). Nevertheless, there is a need to explore other fungal strains such as anaerobic gut fungi that are known to possess a wide range of biomass degrading enzymes (Usmani *et al.*, 2021). For example, Neocallimastigomycota (anaerobic gut fungi) displayed a 300% increase in xylan degradation activity, compared to the commercial

Aspergillus enzyme formulations (Solomon *et al.*, 2016). The lack of a single natural microbial species capable of secreting all the required cellulolytic enzymes in high titers and balanced ratios for efficient enzymatic hydrolysis of lignocellulosic biomass has necessitated use of blends of enzymes from several microorganisms (Maitan-Alfenas *et al.*, 2015). This has led to the development of formulated enzymatic cocktails, with Novozymes and Du-Pont Genencor among the leading commercial producers (Adsul *et al.*, 2020).

Enzymatic cocktails have a pivotal role in improving the efficiency of biomass hydrolysis by reducing the amount of enzymes and time required to convert all the carbohydrates into fermentable sugars and having the possibility to function at high substrate loadings (Adsul *et al.*, 2020). Furthermore, auxiliary enzymes like lytic polysaccharide monooxygenase (LPMO), copper-enzymes that catalyze oxidative cleavage of glycosidic bonds, are often part of commercial enzymatic cocktails. Indeed, the addition of LPMOs to cocktails has led to a significant reduction in the cost of the enzymatic process (Johansen, 2016) and is an effective way of increasing the digestibility of structural carbohydrates (Duque *et al.*, 2021). Rodriguez and co-workers reported an increase in cellulose release from various LCBs (sugarcane bagasse, corn stover, and wheat straw) when LPMOs were used alongside enzymatic cocktail Cellic CTec2, permitting the use of hydrothermal pretreatment rather than organosolv and alkaline ones (Rodríguez-Zúñiga *et al.*, 2015).

The formulation of cocktails can be designed to exploit synergistic effect of enzymes such as cellulase, xylanase and pectinase. Considering the hydrolysis of sugarcane bagasse, it was reported that replacing 20% of cellulase by xylanase led to an increase in glucose yield by 6.6%, 8.8% and 9.5% in sugarcane bagasse pretreated by steam explosion, NaOH and H_2O_2 respectively (Li *et al.*, 2014). These observations suggested that glucose release is positively affected by the degree of synergism between cellulase and xylanase and it is also dependent on hydrolysis time (Li *et al.*, 2014). In a study based on steam treated sweet sorghum bagasse, when the combination of enzymatic cocktails (cellulases - Cellic CTec2 and endoxylanases - Cellic HTec2) was optimized by response surface methodology (RSM), sugars yield increased by 20% (Pengilly *et al.*, 2015). Regarding corn stover pretreated by steam explosion, it was noticed that the use of commercial cellulase (Spezyme CP) in combination with

cellulase from *Aspergillus fumigatus* led to a 26% increase in the conversion of glucan to glucose compared to the use of the sole cocktail (Wang *et al.*, 2012). Furthermore, corn stover pretreated by ammonia fiber expansion (AFEX) resulted in glucose and xylose yields >80% and 70% respectively after enzymatic hydrolysis by a cocktail containing cellulases, xylanses and accessory enzymes, and the addition of accessory hemicellulases further increased xylose yields by 20% (Gao *et al.*, 2011). The use of enzymatic cocktail containing exoglucanase (Cel7A), endoglucanase (Cel5A) and two endoxylanases (XYN10A, XYN11A) in synergism along with swollenin, a non-hydrolytic disruptive protein, led to a significant increase in xylose yields from steam pretreated corn-stover by >300%, by enhancing enzymatic access to the hemicellulose fraction, increasing in turn cellulose accessibility as well. (Gourlay *et al.*, 2013).

Considering the importance of combining pretreatments and cocktail compositions, and the still largely unexplored potential of enzymes biodiversity, it is evident that the next decade will witness large improvements in LCBs exploitation.

Combination of microbial enzymes and cell factories in LCBs-based biorefineries

Together or separate?

Up to now, we introduced diverse LCBs, pretreatment principles and enzymes as elements to be combined for obtaining the desired media to be fermented by microbial cell factories in a bioprocess. However, the timing of hydrolysis and fermentation has not yet been discussed. In the last decades, two main types of processes have been developed: simultaneous saccharification and fermentation (SSF) and separate hydrolysis and fermentation (SHF). These processes, where enzymes and cells are working in a single (SSF) or in separate (SHF) vessels, have pros and cons in terms of efficiency, duration, presence/release of inhibitory molecules, and downstream processing of the final product (Choudhary *et al.*, 2016; Kawaguchi *et al.*, 2016; Haldar and Purkait, 2020). There are several examples in the literature of second-generation bioethanol production by *S. cerevisiae* based on SHF or SSF processes where distinct enzymes are used resulting in different ethanol yields (Figure 3).

An obvious advantage of SHF is the possibility to choose the optimal conditions for both enzymatic hydrolysis and fermentation steps as they are temporally and spatially separated. Temperature, for example, is a key parameter as efficient enzymatic cocktails often have optimum activity at around 50 °C, whereas *S. cerevisiae* fermentation is optimal at 30 °C (Choudhary *et al.*, 2016). In addition, in a SHF process it is possible to eliminate by centrifugation the water insoluble solids (WIS) that cause poor homogenization of the liquid medium and can impair yeast growth (Ask *et al.*, 2012). Although it is not a consideration regarding bioethanol production, separation of WIS is also advantageous if not indispensable when the final product is intracellular (Öhgren *et al.*, 2007). Nevertheless, SHF also displays drawbacks related to both the biological process and the economics. On the biological side, inhibition of enzyme activity caused by the mono- and disaccharides products can reduce yields and, conversely, very successful hydrolysis can deliver medium with sugar concentrations that are problematic for batch fermentation because of osmotic stress. Economically, in SHF, there is higher capital expenditure (CAPEX), because of the need for different

vessels, and the prolonged process time with the additional risk of undesired fermentations by contaminants that may reduce the economical sustainability of the process (Kawaguchi *et al.*, 2016; Haldar and Purkait, 2020).

In SSF, yeast cell factories can metabolize sugars concurrently with their release from the biomass, thus mitigating both the osmotic stress and the inhibitory effect of (simple) sugars on the enzymes (Choudhary *et al.*, 2016; Kawaguchi *et al.*, 2016; Haldar and Purkait, 2020). Another advantage is that the low titer of glucose may facilitate consumption of other saccharides by the yeast, which is often impaired by the catabolic repression caused by glucose. One important factor to consider is the effect of high solids loading (>20% w/v), which can affect cell viability and the action of enzymes: the compromise between the amount of sugars and inhibitors must be considered (Wu *et al.*, 2018; Da Silva *et al.*, 2020). Nevertheless, the major drawback is that it is necessary to carry out the process at temperatures that are far below the optimum for the hydrolytic enzymes. One option to maximize the efficacy of enzyme hydrolysis in a SSF process is to replace *S. cerevisiae* with a more thermophilic yeast that can ferment at a higher temperature (Choudhary *et al.*, 2016).

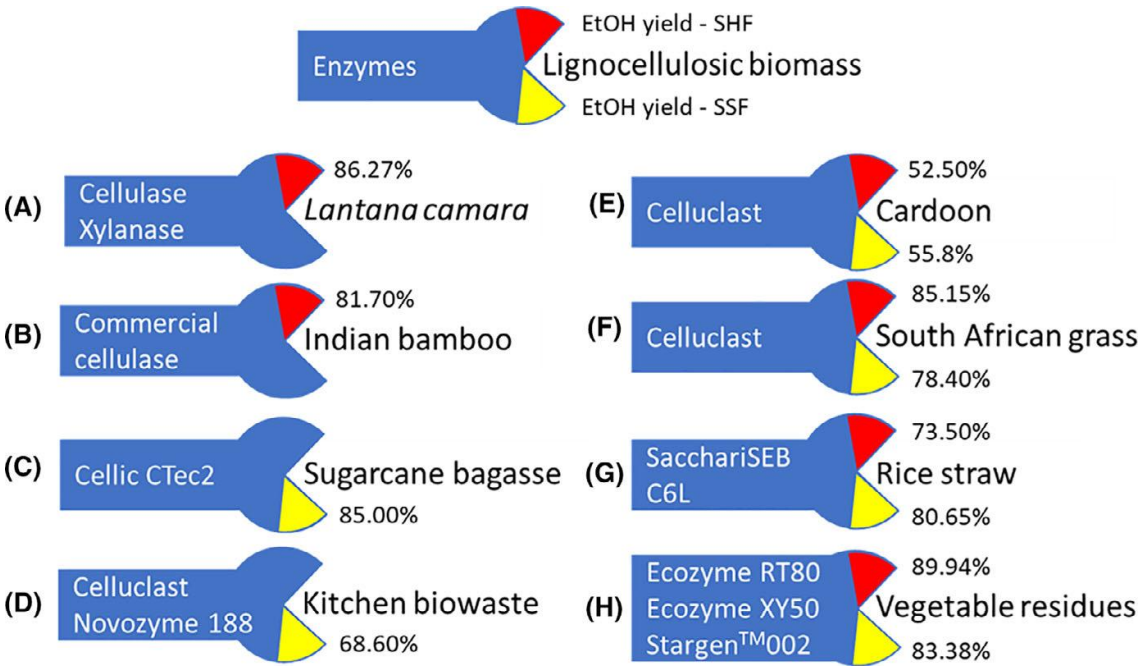


Figure 3. Selected examples of second-generation bioethanol production with S. cerevisiae as cell factory by SHF or SSF processes. The scheme reports the enzyme used, the biomass

utilized as substrate, the different fermentation strategies and the corresponding ethanol yields (reported as percentage, considering 100% the theoretical yield). References: (A) (Kuila and Banerjee, 2014), (B) (Sindhu et al., 2014), (C) (Unrean et al., 2016), (D) (Ntaikou et al., 2018), (E) (Fernandes et al., 2018), (F) (Burman et al., 2019), (G) (Mishra et al., 2016) and (H) (Mithra et al., 2019).

In deciding whether to implement an SHF or an SSF process, it is necessary to consider multiple variables. Several studies have attempted to perform this type of comparison to determine whether SHF or SSF better suited for particular processes (Dahnum *et al.*, 2015; Rodrigues *et al.*, 2016; Wu *et al.*, 2018; Ben Atitallah *et al.*, 2019; Mithra *et al.*, 2019; Bertacchi *et al.*, 2020). For example, a comparison of processes for the production of carotenoids by the yeast *Rhodosporidium toruloides* from *Camelina sativa* meal hydrolysate found that SSF was able to guarantee the highest titer of the final products (Bertacchi *et al.*, 2020). Although generally negative, in this case, it was speculated that the presence of growth inhibitory WIS may actually trigger the production of scavenger molecules, like carotenoids. This illustrates why bespoke analysis of each production process is necessary. Modeling and statistical tools can assist in this decision-making process. For example in bioethanol production, response surface methodology (RSM) was used to infer the optimum conditions for SSF and SHF (Althuri and Banerjee, 2019), whereas empirical equations modeling served to determine glucose and ethanol titers in both processes (Burman *et al.*, 2019).

By implementing these studies, we are constantly learning and therefore designing further optimization, among which it is notable to mention the hybrid solution of starting a process as a suboptimal SHF, followed by SSF with a reduced or optimized enzyme loading (US9187390B2).

Non-Saccharomyces yeasts in LCB-based second generation biorefineries

Despite the widespread application of *S. cerevisiae* in second generation bioethanol production, one of the main roadblocks is its preference to glucose thereby lacking the ability to consume different C5-C6 sugars present in the lignocellulosic biomass. Cultures based on *S. cerevisiae* strains engineered to preferentially and singularly metabolize either glucose, xylose or arabinose have been developed (Verhoeven *et al.*, 2018), but this work is still closer to the proof of concept phase than to production.

Different is the case of a mixed cultivation of *S. cerevisiae* and *S. stipitis* (naturally able to consume pentose sugars), developed for a SHF/SSF processes for the production of ethanol from kitchen biowaste (Ntaikou *et al.*, 2018): here the consortium might be more effective also at industrial scale, with a caveat related to the limited ethanol tolerance of *S. stipitis*.

In the presented scenario and moving beyond bioethanol, yeast biodiversity can offer advantages that are not yet fully exploited and, in some cases, also poorly explored. Prominent examples are the co-consumption of hexose and pentose sugars (which includes both sugar transporters and catalytic enzymes), the native production of enzymes for the hydrolysis of LCB-derived polymers/oligomers, the resistance towards inhibitory compounds arising from the pretreatment and hydrolysis of the biomass, and finally the natural ability to transform substrates into the desired products.

The ability to withstand various growth inhibitors often derives not only from the product but also from the pretreatment and hydrolysis steps of LCBs (Sitepu *et al.*, 2014; Pandey *et al.*, 2019). Oleaginous yeasts often display good resistance towards classic inhibitors present in lignocellulosic hydrolysates (Sitepu *et al.*, 2014; Poontawee *et al.*, 2017; Osorio-González *et al.*, 2019). For example, *S. cerevisiae* growth is inhibited by 0.8 g/L of furfural, whereas several oleaginous yeasts (*e.g. R. toruloides*, *Y. lipolytica*) can withstand 1 g/L of this toxic compound (Sitepu *et al.*, 2014). In this context, different yeasts of the *Yarrowia* clade were tested for their ability to withstand several inhibitory compounds derived from acid-pretreated switchgrass hydrolyzed with a mixture of Cellic Ctec2 and HTec2 cocktails (Quarterman *et al.*, 2017). Acetic acid, which is detached from lignocellulose by pretreatment (Jönsson and Martín, 2016), is an example of such an inhibitor as it can impair metabolism and microbial growth. Whereas *S. cerevisiae* needs to be engineered or evolved for acetic acid tolerance (Martani *et al.*, 2015; Ko *et al.*, 2020), species like *Zygosaccharomyces bailii* are able to withstand acetic acid, with a minimum inhibitory concentration of 375–550 mM (whereas *S. cerevisiae* of 80–150 mM), and to consume it for its own growth, even in the presence of glucose (Kuanyshev *et al.*, 2017; Palma *et al.*, 2018). C*andida tropicalis* also showed the ability to withstand furfural (Wang *et al.*, 2016) up to 1.5 g/L, being therefore more resistant than *S. cerevisiae* (Pandey *et al.*, 2019), and to produce xylitol from several residual biomasses (Eryasar and Karasu-Yalcin, 2016;

Mattam *et al.*, 2016). Furthermore, in respect to the aforementioned limitations of SSF processes, there are studies that compare the performance of species like *Blastobotrys adeninivorans, Pichia kudriavzevii* and *K. marxianus* with that of different *S. cerevisiae* strains (Choudhary *et al.*, 2016).

Figure 4 lists some other recent examples of different non-*Saccharomyces* yeasts deployed to produce valuable products with specific biomass and enzymatic cocktails, employing the natural ability of those species. Indeed, their broader substrate range can be coupled not only to the production of ethanol (which still remains one of the more investigated products when demonstrating industrially relevant products) but also of biodiesel or more generally single cell oil (SCO), as in the case of oleaginous yeasts (Poontawee *et al.*, 2017; Carsanba *et al.*, 2018; Sreeharsha and Mohan, 2020). It is interesting to mention that synthetic consortia can be considered to perform the enzymatic hydrolysis, allowing implementation of a hybrid SHF/SSF without the addition of external enzymes. This is the case of sugarcane bagasse triggering the secretion of endoglucanase, β-glucosidase and xylanase in a co-culture of *S. cerevisiae* and *C. tropicalis* (Qadir *et al.*, 2018).

One important comment relates to pentose and hexose sugars co-consumption: in most cases this is difficult to achieve, and strain engineering or synthetic consortia needs to be accurately assisted by bioprocess engineering, adjusting cultivation parameters, feeding, dimension and ratio of the inoculum to maximize productivity. The concerns related to the fermentation time are not only linked to the overall costs of the process, but also to the fact that the accumulation of the final product is in most of the case detrimental to the cells, if not toxic.

Non-*Saccharomyces* yeasts are therefore an important biological reservoir of biodiversity that can be applied to the development of second-generation biorefineries, in order to widen our horizons beyond the common exploitation of *S. cerevisiae* for bioethanol production, and to maximize the portfolio of enzymatic cocktails available on the market.

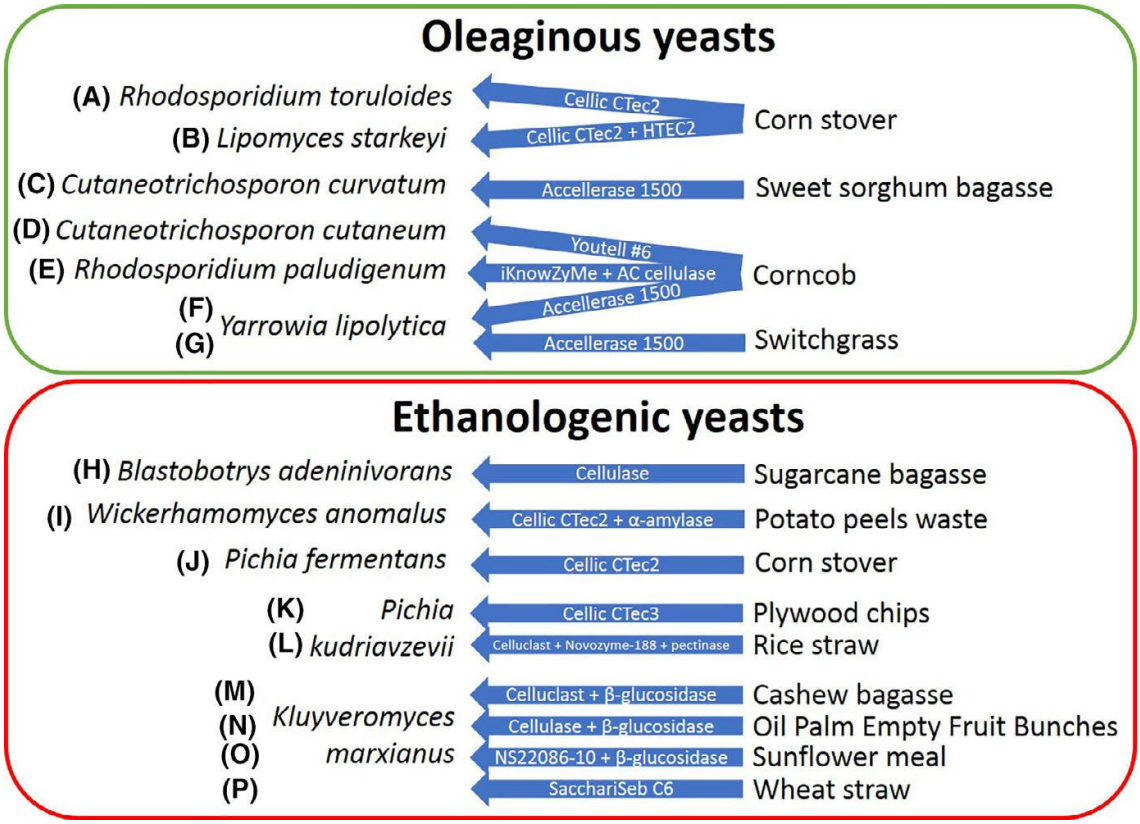


Figure 4: Biodiversity of non-Saccharomyces yeasts in second-generation bioprocesses. Two different and representative yeast metabolisms (lipidogenesis and alcoholic fermentation) are reported together with LCBs and enzymatic cocktails used in some selected examples. References: (A) (Fei et al., 2016), (B) (Pomraning et al., 2019), (C) (Liang et al., 2012), (D) (Gao et al., 2014), (E) (Chaiyaso et al., 2019), (F) (Kahr et al., 2015), (G) (Quarterman et al., 2017), (H) (Antil et al., 2015), (I) (Ben Atitallah et al., 2019), (J) (Mierzejewska et al., 2019), (K) (Yuan et al., 2017), (L) (Oberoi et al., 2012), (M) (Rodrigues et al., 2016), (N) (Sukhang et al., 2020), (O) (Camargo et al., 2014) and (P) (Saini et al., 2015).

Conclusion

Biorefineries will play a pivotal role in the development of a sustainable global bioeconomy. Efficient biorefineries will integrate biomass, bespoke enzyme cocktails, and specific cell factories. Although the traditional focus has been on cell factory design, the critical need to exploit second-generation biomasses to achieve sustainability is now a major driver of research. Indeed, this leads to somewhat of a paradigm shift since, in these scenarios, the substrate achieves equal importance to the product: a process that does not use residual biomasses will struggle to deliver sustainability. This adds a substantial variable that was little considered in traditional fermentations and first-generation bioprocesses. It also creates new opportunities since there is increased scope to exploit microbial diversity, and in the case of yeasts, non-Saccharomyces yeasts because of their properties related to substrate specificity, inhibitor tolerance and growth parameters. The metabolism of these yeasts can then be exploited to produce new products, often with better efficiency that *S. cerevisiae*. The third pillar in second generation biorefineries are the enzymes that link the biomass to the cell factory. This is a crucial area of ongoing investigation that considers the enzymes and how they are applied in either SHF or SSF processes. The use of enzyme cocktails to treat LCBs is now in routine, but considerable work is still required to decide on the best enzyme formulation and the range of options is still too limited. It is important to recognize that the best enzyme cocktail is the one that can deliver the optimum sugar mix to a specific cell factory microbe when starting from a particular LCB source. This is also possible because of the improved knowledge on yeast biodiversity, and the constant develop on enzymes potential. It is implicit in this that considering any of the components in isolation cannot achieve the best outcome. At the moment, each bioprocess needs to be designed from first principles, but it is hoped that, as experience grows, it will become possible to develop framework principles that facilitate rational selection of the components of a biorefinery. In this case, it will be possible to reduce the cost and time needed to develop new second generation biorefineries, including those that operate on a modest scale.

References

Adsul, M., Sandhu, S.K., Singhania, R.R., Gupta, R., Puri, S.K., and Mathur, A. (2020) Designing a cellulolytic enzyme cocktail for the efficient and economical conversion of lignocellulosic biomass to biofuels. Enzyme Microb Technol. https://doi.org/10.1016/j.enzmictec.2019.109442

Agbor, V.B., Cicek, N., Sparling, R., Berlin, A., and Levin, D.B. (2011) Biomass pretreatment: Fundamentals toward application. Biotechnol Adv. https://doi.org/10.1016/j.biotechadv.2011.05.005

Althuri, A. and Banerjee, R. (2019) Separate and simultaneous saccharification and fermentation of a pretreated mixture of lignocellulosic biomass for ethanol production. Biofuels. https://doi.org/10.1080/17597269.2017.1409059

Antil, P.S., Gupta, R., and Kuhad, R.C. (2015) Simultaneous saccharification and fermentation of pretreated sugarcane bagasse to ethanol using a new thermotolerant yeast. Ann Microbiol. https://doi.org/10.1007/s13213-014-0875-2

Ask, M., Olofsson, K., Di Felice, T., Ruohonen, L., Penttilä, M., Lidén, G., and Olsson, L. (2012) Challenges in enzymatic hydrolysis and fermentation of pretreated Arundo donax revealed by a comparison between SHF and SSF. Process Biochem. https://doi.org/10.1016/j.procbio.2012.05.016

Ben Atitallah, I., Antonopoulou, G., Ntaikou, I., Alexandropoulou, M., Nasri, M., Mechichi, T., and Lyberatos, G. (2019) On the evaluation of different saccharification schemes for enhanced bioethanol production from potato peels waste via a newly isolated yeast strain of Wickerhamomyces anomalus. Bioresour Technol. https://doi.org/10.1016/j.biortech.2019.121614

Banerjee, G., Car, S., Scott-Craig, J.S., Borrusch, M.S., Bongers, M., and Walton, J.D. (2010) Synthetic multi-component enzyme mixtures for deconstruction of lignocellulosic biomass. Bioresour Technol. https://doi.org/10.1016/j.biortech.2010.07028

Baruah, J., Nath, B.K., Sharma, R., Kumar, S., Deka, R.C., Baruah, D.C., and Kalita, E.

(2018) Recent trends in the pretreatment of lignocellulosic biomass for value-added products. Front Energy Res. https://doi.org/10.3389/fenrg.2018.00141

Bertacchi, S., Bettiga, M., Porro, D., and Branduardi, P. (2020) Camelina sativa meal hydrolysate as sustainable biomass for the production of carotenoids by *Rhodosporidium toruloides*. Biotechnol Biofuels. https://doi.org/10.1186/s13068-020-01682-3

Bischof, R.H., Ramoni, J., and Seiboth, B. (2016) Cellulases and beyond: The first 70 years of the enzyme producer Trichoderma reesei. Microb Cell Fact. https://doi.org/10.1186/s12934-016-0507-6

Burman, N.W., Sheridan, C.M., and Harding, K.G. (2019) Lignocellulosic bioethanol production from grasses pre-treated with acid mine drainage: Modeling and comparison of SHF and SSF. Bioresour Technol Reports. https://doi.org/10.1016/j.biteb.2019.100299

Camargo, D., Gomes, S.D., and Sene, L. (2014) Ethanol production from sunflower meal biomass by simultaneous saccharification and fermentation (SSF) with *Kluyveromyces marxianus* ATCC 36907. Bioprocess Biosyst Eng. https://doi.org/10.1007/s00449-014-1201-x

Carsanba, E., Papanikolaou, S., and Erten, H. (2018) Production of oils and fats by oleaginous microorganisms with an emphasis given to the potential of the nonconventional yeast Yarrowia lipolytica. Crit Rev Biotechnol. https://doi.org/10.1080/07388551.2018.1472065

Chaiyaso, T., Manowattana, A., Techapun, C., and Watanabe, M. (2019) Efficient bioconversion of enzymatic corncob hydrolysate into biomass and lipids by oleaginous yeast *Rhodosporidium paludigenum* KM281510. Pre Bio & Biotechnol. https://doi.org/10.1080/10826068.2019.1591985

Choudhary, J., Singh, S., and Nain, L. (2016) Thermotolerant fermenting yeasts for simultaneous saccharification fermentation of lignocellulosic biomass. Electron J Biotechnol. https://doi.org/10.1016/j.ejbt.2016.02.007

Dahiya, S., Kumar, A.N., Shanthi Sravan, J., Chatterjee, S., Sarkar, O., and Mohan, S.V.

(2018) Food waste biorefinery: Sustainable strategy for circular bioeconomy. Bioresour Technol. https://doi.org/10.1016/j.biortech.2017.07.176

Dahnum, D., Tasum, S.O., Triwahyuni, E., Nurdin, M., and Abimanyu, H. (2015) Comparison of SHF and SSF processes using enzyme and dry yeast for optimization of bioethanol production from empty fruit bunch. Energy Procedia. https://doi.org/10.1016/j.egypro.2015.03.238

Du, X., Pérez-Boada, M., Fernández, C., Rencoret, J., del Río, J.C., Jiménez-Barbero, J., et al. (2014) Analysis of lignin–carbohydrate and lignin–lignin linkages after hydrolase treatment of xylan–lignin, glucomannan–lignin and glucan–lignin complexes from spruce wood. Planta. https://doi.org/10.1007/s00425-014-2037-y

Duque, A., Álvarez, C., Doménech, P., Manzanares, P., and Moreno, A.D. (2021) Advanced bioethanol production: From novel raw materials to integrated biorefineries. Processes. https://doi.org/10.3390/pr9020206

Van Dyk, J.S. and Pletschke, B.I. (2012) A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy. Biotechnol Adv. https://doi.org/10.1016/j.biotechadv.2012.03.002

Eryasar, K. and Karasu-Yalcin, S. (2016) Evaluation of some lignocellulosic byproducts of food industry for microbial xylitol production by *Candida tropicalis*. 3 Biotech. https://doi.org/10.1007/s13205-016-0521-8

European Commission (2018) Detailed Case Studies on the Top 20 Innovative Bio-Based Products.

Fei, Q., O'brien, M., Nelson, R., Chen, X., Lowell, A., and Dowe, N. (2016) Enhanced lipid production by *Rhodosporidium toruloides* using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source. Biotechnol Biofuels. https://doi.org/10.1186/s13068-016-0542-x

Fernandes, M.C., Ferro, M.D., Paulino, A.F.C., Chaves, H.T., Evtuguin, D. V., and Xavier, A.M.R.B. (2018) Comparative study on hydrolysis and bioethanol production from

cardoon and rockrose pretreated by dilute acid hydrolysis. Ind Crops Prod. https://doi.org/10.1016/j.indcrop.2017.11.037

Galbe, M. and Wallberg, O. (2019) Pretreatment for biorefineries: A review of common methods for efficient utilisation of lignocellulosic materials. Biotechnol Biofuels. https://doi.org/10.1186/s13068-019-1634-1

Gao, D., Uppugundla, N., Chundawat, S.P.S., Yu, X., Hermanson, S., Gowda, K., et al. (2011) Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. Biotechnol Biofuels. https://doi.org/10.1186/1754-6834-4-5

Gao, Q., Cui, Z., Zhang, J., and Bao, J. (2014) Lipid fermentation of corncob residues hydrolysate by oleaginous yeast *Trichosporon cutaneum*. Bioresour Technol. https://doi.org/10.1016/j.biortech.2013.11.044

Goshadrou, A., Karimi, K., and Lefsrud, M. (2013) Characterization of ionic liquid pretreated aspen wood using semi-quantitative methods for ethanol production. Carbohydr Polym. https://doi.org/10.1016/j.carbpol.2013.04.017

Gourlay, K., Hu, J., Arantes, V., Andberg, M., Saloheimo, M., Penttilä, M., and Saddler, J. (2013) Swollenin aids in the amorphogenesis step during the enzymatic hydrolysis of pretreated biomass. Bioresour Technol. https://doi.org/10.1016/j.biortech.2013.05.053

Haldar, D. and Purkait, M.K. (2020) Lignocellulosic conversion into value-added products: A review. Process Biochem. https://doi.org/10.1016/j.procbio.2019.10.001

Hassan, S.S., Williams, G.A., and Jaiswal, A.K. (2019) Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges, and opportunities. Renew Sustain Energy Rev. https://doi.org/10.1016/j.rser.2018.11.041

Hosseini Koupaie, E., Dahadha, S., Bazyar Lakeh, A.A., Azizi, A., and Elbeshbishy, E. (2019) Enzymatic pretreatment of lignocellulosic biomass for enhanced biomethane production - A review. J Environ Manage.

https://doi.org/10.1016/j.jenvman.2018.09.106

Houfani, A.A., Anders, N., Spiess, A.C., Baldrian, P., and Benallaoua, S. (2020) Insights from enzymatic degradation of cellulose and hemicellulose to fermentable sugars – a review. Biomass and Bioenergy. https://doi.org/10.1016/j.biombioe.2020.105481

IEA Bioenergy Task42 (2014) Sustainable and synergetic processing of biomass into marketable food & feed ingredients, chemicals, materials and energy (fuels, power, heat). IEA Bioenergy.

Johansen, K.S. (2016) Lytic Polysaccharide Monooxygenases: Lytic Polysaccharide Mnooxygenases: The Microbial Power Tool for Lignocellulose Degradation. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2016.07.012

Jönsson, L.J. and Martín, C. (2016) Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. Bioresour Technol. https://doi.org/10.1016/j.biortech.2015.10.009

Kahr, H., Pointner, M., Krennhuber, K., Wallner, B., and Jäger, A. (2015) Lipid production from diverse oleaginous yeasts from steam exploded corn cobs. Agron Res. https://doi.org/10.1016/j.copbio.2016.02.031

Kawaguchi, H., Hasunuma, T., Ogino, C., and Kondo, A. (2016) Bioprocessing of bio-based chemicals produced from lignocellulosic feedstocks. Curr Opin Biotechnol.

Kim, D. (2018) Physico-chemical conversion of lignocellulose: Inhibitor effects and detoxification strategies: A mini review. Molecules. https://doi.org/10.1016/j.copbio.2016.02.031

Ko, J.K., Enkh-Amgalan, T., Gong, G., Um, Y., and Lee, S.M. (2020) Improved bioconversion of lignocellulosic biomass by *Saccharomyces cerevisiae* engineered for tolerance to acetic acid. GCB Bioenergy. https://doi.org/10.1111/gcbb.12656

Kuanyshev, N., Adamo, G.M., Porro, D., and Branduardi, P. (2017) The spoilage yeast *Zygosaccharomyces bailii*: Foe or friend? Yeast. https://doi.org/10.1002/YEA.3238

Kuila, A. and Banerjee, R. (2014) Simultaneous saccharification and fermentation of

enzyme pretreated Lantana camara using *Saccharomyces cerevisiae*. Bioprocess Biosyst Eng. https://doi.org/10.1007/s00449-014-1172-y

Lee, I. and Yu, J.H. (2020) The production of fermentable sugar and bioethanol from acacia wood by optimizing dilute sulfuric acid pretreatment and post treatment. Fuel. https://doi.org/10.1016/j.fuel.2020.117943

Li, J., Zhou, P., Liu, H., Xiong, C., Lin, J., Xiao, W., et al. (2014) Synergism of cellulase, xylanase, and pectinase on hydrolyzing sugarcane bagasse resulting from different pretreatment technologies. Bioresour Technol. https://doi.org/10.1016/j.biortech.2013.12.113

Liang, Y., Tang, T., Siddaramu, T., Choudhary, R., and Umagiliyage, A.L. (2012) Lipid production from sweet sorghum bagasse through yeast fermentation. Renew Energy. https://doi.org/10.1016/j.renene.2011.09.035

Maitan-Alfenas, G.P., Visser, E.M., and Guimarães, V. ria M. (2015) Enzymatic hydrolysis of lignocellulosic biomass: Converting food waste in valuable products. Curr Opin Food Sci. https://doi.org/10.1016/j.cofs.2014.10.001

Martani, F., Marano, F., Bertacchi, S., Porro, D., and Branduardi, P. (2015) The *Saccharomyces cerevisiae* poly(A) binding protein Pab1 as a target for eliciting stress tolerant phenotypes. Sci Rep. https://doi.org/10.1038/sreo18318

Mattam, A.J., Kuila, A., Suralikerimath, N., Choudary, N., Rao, P.V.C., and Velankar, H.R. (2016) Cellulolytic enzyme expression and simultaneous conversion of lignocellulosic sugars into ethanol and xylitol by a new *Candida tropicalis* strain. Biotechnol Biofuels. https://doi.org/10.1186/s13068-016-0575-1

McKendry, P. (2002) Energy production from biomass (part 1): Overview of biomass. Bioresour Technol.

Mierzejewska, J., Dąbkowska, K., Chreptowicz, K., and Sokołowska, A. (2019) Hydrolyzed corn stover as a promising feedstock for 2-phenylethanol production by nonconventional yeast. J Chem Technol Biotechnol. https://doi.org/10.1002/jctb.5823

Mishra, A., Sharma, A.K., Sharma, S., Mathur, A.S., Gupta, R.P., and Tuli, D.K. (2016) Lignocellulosic bioethanol production employing newly isolated inhibitor and thermotolerant *Saccharomyces cerevisiae* DBTIOC S24 strain in SSF and SHF. RSC Adv. https://doi.org/10.1039/C6RA00007J

Mithra, M.G., Sajeev, M.S., and Padmaja, G. (2019) Comparison of SHF and SSF Processes under Fed Batch Mode on Ethanol Production from Pretreated Vegetable Processing Residues. Eur J Sustain Dev Res. https://doi.org/10.20897/ejosdr/3950

Nielsen, F., Galbe, M., Zacchi, G., and Wallberg, O. (2020) The effect of mixed agricultural feedstocks on steam pretreatment, enzymatic hydrolysis, and cofermentation in the lignocellulose-to-ethanol process. Biomass Convers Biorefinery. https://doi.org/10.1007/s13399-019-00454-w

Ntaikou, I., Menis, N., Alexandropoulou, M., Antonopoulou, G., and Lyberatos, G. (2018) Valorization of kitchen biowaste for ethanol production via simultaneous saccharification and fermentation using co-cultures of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. Bioresour Technol. https://doi.org/10.1016/j.biortech.2018.04.109

Oberoi, H.S., Babbar, N., Sandhu, S.K., Dhaliwal, S.S., Kaur, U., Chadha, B.S., and Bhargav, V.K. (2012) Ethanol production from alkali-treated rice straw via simultaneous saccharification and fermentation using newly isolated thermotolerant *Pichia kudriavzevii* HOP-1. J Ind Microbiol Biotechnol. https://doi.org/10.1007/s10295-011-1060-2

Öhgren, K., Bura, R., Lesnicki, G., Saddler, J., and Zacchi, G. (2007) A comparison between simultaneous saccharification and fermentation and separate hydrolysis and fermentation using steam-pretreated corn stover. Process Biochem. https://doi.org/10.1016/j.procbio.2007.02.003

Osorio-González, C.S., Hegde, K., Brar, S.K., Kermanshahipour, A., and Avalos-Ramírez, A. (2019) Challenges in lipid production from lignocellulosic biomass using *Rhodosporidium* sp.; A look at the role of lignocellulosic inhibitors. Biofuels, Bioprod Biorefining. https://doi.org/10.1002/bbb.1954

Palma, M., Guerreiro, J.F., and Sá-Correia, I. (2018) Adaptive response and tolerance to acetic acid in *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*: A physiological genomics perspective. Front Microbiol. https://doi.org/10.3389/fmicb.2018.00274

Pandey, A.K., Kumar, M., Kumari, S., Kumari, P., Yusuf, F., Jakeer, S., et al. (2019) Evaluation of divergent yeast genera for fermentation-associated stresses and identification of a robust sugarcane distillery waste isolate *Saccharomyces cerevisiae* NGY10 for lignocellulosic ethanol production in SHF and SSF. Biotechnol Biofuels. https://doi.org/10.1186/s13068-019-1379-x

Pellis, A., Cantone, S., Ebert, C., and Gardossi, L. (2018) Evolving biocatalysis to meet bioeconomy challenges and opportunities. N Biotechnol. https://doi.org/10.1016/j.nbt.2017.07.005

Pengilly, C., García-Aparicio, M.P., Diedericks, D., Brienzo, M., and Görgens, J.F. (2015) Enzymatic hydrolysis of steam-pretreated sweet sorghum bagasse by combinations of cellulase and endo-xylanase. Fuel. https://doi.org/10.1016/j.fuel.2015.03.072

Pomraning, K.R., Collett, J.R., Kim, J., Panisko, E.A., Culley, D.E., Dai, Z., et al. (2019) Transcriptomic analysis of the oleaginous yeast *Lipomyces starkeyi* during lipid accumulation on enzymatically treated corn stover hydrolysate. Biotechnol Biofuels. https://doi.org/10.1186/s13068-019-1510-

Poontawee, R., Yongmanitchai, W., and Limtong, S. (2017) Efficient oleaginous yeasts for lipid production from lignocellulosic sugars and effects of lignocellulose degradation compounds on growth and lipid production. Process Biochem. https://doi.org/10.1016/j.procbio.2016.11.013

Qadir, F., Shariq, M., Ahmed, A., and Sohail, M. (2018) Evaluation of a yeast co-culture for cellulase and xylanase production under solid state fermentation of sugarcane bagasse using multivariate approach. Ind Crops Prod. https://doi.org/10.1016/j.indcrop.2018.07.021

Quarterman, J., Slininger, P.J., Kurtzman, C.P., Thompson, S.R., and Dien, B.S. (2017) A

survey of yeast from the *Yarrowia* clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate. Appl Microbiol Biotechnol. https://doi.org/10.1007/s00253-016-8062-y

Robl, D., Delabona, S., Mergel, C.M., Rojas, J.D., Costa, S., Pimentel, I.C., et al. (2013) The capability of endophytic fungi for production of hemicellulases and related enzymes. BMC Biotech. https://doi.org/10.1186/1472-6750-13-94

Rodrigues, T.H.S., de Barros, E.M., de Sá Brígido, J., da Silva, W.M., Rocha, M.V.P., and Gonçalves, L.R.B. (2016) The Bioconversion of Pretreated Cashew Apple Bagasse into Ethanol by SHF and SSF Processes. Appl Biochem Biotechnol. https://doi.org/10.1007/s12010-015-1936-0

Rodríguez-Zúñiga, U.F., Cannella, D., Giordano, R.D.C., Giordano, R.D.L.C., Jørgensen, H., and Felby, C. (2015) Lignocellulose pretreatment technologies affect the level of enzymatic cellulose oxidation by LPMO. Green Chem. https://doi.org/10.1039/C4GC02179G

Rosales-Calderon, O. and Arantes, V. (2019) A review on commercial-scale high-value products that can be produced alongside cellulosic ethanol. BioMed Central. https://doi.org/10.1186/s13068-019-1529-1

Saini, J.K., Agrawal, R., Satlewal, A., Saini, R., Gupta, R., Mathur, A., and Tuli, D. (2015) Second generation bioethanol production at high gravity of pilot-scale pretreated wheat straw employing newly isolated thermotolerant yeast *Kluyveromyces marxianus*. DBTIOC-35. RSC Adv. https://doi.org/10.1039/C5RA05792B

Santos, R.B., Lee, J.M., Jameel, H., Chang, H.M., and Lucia, L.A. (2012) Effects of hardwood structural and chemical characteristics on enzymatic hydrolysis for biofuel production. Bioresour Technol. https://doi.org/10.1016/j.biortech.2012.01.085

Sarkar, N., Ghosh, S.K., Bannerjee, S., and Aikat, K. (2012) Bioethanol production from agricultural wastes: An overview. Renew Energy. https://doi.org/10.1016/j.renene.2011.06.045

Da Silva, A.S.A., Espinheira, R.P., Teixeira, R.S.S., De Souza, M.F., Ferreira-Leitão, V.,

and Bon, E.P.S. (2020) Constraints and advances in high-solids enzymatic hydrolysis of lignocellulosic biomass: A critical review. Biotechnol Biofuels. https://doi.org/10.1186/s13068-020-01697-w

Sindhu, R., Kuttiraja, M., Binod, P., Sukumaran, R.K., and Pandey, A. (2014) Bioethanol production from dilute acid pretreated Indian bamboo variety (*Dendrocalamus* sp.) by separate hydrolysis and fermentation. Ind Crops Prod. https://doi.org/10.1016/j.indcrop.2013.10.021

Singh, A. and Bishnoi, N.R. (2012) Enzymatic hydrolysis optimization of microwave alkali pretreated wheat straw and ethanol production by yeast. Bioresour Technol. https://doi.org/10.1016/j.biortech.2011.12.084

Sitepu, I., Selby, T., Lin, T., Zhu, S., and Boundy-Mills, K. (2014) Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. J Ind Microbiol Biotechnol. https://doi.org/10.1007/s10295-014-1447-y

Solomon, K. V., Haitjema, C.H., Henske, J.K., Gilmore, S.P., Borges-Rivera, D., Lipzen, A., et al. (2016) Early-branching gut fungi possess large, comprehensive array of biomass-degrading enzymes. Science. https://doi.org/10.1126/science.aad1431.

Sreeharsha, R.V. and Mohan, S.V. (2020) Obscure yet Promising Oleaginous Yeasts for Fuel and Chemical Production. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2020.02.004

Stegmann, P., Londo, M., and Junginger, M. (2020) The circular bioeconomy: Its elements and role in European bioeconomy clusters. Resour Conserv Recycl X. https://doi.org/10.1016/j.rcrx.2019.100029

Sukhang, S., Choojit, S., Reungpeerakul, T., and Sangwichien, C. (2020) Bioethanol production from oil palm empty fruit bunch with SSF and SHF processes using *Kluyveromyces marxianus* yeast. Cellulose. https://doi.org/10.1007/s10570-019-02778-2

Sun, Shaoni, Sun, Shaolong, Cao, X., and Sun, R. (2016) The role of pretreatment in improving the enzymatic hydrolysis of lignocellulosic materials. Bioresour Technol. https://doi.org/10.1016/j.biortech.2015.08.061

Suryawati, L., Wilkins, M.R., Bellmer, D.D., Huhnke, R.L., Maness, N.O., and Banat, I.M. (2008) Simultaneous saccharification and fermentation of Kanlow switchgrass pretreated by hydrothermolysis using *Kluyveromyces marxianus* IMB4. Biotechnol Bioeng. https://doi.org/10.1002/bit.21965

Tarasov, D., Leitch, M., and Fatehi, P. (2018) Lignin-carbohydrate complexes: Properties, applications, analyses, and methods of extraction: A review. Biotechnol Biofuels. https://doi.org/10.1186/s13068-018-1262-1

Unrean, P., Khajeeram, S., and Laoteng, K. (2016) Systematic optimization of fed-batch simultaneous saccharification and fermentation at high-solid loading based on enzymatic hydrolysis and dynamic metabolic modeling of *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol. https://doi.org/10.1007/s00253-015-7173-1

US9187390B2 - Lignin conversion process - Google Patents.

Usmani, Z., Sharma, M., Awasthi, A.K., Sivakumar, N., Lukk, T., Pecoraro, L., et al. (2021) Bioprocessing of waste biomass for sustainable product development and minimizing environmental impact. Bioresour Technol. https://doi.org/10.1016/j.biortech.2020.124548

Verhoeven, M.D., De Valk, S.C., Daran, J.M.G., Van Maris, A.J.A., and Pronk, J.T. (2018) Fermentation of glucose-xylose-Arabinose mixtures by a synthetic consortium of single-sugar-fermenting *Saccharomyces cerevisiae* strains. FEMS Yeast Res. https://doi.org/10.1093/femsyr/foy075

Wang, D., Sun, J., Yu, H.L., Li, C.X., Bao, J., and Xu, J.H. (2012) Maximum saccharification of cellulose complex by an enzyme cocktail supplemented with cellulase from newly isolated *Aspergillus fumigatus* ECU0811. Appl Biochem Biotechnol. https://doi.org/10.1007/s12010-011-9414-9

Wang, S., Cheng, G., Joshua, C., He, Z., Sun, X., Li, R., et al. (2016) Furfural tolerance and detoxification mechanism in *Candida tropicalis*. Biotechnol Biofuels. https://doi.org/10.1186/s13068-016-0668-x

Wu, J., Hu, J., Zhao, S., He, M., Hu, G., Ge, X., and Peng, N. (2018) Single-cell Protein and Xylitol Production by a Novel Yeast Strain *Candida intermedia* FL023 from Lignocellulosic Hydrolysates and Xylose. Appl Biochem Biotechnol. https://doi.org/10.1007/s12010-017-2644-8

Xu, C., Zhang, J., Zhang, Y., Guo, Y., Xu, H., Xu, J., and Wang, Z. (2019) Bioresource Technology Enhancement of high-solids enzymatic hydrolysis e ffi ciency of alkali pretreated sugarcane bagasse at low cellulase dosage by fed-batch strategy based on optimized accessory enzymes and additives. Bioresour Technol. https://doi.org/10.1016/j.biortech.2019.121993

Yennamalli, R.M., Rader, A.J., Kenny, A.J., Wolt, J.D., and Sen, T.Z. (2017) Endoglucanases : insights into thermostability for biofuel applications. Biotechnol Biofuels. https://doi.org/10.1186/1754-6834-6-136

Yuan, S.F., Guo, G.L., and Hwang, W.S. (2017) Ethanol production from dilute-acid steam exploded lignocellulosic feedstocks using an isolated multistress-tolerant *Pichia kudriavzevii* strain. Microb Biotechnol 10: 1581–1590. https://doi.org/10.1111/1751-7915.12712

Zeng, Y., Zhao, S., Yang, S., and Ding, S.Y. (2014) Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. Curr Opin Biotechnol. https://doi.org/10.1016/j.copbio.2013.09.008

Zoghlami, A. and Paës, G. (2019) Lignocellulosic Biomass: Understanding Recalcitrance and Predicting Hydrolysis. Front Chem. https://doi.org/10.3389/fchem.2019.00874

Chapter 2

Adaptive Laboratory Evolution to enhance *Kluyveromyces marxianus* tolerance to sugar beet pulp hydrolysates

Pooja Jayaprakash[1, 2], Francesca Martani [1], John P. Morrissey[2], Paola Branduardi[1]

[1] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, Milano, 20126, Italy.

[2] School of Microbiology, Environmental Research Institute, APC Microbiome Institute, University College Cork, Cork, T12 K8AF, Ireland.

Abstract

The lack of tolerance of *K. marxianus* to inhibitory compounds, particularly weak organic acids, produced during pretreatment, hinders its use when hydrolyzed LCB is used as substrates. In this work, through ALE approach, improved growth of *K. marxianus* in SBP at low pH was achieved. The aim was to isolate *K. marxianus* mutants with increased robustness towards lignocellulose-derived inhibitors, particularly organic acids in their undissociated form, present in SBP at low pH. Preliminary growth kinetics suggested a temperature trade-off when *K. marxianus* parental strain was exposed to SBP pH 4.3, as the growth of this strain was faster with a shorter lag phase at 30 °C than at 40 °C. However, there was lack of growth of this yeast in SBP at pH 3.0, in both the temperatures. Preadaptation of cells in SBP pH 4.3 facilitated the growth of the parental strain in the stressed condition of SBP pH 3.0 at 30 °C and allowed the growth at 40 °C. By measuring mRNA levels of two molecular determinants involved in ROS scavenging we could demonstrate that preadaptation triggers a response involving at least Sod enzymes and glutathione. Moreover, using ALE we obtained *K. marxianus* variants capable of growth in 8% SBP pH 3.0 without the need for preadaptation in SBP at pH 4.3, at least at 30 °C. Among these isolates, some show a significantly reduced lag phase or a higher specific growth rate compared to the initial strain. Further characterizations are required for phenotype-genotype association. In conclusion, *K. marxianus* with its natural capability to utilize a wide range of sugars associated with improved growth in sugar beet pulp hydrolysate gained through adaptive evolution may contribute to rendering it a reliable microbial host for sustainable production of industrially relevant compounds.

Introduction

Growing concerns over the need for a sustainable economy to combat global warming have shifted the attention from the use of fossil fuels to renewable biomass, such as agricultural and industrial waste (Limayem and Ricke, 2012; Kumar *et al.*, 2022; Saleem, 2022). Among the renewable biomasses, lignocellulosic biomass (LCB) is the most abundant renewable resource for the production of biofuels and chemicals. Within a biorefinery this biomass can be valorized in many ways among which the microbial transformation plays an important role, as LCB is rich in carbohydrates composed of sugars such as glucose, mannose, arabinose, and xylose.

The microbial conversion of LCB into final products has several advantages compared to chemical valorization. Exploiting this potential provides a low-cost, natural way to prepare biomass for biofuels production. Furthermore, the taxonomic diversity of microorganisms that modify or degrade lignocellulose provides a vast array of options for adopting such a method (Alessi *et al.*, 2018). The diverse nature of the biomass feedstock, together with the biodiversity of microorganisms, has the potential to lead to the production of a diverse array of molecules of industrial interest (Branduardi, 2021). A detailed description of these aspects is addressed in our review paper titled, 'Interdependence between lignocellulosic biomasses, enzymatic hydrolysis and yeast cell factories in biorefineries' (Bertacchi *et al.*, 2022). The LCB of interest in this study is sugar beet pulp (SBP), a byproduct obtained from processing of sugar beet and considered a potential feedstock for microbial production (Berłowska *et al.*, 2016). SBP contains high fraction of carbohydrates (24-32% hemicellulose, 22-30% cellulose and 38-62% pectin) and its low lignin content (<2%) allows the reduction of pretreatment costs preceding the enzymatic hydrolysis, compared to other lignocellulosic biomasses (Michel *et al.*, 1988; Berlowska *et al.*, 2018). Depending on the type of pretreatment and process parameters, different inhibitory compounds may be released such as weak organics acids like acetic acid, formic acid and levulinic acid; furans including furfural and hydroxymethylfurfural (HMF); and phenolic compounds (Palmqvist and Hahn-Hägerdal, 2000; Taherzadeh and Karimi, 2011). SBP is considered as the most abundant pectin-rich agro-industrial by-product. Based on the enzymatic cocktail used, neutral sugars such as glucose, arabinose, galactose, xylose are released during

enzymatic hydrolysis of SBP. Apart from this, the hydrolysis of pectin fraction releases sugars such as galacturonic acid that negatively impacts microbial growth and fermentation at low pH. Furthermore, when pectin structures are acetylated, their hydrolysis produces considerable quantities of acetic acid (Perpelea *et al.*, 2022), in addition to those derived from acetyl groups on hemicellulose polymers (Almeida *et al.*, 2007; Jönsson and Martín, 2016). Despite that, SBP has been successfully used as a substrate for studying the microbial production of bioethanol and biodiesel (Berlowska *et al.*, 2018; Martani *et al.*, 2020; Martins *et al.*, 2021) exploiting different yeasts.

In this regard, yeasts and particularly *Saccharomyces cerevisiae* have been the elected microorganism in many industrial applications, not only because a number of them have a long history of application within the brewing, bread or dairy industry, but also because of their natural resilience to low pH and ease of manipulation in haploid and diploid forms. Nonetheless, when moving to lignocellulose, not all the yeasts can metabolize pentose sugars, and *S. cerevisiae* is among those (Cai, Zhang and Li, 2012; Zhou *et al.*, 2012). Moreover, the tolerance towards organic acids at low pH is another requirement that needs to be matched with sugar consumption. Organic acids act as a double edge sword in bio-based processes since they are released from the pretreated LCB, inhibiting the fermentative process (Heer and Sauer, 2008), but, at the same time, they can be desirable products obtained by microbial fermentation as valuable building block chemicals with many applications in different industrial sectors (Sauer *et al.*, 2008). Therefore, one of the key instruments in the microbial conversion of lignocellulosic feedstock to advanced biofuels and other commodities is the development and usage of efficient, robust, versatile microbial cell factories with innate or engineered metabolism to consume all the sugars and to tolerate multiple stresses.

In this context, the non-Saccharomyces yeast *Kluyveromyces marxianus* represents a promising microbial cell factory for biorefinery applications. This yeast possesses several advantages such as a short generation time and high growth rate at elevated temperatures (0.86-0.99 h^{-1} at 40 °C), with an upper growth limit at 52 °C for some strains (Lane and Morrissey, 2010; Karim, Gerliani and Aïder, 2020). This trait has been widely exploited in industrial processes requiring high temperatures such as lignocellulosic fermentation for bioethanol production (Madeira-Jr and Gombert, 2018;

Suzuki, Hoshino and Matsushika, 2019). Additionally, *K. marxianus* has the capacity to utilize a broad range of substrates, conferred by plethora of sugar transporters (Donzella *et al.*, 2021) and catabolic pathways (Fonseca *et al.*, 2008; Wang *et al.*, 2018a). Omic studies on *K. marxianu*s are expanding the knowledge on this yeast and in recent years have been successful in correlating certain phenotypes with genotypes, some of which can be found in LCB valorization by fermentation processes. For example, starting from a comparative study based on transcriptomic and proteomic analysis of *K. marxianus*, *S. cerevisiae* and *Y. lipolytica* under different stressors such as low pH, high temperature and high osmotic pressure (Doughty *et al.*, 2020), some of the genes responsive to elevated temperatures were further characterized (Montini *et al.*, 2022). The knowledge acquired from these transcriptomic studies is useful for modifying yeasts for biotechnological processes where growth medium and conditions are sub-optimal. However, the limited tolerance of *K. marxianus* to organic acids at low pH constitutes an obstacle for an efficient conversion of lignocellulosic biomasses into value-added products. The application of rational metabolic engineering approaches to increase robustness of *K. marxianus* is not trivial as it still requires a deeper understanding of the pathways and mechanisms behind the phenotype of interest. The application of non-targeted strategies such as adaptive laboratory evolution (ALE) represents a possible alternative approach. Following ALE, synthetic biology tools can be applied for reverse engineering in order to better understand phenotype-genotype correlation, and this will be possible also in *K. marxianus* thanks to the recent extensive development of genome editing and synthetic biology tools (Rajkumar *et al.*, 2019; Rajkumar and Morrissey, 2020, 2022).

ALE is a common strategy in scientific studies to acquire insights into the underlying principles of molecular evolution and adaptive modifications that accumulate in microbial populations over time and under certain growth circumstances (Dragosits and Mattanovich, 2013). ALE has successfully been performed to increase the robustness of *S. cerevisiae* to hydrolysates (Almario, Reyes and Kao, 2013) as well as specific inhibitors such as furfural (Heer and Sauer, 2008), HMF (Wallace-Salinas and Gorwa-Grauslund, 2013) and for cellulosic biofuels production (Peris *et al.*, 2017). ALE can be performed in a) repeated batch cultivations in the presence of a selective pressure or alternatively, b) prolonged chemostat cultivations under selective

conditions (Mans, Daran and Pronk, 2018). In repeated batch cultivations cells are allowed to pass through lag and exponential growth phases before a small fraction of the culture is transferred into fresh medium. The procedure is repeated until the desired phenotype is obtained. Repetitive progress through the growth stages has been reported to result in a selection favoring certain characteristics such as a shorter lag phase and an increased specific growth rate at sub-limiting substrate concentrations (Mans, Daran and Pronk, 2018). In chemostat cultivations cells are cultivated in the presence of inhibitory compounds for extended periods. The advantages of chemostat cultivations are constant growth rates and population densities. Furthermore, it is possible to control nutrient supply and environmental conditions such as pH and oxygenation. In continuous cultures, the growth rate is kept constant by the limitation of a major growth nutrient, such as glucose, nitrogen or phosphate. The nutrient limitation results in a selection pressure for phenotypes with an increased affinity for the limited substrate and consequently a higher specific growth rate at suboptimal concentrations of the limiting nutrient (Wright *et al.*, 2011). In the literature, an ALE approach involving continuous cultivation for improving *K. marxianus* tolerance to ethanol was reported (Mo *et al.*, 2019). However, in this study we opted for an ALE protocol exploiting repeated batch cultivation in shake flasks, considering the advantages such as easy set up and ease of operating parallel cultures (Mans, Daran and Pronk, 2018).

The aim was to isolate *K. marxianus* mutants with increased robustness towards lignocellulose-derived inhibitors, particularly organic acids, present in SBP at low pH. We used SBP-based media at two different pH values, 4.3 and 3.0. The pH value 4.3 was attained by SBP preparation after the enzymatic hydrolysis. The second pH value 3.0 is the minimum desirable requirement in the case of organic acids with low pKa, such as lactic acid (pKa: 3.78), are the final products of interest. At pH 3.0, the acetic acid (pKa: 4.76) present in SBP is mainly in its undissociated form, therefore having a more detrimental effect than at higher pH. In industrial production, low pH is also an important parameter to minimize bacterial contamination. Furthermore, ALE was performed at two different temperatures, 30 ^{o}C and 40 ^{o}C, the second to favor the maintenance of the thermotolerance of *K. marxianus*.

Growth analysis in different media such as YPD, Verduyn and sugar beet pulp hydrolysate at 30 °C and 40 °C led to a temperature trade-off when *K. marxianus* parental strain was cultivated in SBP at pH 4.3, with cells being able to grow faster at 30 °C rather than at higher temperature as seen in YPD and Verduyn. Moreover, there was a lack of growth of this yeast in SBP at pH 3.0, in both the temperatures. As the aim was to select yeast variants able to grow in SBP at pH 3.0, we successfully tested the possibility of pre-adapting cells by pre-culturing in SBP at pH 4.3 prior to starting the ALE in SBP pH 3.0. The evolution was measured in terms of lag phase and specific growth rate of isolates from different days of evolution compared to the same parameters measured in the parental strain at T0. Over 105 days of evolution, corresponding to 603 generations at 30 °C and 507 generations at 40 °C, several isolates were checked for the reproducibility of the acquired phenotype and for selecting those that will be used for whole genome re-sequencing.

Materials and Methods

Raw biomass and pretreatment

SBP was provided by Cooperativa Produttori Bieticoli (CoProB), Minerbio (BO, Italy). The total solid content in the fresh SBP was 26%. SBP was stored at -20 °C and thawed prior to use. SBP was mixed with distilled water at the concentration of 250 g/L such that the solids were 65.5 g/L. The SBP with water was pretreated by autoclaving in a lab-scale autoclave at 121 °C for 20 min at a pressure of 1.2 bar to breakdown the SBP biomass rather than to sterilize it. After treatment, excess of water was removed from the pulp.

Enzymatic hydrolysis of SBP

Enzymatic hydrolysis was performed using the cocktail mixture NS22119 (provided by Novozymes), consisting of carbohydrases, including arabinase, β-glucanase, cellulase, hemicellulase, pectinase, and xylanase. Autoclaved SBP was mixed with sodium citrate 100 mM at pH 5.5 at total solids loading ranging from 10 to 70 g/L dry weight. The hydrolyses were carried out in Erlenmeyer flasks in an orbital shaker at 50 °C and 160 rpm for 72 h, as previously reported (Martani *et al.*, 2020). The samples were centrifuged at 10,000 rpm for 10 min and the supernatant was recovered and stored at 4 °C. The supernatant obtained was 10% g/L dry weight SBP hydrolysate. To perform growth kinetics in 6% and 8% SBP, dilutions were carried in water starting from the 10% SBP hydrolysate stock. The pH of the SBP hydrolysate after enzymatic hydrolysis was 4.3 and 1M HCl was used to lower the pH of the SBP hydrolysate to 3.0. The nitrogen content in 3% SBP was determined to be 27.0 mg/L (Martani *et al.*, 2020).

Strains and media

The strain used in this study was *K. marxianus* NBRC1777 (Biological Resource Centre, NITE (NBRC), Tokyo, Japan). The yeast was stored in cryotubes at -80 °C, in 20% glycerol v/v. Shake flask experiments were carried out in a range of different media; namely, YP medium (1% w/v yeast extract, 2% w/v peptone) with 2% w/v glucose (D) as carbon source; Verduyn minimal medium (Verduyn *et al.*, 1992) containing (per liter): 1 g of yeast extract (0.114 g of nitrogen), 1.31 g of $(NH_4)_2SO_4$ (0.278 g of nitrogen), 0.95 g of Na_2HPO_4, 2.7 g of KH_2PO_4, 0.2 g of $Mg_2SO_4{\cdot}7H_2O$, 0.04g. After the

pH was adjusted to 5.5 using NaOH 4 M, the medium was supplemented with glucose 2% , 10 mL of 100X trace mineral stock solution consisting of (per liter): 4 g $CaCl_2 \cdot 2H_2O$; 0.55 g $FeSO_4 \cdot 7H_2O$; 0.52 g citric acid; 0.10 g $ZnSO_4 \cdot 7H_2O$; 0.076 g $MnSO_4 \cdot H_2O$; and 100 µL of 18 M H_2SO_4 and 1 mL of 1000x vitamin stock.

Shake flask evolution experiment: cultivation conditions and strategy

The ALE experiment was performed as batch cultivation in shake flasks. As illustrated in Figure 1, starting from a fresh YPD agar plate, a single colony of the parental *K. marxianus* strain was used to inoculate a tube containing 2 mL YPD medium and was grown overnight at 30 °C or 40 °C, in an orbital shaker at 220 rpm. A second pre-inoculum of the cells was performed in 5 mL YPD and grown at 30 °C and 40 °C, 220 rpm until it reached exponential phase. Then, the appropriate volume of cells was used to inoculate a 100 ml shake flask containing 20 mL of 8% SBP hydrolysate at pH 4.3 to obtain an initial OD_{660nm} of 0.1. When the cells reached again exponential phase (OD_{660nm} 0.8-1.0), they were used to start the evolution by inoculating the 100 mL shake flasks containing 20 mL 8% SBP hydrolysate at pH 3.0 (limiting condition) at OD_{660nm} 0.1.

The ALE strategy consisted of consecutive serial transfers of the appropriate volume of culture in fresh medium with 8% SBP at pH 3.0 to a target OD_{660nm} of 0.1 whenever cells reached the early exponential growth phase. Consecutive transfer of the cells in limiting conditions was performed until a reduced lag phase and/or increased specific growth rate was observed. The experimental setting comprised of three independent biological replicates (shake flasks a, b and c) and two different temperatures, 30 °C and 40 °C (Figure 1). Cell growth was followed by measuring the OD_{660nm}. After each transfer of cells into fresh medium, samples were withdrawn and stored in glycerol vials at -80 °C for further evaluation. For pre-adaptation of the parental strain, the cells were grown in the similar way with two pre-inoculums in YPD followed by growth in 8% SBP at pH 4.3 until exponential phase and inoculated in 8% SBP pH 3.0 and growth was monitored.

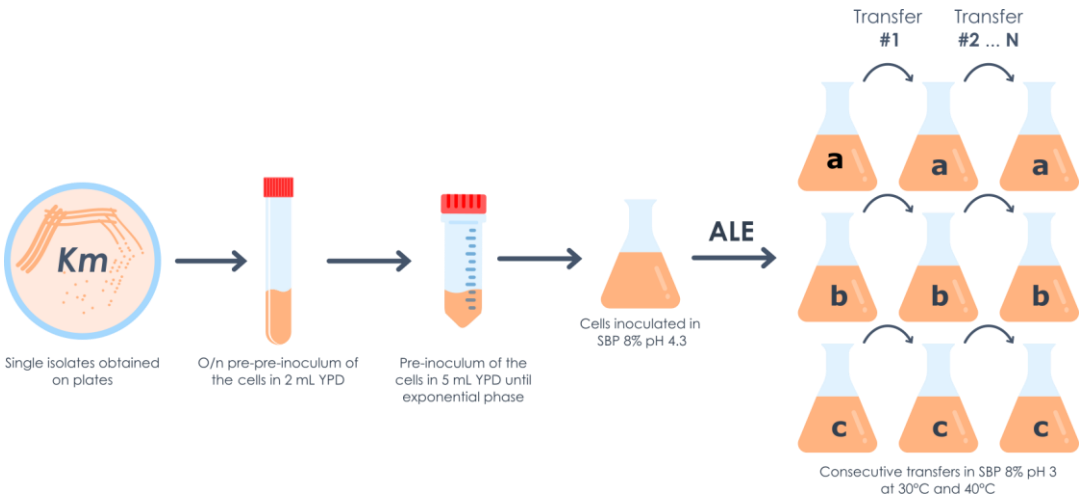


Figure 1: Schematic representation of the ALE experiment. Single isolates of the parental strain were pre-grown in 2 mL YPD overnight and after that in 5 mL YPD until exponential phase. From these cultures cells were inoculated in shake flasks with 8% SBP pH 4.3 until exponential phase, and from here the consecutive transfers in the same medium but at pH 3.0 (30 °C) started. "n" refers to the number of transfers performed; 102 transfers at 30 °C and 109 transfers at 40 °C.

Characterisation of evolved population

Single colonies of evolved isolates were obtained on YPD agar plates by plating sample collected at different days of the evolution experiment. As illustrated in Figure 2, to test the performances of the different isolates, they were pre-cultivated in 5 mL YPD until exponential phase and then transferred into 8% SBP pH 3.0 at OD_{660nm} 0.1. The growth of different isolates was monitored by OD_{660nm} measurements and lag phase and specific growth rate were calculated.

Real Time quantitative PCR

For checking the expression levels of superoxide dismutase (SOD) and glutathione (*GSH)* genes in parental strain in control condition (YPD) and in 8% SBP pH 4.3, total RNA was extracted using the ZR Fungal/Bacterial RNA Miniprep following the manufacturer's protocol (Zymoresearch). The retro transcription to obtain cDNA was performed using iScriptTM cDNA Synthesis kit (Bio-Rad Laboratories, Inc.). The obtained cDNA used to perform a real-time quantitative PCR using SsoFastTM EvaGreen® Supermix with Low ROX (Bio-Rad Laboratories, Inc.) with specific primers (table 1), according to the manufacturer's instructions. Results obtained were

normalized using actin as housekeeping gene (gene copy number and expression ratio=1).

Table 1: List of primers used in this work

Name	Sequence (5'-3')
GSH1_fw	TTGACCGACTTTGAGAACGC
GSH1_rv	TGTCGGCAAAGTCCGATTTC
SOD1_fw	CGATGCTAACGCCTTGAGAG
SOD1_rv	CACGCTACCCTTAGCAACAC
ACT1_fw	TCCAATCTACGCCGGTTTCT
ACT1_rv	GTTCGAAGTCCAAAGCGACA

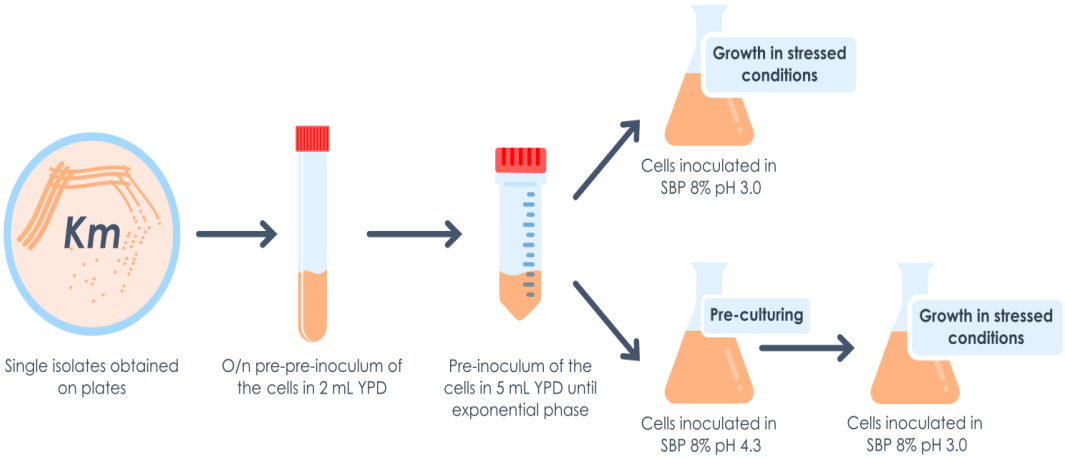


Figure 2: Schematic representation of steps involved in screening, pre-cultured *K. marxianus* NBRC1777 parental strain and non-precultured evolved isolates for growth in the stressed condition of 8% SBP at pH 3.0.

Results and Discussion

Effect of pH, temperature and SBP hydrolysate on growth of *K. marxianus*

The *K. marxianus* NBRC1777 yeast strain employed in this study is known for its potential for industrial but its tolerance towards organic acids is limited. To test the toxic effects of SBP hydrolysate, comprising galacturonic acid, acetic and lactic acid among other inhibitors, we carried out growth tests of the yeast in different concentrations of SBP (6%, 8% and 10%) at 30 °C. The concentrations selected were in accordance with previous work from our group, although the data was obtained from studies with a different yeast species (Martani *et al.*, 2020).

6% and 8% SBP were prepared from 10% SBP hydrolysates by dilution, which did not affect the initial pH value of 4.3. For clarity, the composition of 10%, 8% and 6% SBP hydrolysate is reported in Table 1. These were the media used for the growth kinetics, showing that the higher biomass of *K. marxianus* strain was obtained in 8% SBP compared to 6% and 10% SBP (Figure 3). Therefore, 8% SBP was selected for the study. The differences in growth can be attributed to the balance between the amounts of sugars and the concentration of inhibitory compounds. A similar observation was noticed in a previous study where *L. starkeyi* was cultivated in SBP hydrolysates from 1 to 7% TS. At low (1 and 2% TS) and high (6 and 7% TS) concentrations of SBP hydrolysates, the growth of *L. starkeyi* was poor or null, because of the limited availability of sugars or the inhibitory concentration of toxic compounds, respectively (Martani *et al.*, 2020). Comparing the studies, it is relevant to notice that the SBP concentration that is detrimental to *L. starkeyi* growth is the best among those tested for *K. marxianus* growth.

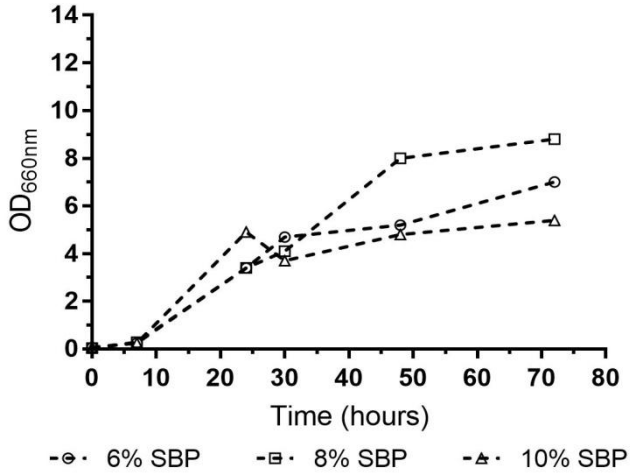


Figure 3: Growth profile of *K. marxianus* NBRC177 in SBP 6%, 8% and 10% pH 4.3 at 30 ^{o}C. Growth of the parental strain was monitored for 70 hours and OD_{660nm} was measured.

Table 1: Composition of SBP hydrolysate at 10%, 8% and 6% as determined by HPLC analysis. pH was 4.3 for all the preparations

Substance	**Concentration (g/L)**		
	10% SBP	8% SBP	6% SBP
Glucose	8.66	7.71	6.59
Galacturonic acid	3.43	2.82	2.05
Arabinose	8.75	7.06	6.65
Acetic acid	6.35	5.88	5.21
Xylose	3.28	3.04	2.86
Lactic acid	1.86	1.48	1.11

We analysed the growth kinetics of *K. marxianus* NBRC1777 at 30 °C and 40 °C (Figure 4 panels a and b, respectively) in rich medium (YPD pH 5.0), synthetic minimal medium (Verduyn + 2% glucose pH 5.0) and crude substrate (SBP) at pH 4.3 and pH 3.0 (considered an ideal condition for weak organic acid production, as explained in the Introduction). The specific growth rate and duration of the lag phase were calculated and compared (Figure 4 panels c and d, respectively). At 40 °C the kinetics were followed for 56 hours.

The growth of *K. marxianus* parental strain differs with the media used and temperatures tested. Panel a and b of Figure 1, reveals a similar growth in YPD at 30°C and 40 °C whereas, in synthetic minimal medium, the OD_{660nm} is higher at 30°C. In SBP pH 4.3, growth is faster at 30 °C compared to 40 °C. From Figure 4, panel c, we can

conclude that the specific growth rate (SGR) of *K. marxianus* was higher in YPD and Verduyn at 40 °C whereas, in SBP pH 4.3, the SGR was higher at 30 °C. Further, the lag phase was longer by 30hours at 40 °C for SBP pH 4.3 compared to 30 °C as seen in Figure 4, panel d. Therefore, in the presence of crude substrate, there is a trade-off for temperature and the lower temperature results as the more favorable condition for the growth of this yeast strain, in comparison to the growth in YPD and Verduyn + 2% glucose. This could be due to the synergistic stress of organic acids present in the SBP hydrolysate and the higher temperature, a condition that is exacerbated at lower pH where regardless of the temperature we do not observe growth.

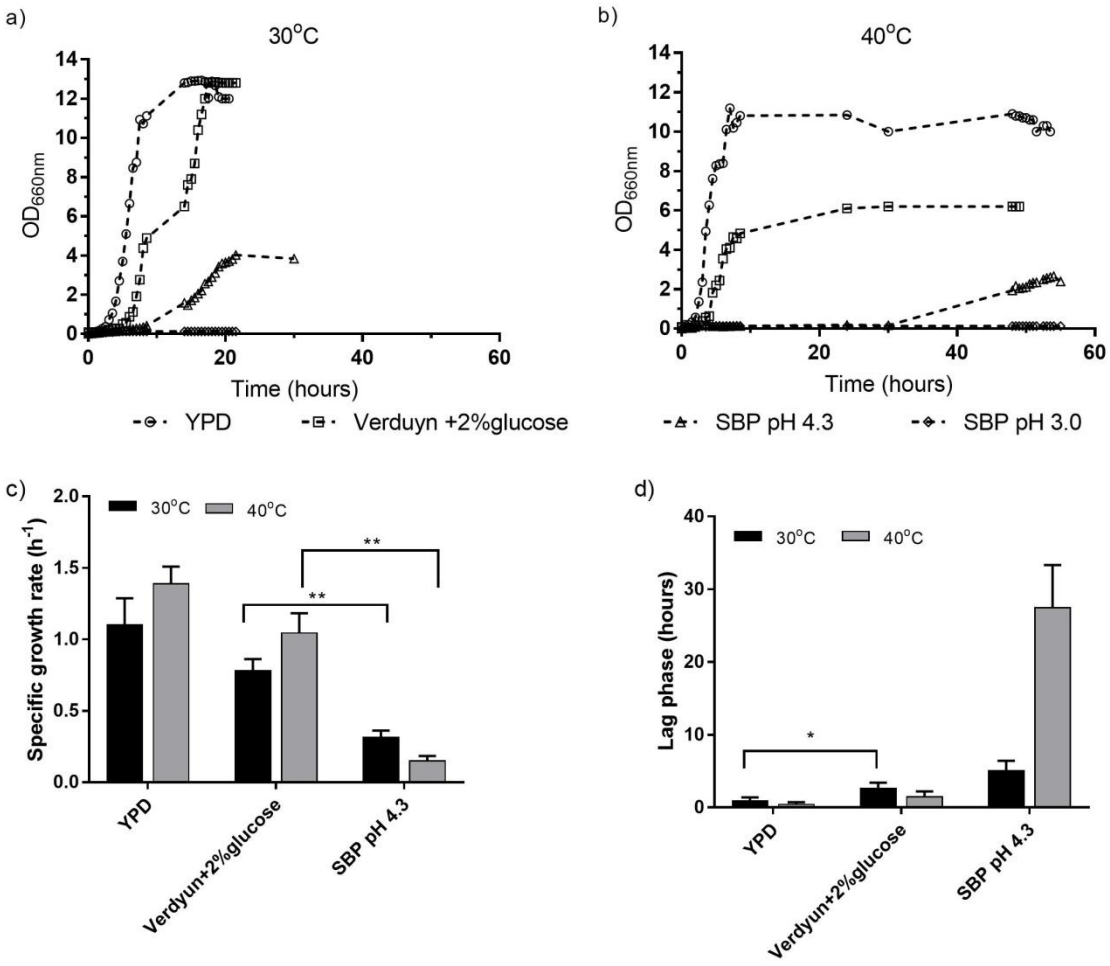


Figure 4: Growth profile of *K. marxianus* NBRC1777 in YPD, Verduyn+2% glucose and SBP hydrolysate at 30 °C (a) and 40 °C (b). Specific growth rate (c) and lag phase (d) of *K. marxianus* NBRC1777 in YPD, Verduyn+2% glucose and SBP hydrolysate at 30°C and 40°C. Values are the means of three independent experiments. * $p \leq 0.05$; ** $p \leq 0.005$; *** $p \leq 0.0005$

Overall, we observed a reduced growth, in terms of elongated lag phases and lower growth rates, in SBP, particularly at low pH. Considering the data from literature, we can hypothesize that the detrimental effect correlates with an increased passive diffusion of undissociated acetic and lactic acid into the cells, leading to acidification of the cytoplasm (Mira, Teixeira and Sá-Correia, 2010). Further, galacturonic acid, present in SBP although in minimum concentrations, can also enter the cell at pH below 3.5 and play a role in the synergistic stress. Also, this acid is involved in the inhibition of xylose fermentation (Huisjes *et al.*, 2012). In this condition, cells employ plasma membrane ATPases and multidrug resistance transporters to pump the protons and counter anions out of the cells, leading to an energy drain (Taherzadeh and Karimi, 2011; Piotrowski *et al.*, 2014; Caspeta, Castillo and Nielsen, 2015). Considering these preliminary analyses, we defined the suitable conditions to carry out the ALE experiment in the limiting / detrimental condition of SBP pH 3.0.

Effect of short term pre-culturing in SBP pH 4.3

It is known that cells react differently, if a stress is imposed abruptly or if they can adapt, rearranging their molecular and metabolic fluxes, under sub-limiting stressful conditions. This adaptation can be done evoking a general stress responses pathway such as a transient heat shock (Morano, Grant and Moye-Rowley, 2012) or can be done by simply reducing the stressor(s) concentrations or levels. In a work based on improving the tolerance of *S. cerevisiae* brewing strain to lignocellulosic inhibitors, cells were better able to cope at pH 3.7 if they were pre-grown at pH 5.0 in defined media formulated with addition of lignocellulosic inhibitors at pH 5.0 prior to inoculating them in media with same inhibitor cocktail at pH 3.7 (Narayanan *et al.*, 2016). Similarly, in this work we tested the effect of pre-culturing *K. marxianus* in 8% SBP pH 4.3% until exponential phase prior to moving cells in the same medium at pH 3.0 (Figure 5). As hypothesized, preadapted cells were able to grow at pH 3.0 in SBP at both temperatures, 30 °C and 40°C, whereas the unadapted cells again failed to grow at pH 3.0 at either temperature.

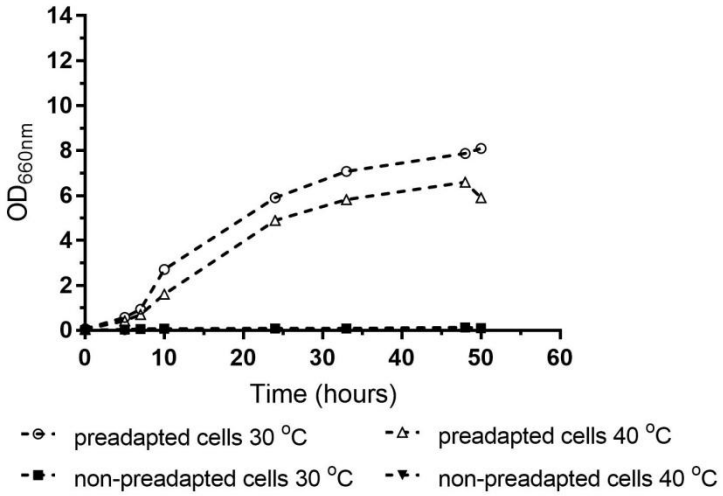


Figure 5: Growth profile of pre-adapted (empty symbols) and non-preadapted (full symbols) *K. marxianus* NBRC1777 cells at 30 °C and 40 °C. The pre-adapted cells were grown in 8% SBP pH 4.3 until exponential phase prior to growth them in 8% SBP pH 3.0.

The effect of short term pre-adaptation to inhibitors at higher pH might be to trigger a cellular response including the expression of various stress response genes and transcription factors/activators as described for *S. cerevisiae* (Modig *et al.*, 2008; Wright *et al.*, 2011) but more recently also for *K. marxianus* (Wang *et al.*, 2018), and that subsequently increases the tolerance to organic acids at lower pH.

Unveiling the molecular basis of the acquired robustness in preadapted *K. marxianus* cells

Lignocellulosic inhibitors such as acetic acid, HMF, furfural are known to lead to the generation of reactive oxygen species (ROS) that cause oxidative stress (Narayanan *et al.*, 2016). Indeed, the consequent response has been recently proven to be a good indicator of the acquisition of resistance to multiple stresses (Gurdo *et al.*, 2018). One of the defense mechanisms used by the cells to protect against oxidative stress is to increase the activity or the level of detoxifying enzymes or of antioxidants, such as superoxide dismutase (SOD) and glutathione (*GSH*), respectively (Wang *et al.*, 2018). *GSH1* encodes the rate-limiting step in yeast glutathione biosynthesis and *SOD1* encodes a Cu-Zn superoxide dismutase that plays a role in oxygen radical detoxification and in copper ion buffering (Perrone, Tan and Dawes, 2008). In a transcriptomic analysis study by Wang and co-workers involving *K. marxianus,*

tolerance to multiple inhibitors, *SOD1* and *SOD2* were up-regulated. Hypothesizing a similar triggering in *K. marxianus* cells pre-cultured in 8% SBP at pH 4.3, we investigated mRNA levels of *SOD1* and *GSH1* by RT-qPCR and compared to the levels measured in control condition (YPD). Notably, there is an increase in the expression levels of both genes in cells cultivated in SBP pH 4.3 (Figure 6). The results obtained provide insights for future metabolic engineering approach to obtain more robust strains for industrial fermentation with cellulosic biomass.

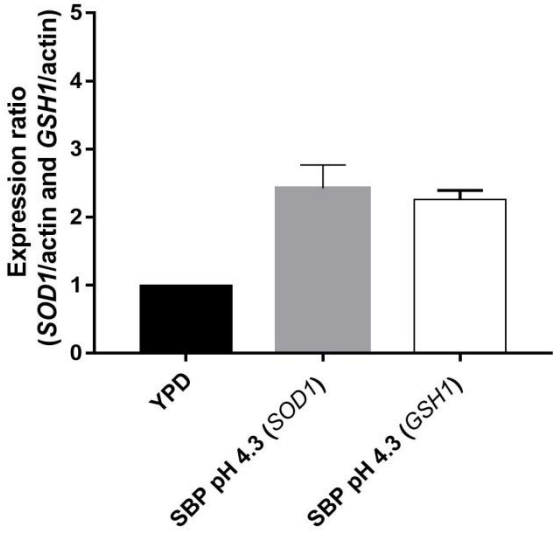


Figure 6: Expression levels of *SOD1* (grey bar) and *GSH1* (white bar) genes. Expression levels of *SOD1* and *GSH1* analysed during exponential phase in *K. marxianus* NBRC177 cells grown in control condition YPD pH 4.3 (black bar) and cells pre-adapted in SBP pH 4.3. The expression level of the actin gene was stable in the different conditions and defined as unity. Error bars indicate standard deviation, n=3.

ALE of *K. marxianus* in SBP

Considering the selection of *K. marxianus* strains ready to take on the harshness of industrial conditions without requiring prior adaption, we set up an ALE experiment. As illustrated in Figure 1, the parental strain was pre-grown in 8% SBP at pH 4.3 until the exponential phase, prior to transferring the cells to 8% SBP pH 3.0. Because evolution can proceed starting from casual initial events and might lead to very distant evolved strains, three independent evolution culture lines were initiated. Consecutive transfers of cells were done into fresh medium of 8% SBP pH 3.0 when the exponential phase was reached. After a total of 603 and 507 generations at 30 °C and 40 °C, respectively,

the experiment ended and the pertinence and progress of the adaptation was evaluated.

The evolved isolates were obtained from different days of evolution, as indicated in Figure 7. The evolved isolates were re-streaked on YPD plates for obtaining single colonies to be screened for their performance in the stressed condition. The parental strain and evolved isolates were tested for growth in 8% SBP pH 3.0 without a pre-adaptation step, at 30 °C and 40 °C. As before, the parental strain failed to grow at both temperatures in 8% SBP pH 3.0. At 40 °C even the evolved isolates failed to grow in 8% SBP pH 3.0, whereas they could grow at 30 °C. The growth of evolved isolate obtained at 30 °C when tested at 40 °C, showed growth inhibition. Therefore, the phenotypic robustness against the inhibitors in SBP at lowest pH was not retained at high temperature.

Finally, we wanted to answer to our initial question quantifying growth in terms of lag phase and specific growth rate. During an ALE experiment some traits can keep on improving, but others can be also lost: for that reason, we selected *line b* among the three independent evolution lines, and for each of the 21 days of sampling we selected three independent clones (namely *evob1, evob2 and evob3*) to calculate lag phase length and specific growth rate (Figure 7 panel a and b, respectively). Overall, it can be concluded that that apart from a couple of exceptions, all the independent clones could grow in 8% SBP pH 3.0 without preadaptation, but they are always very dissimilar, demonstrating a high degree of phenotypic heterogeneity. Moreover, it is not possible to trace a trend, either positive or negative starting from day 7 to day 117, for lag phase length or for specific growth rate.

Nevertheless, it is possible to spot the most interesting clones: the lowest lag phase (13 hours) was measured in isolate *83evob2*, but its specific growth rate (0.13 h^{-1}) is similar to most of the other isolates (Figure 7 panel b), exhibiting longer lag phases. Another evolved isolate *20evob3*, exhibited the highest growth rate as in panel b of Figure 7, but in parallel did not have a shorter lag phase. This could be attributed to: 1) The pre-culturing step in SBP pH 4.3 at the beginning of ALE when cells could have undergone a brief period of pre-adaptation to the inhibitors, thereby acquiring a

growth advantage to perform, in a similar way, at pH 3. 2) Lack of gradual increase in selection pressure such as concentration of organic acids in SBP over a period of time.

To determine the molecular mechanism behind the improved robustness of the evolved isolates, whole genome re-sequencing of these isolates is required to identify and test the causal single nucleotide polymorphisms (SNPs). Reverse engineering can be used to confirm the role of the mutations involved in the robust phenotype.

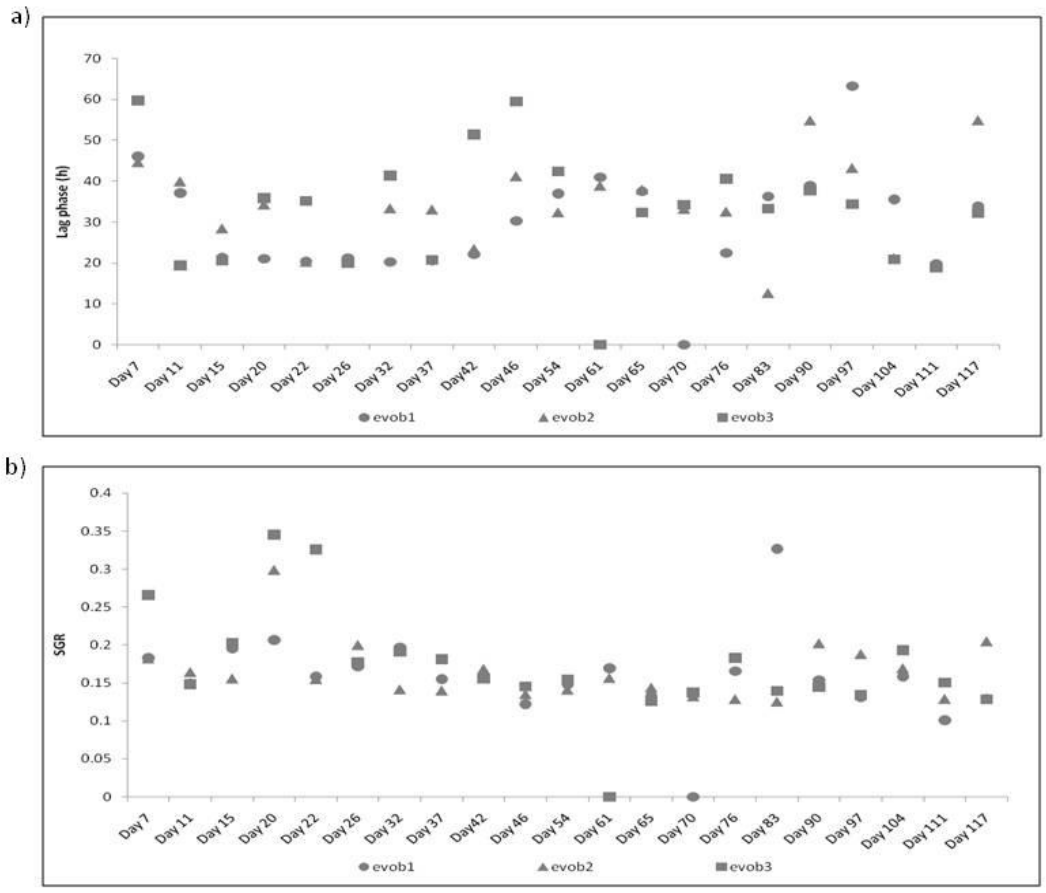


Figure 7: Evaluation of best performing isolates by screening for growth in 8% SBP pH 3.0 at 30 °C without a pre-culturing step; calculation of lag phase (a) and specific growth rate (b) of three different isolates 1, 2 and 3 taken from one of the triplicate shake flask, referred to as line 'b', during different days of the evolution experiment.

References

Alessi, A.M., Bird, S.M., Oates, N.C., Li, Y., Dowle, A.A., Novotny, E.H., Deazevedo, E.R., Bennett, J.P., Polikarpov, I., Young, J.P.W., McQueen-Mason, S.J., Bruce, N.C., 2018. Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-018-1164-2

Almario, M.P., Reyes, L.H., Kao, K.C., 2013. Evolutionary engineering of *Saccharomyces cerevisiae* for enhanced tolerance to hydrolysates of lignocellulosic biomass. Biotechnol. Bioeng. https://doi.org/10.1002/bit.24938

Almeida, J.R.M., Modig, T., Petersson, A., Hähn-Hägerdal, B., Lidén, G., Gorwa-Grauslund, M.F., 2007. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. J. Chem. Technol. Biotechnol. https://doi.org/10.1002/jctb.1676

Berlowska, J., Cieciura-Włoch, W., Kalinowska, H., Kregiel, D., Borowski, S., Pawlikowska, E., Binczarski, M., Witonska, I., 2018. Enzymatic conversion of sugar beet pulp: A comparison of simultaneous saccharification and fermentation and separate hydrolysis and fermentation for lactic acid production. Food Technol. Biotechnol. https://doi.org/10.17113/ftb.56.02.18.5390

Berłowska, J., Pielech-Przybylska, K., Balcerek, M., Dziekońska-Kubczak, U., Patelski, P., Dziugan, P., Kręgiel, D., 2016. Simultaneous Saccharification and Fermentation of Sugar Beet Pulp for Efficient Bioethanol Production. Biomed Res. Int. https://doi.org/10.1155/2016/3154929

Bertacchi, S., Jayaprakash, P., Morrissey, J.P., Branduardi, P., 2022. Interdependence between lignocellulosic biomasses, enzymatic hydrolysis and yeast cell factories in biorefineries. Microb. Biotechnol. https://doi.org/10.1111/1751-7915.13886

Branduardi, P., 2021. Closing the loop: the power of microbial biotransformations from traditional bioprocesses to biorefineries, and beyond. Microb. Biotechnol. https://doi.org/10.1111/1751-7915.13713

Cai, Z., Zhang, B., Li, Y., 2012. Engineering *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: Reflections and perspectives. Biotechnol. https://doi.org/10.1002/biot.201100053

Caspeta, L., Castillo, T., Nielsen, J., 2015. Modifying yeast tolerance to inhibitory conditions of ethanol production processes. Front. Bioeng. Biotechnol. https://doi.org/10.3389/fbioe.2015.00184

Donzella, L., Varela, J.A., Sousa, M.J., Morrissey, J.P., 2021. Identification of novel pentose transporters in *Kluyveromyces marxianus* using a new screening platform. FEMS Yeast Res. https://doi.org/10.1093/femsyr/foab026

Doughty, T.W., Domenzain, I., Millan-Oropeza, A., Montini, N., de Groot, P.A., Pereira, R., Nielsen, J., Henry, C., Daran, J.M.G., Siewers, V., Morrissey, J.P., 2020. Stress-induced expression is enriched for evolutionarily young genes in diverse budding yeasts. Nat. Commun. https://doi.org/10.1038/s41467-020-16073-3

Dragosits, M., Mattanovich, D., 2013. Adaptive laboratory evolution - principles and applications for biotechnology. Microb. Cell Fact. https://doi.org/10.1186/1475-2859-12-64

Fonseca, G.G., Heinzle, E., Wittmann, C., Gombert, A.K., 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-008-1458-6

Gurdo, N., Novelli Poisson, G.F., Juárez, B., Rios de Molina, M.C., Galvagno, M.A., 2018. Improved robustness of an ethanologenic yeast strain through adaptive evolution in acetic acid is associated with its enzymatic antioxidant ability, Journal of Applied Microbiology. https://doi.org/10.1111/jam.13917

Heer, D., Sauer, U., 2008. Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. Microb. Biotechnol. https://doi.org/10.1111/j.1751-7915.2008.00050.x

Huisjes, E.H., de Hulster, E., van Dam, J.C., Pronk, J.T., van Maris, A.J.A., 2012. Galacturonic acid inhibits the growth of *Saccharomyces cerevisiae* on galactose, xylose, and arabinose. Appl. Environ. Microbiol.

https://doi.org/10.1128/AEM.07617-11

Jönsson, L.J., Martín, C., 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2015.10.009

Karim, A., Gerliani, N., Aïder, M., 2020. *Kluyveromyces marxianus*: An emerging yeast cell factory for applications in food and biotechnology. Int. J. Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2020.108818

Kumar, P., Sanjukta, S., Latika, S., Koel, B., Divya, S., 2022. Utilization of agricultural waste biomass and recycling toward circular bioeconomy United State of America. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-022-20669-1

Lane, M.M., Morrissey, J.P., 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. Fungal Biol. Rev. https://doi.org/10.1016/j.fbr.2010.01.001

Limayem, A., Ricke, S.C., 2012. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. Prog. Energy Combust. Sci. https://doi.org/10.1016/j.pecs.2012.03.002

Madeira-Jr, J.V., Gombert, A.K., 2018. Towards high-temperature fuel ethanol production using *Kluyveromyces marxianus*: On the search for plug-in strains for the Brazilian sugarcane-based biorefinery. Biomass and Bioenergy. https://doi.org/10.1016/j.biombioe.2018.09.010

Mans, R., Daran, J.M.G., Pronk, J.T., 2018. Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2017.10.011

Martani, F., Maestroni, L., Torchio, M., Ami, D., Natalello, A., Lotti, M., Porro, D., Branduardi, P., 2020. Conversion of sugar beet residues into lipids by Lipomyces starkeyi for biodiesel production. Microb. Cell Fact. https://doi.org/10.1186/s12934-020-01467-1

Martins, L.C., Palma, M., Angelov, A., Nevoigt, E., Liebl, W., Sá-Correia, I., 2021. Complete utilization of the major carbon sources present in sugar beet pulp

hydrolysates by the oleaginous red yeasts *Rhodotorula toruloides* and R. Mucilaginosa. J. Fungi. https://doi.org/10.3390/jof7030215

Michel, F., Thibault, J. -F, Barry, J. -L, de Baynast, R., 1988. Preparation and characterisation of dietary fibre from sugar beet pulp. J. Sci. Food Agric. https://doi.org/10.1002/jsfa.2740420109

Mira, N.P., Teixeira, M.C., Sá-Correia, I., 2010. Adaptive response and tolerance to weak acids in *Saccharomyces cerevisiae*: A genome-wide view. Omi. A J. Integr. Biol. https://doi.org/10.1089/omi.2010.0072

Mo, W., Wang, M., Zhan, R., Yu, Y., He, Y., Lu, H., 2019. *Kluyveromyces marxianus* developing ethanol tolerance during adaptive evolution with significant improvements of multiple pathways. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-019-1393-z

Modig, T., Almeida, J.R.M., Gorwa-Grauslund, M.F., Lidén, G., 2008. Variability of the response of *Saccharomyces cerevisiae* strains to lignocellulose hydrolysate. Biotechnol. Bioeng. https://doi.org/10.1002/bit.21789

Montini, N., Doughty, T.W., Domenzain, I., Fenton, D.A., Baranov, P. V., Harrington, R., Nielsen, J., Siewers, V., Morrissey, J.P., 2022. Identification of a novel gene required for competitive growth at high temperature in the thermotolerant yeast *Kluyveromyces marxianus*. Microbiology. https://doi.org/10.1099/mic.0.001148

Morano, K.A., Grant, C.M., Moye-Rowley, W.S., 2012. The response to heat shock and oxidative stress in *Saccharomyces cerevisiae*. Genetics. https://doi.org/10.1534/genetics.111.128033

Narayanan, V., Sànchez i Nogué, V., van Niel, E.W.J., Gorwa-Grauslund, M.F., 2016. Adaptation to low pH and lignocellulosic inhibitors resulting in ethanolic fermentation and growth of *Saccharomyces cerevisiae*. AMB Express. https://doi.org/10.1186/s13568-016-0234-8

Palmqvist, E., Hahn-Hägerdal, B., 2000. Fermentation of lignocellulosic hydrolysates. II: Inhibitors and mechanisms of inhibition. Bioresour. Technol. https://doi.org/10.1016/S0960-8524(99)00161-3

Peris, D., Moriarty, R. V., Alexander, W.G., Baker, E.C., Sylvester, K., Sardi, M., Langdon, Q.K., Libkind, D., Wang, Q.M., Bai, F.Y., Leducq, J.B., Charron, G., Landry, C.R., Sampaio, J.P., Gonçalves, P., Hyma, K.E., Fay, J.C., Sato, T.K., Hittinger, C.T., 2017. Hybridization and adaptive evolution of diverse *Saccharomyces species* for cellulosic biofuel production. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-017-0763-7

Perpelea, A., Wijaya, A.W., Martins, L.C., Rippert, D., Klein, M., Angelov, A., Peltonen, K., Teleki, A., Liebl, W., Richard, P., Thevelein, J.M., Takors, R., Sá-Correia, I., Nevoigt, E., 2022. Towards valorization of pectin-rich agro-industrial residues: Engineering of *Saccharomyces cerevisiae* for co-fermentation of D-galacturonic acid and glycerol. Metab. Eng. https://doi.org/10.1016/j.ymben.2021.10.001

Perrone, G.G., Tan, S.X., Dawes, I.W., 2008. Reactive oxygen species and yeast apoptosis. Biochim. Biophys. Acta - Mol. Cell Res. https://doi.org/10.1016/j.bbamcr.2008.01.023

Piotrowski, J.S., Zhang, Y., Bates, D.M., Keating, D.H., Sato, T.K., Ong, I.M., Landick, R., 2014. Death by a thousand cuts: The challenges and diverse landscape of lignocellulosic hydrolysate inhibitors. Front. Microbiol. https://doi.org/10.3389/fmicb.2014.00090

Rajkumar, A.S., Morrissey, J.P., 2022. Protocols for marker-free gene knock-out and knock-down in *Kluyveromyces marxianus* using CRISPR/Cas9. FEMS Yeast Res. https://doi.org/10.1093/femsyr/foab067

Rajkumar, A.S., Morrissey, J.P., 2020. Rational engineering of *Kluyveromyces marxianus* to create a chassis for the production of aromatic products. Microb. Cell Fact. https://doi.org/10.1186/s12934-020-01461-7

Rajkumar, A.S., Varela, J.A., Juergens, H., Daran, J.M.G., Morrissey, J.P., 2019. Biological parts for *Kluyveromyces marxianus* synthetic biology. Front. Bioeng. Biotechnol. https://doi.org/10.3389/fbioe.2019.00097

Saleem, M., 2022. Possibility of utilizing agriculture biomass as a renewable and sustainable future energy source. Heliyon.

https://doi.org/10.1016/j.heliyon.2022.e08905

Sauer, M., Porro, D., Mattanovich, D., Branduardi, P., 2008. Microbial production of organic acids: expanding the markets. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2007.11.006

Suzuki, T., Hoshino, T., Matsushika, A., 2019. High-temperature ethanol production by a series of recombinant xylose-fermenting *Kluyveromyces marxianus* strains. Enzyme Microb. Technol. https://doi.org/10.1016/j.enzmictec.2019.109359

Taherzadeh, M.J., Karimi, K., 2011. Fermentation inhibitors in ethanol processes and different strategies to reduce their effects, 1st ed, Biofuels. Elsevier Inc. https://doi.org/10.1016/B978-0-12-385099-7.00012-7

Verduyn, C., Postma, E., Scheffers, W.A., van Dijken, J.P., 1992. Effect of Benzoic Acid on Metabolic Fluxes in Yeasts: A continuous-culture study on the regulation of respiraton and alcoholic fermentation. Yeast. https://doi.org/10.1002/yea.320080703

Wallace-Salinas, V., Gorwa-Grauslund, M.F., 2013. Adaptive evolution of an industrial strain of *Saccharomyces cerevisiae* for combined tolerance to inhibitors and temperature. Biotechnol. Biofuels. https://doi.org/10.1186/1754-6834-6-151

Wang, D., Wu, D., Yang, X., Hong, J., 2018b. Transcriptomic analysis of thermotolerant yeast: *Kluyveromyces marxianus* in multiple inhibitors tolerance. RSC Adv. https://doi.org/10.1039/c8ra00335a

Wright, J., Bellissimi, E., de Hulster, E., Wagner, A., Pronk, J.T., Van Maris, A.J.A., 2011. Batch and continuous culture-based selection strategies for acetic acid tolerance in xylose-fermenting *Saccharomyces cerevisiae*. FEMS Yeast Res. https://doi.org/10.1111/j.1567-1364.2011.00719.x

Zhou, H., Cheng, J. sheng, Wang, B.L., Fink, G.R., Stephanopoulos, G., 2012. Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by *Saccharomyces cerevisiae*. Metab. Eng.https://doi.org/10.1016/j.ymbe.2012.07.011

Chapter 3

Contribution of *PDR12* to weak organic acids resistance in the yeast *Zygosaccharomyces parabailii*

Pooja Jayaprakash[1, 2], John P. Morrissey[2], Paola Branduardi[1]

[1] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, Milano, 20126, Italy.

[2] School of Microbiology, Environmental Research Institute, APC Microbiome Institute, University College Cork, Cork, T12 K8AF, Ireland.

Abstract

The ability of the hybrid yeast *Zygosaccharomyces parabailii* to thrive in the presence of organic acids can be greatly exploited for industrial production. Understanding the mechanism involved in tolerance to organic acids can be used for avoiding the growth of this yeast in food production pipelines as well as to promote its use as cell factory for the production of organic acids and other bio-products. To understand the phenotype-genotype correlation involved in the organic acid tolerance trait, we investigated the role of *PDR12*. *PDR12* is a member of the ATP binding cassette transporter family, known to have a role in tolerance towards sorbic, benzoic and propionic acids in *Saccharomyces cerevisiae*. In this study, we constructed and characterized a set of single and double *Z. parabailii pdr12* mutants. As comparison, *S. cerevisiae* wild-type and *pdr12* mutant strains were also used. The phenotypic characterization revealed that Pdr12p in *Z. parabailii* is not involved in tolerance towards longer chain organic acids stress such as sorbic and benzoic, while it is involved in mediating tolerance to butyric and acetic acids. Interestingly, individual mutants displayed phenotypes consistent with differential roles for each allele as only the *PDR12* allele on chromosome 8 contributed to tolerance towards butyric acid. The structural analysis of the two Pdr12p isoforms showed 93% identity and 97% similarity. Further investigations will concentrate on the amino acidic differences that could justify the different contribution of the two isoforms to butyric acid tolerance.

Introduction

Microbial bio-based production of organic acids as platform chemicals via fermentation offers diverse advantages related to the cycle of circular bioeconomy (Becker and Wittmann, 2019; Heer and Sauer, 2008). Microbial production using yeasts as hosts have received considerable interest as they can be more cost effective when dealing with downstream processing and purity of the product, owing to the advantages of growth in low pH and defined media (Di Lorenzo et al., 2022). Likewise, engineered *S. cerevisiae* is used for the production of succinic acid (Ito et al., 2014; Raab et al., 2010), lactic acid (Sauer et al., 2008), and fumaric acid (Xu et al., 2012), which are used as versatile chemical platforms in several industrial sectors. Microbial citric, succinic, lactic, itaconic, lactobionic, gluconic, fumaric, propionic, and acetic acids have approached the market, with varying degrees of success. Additionally, the use of lignocellulosic material for production of biochemicals offers several advantages such as low cost, ubiquitous in nature, rich in sugars and does not compete with the food chain (Sun and Cheng, 2002). However, lignocellulosic biomass must undergo a pretreatment step followed by enzymatic hydrolysis to release monomeric sugars. Depending on the applied methods, different sugars can be further degraded into inhibitory compounds such as furans, phenols, furfural and weak organic acids (WOA) (Limayem and Ricke, 2012). Among these, acetic acid is the most abundant (Jönsson and Martín, 2016). In addition to other inhibitory compounds, the concentration of acetic acid can be detrimental to the overall fermentation process. Therefore, understanding the mechanism of weak organic acid resistance is important to develop robust microbial hosts which can use lignocellulosic biomass as substrate to produce bulk chemicals such as weak organic acids.

Most of the studies based on weak acid response in yeasts are derived predominantly from *S. cerevisiae* and the response has not been widely studied beyond this yeast, especially considering non-conventional yeasts that evolved in specific habitats and possess beneficial traits that can have significance in yeast based biorefineries. *Zygosaccharomyces bailii* is one such yeast that has been studied over the past years and is described as both friend and foe. This yeast is extremely tolerant to conditions that are usually detrimental for cell growth such as high osmotic pressure, low pH, high

ethanol concentrations and the presence of weak organic acids and/or various food preservatives such as acetic, sorbic, benzoic and propionic acids (Kuanyshev et al., 2017). In the past, *Z. bailii* has been engineered for the production of L-ascorbic acid (vitamin C)(Sauer et al., 2004), lactic acid (Branduardi et al., 2004), as well as used for bioethanol production (Paixão et al., 2013). Previous work from our lab suggested that the interspecies hybrid yeast, *Zygosaccharomyces parabailii*, is more robust than some of the haploid *Z. bailii* strains, particularly in its tolerance to weak organic acids such as lactic acid, which is one of the important commodity chemicals that can be produced by microbial fermentation (Kuanyshev et al., 2016). This hybrid is formed by mating between *Z. bailii sensu stricto* and an unidentified *Zygosaccharomyces* species. Further, the annotation and assembly of *Z. parabailii* genome suggested that two copies of almost every gene, differing by 7% in nucleotide sequence on average, was present (Ortiz-Merino et al., 2017). Understanding the mechanism involved in tolerance to organic acids can be used for avoiding the growth of this yeast in food production pipelines as well as to promote its use as cell factory for the production of organic acids and other bio-products.

Studies on *S. cerevisiae* demonstrated that it employs different cellular mechanisms in response to WOAs stress such as expression of transporters to extrude protons and toxic counter ions, modification of the cell membrane and wall to restrict WOA diffusion, and activation of antioxidant mechanisms to counter intracellular damage. Additionally, *S. cerevisiae* can also modify its phospholipids and ergosterol content in order to reduce cell membrane permeability, thereby reducing the overall diffusion of WOAs (Berterame et al., 2016). The primary response of the yeast cell to WOA stress is to maintain intracellular pH homeostasis via proton extrusion from the cytoplasm (Piper et al., 2001). To avoid toxic counter ion buildup, yeast activates multidrug resistance transporters. In *S. cerevisiae* the expression of WOA-specific transporters is controlled by the transcription factors War1p and Haa1p. Haa1p controls *TPO2* and *TPO3*, both encoding H^+ antiporters and responsible for weak acid anion extrusion, namely acetic, propionic, and benzoic acids, and were formerly considered to be involved in polyamine transport (Mira et al., 2010b). War1p is rapidly phosphorylated upon weak acid stress and thus activated to induce *PDR12*, a member of the ATP binding cassette transporter family known to have a role in tolerance towards sorbic,

benzoic and propionic acids by exporting the counter-anions out of cells. *PDR12* is rapidly induced in response to sorbic and benzoic acids stress, resulting in cell membrane levels of Pdr12p that are comparable to Pma1p, the most prevalent plasma membrane protein (Kren et al., 2003; Piper et al., 2001). On the other hand, *PDR12* is not expressed in response to acetic or formic acids (Hatzixanthis et al., 2003), as its absence improves strain robustness in this situation (Nygard et al., 2014). Other investigations have found that Pdr12p is responsible for active extrusion of longer chain organic acids like propionate, sorbate, and benzoate (Piper et al., 1998). *PDR12* gene was found to be strongly induced in *S. cerevisiae* strains isolated from spoiled foods when sorbic acid is present in culture medium but in *Z. bailii* only weak induction of Pdr12p was seen in the presence of sorbic acid (Papadimitriou et al., 2007; Piper et al., 2001). However, there is a lack of studies examining the wider role of *PDR12* in *Z. bailii* in response to acetic, butyric, lactic, and sorbic and benzoic acid stress.

The goal of our study was to determine whether Pdr12p is involved in tolerance to different weak organic acids in *Z. parabailii*. In addition, as the hybrid *Z. parabailii* carries two alleles of *PDR12* that derive from distinct parental genomes, we asked whether there could be a differential contribution of an allele from one or other of the sub-genomes. To do this, we constructed and characterized a set of single and double *Z. parabailii pdr12* mutants. As comparison, *S. cerevisiae* wild-type and *pdr12* mutant strains were used. These analyses determined that Pdr12p is not involved in tolerance towards longer chain organic acids stress such as sorbic and benzoic, while it is involved in mediating tolerance to butyric and acetic acids in *Z. parabailii*, though individual mutants displayed phenotypes consistent with differential roles for each parental allele. Furthermore, analysis of the Pdr12p sequence provided insights in the amino acids differences and the percentage of non-conservative substitutions of amino acids.

Materials and Methods

Strains and growth conditions

Z. parabailii ATCC60483 (Zygo2) strain was used in this study (Suh et al., 2013), and its complete genome sequence is available (GCA_001984395.2). Other strains used and constructed over the course of this work are listed in Table 1. Yeast cultures were grown in YP medium (1% w/v yeast extract, 2% w/v peptone) with 2% w/v glucose (D) as carbon source, or minimal synthetic medium (0.67% w/v YNB Biolife without amino acids) with 2% w/v glucose as carbon source. Solid media was prepared by adding 2% (w/v) agar. To test the growth of WT and mutant strains in the presence of different organic acids, Verduyn minimal media containing the potassium salt of sorbic acid (potassium sorbate), sodium salt of benzoic acid (sodium benzoate) at pH 4 and butyric acid and acetic acid at pH 3 were used. Yeast cultures were routinely grown at 30 °C in conical flasks (with a 1:5 ratio of medium to flask volume) on a rotary shaker set to 160 rpm. Where required hygromycin B (hyg), and geneticin (G418) were added to the media to a final concentration of 200 mg L^{-1} each. Frozen stock cultures of mutants were prepared from overnight cultures by addition of 20% (v/v) glycerol, and aseptically stored at -80°C.

Construction of plasmids and disruption cassettes

The disruption of both the alleles of *PDR12* was achieved by homologous recombination with 500bp flanking regions using two different deletion cassettes. The 500bp region upstream of each *PDR12* gene, together with the first 2000bp of each ORF was amplified with FW_ampl_gDNA_chr2 and Rev_ampl_gDNA_chr2 for Zygo2 *PDR12*_chr2 (Fig.1ai) and FW_ampl_gDNA_chr8 and Rev_ampl_gDNA_chr8 for Zygo2 *PDR12*_chr8 (Fig. 1aii). The amplicon was ligated into pSTBlue vector at the multiple cloning sites digested by EcoRV (pSTBlue-1 Perfectly Blunt® Cloning Kit – Novagen) to create plasmids pSTblue_*PDR12*chr2 (Fig1bi) and pSTblue_*PDR12*chr8 (Fig1bii) for restriction digestion to insert the antibiotic marker. In order to create disruption cassettes, an antibiotic selection marker was inserted into each *PDR12* amplicon in the respective vector. This was achieved by amplifying the marker from the plasmids pUG6 (Kan^R) and pUG026 (hph^R) and cloning by Gibson assembly after digestion of the plasmid with the appropriate enzyme(s). The kanMX markers amplified by Gibson

compatible primers were inserted into Zygo2 *PDR12*_chr2 amplicon digested by BstBI as illustrated in Figure 1ci, to create the plasmid pSTblue_ *PDR12*chr2_KANMX. The hphMX6 marker amplified by Gibson compatible primers were inserted into the Zyog2 *PDR12*_chr8 amplicon digested by BsaBI and EcoRV as seen in Figure 1cii to create the plasmid pSTblue_*PDR12*chr8_hph. The CEN.PK 113-7D *pdr12* mutant was constructed using the kanMX cassette system amplified from the YTK077 plasmid with primers Cen.PK_*PDR12*_kanmx_FW and Cen.PK_*PDR12*_Kanmx_Rev, harbouring 50bp homology arms.

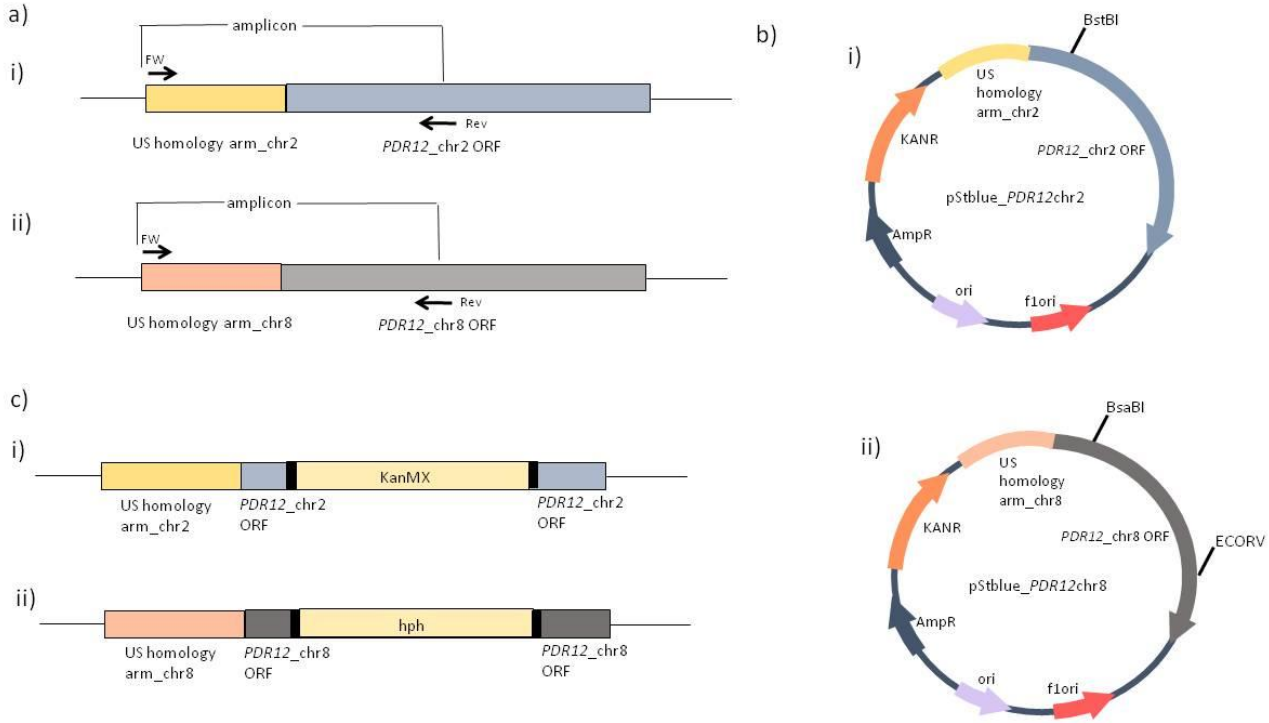


Figure 1: Schematic representation of the creation of disruption cassettes. a) *PDR12* genes and upstream homology arm of 500bp in wild type genomic locus indicating the positions where the primers anneal for the amplification of i) Zygo2 *PDR12*_chr2 and ii) Zygo2 *PDR12*_chr8 amplicons. b) Cloning of the *PDR12* amplicons into two pSTblue plasmids for creating i) pSTBlue_*PDR12*chr2 to be digested by BstBI for the insertion of KANMX cassette and ii) pSTBlue_*PDR12*chr8 to be digested by BsaBI and EcoRV for the insertion of hph cassette c) Disruption cassettes constructed by Gibson assembly specific to each allele of the gene containing the i) KanMX marker in the Zygo2 *PDR12*_chr2 amplicon and ii) hphMX6 marker in the Zygo2 *PDR12*_chr8 amplicon: in black the Gibson flanking sites are highlighted.

Yeast transformation

Zygo2 WT and mutants were transformed using the LiAc/PEG/ss-DNA method (Branduardi et al., 2014). Overnight pre-cultures in YPD medium were used to

inoculate a shake flask containing YPD medium to an initial OD_{660nm} of 0.2. Cultures were then incubated at 30 °C until OD_{660nm} of 0.8-1 was reached, harvested, and transformed with the DNA. For construction of Zygo2Δ*pdr12*_chr2, Zygo2 WT was transformed with PCR amplified deletion cassette from plasmid pSTblue_Z2*PDR12*chr2_KANMX. Similarly, for the construction of Zygo2Δ*pdr12*_chr8, Zygo2 WT was transformed with PCR amplified deletion cassette from plasmid pSTblue_Z2*PDR12*chr8_hph. Further, the mutant with single allele of *PDR12* disrupted was further transformed with PCR amplified deletion cassette from plasmid pSTblue_Z2*PDR12*chr8_hph to obtain Zygo2Δ/Δ*pdr12*. After heat shock, cells were recovered in YPF (fructose) medium supplemented with sorbitol (1 M) overnight and plated in selective rich media containing geneticin (G418) or hygromycin (hph) or both. Plates were kept at 30 °C for 3-4 days before analyzing the colonies for disruption of the *PDR12* alleles.

Table 1: List of strains used in this study

Strains	**Genotype**	**Source**
Z. parabailii ATCC60483 (Zygo2)	WT	(Suh et al., 2013)
Zygo2 Δ*pdr12*_chr2	*pdr12Δ*::kan/*pdr12*	This study
Zygo2 Δ*pdr12*_chr8	*pdr12/ pdr12Δ::hph*	This study
Zygo2 Δ/Δ*pdr12*	*pdr12Δ::kan/ pdr12Δ::hph*	This study
CEN.PK 113-7D	WT	(Nijkamp et al., 2012)
CEN.PK Δ*pdr12*	*pdr12Δ*::kan	This study
BY4741	WT	Lab strain
BY4741 Δ*pdr12*	*pdr12Δ*	Lab strain

Table 2: List of primers used in this study

Gene	Primer name	Sequence
Z2*PDR12*_chr2	FW_ampl_gDNA_chr2	5' TAGACGAACCGGTTTGAAAC 3'
	Rev_ampl_gDNA_chr2	5' GACAGCGTGTGCAGTAAC 3'
Z2*PDR12*_chr8	FW_ampl_gDNA_chr8	5' GAATTTTTTATTTTAGGAATCCCAGTTGGC 3'
	Rev_ampl_gDNA_chr8	5' CACATAGCATTAGAATGGTCAAACC 3'
Z2*PDR12*_chr2	ext_FW_chr2	5' TGAACAGCCTTGTGGTTAC 3'
Z2*PDR12*_chr8	ext_FW_chr8	5' TGGTTCTAATGACAGAACCTCATG 3'
KanR	KanMX_FW	5' ATGGGTAAGGAAAAGACTCACGTT 3'
hph^R	hph_Rev	5' TTATTCCTTTGCCCTCGGACG 3'
CenPK_*PDR12*	CenPK_PDR12_KANMX FW	5'TTAAAAAAAAGGTTTACAGATTTATTGTTATTGTTCTTATTAATAAAAAAagcttgcctcgtcccc 3'
	Cen.Pk PDR12_KANMX REV	5'ATAAAAATTGTGTGTTAAACCACGAAATACAAATATATTTGCTTGCTTGTcagtatagcgaccagcattcaca 3’
CenPK_*PDR12*	dg_CENPK PDR12 del_FW ext	5' CTCTCTGTTTTCCCAGTTACTAATTTTCACT 3'
	dg_CENPK PDR12_Rev ext	5'GGACGCCAAAATTTGTGAAAAAAAATTGAAAATAAAAATT 3'
	dg_CENPK PDR12 del_KANMX Rev	5' GCATTTCTTTCCAGACTTGTTCAACAG 3'
pSTblue plasmid	T7 primer	5' TAATACGACTCACTATAGGG 3'
	SP6 primer	5' ATTTAGGTGACACTATAG 3'

Table 3: List of plasmids used in this study

Name	Relevant Characteristics	Source
pSTBlue	ori ampR KanR panARS(OPT) MCS	Novagen
pSTBlue_PDR12chr2	US homology arm and PDR12chr2 ORF cloned into the MCS of pSTBlue	This study
pSTBlue_PDR12chr8	US homology arm and PDR12chr8 ORF cloned into the MCS of pSTBlue	This study
pSTBlue_PDR12chr2_KANMX	ori ampR panARS(OPT) US PDR12chr2 AgTEF1p-KANMX-AgTEF1t PDR12chr2	This study
pSTBlue_PDR12chr8_hph	ori ampR panARS(OPT) US PDR12chr8 AgTEF1p-hph-AgTEF1t PDR12chr8	This study
pUG6	ori ampR panARS(OPT) AgTEF1p-KANMX-AgTEF1t	(Gueldener et al., 2002)
pAG26	ori ampR panARS(OPT) AgTEF1p-hph-AgTEF1t	Addgene
YTK077	ori ampR panARS(OPT) AgTEF1p-KANMX-AgTEF1t	Addgene

Molecular analysis

PCR amplification with Q5® High-Fidelity DNA Polymerase (New England Biolabs) was performed according to the manufacturer's instructions. Diagnostic PCRs were performed using WonderTaq polymerase (EuroClone S.p.A.). All primer sequences are shown in Table 2. DNA fragments obtained by PCR were separated by gel electrophoresis. Gel and PCR purifications were carried out using NucleoSpin® Gel and PCR Clean-up columns (Macherey-Nagel). Gibson assembly was performed according to the manufacturer's recommendations (New England Biolabs). All DNA sequencing was performed by Sanger sequencing. *Escherichia coli* strain DH5α was used for plasmid transformation, amplification, and storage.

To assess the disruption of *PDR12*_chr2 and correct integration of the disruption cassette, genomic DNA were extracted from transformants growing on rich media containing G418 and PCR was performed using primers ext_FW_chr2 and KANMX_rev and sequenced. To assess the disruption of *PDR12*_chr8 in the WT strain, genomic DNA were extracted from transformants growing on rich media containing hygromycin and

PCR was performed using primers ext_FW_chr8 and hph_rev. For the selection of *pdr12* double mutants, transformants growing on rich media containing both G418 and hygromycin antibiotics were selected and PCR was performed with primers ext_FW_chr8 and hph_rev to check the disruption of *PDR12*_chr8. The mutated *PDR12* alleles were further confirmed by Sanger sequencing and aligned to the wild type gene sequences retrieved from the *Z. parabailii* ATCC60483 genome. Mutant with confirmed disruption of single or double alleles of *PDR12* were selected and stored for future purposes.

Phenotypic characterisation of the strains

For the phenotypic characterisation of the strains, drop tests were performed in plates containing Verduyn minimal media (2% glucose) with different organic acids at pH below their dissociation constants (pKa). The organic acids tested were sorbic acid (0.5mM-5mM) and benzoic acid (0.5mM-3mM) at pH 4.0, while acetic acid (100mM-200mM) and butyric acid (10mM-20mM) were tested at pH 3.0. The range of concentrations of the acids was selected based on previous literature studies. Stock solutions of the acids, namely sodium benzoate (50mM) and potassium sorbate (10mM) were prepared at pH 4.0 and acetic acid (1M) and butyric acid (50mM) at pH 3.0 and autoclaved. Verduyn minimal media (at pH 3.0 and pH 4.0) and agar were prepared separately at 4X concentration and autoclaved, since agar does not solidify at low pH. In order to prepare the plates, Verduyn minimal media, agar and the required concentration of the organic acids were diluted with water to make up to volume of 25ml in falcon tube and poured into a petri dish. The growth medium with acids was not buffered since the minimal media as well as the acid stocks were prepared at the same pH of 3.0 or pH 4.0 as required. As control plates, minimal media at pH 3.0 and 4.0 without acids were prepared and growth of the strains was tested. Strains were grown overnight in 5ml YPD, then refreshed in 5ml YPD for 4 hours at OD_{660nm} 0.5 until the cells reached exponential phase (OD_{660nm} 0.8-1.0). After three washing steps all the strains were diluted for a final OD_{660nm} of 1.0 and cells up to 10^{-4} dilutions were spotted on control plates as well as on plates containing organic acids and incubated at $30^{o}C$ for period of 72 hours.

Structural analysis of Pdr12p

The deduced protein sequences of *PDR12* of *S. cerevisiae*, *Z. parabailii* chr2 and *Z. parabailii* chr8 were retrieved from NCBI. Multiple sequence alignment of the different Pdr12p isoforms was performed using the MUSCLE algorithm. Using BLAST, we selected the closest homolog of *PDR12* that has experimentally resolved structures. The protein Pdr5 was suggested to have the closest homology model with an identity of 39% to Pdr12p. The Cryo-EM structure of Pdr5p from *S. cerevisiae* is available on PDB with the code 7P04 (https://www.rcsb.org/structure/7P04). Swiss model tool was used to build the structures of Pdr12p of *S. cerevisiae* and *Z. parabailii* using the methodology of homology modeling starting from Pdr5p structure as the template. Further, PyMOL was used to align the protein structure of Pdr12p in three combinations: 1) *S. cerevisiae* with *Z. parabailii* Pdr12p_chr2 2) *S. cerevisiae* with *Z. parabailii* Pdr12p_chr8 and 3) *Z. parabailii* Pdr12p_chr2 with *Z. parabailii* Pdr12p_chr8 to spot the differences in the amino acid on the protein structures. The online tool Deep TMHMM was used to predict the position of the transmembrane helixes on the protein structures and to validate the information obtained from the homology modeling of Pdr12p.

Results and Discussion

Construction of *Z. parabailii pdr12* mutants

In order to elucidate the role of *PDR12* in tolerance to organic acids in the hybrid yeast *Z. parabailii* with emphasis on the differential role of the two alleles, mutants with single and double disruption of this gene were created. This disruption was achieved with deletion cassettes containing long homology arms, a method successfully used in the creation of *leu2* mutants in *Z. parabailii* strain at that time not characterized as hybrid (Dato et al., 2010).

The disruption of *PDR12*_chr2 was obtained by transforming the Zygo2 WT strain with the KanMX cassette and confirmed by PCR using primers that anneals to the marker (KANMX_Rev) and to the homologous region upstream to the ORF (ext_FW_chr2). The disruption of *PDR12*_chr8 in Zygo2 WT was obtained by transformation with the hph deletion cassette. This disruption was confirmed by PCR using primers that anneals to the marker (hph_Rev) and to the homologous region upstream to the ORF (ext_FW_chr8). Furthermore, for the disruption of the *PDR12*_chr8 in the mutant already disrupted in *PDR12*_chr2, the cassette containing hphMX6 marker and 500bp upstream and downstream homology arms was used. Out of the six mutants obtained from this transformation, only one (mutant 'a') showed the expected band size when assessed by PCR (see panel a(i) and b (i) of figure 2). The mutant 'a' was a candidate strain further assessed by PCR to determine whether the marker genes had integrated into the correct loci by using primers that anneals to the antibiotic marker and to the homologous region upstream to the ORF (figure 2 panel a(ii) and b(ii). To determine the precise insertion of each marker gene, the region overlapping the insertion site was amplified by PCR, sub-cloned into two different plasmid vectors, and then sequenced with primers T7and SP6 present on the plasmid backbone. The results confirmed that a double knock-out had been obtained as both the marker genes had correctly inserted into the 5' region of the *PDR12* homologues on both alleles.

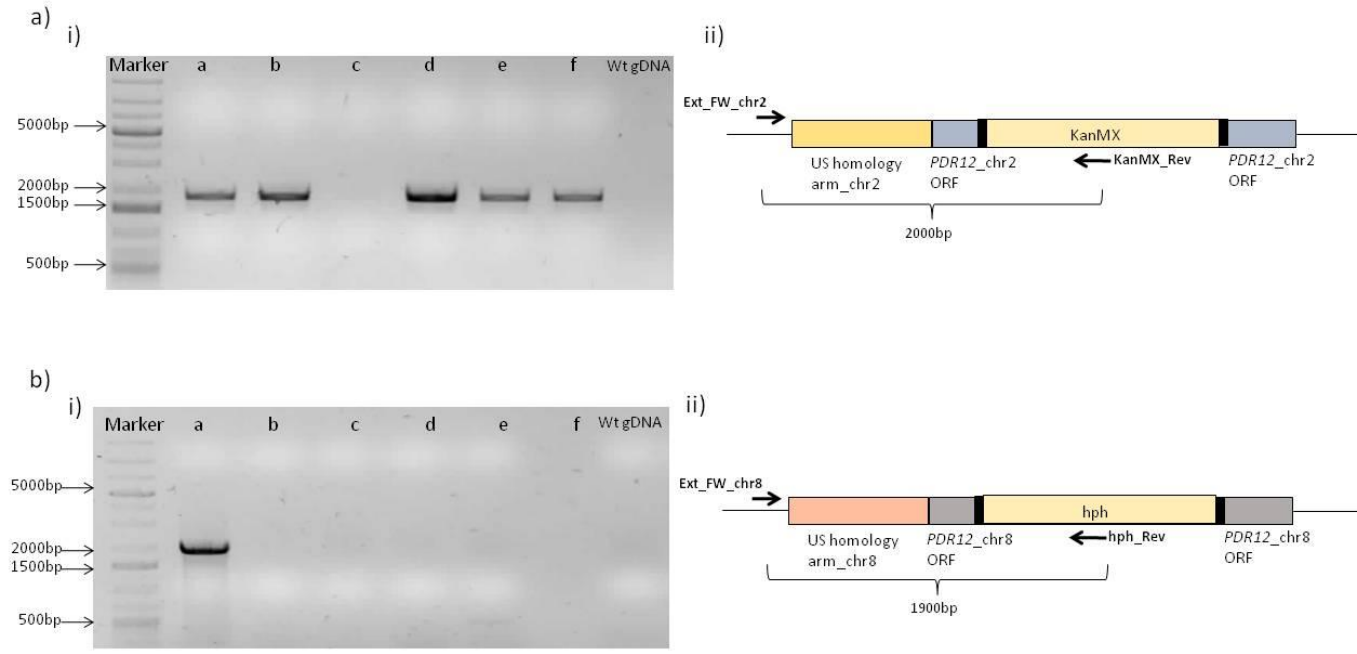


Figure 2: Verification of *PDR12* double disruption in *Z. parabailii*. Control PCRs showing the amplification of mutated loci in mutants (a-f) obtained after transforming the mutant with single allele disrupted. To check the disruption of a) i) Zygo2 *PDR12*_chr2 with the kanMX disruption cassette ii) using primers external to the upstream region (ext_FW_chr2) and in the antibiotic marker (KANMX_Rev) as indicated in the schematic representation. b) i) Zygo2 *PDR12*_chr8 with the hph disruption cassette ii) using primers external to the upstream region (ext_FW_chr8) and in the antibiotic marker (hph_Rev) as indicated in the schematic representation.

Z. parabailii pdr12 mutants exhibit growth inhibition in the presence of short chain organic acids

The growth of Zygo2 and *S. cerevisiae* WT and *pdr12* mutants were assessed by drop tests in the presence of different organic acids at a range of concentrations. In control conditions (Figure 3 panel a), all strains grew similarly, at both pH values. In the presence of butyric acid (panel b), the growth of Zygo2 *pdr12* double mutant was reduced at concentration of 10 Mm, with differences in growth also observed between mutants with single alleles disrupted at 15 mM. The Zygo2 *pdr12* double mutant and mutant with the *PDR12*_chr8 allele disrupted exhibited reduced growth compared to the WT strain and the mutant with *PDR12*_chr2 disruption.

In panel c it can be observed that, in the presence of acetic acid, the Zygo2 *pdr12* double mutant exhibited reduced growth above 100 mM concentration, unlike the mutants with single alleles disrupted and the WT strain. In fact, no differences in growth were observed between mutants of Zygo2 with single alleles disrupted, indicating that both alleles contribute to the tolerance towards acetic acid. Further, no growth was observed for CEN.PK WT and mutants, because the minimum inhibitory concentration of acetic acid for CEN.PK is much lower than that of *Z. parabailii*. Differently, the BY4741 WT strain exhibited growth. Overall, the disruption of *PDR12* alleles in Zygo2 WT reduced tolerance to acetic and butyric acid.

In contrast, in the presence of longer chain organic acids such as sorbic and benzoic, the growth of Zygo2 *pdr12* mutants as well as WT strain was not affected with increasing concentrations of the acids from 0.5mM to 2mM as seen in panel d and e, suggesting that no involvement of *PDR12* was seen in tolerance to these acids. On the other hand, the growth of *S. cerevisiae pdr12* mutants was affected in the presence of sorbic and benzoic acids from 1 mM and 0.5 mM respectively, compared to its WT strain (panel d and e), differently to what was observed with *Z. parabailii*.

The phenotypic characterization revealed that *PDR12* in *Z. parabailii* is required for counteracting acetic and butyric acid stress rather than sorbic and benzoic acid stress. From previous studies on *Z. bailii*, loss of tolerance towards sorbic and benzoic acids and the inability to degrade sorbate and benzoate was attributed to the loss of *YME2* gene (Piper et al., 1998). On the other hand, *S. cerevisiae pdr12* mutants were sensitive to sorbic and benzoic acid as observed also in previous literature studies. In *S. cerevisiae*, Haa1 is considered as the primary regulator of transcriptional response to acetic acid stress (Almario et al., 2013) and has also been linked to the lactic acid response (Fernandes et al., 2005).

Differential impact of individual alleles of *Z. parabailii PDR12* in the presence of butyric acid

The growth of Zygo2 *pdr12* mutants with single alleles disrupted was assessed together with the WT and the double mutants to characterize the differential impact of the individual alleles in the presence of different organic acids. In the presence of butyric acid stress, a differential impact of individual alleles of *PDR12* in *Z. parabailii* was uncovered. While disrupting both the *PDR12* gene copies resulted in a more severe phenotype as observed at 10mM concentration than in single deleted strains, we also noticed differences in growth between the mutants with individual alleles disrupted at 15mM concentration (panel b, figure 3). Interestingly, it was observed that Zygo2 *PDR12* allele on chromosome 8 contributes to tolerance towards butyric acid since its disruption negatively affected its tolerance, unlike the allele on chromosome 2. This difference could be due to variations in the expression or function of the gene considering that the two alleles of *PDR12* derive from two different parental genomes in the hybrid *Z. parabailii*.

The first hypothesis refers to a transcriptomic analysis study on *Z. parabailii* that revealed how in response to lactic acid stress there was a statistically significant divergence in expression responses between the homeologous gene pairs (Ortiz-Merino et al., 2018). In a work based on polygenic analysis of very high acetic acid tolerance in the yeast *S. cerevisiae*, several novel genes linked to high acetic acid tolerance were identified. These genes complemented the defective alleles present in the other parental strain, indicating that eliminating inferior mutant alleles may be just as crucial as introducing rare superior alleles for achieving very high tolerance to acetic acid (Stojiljkovic et al., 2020). Therefore, the trait of superior alleles can be useful for further improving the organic acid tolerance in specific industrial yeast strains.

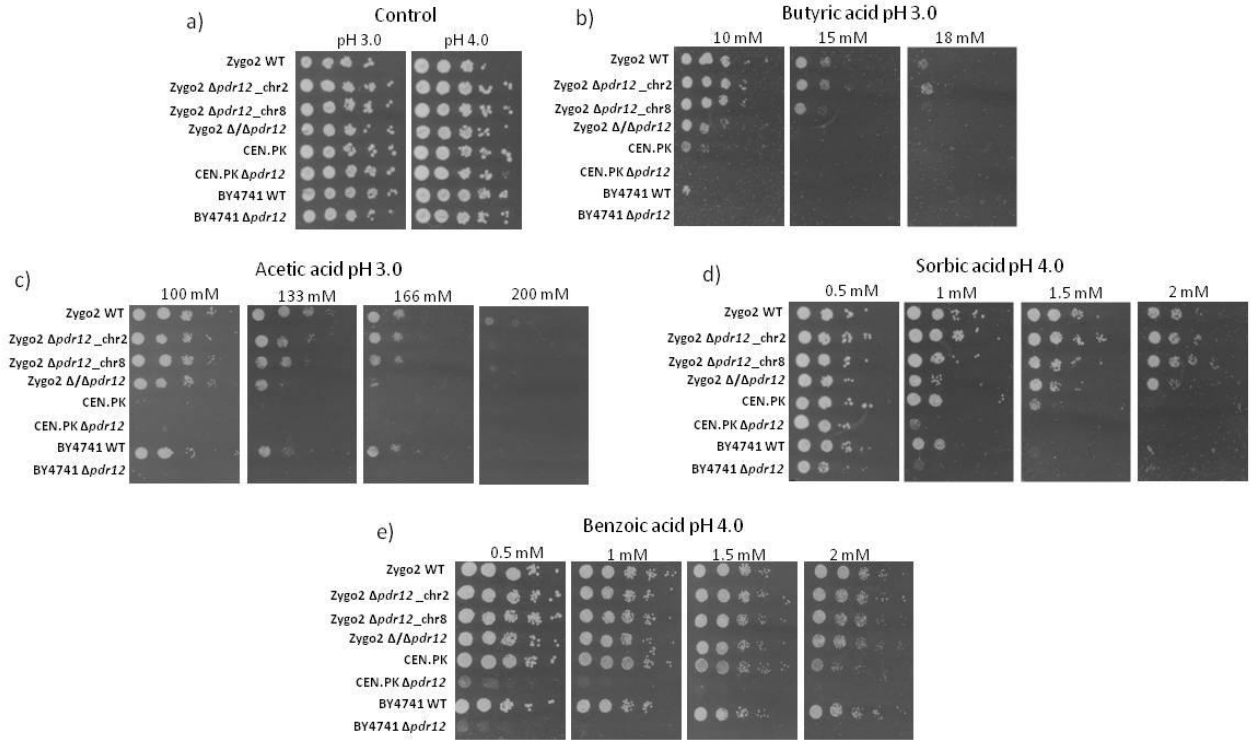


Figure 3: Phenotypic characterisation of *Z. parabailii* and *S. cerevisiae* WT and *pdr12* mutants. Drop tests performed on minimal media plates at OD_{660} 0.8-1.0, with different concentrations of organic acids. a) Control plates containing minimal media at pH 3.0 and 4.0 b) Butyric acid plates at pH 3.0 c) Acetic acid plates at pH 3.0 d) Sorbic acid plates at pH 4.0 and e) Benzoic acid plates at pH 4.0. Highlighted in red are the differences in growth between the two alleles of Zygo2 *PDR12*, double mutant and *S. cerevisiae* mutants.

Structural analysis of *PDR12* of *S. cerevisiae* and *Z. parabailii*

Analysis of the protein structures of Pdr12p of *S. cerevisiae* and *Z. parabailii* could help in providing insights for the differences seen in the role of organic acid tolerance in the two yeasts as well as between the two alleles of *Z. parabailii*. Pdr12p belongs to the ABC transporters (ATP-binding cassette) family. The ABC transporters are membrane proteins that transport solutes across lipid bilayers. These transporters are consisted of four domains: two cytoplasmic nucleotide-binding domains (NBDs) and two transmembrane domains (TMDs). The NBDs bind and hydrolyze ATP to fuel the transport cycle, and their main sequence is largely conserved throughout the superfamily. TMDs, which can have a wide range of primary sequences and protein folds, define the translocation route across the membrane and, in most cases (but not

always), dictate solutes specificity (Wilkens, 2015). An example of a Pdr12p structure is illustrated in Figure 4.

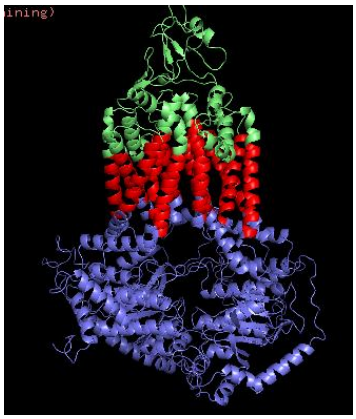
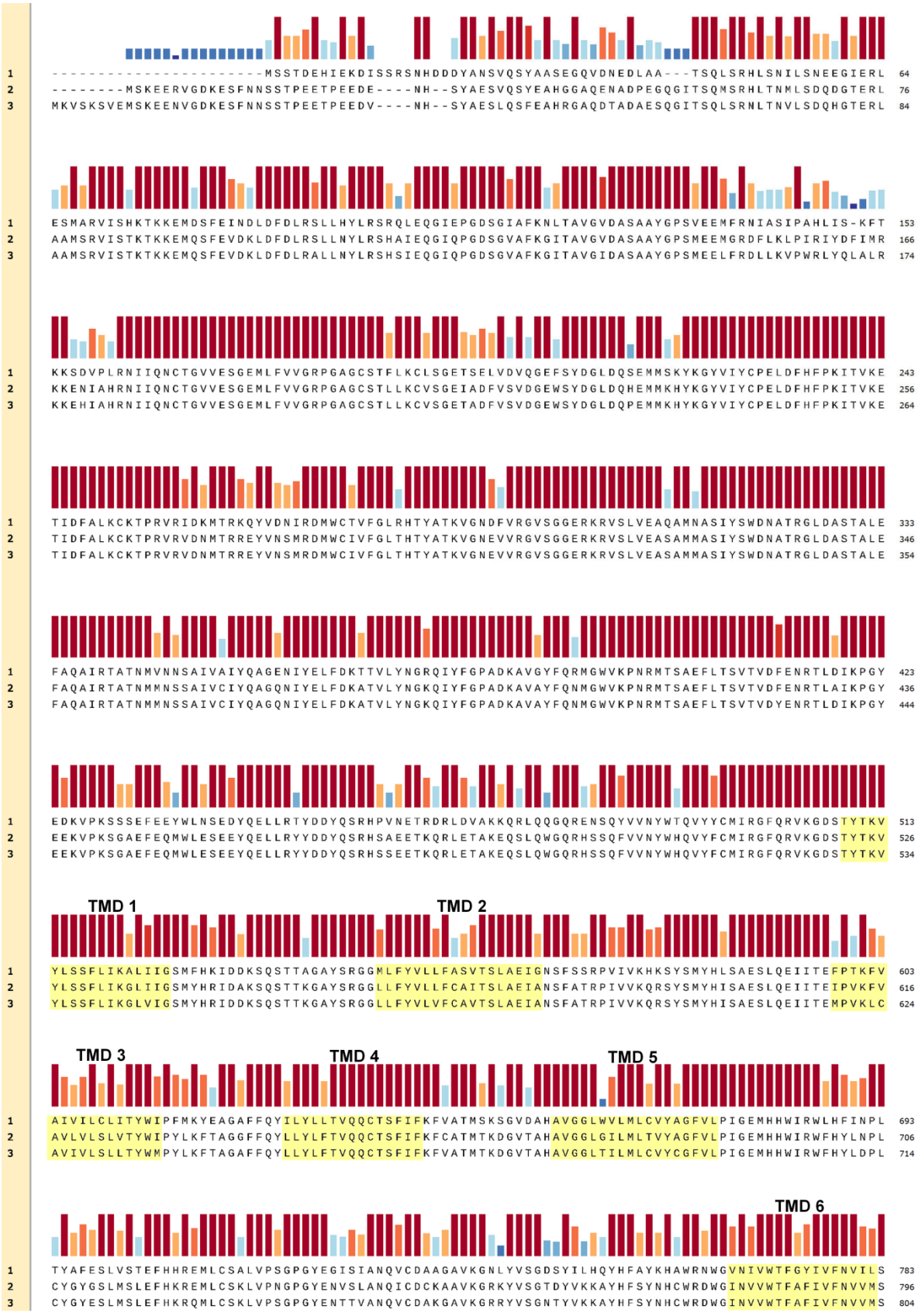


Figure 4: The Pdr12 transporter structure obtained from Swiss model, an automated web based protein structure homology-modeling expert system. The region colored in blue represents the part in the cytosol; region in red represents the transmembrane helices and in green the extracellular region.

When simulating the homology modeling for all the three Pdr12p isoforms, similar structures were obtained, with no obvious visible differences. Hence, the Pdr12p sequences of *S. cerevisiae* and *Z. parabailii* were aligned and the results can be seen in Figure 5. When comparing the complete amino acid sequence of the three Pdr12p isoforms, similarities and differences were revealed. In particular, regardless of which isoform from *Z. parabailii* is being compared to their *S. cerevisiae* counterpart, the alignment of the amino acids sequences reveals an approximate 78% identity and 87% similarity. When aligning the Pdr12p isoforms from *Z. parabailii* against each other, a 93% identity and 97% similarity is revealed. As can be seen in Figure 5, the majority of the amino acids is identical between the isoforms and is represented by tall dark red bars on top. More interestingly though, a gradient of differences in physico-chemical properties where amino acids were substituted between the isoforms can be seen by the decreasing height of the bars and a color gradient ranging from red through orange to blue, i.e. the closer to dark blue, shorter is the bar and higher is the differences in physico-chemical properties.


TMD 1
TMD 2
TMD 3
TMD 4
TMD 5
TMD 6

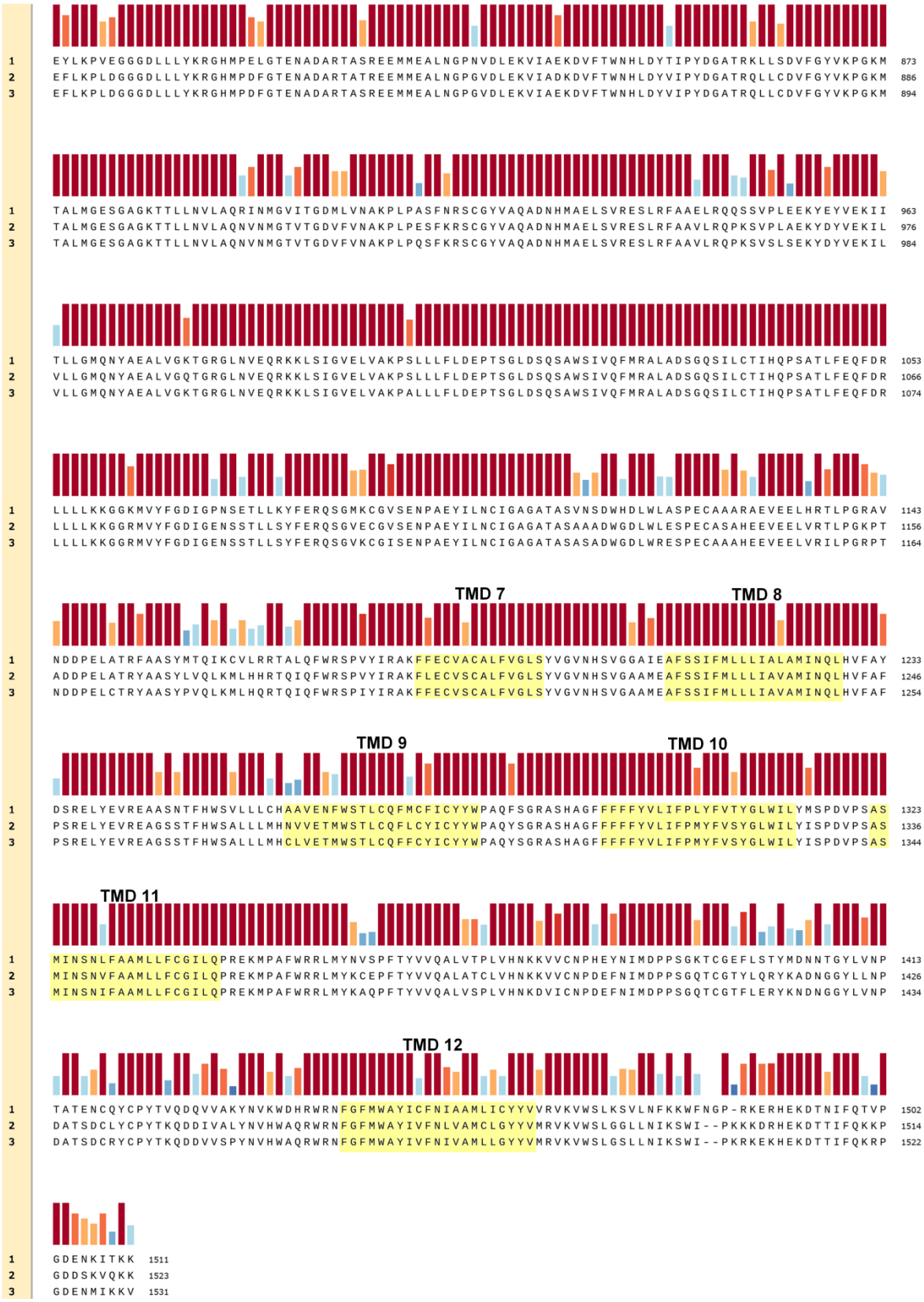


Figure 5: Alignment of yeast Pdr12p. The protein sequences of Sc Pdr12p, Zb Pdr12p_chr2, Zb Pdr12p_chr8 were aligned using the MUSCLE algorithm. 1, 2 and 3 refers to the Pdr12p sequences of *S. cerevisiae, Z. parabailii* chromosome 2 and *Z. parabailii* chromosome 8,

respectively. Highlighted in yellow are the 12 transmembrane domains in the three Pdr12 sequences. The bars ranging from dark red (tallest) to dark blue (shortest) indicate gradient of similarity in physico-chemical properties between amino acids exchange, i.e. from high to low, respectively.

Considering the active involvement of TMDs in export and import of solutes, particularly organic acids, differences in these regions were analyzed with more depth. The simulation for the structure of the three Pdr12p isoforms revealed 12 TMDs, which are highlighted in yellow in Figure 5. The alignment of the amino acids sequences related to these regions revealed which amino acids differ between the isoforms. These substitutions were counted and classified into either conservative (similar physico-chemical properties) or non-conservative (distinct physico-chemical properties). The comparison was made by pairs of isoforms, resulting in three sets of data that can be seen in panels a, b and c of Figure 6. Panels a and b show that TMDs 2, 3, 6, 9 and 12 differs from *Z. parabailii* to *S. cerevisiae,* with degrees of identity ranging from 44%-72%. The other seven TMDs show degree of identity higher than 80%. In a global perspective, grouping the mutations from all TMDs together as seen in panel d, regardless of the Pdr12p isoform from *Z. parabailii* being compared to *S. cerevisiae*, approximately 19% of the amino acids (42 or 43) are substituted, while only 9% (19 or 20) represent non-conservative substitutions. Since the channel involved in the export of organic acids is located within the 12 TMDs, we can hypothesize that among the non-conservative substitutions we can find the causal substitutions responsible for the different organic acid affinities of the Pdr12p isoforms of *S. cerevisiae* and *Z. parabailii*. The comparison between the two Pdr12p isoforms from *Z. parabailii* resulted in approximately 91% identity (panel d in Figure 6). In other words, 9% of amino acids (19) are substituted, while only 4% (9) represents non-conservative substitutions. With an in-depth analysis, among these 9 amino acids it could be possible to find the reason why the Pdr12p isoform from chromosome 8 is involved in butyric acid tolerance while the other isoform from chromosome 2 is not.

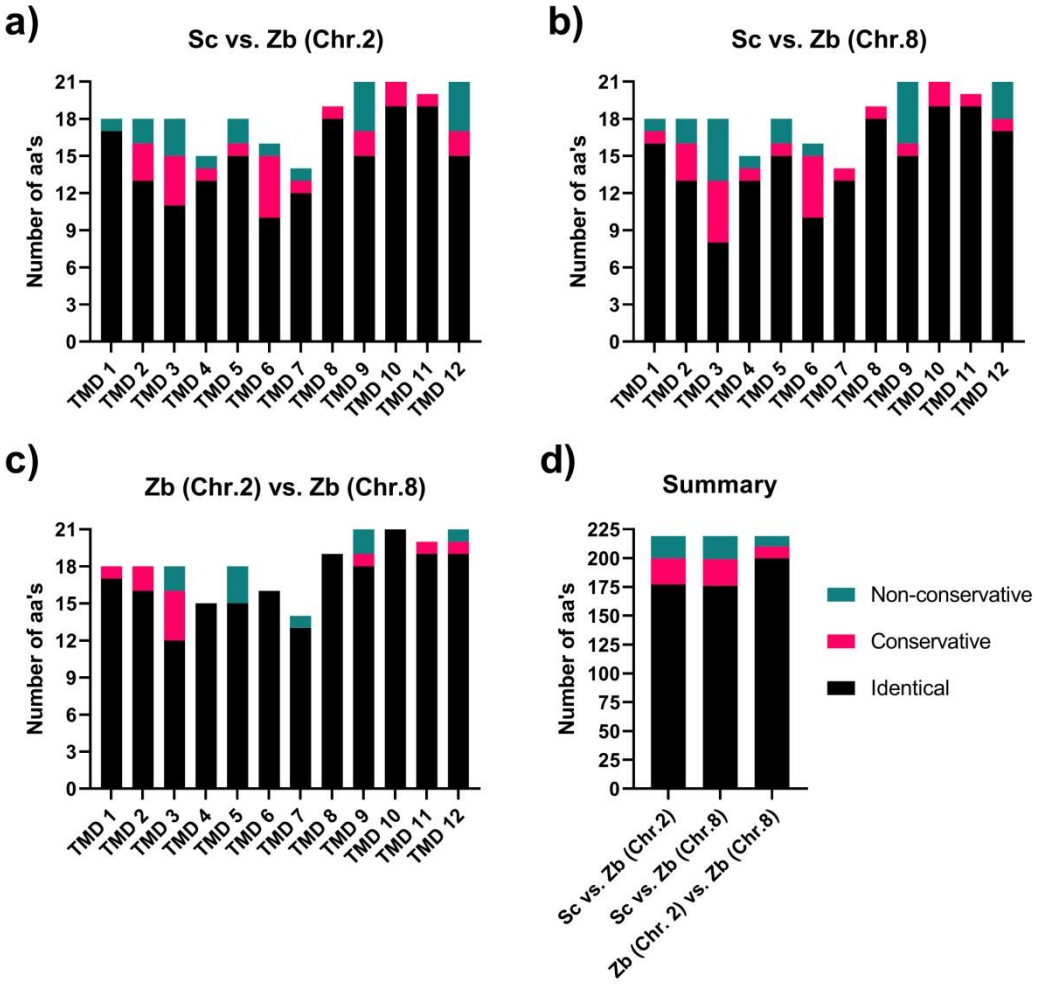


Figure 6: Comparison between the degrees of similarity of the transmembrane domains (TMDs) for the different Pdr12p isoforms. The sequence alignments result in identical (black - no differences between isoforms), conservative (pink-substitution for amino acids with similar physico-chemical properties) or non-conservative (green-substitution for amino acids with different physico-chemical properties) degree of similarity for each amino acid. Panel a) represents the comparison between Pdr12p isoforms of *S. cerevisiae* and *Z. parabailii* chromosome 2, panel b) of *S. cerevisiae* and *Z. parabailii* chromosome 8, and panel c) of *Z. parabailii* chromosome 2 and 8. Panel d) shows the summary of the comparisons from a) to c) by grouping the degree of similarity of all the TMDs together.

References

Almario, M.P., Reyes, L.H., Kao, K.C., 2013. Evolutionary engineering of Saccharomyces cerevisiae for enhanced tolerance to hydrolysates of lignocellulosic biomass. Biotechnol. Bioeng. https://doi.org/10.1002/bit.24938

Becker, J., Wittmann, C., 2019. A field of dreams: Lignin valorization into chemicals, materials, fuels, and health-care products. Biotechnol. Adv. https://doi.org/10.1016/j.biotechadv.2019.02.016

Berterame, N.M., Porro, D., Ami, D., Branduardi, P., 2016. Protein aggregation and membrane lipid modifications under lactic acid stress in wild type and *OPI1* deleted *Saccharomyces cerevisiae* strains. Microb. Cell Fact. https://doi.org/10.1186/s12934-016-0438-2

Branduardi, P., Dato, L., Porro, D., 2014. Molecular tools and protocols for engineering the acid- tolerant yeast *Zygosaccharomyces bailii* as a potential cell factory. Methods Mol. Biol. https://doi.org/10.1007/978-1-4939-0563-8_4

Branduardi, P., Valli, M., Brambilla, L., Sauer, M., Alberghina, L., Porro, D., 2004. The yeast *Zygosaccharomyces bailii*: A new host for heterologous protein production, secretion and for metabolic engineering applications. FEMS Yeast Res. https://doi.org/10.1016/S1567-1356(03)00200-9

Dato, L., Branduardi, P., Passolunghi, S., Cattaneo, D., Riboldi, L., Frascotti, G., Valli, M., Porro, D., 2010. Advances in molecular tools for the use of *Zygosaccharomyces bailii* as host for biotechnological productions and construction of the first auxotrophic mutant. FEMS Yeast Res. https://doi.org/10.1111/j.1567-1364.2010.00668.x

Fernandes, A.R., Mira, N.P., Vargas, R.C., Canelhas, I., Sá-Correia, I., 2005. *Saccharomyces cerevisiae* adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. Biochem. Biophys. Res. Commun. https://doi.org/10.1016/j.bbrc.2005.09.010

Gueldener, U., Heinisch, J., Koehler, G.J., Voss, D., Hegemann, J.H., 2002. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding

yeast. Nucleic Acids Res. https://doi.org/10.1093/nar/30.6.e23

Hatzixanthis, K., Mollapour, M., Seymour, I., Bauer, B.E., Krapf, G., Schüller, C., Kuchler, K., Piper, P.W., 2003. Moderately lipophilic carboxylate compounds are the selective inducers of the *Saccharomyces cerevisiae* Pdr12p ATP-binding cassette transporter. Yeast. https://doi.org/10.1002/yea.981

Heer, D., Sauer, U., 2008. Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. Microb. Biotechnol. https://doi.org/10.1111/j.1751-7915.2008.00050.x

Ito, Y., Hirasawa, T., Shimizu, H., 2014. Metabolic engineering of *Saccharomyces cerevisiae* to improve succinic acid production based on metabolic profiling. Biosci. Biotechnol. Biochem. https://doi.org/10.1080/09168451.2014.877816

Jönsson, L.J., Martín, C., 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. Bioresour. Technol. https://doi.org/10.1016/j.biortech.2015.10.009

Kren, A., Mamnun, Y.M., Bauer, B.E., Schüller, 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. Mol. Cell. Biol. https://doi.org/10.1128/mcb.23.5.1775-1785.2003

Kuanyshev, N., Adamo, G.M., Porro, D., Branduardi, P., 2017. The spoilage yeast *Zygosaccharomyces bailii* : Foe or friend? Yeast. https://doi.org/10.1002/yea.3238

Kuanyshev, N., Ami, D., Signori, L., Porro, D., Morrissey, J.P., Branduardi, P., 2016. Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation. FEMS Yeast Res. https://doi.org/10.1093/femsyr/fow058

Limayem, A., Ricke, S.C., 2012. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. Prog. Energy Combust. Sci. https://doi.org/10.1016/j.pecs.2012.03.002

Nijkamp, J.F., van den Broek, M., Datema, E., de Kok, S., Bosman, L., Luttik, M.A.,

Daran-Lapujade, P., Vongsangnak, W., Nielsen, J., Heijne, W.H.M., Klaassen, P., Paddon, C.J., Platt, D., Kötter, P., van Ham, R.C., Reinders, M.J.T., Pronk, J.T., de Ridder, D., Daran, J.M., 2012. De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. Microb. Cell Fact. https://doi.org/10.1186/1475-2859-11-36

Ortiz-Merino, R.A., Kuanyshev, N., Braun-Galleani, S., Byrne, K.P., Porro, D., Branduardi, P., Wolfe, K.H., 2017. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. PLoS Biol. https://doi.org/10.1371/journal.pbio.2002128

Ortiz-Merino, R.A., Kuanyshev, N., Byrne, K.P., Varela, J.A., Morrissey, J.P., Porro, D., Wolfe, K.H., Branduardi, P., 2018. Transcriptional response to lactic acid stress in the hybrid yeast *Zygosaccharomyces parabailii*. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.02294-17

Paixão, S.M., Teixeira, P.D., Silva, T.P., Teixeira, A. V., Alves, L., 2013. Screening of novel yeast inulinases and further application to bioprocesses. N. Biotechnol. https://doi.org/10.1016/j.nbt.2013.02.002

Papadimitriou, M.N.B., Resende, C., Kuchler, K., Brul, S., 2007. High Pdr12 levels in spoilage yeast (*Saccharomyces cerevisiae*) correlate directly with sorbic acid levels in the culture medium but are not sufficient to provide cells with acquired resistance to the food preservative. Int. J. Food Microbiol. https://doi.org/10.1016/j.ijfoodmicro.2006.06.035

Piper, P., Calderon, C.O., Hatzixanthis, K., Mollapour, M., 2001. Weak acid adaptation: The stress response that confers yeasts with resistance to organic acid food preservatives. Microbiology. 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 J. https://doi.org/10.1093/emboj/17.15.4257

Raab, A.M., Gebhardt, G., Bolotina, N., Weuster-Botz, D., Lang, C., 2010. Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. Metab. Eng. https://doi.org/10.1016/j.ymben.2010.08.005

Sauer, M., Branduardi, P., Valli, M., Porro, D., 2004. Production of L-ascorbic acid by metabolically engineered *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.70.10.6086-6091.2004

Sauer, M., Porro, D., Mattanovich, D., Branduardi, P., 2008. Microbial production of organic acids: expanding the markets. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2007.11.006

Stojiljkovic, M., Foulquié-Moreno, M.R., Thevelein, J.M., 2020. Polygenic analysis of very high acetic acid tolerance in the yeast *Saccharomyces cerevisiae* reveals a complex genetic background and several new causative alleles. Biotechnol. Biofuels. https://doi.org/10.1186/s13068-020-01761-5

Suh, S.O., Gujjari, P., Beres, C., Beck, B., Zhou, J., 2013. Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. Int. J. Syst. Evol. Microbiol. https://doi.org/10.1099/ijs.0.048058-0

Sun, Y., Cheng, J., 2002. Hydrolysis of lignocellulosic materials for ethanol production: A review. Bioresour. Technol. https://doi.org/10.1016/S0960-8524(01)00212-7

Wilkens, S., 2015. Structure and mechanism of ABC transporters. F1000Prime Rep. https://doi.org/10.12703/P7-14

Xu, G., Zou, W., Chen, X., Xu, N., Liu, L., Chen, J., 2012. Fumaric Acid Production in *Saccharomyces cerevisiae* by In Silico Aided Metabolic Engineering. PLoS One. https://doi.org/10.1371/journal.pone.0052086

Chapter 4

Establishment of a novel CRISPR-Cas9 system in the hybrid yeast *Zygosaccharomyces parabailii* reveals allele exchange mechanism

Pooja Jayaprakash[1,2,*], Liliane Barroso[1,3,*], Matteo Vajente[1] , Edward J. Louis[3], John P. Morrissey[2], Paola Branduardi[1]

[1] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, Milano, 20126, Italy.
[2] School of Microbiology, Environmental Research Institute, APC Microbiome Institute, University College Cork, Cork, T12 K8AF, Ireland.
[3] Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, United Kingdom

* These authors contributed equally to this work.

Manuscript in preparation for submission.

Contribution to the work:
P.J. was involved in designing the project, carried out part of the experiments, was involved in results analysis and contributed in the writing and preparation of the final manuscript.

Abstract

The non-*Saccharomyces* hybrid yeast *Zygosaccharomyces parabailii* holds potential as cell factory mainly because of its robustness in withstanding stressors that are usually characterizing bio-based processes. However, genetic engineering is hindered due to the complex genome of the hybrid and lack of optimized gene editing tools such as CRISPR-Cas9. In this work, we developed a CRISPR-Cas9 gene editing system in *Z. parabailii* which allows simultaneous disruption or deletion of two alleles of a gene. We evaluated four different gRNA expression systems consisting of combinations of tRNAs, tRNA and ribozyme or ribozymes as self-cleaving flanking elements. The functionality of the gRNA systems was tested by analysing the inactivation of *ADE2* in the wild type *Z. parabailii.* Disruption efficiency of 70% by non-homologous end joining was achieved using the gRNA system flanked by two tRNA. Moreover, this gRNA system was successfully used in the construction of a *Z. parabailii dnl4* mutant. This mutant exhibited homology directed repair efficiency of 15% in the deletion of both *ADE2* alleles compared to the lack of double deletion by HDR observed in the wild type strain. Analysis of mutations in the gRNA target regions of both *ADE2* and *DNL4* genes suggested inter-allelic rearrangements between the two gene loci, as well as absence of large regions of chromosomes. The findings and methods described in this manuscript enable avenues of research both in genome engineering and in exploitation of *Z. parabailii* for industrial uses.

Introduction

Yeasts play crucial roles in human society. Traditional foods and beverages use yeasts such as *Saccharomyces cerevisiae*, *Kluyveromyces marxianus* and *Debaryomyces hansenii* and modern biotechnology adds species such as *Komagataella phaffii*, *Yarrowia lipolytica* and *Rhodotorula toruloides*; there are also important yeast pathogens such as *Candida albicans* (Karaalioğlu and Yuceer, 2021). In addition, for many years *S. cerevisiae* has been the pre-eminent model organism for studies on eukaryotic genetics, cell biology and molecular biology. Up to relatively recently, with the notable exception of *S. cerevisiae,* it was difficult to genetically manipulate yeasts, and this limited the capacity to understand their biology, to construct mutants, or to re-engineer their pathways and metabolism. This changed with the development of CRISPR (clustered regularly interspaced short palindromic repeat) -Cas9 (CRISPR associated protein 9) as a tool for introduction of targeted double stranded breaks that can be repaired using endogenous DNA repair machinery to introduce inactivation mutations by non-homologous end joining (NHEJ) repair or deletions via homology dependent repair (HDR) (Stovicek, *et al* 2017). Currently, it is increasingly straightforward to genetically engineer many species and there are multiple examples of this, and of complex genome engineering in so-called non-conventional yeasts. This approach has been adapted to a variety of yeast species, including *S. cerevisiae*, *Y. lipolytica*, *K. phaffi*, *Kluyveromyces lactis, Schizosaccharomyces pombe* as well as pathogenic yeast species, such as *Candida albicans* and *Cryptococcus neoformans* (Stovicek, *et al* 2017). There is, however, another type of yeast species, the importance of which is increasingly being recognised, namely, hybrids. These arise from the union of two different parental species and the canonical example is *Saccharomyces pastorianus,* responsible for production of lager-style beer, which can trace its ancestry to the mating between *S. cerevisiae* and *Saccharomyces eubayanus* several hundred years ago in a German brewery. Population genomics studies have uncovered multiple other examples, especially amongst yeasts used in traditional biotechnology (Krogerus *et al*., 2017). Several natural hybrids formed between *S. cerevisiae* and *Saccharomyces kudriavzevi* have been isolated from Belgian trappist beers (González, Barrio and Querol, 2008), and natural hybrids between *S. cerevisiae* and *Saccharomyces uvarum*

are frequently used in wine making (Christine *et al.*, 2007; Alsammar and Delneri, 2020). Additionally, hybrid yeasts have recently been reported also in clinical isolates, particularly from the *Candida* clade such as *Candida inconspicua*, *Candida orthopsilosis*, and *Candida metapsilosis* (Mixão and Gabaldón, 2020; O'Brien *et al*., 2021). The *Z. bailii sensu lato* species holds significance in the biorefinery era and is widely known for its remarkable characteristics such as the ability to withstand low pH, high osmotic pressure, high ethanol concentrations and weak organic acids. However, its ability to cause food spoilage in highly acidic and high sugar content foods is less desirable (Kuanyshev *et al.*, 2017). The genus is comprised of 12 species, of which at least three are interspecies hybrids such as *Zygosaccharomyces parabailii* formed by the mating between *Zygosaccharomyces bailii* CLIB213 and an unidentified *Z. bailii* (Hulin and Wheals, 2014). To expand the industrial potential of *Z. parabailii* and to understand the mechanism underlying its unique characteristics, the development of optimized genome-editing tools, such as CRISPR-Cas9, is desirable.

The key steps in successful engineering using CRISPR-Cas9 includes design, expression, and delivery of the gRNA. In different yeast systems this is achieved using an RNA polymerase III (Pol III) promoter that provides a leader sequence to be cleaved after molecular maturation or a polymerase II (Pol II) promoter flanked with ribozyme cleaving systems, or even tRNAs as summarized in Figure 1. Apart from these, in some yeast such as *S. cerevisiae, Y. lipolytica* and *K. lactis,* T7 polymerase-dependent artificial promoters are used (Morse *et al.*, 2018). The RNA(s) transcribed by Pol III promoters such as *SNR52* and *RPR1* remains in the nucleus and are thus considered most suitable for gRNA expression (Shan, Dai and Wang, 2021). The gRNA expression driven by *SNR52* Pol III promoter was implemented for gene editing in *S. cerevisiae* and was further extended to *K. lactis* (Horwitz *et al.*, 2015), *Candida albicans* (Vyas, Barrasa and Fink, 2015)*, K. marxianus* (Nambu-Nishida *et al.*, 2017)*, Scheffersomyces stipitis* (Cao et al., 2021), and other species. In yeast species such as *K. marxianus,* highest gene editing efficiency was achieved using RNA pol III promoter *RPR1*-tRNA^{gly} hybrid (Löbs *et al.*, 2017), and highest gene editing efficiency was observed when gRNA was expressed using the *S. cerevisiae* gene for tRNA^{Phe} and the hepatitis D ribozyme (HDV) sequence (Cernak, Estrela and Poddar, 2018). In some yeasts, such as *Y. lipolytica* and *Ogataea polymorpha*, a fusion of RNA pol III promoters with tRNAs is used because the

binding sites of these promoters are located within their mature transcripts, which leads to the addition of nucleotides to the gRNA (Shan, Dai and Wang, 2021). In *O. polymorpha,* the mutation efficiency was relatively low when the sgRNA was expressed under the control of *Op*SNR6 promoter whereas the use of $tRNA^{CUG}$ significantly improved the editing efficiencies (Numamoto, Maekawa and Kaneko, 2017). Further, due to the absence of *SNR52* promoter or its homologs in *R. toruloides*, gRNA expression has been achieved by different Pol III promoters such as U6 (Jiao *et al.*, 2019) or 5S rRNA promoter in combination with $tRNA^{arg}$ which has led to more effective CRISPR activity from the produced transcripts (Schultz, Cao and Zhao, 2019). Despite successful gene deletions in different yeast species using pol III promoters with high efficiencies, the identification of these promoters in non-conventional yeasts still remains a challenge (Shan, Dai and Wang, 2021). In gRNA expression using RNA pol II promoters flanked with hammerhead (HH) and HDV ribozyme sequences, successful gene editing has been observed in different yeast species, such as *C. neoformans* (Arras *et al.*, 2016) and *K. lactis* (Juergens *et al.*, 2018), targeting *ADE2* as proof of concept. In *O. parapolymorpha* the use of two ribozyme flanked spacers facilitated the expression of gRNA to a greater extent, thereby leading to double gene editing in this yeast (Juergens *et al.*, 2018). However, pol II promoters are rarely used in comparison to pol III promoters due to their lower genome editing efficiency and the need for self-cleaving elements attributed to 5' capping and 3' polyadenylation of Pol II transcripts following transcription.

There are additional challenges when it comes to engineering non-conventional yeasts, particularly hybrids, such as the availability of stable plasmids and low copy number functional plasmids with variable expression in cells across a single population (Löbs, Schwartz and Wheeldon, 2017). Additionally, the genetic manipulation of hybrid yeasts is challenging due to a complex genome consisting of at least two copies of almost every gene and the scarcity of molecular tools to modify multiple alleles. In *S. cerevisiae*, HR is the predominant DNA repair process thereby making genome engineering more effective by facilitating the development of several *in vivo* DNA assembly tools. Diversely, in most of the non-Saccharomyces yeasts described, gene editing by HR is inefficient. Integration by HR is often preferred for gene editing because it avoids disruption of essential genes, enables control over the integration

loci and allows integration into a site with a consistent expression profile (Löbs, Schwartz and Wheeldon, 2017). In the case of the hybrid yeast *Z. parabailii,* the inability to obtain stable haploid strains is one of the key restrictions and in order to disrupt a gene function, two or more alleles of the gene differing by 7% in the nucleotide sequence on average (Ortiz-Merino *et al.*, 2017) should be simultaneously non-functional. Furthermore, the lack of available auxotrophic strains accessible for long time forced researchers to rely on dominant markers. Finally, because the ultimate objective is to use this yeast in industrial operations, marker-free stable modified strains are required.

The only yeast hybrid successfully edited using the CRISPR-Cas9 system is *S. pastorianus*. For the precise editing of this aneuploid hybrid, an efficient gRNA expression system flanked with HH and HDV using an RNA polymerase II was designed, following failure of the system using the RNA polymerase III. This system was used to precisely delete the four existing alleles of the *SeILV6* gene in *S. pastorianus*. Additionally, with a similar construct, multiplexing activity for deletion of different genes located in different chromosomes was also successfully demonstrated (Gorter de Vries *et al.*, 2017).

In a more recent work by Kuanyshev and co-workers, CRISPR-Cas9 system was efficiently used in the haploid species *Z. bailii* with gRNA expressed using Pol II promoter along with tRNA (Kuanyshev *et al.*, 2021). Interestingly, the targeted editing allowed the concomitant deletion of a pyruvate decarboxylase (*PDC*) gene and introduction of a heterologous bacterial lactate dehydrogenase (LDH), resulting in a heterolactic fermenting strain. However, the authors did not specify if the gRNA cassette consisted of pol II promoters fused with one or two tRNAs, or a combination of tRNA and ribozymes for the successful expression of the gRNA.

Considering that *Z. parabailii* is more robust than its haploid counterparts (Kuanyshev *et al.*, 2016) and the need to better understand its specific interdependence of phenotype-genotype association, the goal of this study was to develop an efficient CRISPR-Cas9 system for genome editing. In order to determine the most efficient gRNA expression system in this hybrid yeast, we employed four different expression cassettes consisting of the spacer and structural RNA between self-cleaving elements, $tRNA^{Gly}$-$tRNA^{Asp}$, $tRNA^{Gly}$-HDV, HH-$tRNA^{Gly}$ and HH-HDV targeting *ADE2* as a proof of

concept. Similar to other non-conventional yeasts, *Z. parabailii* has lower HR efficiency relative to *S. cerevisiae,* and to overcome this limitation we constructed a *Z. parabailii dnl4* double mutant using previously optimized Cas9-gRNA system. Further, we report the presence of allelic exchange mechanisms in *Z. parabailii dnl4* and *ade2* mutants. Hence, this work highlights the use of a potential CRISPR-Cas9 gRNA expression system for genome engineering in the robust hybrid yeast *Z. parabailii* and the successful HDR for double deletions in *dnl4* mutants with insights into allele exchange mechanisms.

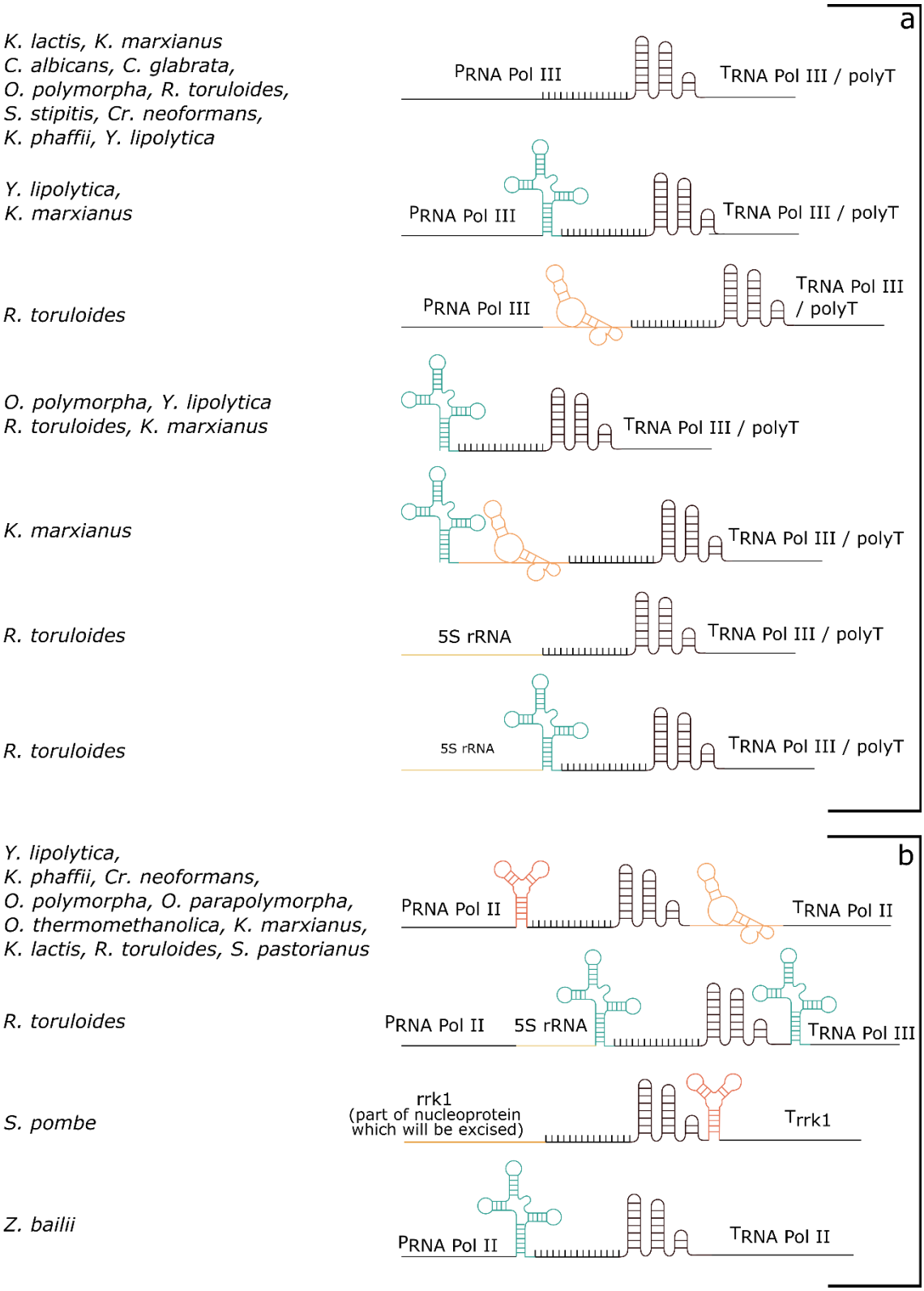


Figure 1: Graphical representation of different gRNA expression systems used in various conventional and non-conventional yeasts. In red hammerhead ribozyme, orange HDV ribozyme, blue tRNA and black the spacer sequence along with structural tRNA. a) gRNAs designed for transcription by RNA Pol III. b) gRNA designed for transcription by RNA Pol II.

Materials and Methods

Strains and growth conditions

Z. parabailii ATCC60483 strain was used in this study (Suh et al. 2013). This hybrid has been used in several previous studies (Kuanyshev *et al.*, 2016; Ortiz-Merino *et al.*, 2017; Ortiz-Merino *et al*., 2018), and its complete genome sequence is available (GCA_001984395.2) Other strains constructed over the course of this work are listed in Table 1. Yeast cultures were grown in YP medium (1% w/v yeast extract, 2% w/v peptone) with 2% w/v glucose (D) or fructose (F) as carbon source, or minimal synthetic medium (0.67% w/v YNB Biolife without amino acids) with 2% w/v glucose as carbon source. Solid media was prepared by adding 2% (w/v) agar. Where required to detect the red colour of *ade2*-disrupted strains, minimal media was supplemented with 10 mg L^{-1} adenine. Yeast cultures were routinely grown at 30 °C in conical flasks (with a 1:5 ratio of medium to flask volume) on a rotary shaker set to 160 rpm. Where required hygromycin B (hyg), nourseothricin (clonNAT) and geneticin (G418) were added to YPF/YPD medium to a final concentration of 200 mg L^{-1}, 100 mg L^{-1} and 200 mg L^{-1} respectively. Frozen stock cultures of mutants were prepared from overnight cultures by addition of 20% (v/v) glycerol, and aseptically stored at -80 °C.

Plasmid construction and design of gRNA expression cassettes

The broad-host-range Cas9/gRNA co-expression backbone plasmid pUDP002 (Juergens *et al.*, 2018) was used. To integrate the synthetic gRNA constructs into pUDP002 they were flanked by inward-facing BsaI restriction sites generating sticky ends (underlined) 'GGTCTCGCAAA' and 'ACAGCGAGACC' at their 5' and 3' ends, respectively, compatible with BsaI-digested pUDP002 (Figure 2). The final gRNA expression cassettes were synthesized as gene fragments with Twist Bioscience (Table 3).

The gRNA expression cassettes were composed of (i) 5' self-cleaving element (HH or $tRNA^{Gly}$), followed by (ii) 20-nucleotide spacer, by the (iii) gRNA scaffold, and by a (iv) 3' self-cleaving element (HDV, $tRNA^{Gly}$ or $tRNA^{Asp}$) (Figure 2). All plasmids used in this study are described in Table 2 and the synthetic gRNA constructs are available in Table 3.

Sequences for the HH and HDV were taken from (Gao and Zhao, 2014). When using HH as a 5' self-cleaving element, the first six nucleotides were the reverse complement of the first six nucleotides of the spacer sequence. Sequences of tRNAs were identified from the genome of *Z. parabailii* (Ortiz-Merino *et al.*, 2017) using the software tRNAscan-SE (Lowe and Eddy, 1997). Among these sequences, $tRNA^{Gly}$ and $tRNA^{Asp}$ were selected avoiding tRNAs with internal BsaI and Esp3I recognition sites. Different tRNAs for the 5' and 3' ends of the gRNA were selected to avoid undesired secondary structure formation. The structural gRNA sequence was taken from (Gorter de Vries *et al.*, 2017). The 20-nucleotide spacer sequences were identified using the software sgRNAcas9 (Xie *et al.*, 2014).

To design the universal gRNA expression cassette tRNA-tRNA-U1, the 20-nucleotide spacer sequence was replaced by a fragment with outward-facing BsaI restriction sites, for easy insertion of spacer sequences using Golden Gate Assembly and annealed oligonucleotides. This cassette was designed with flanking Esp3I recognition sites (respectively CGTCTCACAAA at 5' and ACAGCGAGACG at 3') which create ends compatible with BsaI-digested pUDP002 (underlined). pUDPZb_U1 was obtained by BsaI restriction of pUDP002, Esp3I restriction of tRNA-tRNA-U1, purification of both fragments and ligation using Quick Ligase (Supp Figure 1).

Spacer sequences able to target both alleles of *ADE2* or *DNL4* genes were selected. To inactivate *ADE2*, a spacer sequence already tested in the haploid yeast *Z. bailii* was used (Kuanyshev *et al.*, 2020) (Table 2). For the *DNL4* disruption, five different spacer sequences targeting both alleles were designed and tested (Table 2). Before synthesis, the designed gRNA expression cassettes were checked for the correct secondary structure formation using SPOT-RNA (Singh *et al.*, 2019). After the hybridization of oligonucleotides with the selected target sequences flanked by sticky ends compatible with the digested pUDPZb_U1 (CGCA at 5' and GTTT at 3'), the hybridized nucleotides were assembled with the pUDPZB_U1 backbone using Golden Gate Assembly. The integration of target sequences in all the gRNA constructs were confirmed by Sanger sequencing.

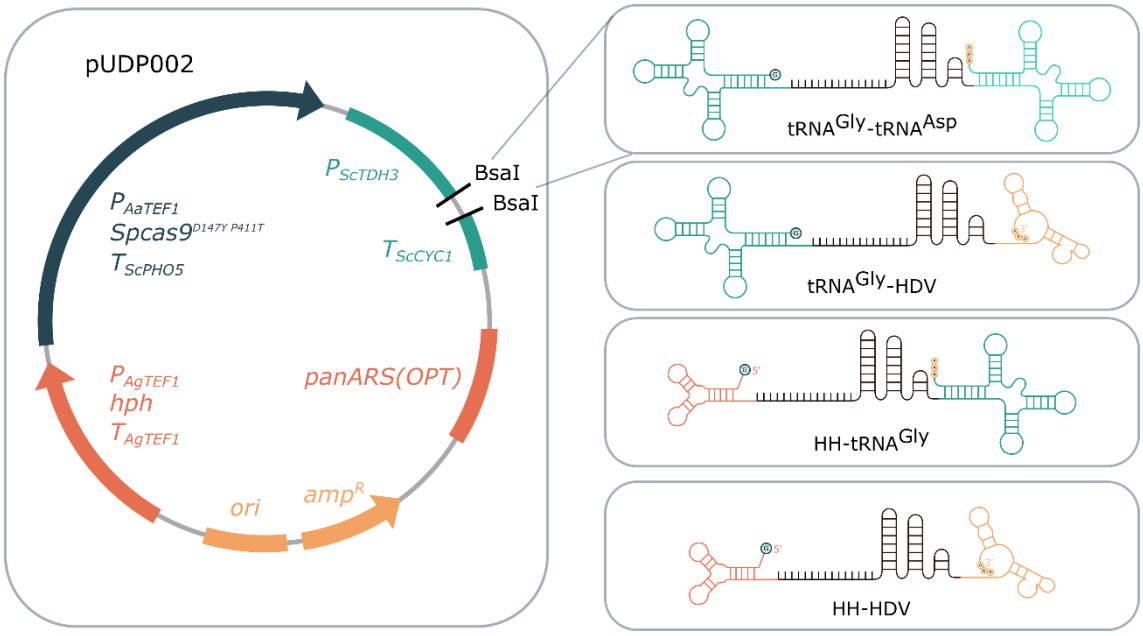


Figure 2: Steps involved in the construction of gRNA plasmids (pUDPZb) in pUDP002 plasmid backbone. The constructed gRNA cassettes were assembled by Golden Gate using the BsaI highlighted restriction sites, between the *ScTDH3* promoter and *ScCYC1* terminator; gRNA expression cassettes with the spacer and structural RNA are between self-cleaving elements, $tRNA^{Gly}$ and $tRNA^{Asp}$, $tRNA^{Gly}$ and HDV, HH-$tRNA^{Gly}$, HH-HDV.

Repair DNA fragments construction

Two different repair fragments were constructed for targeted deletion of each *ADE2* allele. Specific homology regions upstream and downstream of the two loci of *ADE2* were selected and amplified from the gDNA of *Z. parabailii*. Further, two different dominant markers *(kanMX* for chromosome 1 and *natMX* for chromosome 2) were cloned between the homologous sequences. Both fragments were assembled using Gibson Assembly in the backbone plasmid pSTBlue-1.

The kanMX_Chr1 cassette was built by amplifying from the genome of *Z. parabailii* an upstream region (537 nt) and a downstream region (519 nt) using primers Chr1_US_fw/Chr1_US_rev and Chr1_DS_fw/Chr1_DS_rev. The dominant marker *kanMX* (coding for aminoglycoside phosphotransferase) was amplified from the plasmid pZ3 (Branduardi *et al.*, 2004) using the primers kanMX_fw/kanMX_rev. The natMX_Chr2 cassette was built by amplifying from the genome of *Z. parabailii* an upstream region (539 nt) and a downstream region (522 nt) using primers Chr2_US_fw/Chr2_US_rev and Chr2_DS_fw/Chr2_DS_rev. The dominant marker

natMX (coding for aminoglycoside phosphotransferase) was amplified from the plasmid pZ5 (Branduardi *et al.*, 2004) using the primers NAT_fw/NAT_rev.

Positive clones were identified using blue/white screening and confirmed by sequencing.

To transform the repair fragments, linear DNA was amplified by PCR using Q5® High-Fidelity DNA Polymerase, the primers Chr1_US_fw/Chr1_DS_rev and Chr2_US_fw/Chr2_DS_rev and the previous constructed plasmids as template. The PCR product was further purified, quantified, and used in *Z. parabailii* transformation.

Yeast transformation

Z. parabailii was transformed by using an optimized version of the LiAc/PEG/ss-DNA method (Branduardi, Dato and Porro, 2014). Overnight pre-cultures in YPD medium were used to inoculate a shake flask containing YPD medium to an initial OD_{660nm} of 0.2. Cultures were then incubated at 30 °C until OD_{660nm} of 0.8-1 was reached, harvested, and transformed with the DNA. 3 µg of purified plasmid were used in each transformation. To transform the repair fragments, double the molar amount was used compared to the plasmid, respectively 1.45 µg and 1.31 µg of purified amplified DNA for *kanMX*_Chr1 and *natMX*_Chr2. After heat shock, cells were recovered in YPF medium supplemented with sorbitol (1 M) overnight and plated in selective rich media. Plates were kept at 30 °C for 3-4 days and transformants were further plated in minimum media with 10 mg l^{-1} adenine and incubated at 30 °C for 2 days and at 4 °C for 3 days before assessing the percentage of red colonies. These plates were then replica plated in rich media with the addition of geneticin (G418) or nourseothricin (clonNAT) to assess the percentage of single and double insertion of repair fragments.

Molecular analysis

PCR amplification with Q5® High-Fidelity DNA Polymerase (New England Biolabs) was performed according to the manufacturer's instructions. Diagnostic PCRs were performed using WonderTaq polymerase (EuroClone S.p.A.). All primer sequences are shown in Table 4. DNA fragments obtained by PCR were separated by gel electrophoresis. Gel and PCR purifications were carried out using NucleoSpin® Gel and PCR Clean-up columns (Macherey-Nagel). Golden Gate Assembly was performed using BsaI-HF®v2 and T4 ligase (New England Biolabs) according to the manufacturer's

recommendations (75 ng destination plasmid, amplicon fragment in a 2:1 molar ratio, 2.5 µL T4 ligase buffer, 500 U T4 ligase, 15 U BsaI-HFv2, water to 25 µL). All DNA sequencing were performed by Sanger sequencing. *Escherichia coli* strain DH5α was used for plasmid transformation, amplification, and storage.

To confirm the disruption of *ADE2*, red colonies were selected, and both alleles of the gene were amplified by colony PCR (using the primers Fw_NHEJ_gRNA_Chr1/Rev_NHEJ_gRNA_Chr1, Fw_NHEJ_gRNA_Chr2/Rev_NHEJ_gRNA_Chr2) and sequenced to confirm mutations in the target region.

For the selection of *dnl4* double mutants, transformants were selected and both loci of *DNL4* were amplified (using the primers DG_ct1 dnl4 chr2_Fw/DG_ct1 dnl4 chr2_Rv, DG_cts dnl4 chr2_Fw/DG_cts dnl4 chr2_Rv, DG_ct1 dnl4 chr7_Fw/ DG_ct1 dnl4 chr7_Rv, DG_cts dnl4 chr7_Fw/ DG_cts dnl4 chr7_Rv) and sequenced. Sequencing results were aligned to the original gene sequences retrieved from the *Z. parabailii* ATCC60483 genome. Colonies with confirmed disruption of both alleles of *ADE2* and *DNL4* were selected and stored for future purposes.

To confirm the correct integration of the repair fragments into the target allele in both wild type and *dnl4* mutants, diagnostic PCRs were performed on the colonies growing on either G418 or clonNAT using specific primers for each allele and antibiotic marker (Fw_HDR_Chr1_UP/ Rev_HDR_Chr1_kanMX and Fw_HDR_Chr2_UP/ Rev_HDR_Chr2_*natMX*).

Table 1. Strains used in this study

Name of the strain	Genotype	Source
Z. parabailii ATCC60483	WT	ATCC collection (Suh *at al*., 2013)
Zp_ade2	*Δ/ Δade2*	This study
*Zp_dnl4 #*1	*Δ/Δdnl4*	This study
*Zp_dnl4 #*2	*Δ/ Δdnl4*	This study
*Zp_dnl4#*1_a*de2*	*Δ/Δdnl4 Δ/Δade2*	This study
*Zp_dnl4#*1_a*de2_KAN*	*Δ/Δdnl4* *ade2*chr1::*kanMX Δade2*chr2	This study
*Zp_dnl4#*1_a*de2_NAT*	*Δ/Δdnl4* *Δade2*chr1 *ade2*chr2::*natMX*	This study

Table 2. Plasmids used in this study

Name	Relevant characteristics	Source
pUDP002	ori amp^{R} panARS(OPT) AgTEF1p-hph-AgTEF1t ScTDH3p $_{BsaI\ BsaI}$ ScCYC1t AaTEF1p Spcas9-ScPHO5t	Gorter de Vries *et al*, 2017
pUDPZb_ Gly_*ADE2*c_Asp	tRNA^{Gly} – *ADE2* target - tRNA^{Asp} cloned into pUDP002	This study
pUDPZb_Gly_*ADE2*c_HDV	tRNA^{Gly} – *ADE2* target - HDV cloned into pUDP002	This study
pUDPZb_ HH_*ADE2*c_Gly	HH – *ADE2* target - tRNA^{Gly} cloned into pUDP002	This study
pUDPZb_ HH_*ADE2*c_HDV	HH – *ADE2* target - HDV cloned into pUDP002	This study
pSTBlue_repair chr1_*kanMX*	ori amp^{R} panARS(OPT) US *ADE2* AgTEF1p-*kanMX*-AgTEF1t DS *ADE2*	This study
pSTBlue_repair chr2_NAT	ori amp^{R} panARS(OPT) US *ADE2* AgTEF1p-*natMX*-AgTEF1t DS *ADE2*	This study
pZ3	ori amp^{R} panARS(OPT) AgTEF1p-*kanMX*-AgTEF1t	Branduardi et al., 2004
pZ5	ori ampR panARS(OPT) AgTEF1p-NAT-AgTEF1t	Branduardi et al., 2004
pUDPZb_U1	pUDP002 containing tRNA^{Gly}–BsaI restriction site– tRNA^{Asp}	This study
pUDPZb_U1_*DNL4* CT1	*DNL4* target1 inserted in the BsaI restriction site in pUDPZb_U1	This study
PUDPZb_U1_*DNL4* CT2	*DNL4* target2 inserted in the BsaI restriction site in pUDPZb_U1	This study
PUDPZb_U1_*DNL4* CT3	*DNL4* target3 inserted in the BsaI restriction site in pUDPZb_U1	This study
PUDPZb_U1_*DNL4* CT4	*DNL4* target4 inserted in the BsaI restriction site in pUDPZb_U1	This study
PUDPZb_U1_*DNL4* CT5	*DNL4* target5 inserted in the BsaI restriction site in pUDPZb_U1	This study

Table 3. Synthesized gRNAs constructs

Synthesized sequences	Sequence (5'-3')
Self-cleaving elements	
HH	ATATTGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC
HDV	GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCA ACATGCTTCGGCATGGCGAATGGGAC
$tRNA^{Gly}$	GCGCAAGTGGTTTAGTGGTAAAATTCATCGTTGCCATCGATGA GCCCCCGGTTCGATTCCGGG CTTGCGCA
$tRNA^{Asp}$	TCCGTGATAGTTTAATGGTCAGAATGGGCGCTTGTCGCGTGCC AGATCGGGGTTCAATTCCCCGTCGCGGAG
Structural gRNA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT ATCAACTTGAA AAAGTGGCACCGAGTCGGTGC
gRNA constructs	
$tRNA^{Gly}$ - *ADE2* target - $tRNA^{Asp}$	CCCTTTAA**GGTCTC**GCAAAGCGCAAGTGGTTTAGTGGTAA AATTCATCGTTGCCATCGATGAGCCCCCGGTTCGATTCCGG GCTTGCGCACAATATGATAGCGTCGTCCAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT GAAAAAGTGGCACCGAGTCGGTGCTTTTTCCGTGATAGTT TAATGGTCAGAATGGGCGCTTGTCGCGTGCCAGATCGGGG TTCAATTCCCCGTCGCGGAGACAGC**GAGACC**CCCTTTAA
$tRNA^{Gly}$ - *ADE2* target - HDV	CCCTTTAA**GGTCTC**GCAAAGCGCAAGTGGTTTAGTGGTAA AATTCATCGTTGCCATCGATGAGCCCCCGGTTCGATTCCGG GCTTGCGCACAATATGATAGCGTCGTCCAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT GAAAAAGTGGCACCGAGTCGGTGCTTTTGGCCGGCATGG TCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTC GGCATGGCGAATGGGACACAGC**GAGACC**CCCTTTAA
HH - *ADE2* target - $tRNA^{Asp}$	CCCTTTAA**GGTCTC**GCAAAATATTGCTGATGAGTCCGTGAG GACGAAACGAGTAAGCTCGTCCAATATGATAGCGTCGTCC AGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT GCGCAAGTGGTTTAGTGGTAAAATTCATCGTTGCCATCGAT GAGCCCCCGGTTCGATTCCGGGCTTGCGCAACAGC**GAGA CC**CCCTTTAA
HH- *ADE2* target - HDV	CCCTTTAA**GGTCTC**GCAAAATATTGCTGATGAGTCCGTGAG GACGAAACGAGTAAGCTCGTCCAATATGATAGCGTCGTCC AGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGG CAACATGCTTCGGCATGGCGAATGGGACACAGC**GAGACC**C CCTTTAA
$tRNA^{Gly}$ - $tRNA^{Asp}$ -U1	AAGGAAG**CGTCTC**ACAAAGCGCAAGTGGTTTAGTGGTAA AATTCATCGTTGCCATCGATGAGCCCCCGGTTCGATTCCGG GCTTGCGCATGAGACCTCACACTTCACGGGTAGTATGGTAG GGTCTCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GCTTTTTCCGTGATAGTTTAATGGTCAGAATGGGCGCTTGT

	CGCGTGCCAGATCGGGGTTCAATTCCCCGTCGCGGAGACA GC**GAGACG**CTTCCAGGG

Table 4. Primers used in this study

Name	Sequence (5'-3')
Construction of *Z. parabailii* repair fragment for *ADE2* chromosome 1 copy	
Chr1_DS_fw	CTGGTCGCTATACTGTTTGATGTTTTCATATTTTGAACGTAC
Chr1_DS_rev	GGCTCGAGAAGCTTGTCGACGAATTCAGATTTATATCGTTGGATTTTTACTTGGG
Chr1_US_fw	GCGTTACGTATCGGATCCAGAATTCGTGATTGCCTATTCCTCCCATTTTTACTAGT
Chr1_US_rev	CTGGGCCTCCATGTCGACAAGAACAGAGGGTTTTGCG
*kanMX*_fw	CCCTCTGTTCTTGTCGACATGGAGGCCCAGAATAC
*kanMX*_rev	TATGAAAACATCAAACAGTATAGCGACCAGCATTC
Construction of *Z. parabailii* repair fragment for *ADE2* chromosome 2 copy	
Chr2_DS_fw	CTGGTCGCTATACTGTGATGTTTTCATATTTTTTTGAGCGTAT
Chr2_DS_rev	GGCTCGAGAAGCTTGTCGACGAATTCAGATGAATTATATTGTTGGGTTTTTACTTGGA
Chr2_US_fw	GCGTTACGTATCGGATCCAGAATTCGTGATTCTAAAATAAATTTCCGTTTTACTAACCC
Chr2_US_rev	CTGGGCCTCCATGTCACCTATGCGACACCATCAG
NAT_ fw	TGGTGTCGCATAGGTGACATGGAGGCCCAGAATAC
NAT_ rev	AATATGAAAACATCACAGTATAGCGACCAGCATTC
Diagnostic primers for molecular analysis of *ade2* mutants	
Fw_NHEJ_gRNA_Chr1	CCAAAGAACTTTTCATCTCTGTCG
Rev_NHEJ_gRNA_Chr1	GCGTATTTGGTCATTTTTTGTGGC
Fw_NHEJ_gRNA_Chr2	GCCAAAGAACTTTTCCTCTCTCTCA
Rev_NHEJ_gRNA_Chr2	GCATATTTAGCCATTCTTTGCGGT
Fw_HDR_Chr1_UP	AACTAGAACTCGATGGGCTA
Rev_HDR_Chr1_kanMX	TAAATCAGCATCCATGTTGG
Fw_HDR_Chr2_UP	CAACGAACAGTATTCTAGGACAC
Rev_HDR_Chr2_*natMX*	GTGTCGTCAAGAGTGGTAC
ARO10_Chr1_fw	CAACATAAGCTGTAGTTGTACCGGG
ARO10_Chr1_rv	TGAGAGAACCCCTCGTAATATTCTA
ARO10_Chr2_fw	CAGCATAAGCTGTAACTGTACCGGA
ARO10_Chr2_rv	AGAGGACCCCACACAATATTCTC
RGA1/RGA2_DS new_FW	TAGGCCAGCTTCAAAACC
RGA1/RGA2_DS new_FW	GGTTTTGAAGCTGGCCTA
UBP10 US_FW	GACAAGCTGATGACGCTGTA
UBP10_chr2_rev	CTTGAGGTGTACAAGCAGT
DIS3_chr2_fw	CACGATGTAGAATACAGACCA
DIS3_chr2_rev	TGGATCCAAAACAGACCTCAA
Diagnostic primers for gRNA insertion in pUDP002/pUDPZb_U1	
pUDP002_GPDpro-F	CGGTAGGTATTGATTGTAATTCTG
pUDP002_CYC1	GCGTGAATGTAAGCGTGAC
Diagnostic primers to confirm mutations in *DNL4*	
DG_ct1 dnl4 chr2_Fw	GAACATAAATTTCTGAGCTGT
DG_ct1 dnl4 chr2_Rv	GTTTCTGTTAGTGTCAAACGAG
DG_cts dnl4 chr2_Fw	CGCCTCTACACATTAGAAAGC

DG_cts dnl4 chr2_Rv	GAAGTTTCCTCTTACAGGGAA
DG_ct1 dnl4 chr7_Fw	GAACATAAATTTCTGAACTGC
DG_ct1 dnl4 chr7_Rv	TCTGTTAGTGTCAAATGAT
DG_cts dnl4 chr7_Fw	TGCCTTTACACATTAGAAAACA
DG_cts dnl4 chr7_Rv	GAAGTTTCCCGTTACGGG
Golden gate assembly of *DNL4* targets in pUDPZb_U1	
DNL4_ct1_fw	CGCATGTATCCGATCTAAAGACTG
DNL4_ct1_Rv	AAACCAGTCTTTAGATCGGATACA
DNL4_ct2_fw	CGCAGCAATTGCAATGGGCTCTGA
DNL4_ct2_Rv	AAACTCAGAGCCCATTGCAATTGC
DNL4_ct3_fw	CGCAATTGTGATTGGAAGAACTCC
DNL4_ct3_Rv	AAACGGAGTTCTTCCAATCACAAT
DNL4_ct4_fw	CGCAAATCGCTAATGGCATTTCCC
DNL4_ct4_Rv	AAACGGGAAATGCCATTAGCGATT
DNL4_ct5_fw	CGCAAGCGGCACAAAGTTCAAGAC
DNL4_ct4_Rv	AAACGTCTTGAACTTTGTGCCGCT
ADE2 and *DNL4* target sequences	
ADE2 common target	CAATATGATAGCGTCGTCCA
DNL4 Target 1	TGTATCCGATCTAAAGACTG
DNL4 Target 2	GCAATTGCAATGGGCTCTGA
DNL4 Target 3	ATTGTGATTGGAAGAACTCC
DNL4 Target 4	AATCGCTAATGGCATTTCCCA
DNL4 Target 5	AGCGGCACAAAGTTCAAGAC

Results

Efficient gRNA expression system for gene disruption in *Z. parabailii*

To evaluate the efficiency of different combinations of self-cleaving elements in gRNA constructs in *Z. parabailii*, we developed a CRISPR-Cas9 system based on the pUDP002 plasmid, which was shown to be effective in different yeast species (Juergens et al., 2018). This plasmid has the *panARS* element for replication in multiple yeasts, *spCas9* expressed from the *TEF1* promoter, and a cassette where a gRNA construct can be inserted between a *ScTDH3* promoter and *ScCYC1* terminator (Figure 2). Using combinations of two tRNAs ($tRNA^{Gly}$ and $tRNA^{Asp}$) and two ribozymes HH and HDV, we designed four different gRNA constructs and inserted these into the pUDP002 backbone (Figure 2). The target gene was *ADE2* as its inactivation leads to intracellular accumulation of a red pigment and thus mutants are readily identified on agar plates (Ugolini and Bruschi, 1996). *Z. parabailii* carries two copies of *ADE2*, present on chromosome 1 (*ADE2*chr1) and chromosome 2 (*ADE2*chr2), that are 96% identical at the nucleotide level. To compare the efficiency of the four gRNA constructs, a synthetic construct comprising a common target sequence for the two *ADE2* alleles of *Z. parabailii* ATCC60483, flanked by different gRNA expression systems, was inserted into the pUDP002 backbone, creating four pUDPZb plasmids (Table 2). While transformation efficiency varied between experiments, consistent patterns were seen (Supp Table 1). The average number of transformants ranged from 5 – 469, with more transformants consistently obtained from plasmids HH-tRNA and HH-HDV, which carried a HH ribozyme at the 5' end than plasmids tRNA-tRNA, and tRNA-HDV, which had a $tRNA^{Gly}$ in that position. Between 5 and 52 transformants carrying each plasmid were screened to determine the efficiency of *ADE2* inactivation, as determined by the ratio of red (*ade2*) to white (*ADE2*) colonies (Figure 3a). This time, the inverse was seen and whereas gRNA constructs carrying a 5' self-cleaving $tRNA^{Gly}$ gave rise to *ade2* mutants at a frequency of 68 - 75%, almost no *ade2* mutants were generated by gRNA constructs using a 5' HH ribozyme. No significant differences in efficiency were observed between having a $tRNA^{Asp}$ or a HDV ribozyme at the 3' end of the gRNA construct.

Allele-specific primers were used to amplify a region of both *ADE2* loci to determine the genetic lesions in the mutants. To that end, DNA amplification of *ADE2* target region was performed using specific primers designed for each locus (Table 4), sequenced by Sanger sequencing, and aligned to the wild type gene sequence. Amplification of both *ADE2* loci of red colonies was mostly not successful, indicating a probable more extensive mutation affecting the annealing sequence of the oligonucleotides. An example of a mutated red colony is shown in Figure 3b, where the double strand break was repaired by the insertion of a thymine in both the alleles. This type of repair, inserting a thymine in the fourth position at 5' before the PAM region, was observed also in other three analysed red colonies, in at least one of the targeted loci. This target specific mutation of a nucleotide insertion is described as recurrent and generally predictable for mutations generated by the CRISPR-Cas9 system (Allen et al., 2019). Based on these results, generation of efficient gRNAs in *Z. parabailii* should use a 5' $tRNA^{Gly}$ rather than a 5' HH ribozyme and the 3' element can be either $tRNA^{Asp}$ or a HDV ribozyme.

For the analysis of the disruption of the gene by HDR, two DNA repair fragments were designed. Each of the fragments contains either *kanMX* or *natMX* hybrid genes, conferring resistance to G418 or clonNAT, respectively. Additionally, since efficiency of HDR is known as low in *Z. parabailii*, long homologous arms (around 500 bp), specific for each allele upstream and downstream to the gene, were added to the 5' and 3' ends of the DNA repair fragments (Figure 3c). The average number of transformants ranged from 7–698, and more transformants were obtained from plasmids HH-tRNA and HH-HDV than from plasmids tRNA-tRNA and tRNA-HDV. To calculate the efficiency of integration of DNA repair fragments into the targeted loci (*kanMX* into *ADE2*chr1 and *natMX* into *ADE2*chr2), all the transformants obtained were replica plated on YPF solid media containing G418 or clonNAT to select for resistant transformants. The colonies growing on each selective media were analysed by PCR for the integration of the fragment to the targeted loci using specific primers for each locus and repair fragment. From this analysis, it was observed that the frequency of either ade2chr1::*kanMX* or ade2chr2::*natMX* strains range widely from 0.5% to around 2% according to the used gRNA construct and DNA repair fragment (Supp Table 2). However, ade2chr1::*kanMX* ade2chr2::*natMX* in the same strains were not observed in

the analysed transformants (Supp Table 2), meaning that the efficiency of *ADE2* inactivation is high, based on the obtained ratio of red colonies. However, efficiency of gene deletion by double repair by HDR was not observed.

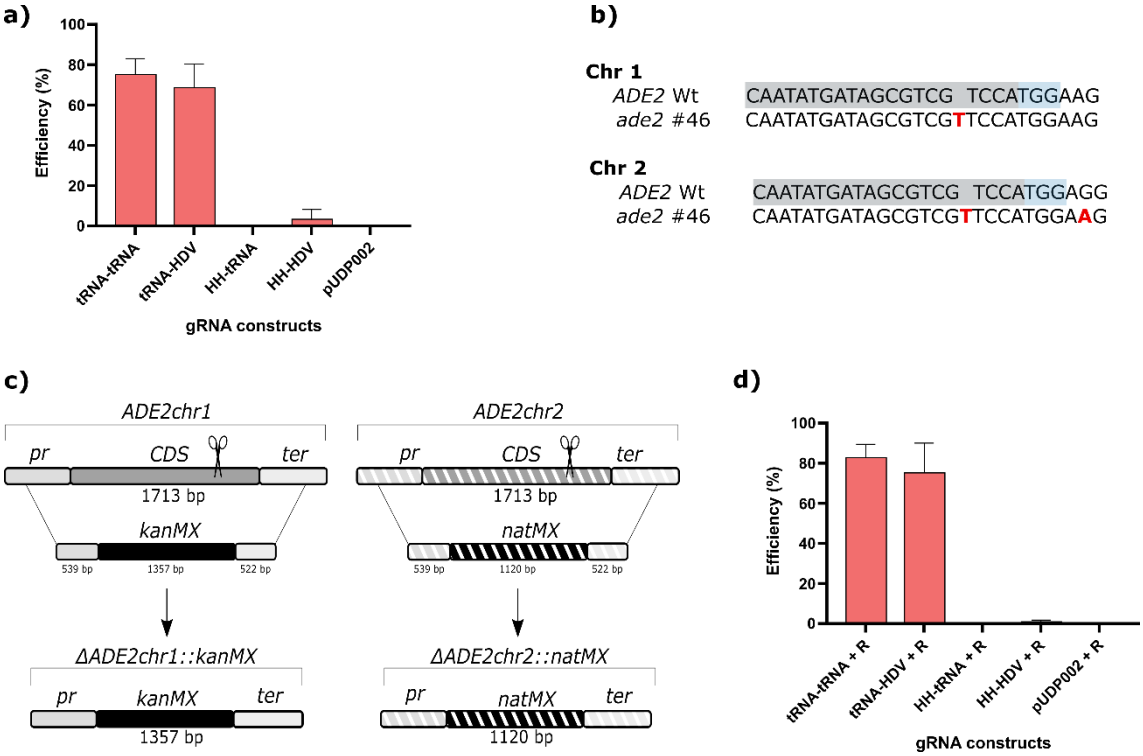


Figure 3: Evaluation of different gRNA constructs for the disruption of *ADE2* in *Z. parabailii*. a) Efficiency of red colonies obtained using Cas9 plasmids (pUDPZb) containing different gRNA constructs targeting *ADE2.* b) example of the mutations obtained in *Z. parabailii ADE2* alleles when targeting the common region highlighted in grey with the PAM sequence highlighted in blue. c) DNA cassettes bearing *kanMX* or *natMX* and a 500 bp homology arms, specific to each allele targeting the deletion of each of the two alleles of the gene. d) Efficiency of red colonies obtained using Cas9 plasmids (pUDP002) containing different gRNA constructs targeting *ADE2* with co-transformation of two DNA cassettes bearing a selective marker.

Construction of *Z. parabailii dnl4* strain

As HDR of both loci of *ADE2* in *Z. parabailii* was not observed, a mutant defective in one component of the NHEJ-related pathway was created, by disrupting *DNL4,* which encodes a ligase required in the NHEJ (Nambu-Nishida *et al.*, 2017).

For the targeted disruption of the *DNL4* alleles in *Z. parabailii* in chromosomes 2 and 7 (*DNL4*chr2, *DNL4*Chr7) that are 93% identical at the nucleotide level, pUPDZb_U1 was used as the *Cas9* delivery system. For the construction of this plasmid, pUDP002 was used as backbone and a gRNA construct (tRNA-tRNA) was inserted between the *ScTDH3* promoter and *ScCYC1* terminator. This gRNA construct consists of a BsaI restriction site for the insertion of the DNA target (Supp Figure 1 and Table 3),

resulting in a universal Cas9 plasmid bearing an efficient tRNA systems for the disruption of genes in *Z. parabailii*. Following, five different common DNA targets for *DNL4* (Table 4) were cloned into the pUDPZb_U1 (Supp Figure 1), obtaining five different *Cas9* plasmids. Transformations were performed using the constructed plasmids (Supp Table 3), and three colonies from each transformation were selected for DNA amplification and Sanger sequencing of the targeted loci. Initial analysis indicated that only common target 3 gave rise to mutations at the *DNL4* locus, therefore further analysis focused on this set of transformants. Five more transformants were selected for analysis by sequencing of the targeted region of the *DNL4* alleles. Out of a total of eight analysed transformants, two had a double disruption of the gene by a nucleotide insertion in each *DNL4* loci in the targeted region (Figure 4a), whereas the remaining six transformants showed intact *DNL4* target sequences. The insertion of a single nucleotide occurred a few nucleotides before the PAM sequence, with an insertion of a cytosine in both alleles of mutant 1 and a thymine in both alleles of mutant 2 (Figure 4a). This *dnl4* mutant #1 was the selected mutant for further use.

Sequence analysis of the mutated *dnl4* loci in mutant #2 revealed an unexpected result, as sequence immediately 3' of the target (mutation) in the *DNL4*chr7 allele actually matched that of the *DNL4*chr2 allele. In fact, we could see that there was a short (127-195 bp) allelic exchange between the two *DNL4* loci (Figure 4a). The precise cross-over point could not be determined because the sequence is identical for part of the region. Allelic exchange has been associated to the use of CRISPR-Cas9 system for genome editing in yeast, leading to the loss of heterozygosity in diploid yeast (Gorter de Vries 2018), and it can be explained as sequential events of cut and repair mechanisms after the double strand break caused by the CRISPR-Cas9 system. As schematized in Figure 4b, the hypothesis of this allelic exchange relies on the repair of the double stranded break of the *DNL4*chr7 using the same region of *DNL4*chr2 as a template for HDR after the repair of the first locus by NHEJ and the insertion of a thymine. This result caused us to reassess the *ade2* mutants generated previously (Figure 3), and we could see that in the *ade2* mutant shown in Figure 3b, a short region of *ADE2*chr1 was used as template for the repair of the double strand break in

*ADE2*chr2. Although we did not analyse all the mutants in detail, it is clear that this mechanism also gave rise to at least some of the *ade2* double mutants.

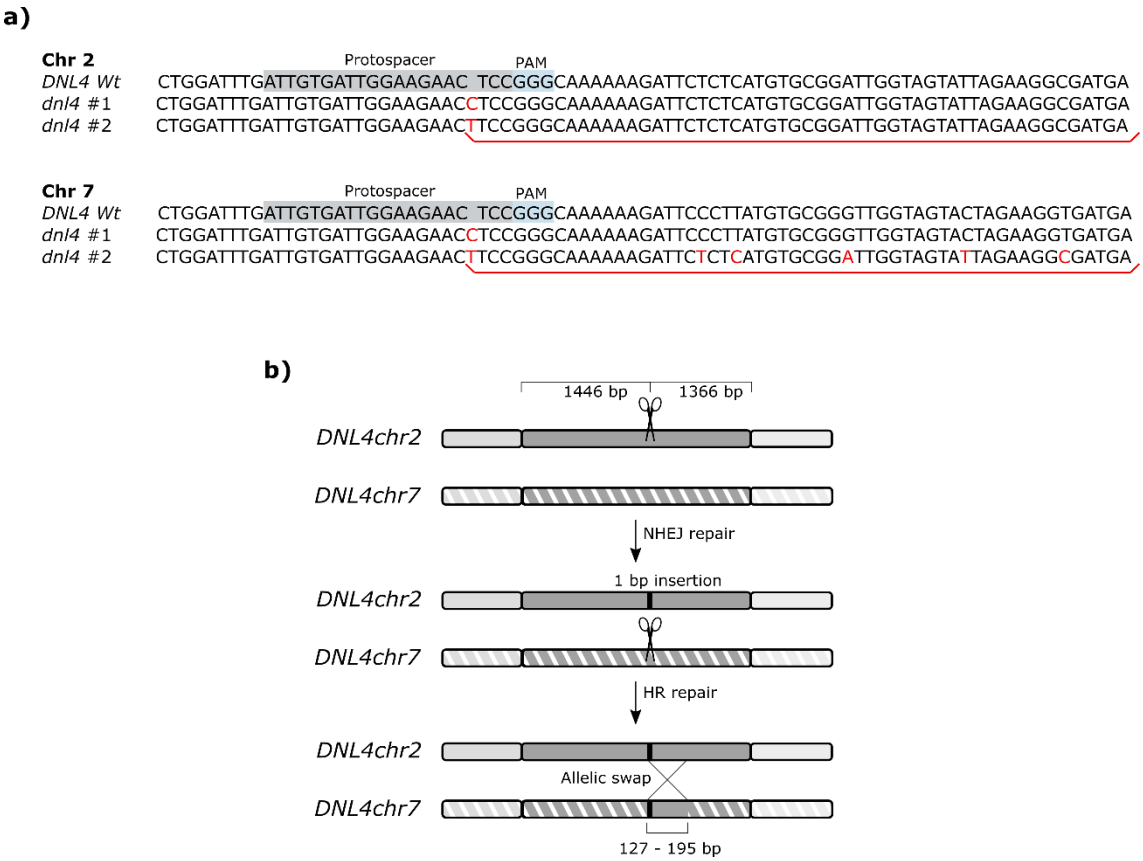


Figure 4: Construction of *Z. parabailii dnl4* mutants 1 and 2 obtained using the pUDPZb plasmid. a) Sequencing analysis of the *DNL4* target regions of two mutants with the double inactivation. Insertions of nucleotides in red in the target region. Allele in chromosome 7 of mutant 2 shows allelic exchange with chromosome 2. b) Representation of the repair mechanism observed in mutant 2: *DNL4*chr 2 is cut by Cas9 in the targeted sequence, and the double strand break is repaired by NHEJ (as observed in most cases), leading to a frameshift mutation; Cas9 cuts *DNL4*chr7, and the first allele is used as template for HDR, resulting in two disrupted alleles due to the same frameshift mutation; in chromosome 7 a hybrid protein is created.

Construction of targeted mutants in *Z. parabailii dnl4*

To understand if the *dnl4* double mutation in *Z. parabailii* favours the HR-mediated repair system in this yeast, a comparison between the efficiency of this system in *Z. parabailii* wt and *Z. parabailii dnl4 #1* strains was performed. To this end, two independent transformation experiments were performed with the pUDPZb (tRNA-tRNA) targeting *ADE2* with and without DNA repair fragments on *Z. parabailii* WT and *Z. parabailii dnl4 #1* strains (Supp Table 4). Transformants growing on YPD

supplemented with hygromycin were screened for the red phenotype on minimal medium enriched with adenine, and the ratio of red/white colonies was calculated. When no repair fragment was included, the ratio of red colonies was between 27.5 and 60 % for the WT strain, while no red colonies were observed for the *dnl4* mutant (Supp Table 4). When co-transformed with DNA repair fragments, 49 – 60 % of the WT and 63 – 83 % of the *dnl4* mutant transformants were red. To understand the actual efficiency of repair of DSB by HDR, red transformants (hence *ade2* mutants) from the first transformation of WT and *dnl4* mutants with repair fragments were assessed for antibiotic resistance (indicating integration of the repair fragment) by replica plating onto YPD supplemented with G418 or clonNAT. Out of the 31 putative *ade2* mutants (red colonies) of the WT transformation, only three grew on G418 and one on clonNAT. Molecular analysis with primers designed to detect insertion of the DNA repair fragments showed that only one of the transformants had integrated *kanMX* at the *ADE2*chr1 locus and none had integrated clonNAT at the corresponding *ADE2*chr2 locus. These data are consistent with those for previous experiments (Figure 3) and indicate that targeting of repair fragments is inefficient, and mutations in the WT strain arise by NHEJ repair and not by HDR. For the *dnl4* mutant, 17 out of 27 transformants were red and these were further analysed. Firstly, the growth on G418 and clonNAT was assessed, and secondly PCR analysis to detect insertion of DNA repair fragments was performed. Of the 17 *ade2* mutants, either 10 or 11 grew on clonNat and G418, respectively (Figure 5a). The molecular analysis corroborated with these data, with all G418+ strains amplifying the *kanMX* gene at the *ADE2chr1* locus and all clonNat+ strains amplifying *natMX* at the *ADE2chr2* locus (Figure 5b). Only four strains had the genotype *ade2chr1::kanMX ade2chr2::natMX* indicating integration of each repair fragment at the cognate locus (Figure 5, lanes 2, 6, 8, 12). The remaining thirteen transformants had integrated either *kanMX* (Figure 5, lanes 5, 7, 9, 13, 15, 16, 17) or *natMX* (Figure 5, lanes 1, 3, 4, 10, 11, 14) at the correct *ADE2* locus, but not both. These thirteen mutants were examined in more detail to determine what molecular events had taken place to generate these mutants, which by definition lacked functional *Ade2*, and so must have lesions at both *ADE2* loci. The six *ade2chr2::natMX* strains all had thymine insertion in the targeted region in *ADE2*chr1 before the PAM region. Thus, despite this being a *dnl4* mutant, *ADE2*chr1 was

inactivated by point mutation. The situation with the seven *ade2chr1::kanMX* strains was more complex, and efforts to amplify the *ADE2chr2* locus were not successful. One possibility was that recombination between homeologous chromosomes replaced *ADE2*chr2 with the *ADE2*chr1 locus (including the integrated *kanMX* gene) but PCR analysis ruled this out. In fact, we found that all seven strains carried substantial chromosomal deletions at the *ADE2*chr2 locus. PCR amplification to detect various genes upstream and downstream of the locus (Supp Figure 2) identified two classes of mutants (Figure 5c). Three of the mutants had a deletion of <1.65 MB as the flanking genes *UBP10* and *DIS3* were present. The other four mutants appear to have lost an additional region downstream of *DIS3* and probably are deleted to the right telomere of chr 2.

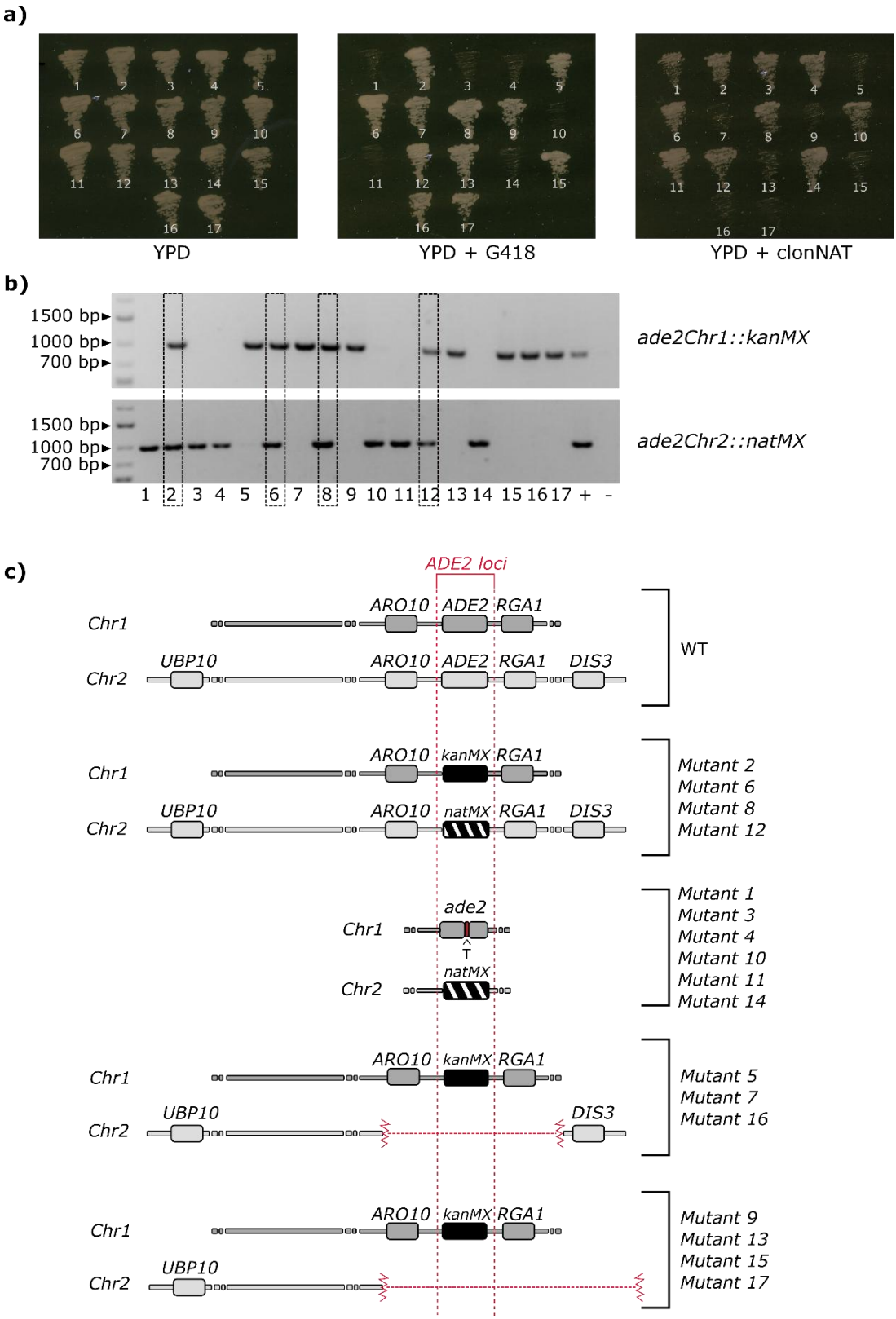


Figure 5: Confirmation of integration of the *kanMX* and *natMX* repair fragments in *Z. parabailii dnl4* mutant. a) Growth of same 17 colonies on YPD, YPD+G418 and YPD+clonNAT. b) Molecular confirmation of the integration of the repair fragments with *kanMX* or *natMX* (~1000 bp) into the targeted region of the *ADE2* allele in chromosome 1 and chromosome 2, respectively. Mutants analysed in lanes 2, 6, 8 and 12 have both fragments integrated in the

targeted region. c) Nature of the mutations in the *ADE2* locus in chromosome 1 and chromosome 2 of the seventeen analysed *Z. parabailii dnl4 ade2* mutants.

Discussion

To our knowledge the only hybrid where efficient gene editing system was used CRISPR-Cas9 is *S. pastorianus* and de novo *S. pastorianus* strains constructed in the lab (Gorter de Vries, et al. 2017, Mertens et al. 2019). This work is an expansion of establishing targeted gene editing in the yeast hybrid *Z. parabailii* that has promising traits such as resistance to weak acids and low pH. The Cas9-gRNA co-expression plasmid backbone pUDP002, previously established for gene editing in several yeasts in the work of Juergens et al., was used in this study for insertion of the tested gRNA systems. To establish the most efficient gene targeting system in *Z. parabailii* using Pol II promoters, four different gRNA constructs with combinations of tRNAs and ribozyme systems were designed and tested. In this study we used self-cleaving elements HH and HDV ribozymes as well as $tRNA^{gly}$ considering that tRNA glycine cassettes have been widely used for gRNA expression in yeasts and plants (Qi *et al*., 2016) and $tRNA^{asp}$, when two tRNA's are used in combination the formation of secondary structures can be avoided. We established that the gRNA systems consisting of 5' tRNA (tRNA-tRNA and tRNA-HDV) were the most efficient for *ADE2* disruptions and no significant differences in efficiencies were observed between the two systems, with a disruption efficiency of 68 – 75%. In this work the ribozyme systems failed to produce red colonies, similar to what was described in a previous work (Kuanyshev *et al.*, 2021). Similar to other non-conventional yeasts, *Z. parabailii* has lower HDR efficiency relative to *S. cerevisiae*, as previously shown in works with classical deletion of genes in *Z. bailii* (Mollapour and Piper, 2001, Dato et al., 2010, Passolunghi et al., 2010) and this was also observed in the deletion of *ADE2* in the present work using DNA repair fragments. In WT *Z. parabailii*, efficiency of HDR by integration of single DNA repair fragment in the targeted locus was 2% or lower. Similar low efficiency was observed in the deletion of *LEU2* in *Z. parabailii* using classical method of deletion with cassettes containing long homology arms (Dato et al., 2010), where 1 to 5% single integration efficiency was achieved by HDR. Nevertheless, deletion of both alleles of *ADE2* using DNA repair fragments by CRISPR-Cas9 was not successful in *Z. parabailii* WT strain. To overcome this limitation a *dnl4* double mutant was constructed using the CRISPR-Cas9 gRNA

system with 5' tRNA-tRNA 3' which proved to be functional with high efficiency. However, the efficiency of *DNL4* disruptions in *Z. parabailii* was lower than that of *ADE2*, possibly due to gRNA target selections or target DNA sequence accessibility of *DNL4*. Target specificity has been described as a critical issue of applying a successful CRISPR-Cas9 system in different organisms. To select the most suitable target for *DNL4* alleles in *Z. parabailii* and avoid off-targets, the software sgRNACas9 was used (Xie *et al.*, 2014), which allows for detailed analysis of the sequences mismatches and off-targets in the genome. However, other issues such as accessibility of Cas9 to recognise PAM and seed sites in the genome might have affected the efficiency of the designed system (Wu, et al., 2014). Additionally, the low total number of screened transformants eight, may bias the obtained efficiency.

Improved efficiency of HDR was confirmed in the *Z. parabailii dnl4* mutant for the deletion of *ADE2.* The double disruption efficiency was 63%, out of which 15% had correctly integrated both repair fragments in the *ADE2* loci; 22% showed a single integrated fragment and a point mutation on the other locus; 26% had a single integrated fragment and a substantial chromosomal deletion. Since no random integration of repair fragments in the genome was observed when using the *dnl4* mutant, marker genes can be efficiently used to screen for correct mutants. This efficiency might probably be further improved with the use of even longer homology arms, as observed in the *LEU2* deletion using classical deletion method (Dato et al., 2010).

In this work on *ADE2* and *DNL4* gene disruptions, targeting both alleles of the gene led to interallelic chromosomal rearrangements at the target site. Recently, similar events when using CRISPR-Cas9 gene editing system was described as the loss of heterozygosity in diploid yeasts by the exchange of sequences between the targeted allele and its homolog when single alleles are targeted (Gorter de Vries 2019). Deletion of large sequences or chromosome arms in the targeted regions has also been reported as an effect of the use of CRISPR-Cas9 (Kosicki et al., 2018) and may explain the failure in the amplification of certain loci of the obtained mutants. Both, the replacement and deletion of large sequences of one of the targeted alleles have been related to the loss of heterozygosity when using CRISPR-Cas9 gene editing system (Kosicki et al., 2018, Gorter de Vries 2019). In the case of hybrid yeast with industrial

importance, the loss of heterozygosity might have a relevant impact (Gorter de Vries 2017b), and for this reason finding alternatives to minimize this event is of interest.

Conclusion

The CRISPR-Cas9 gRNA expression system developed in this study enabled targeted gene engineering in the hybrid *Z. parabailii*. This system was successfully applied for the deletion of *ADE2* as proof of concept and was further used for construction of a *Z. parabailii DNL4* mutant to improve the efficiency of HDR. The allelic exchange mechanism seen in the *Z. parabailii* WT strain as well as *DNL4* mutant gives a starting point for assessing inter allelic rearrangements and to understand if one DNA repair fragment can suffice the deletion of two alleles of a gene. Since *Z. parabailii* is resilient to genetic modifications, these advances greatly increase its genetic accessibility thereby facilitating future research into the complex hybrid genome of *Z. parabailii*.

Supplementary Material

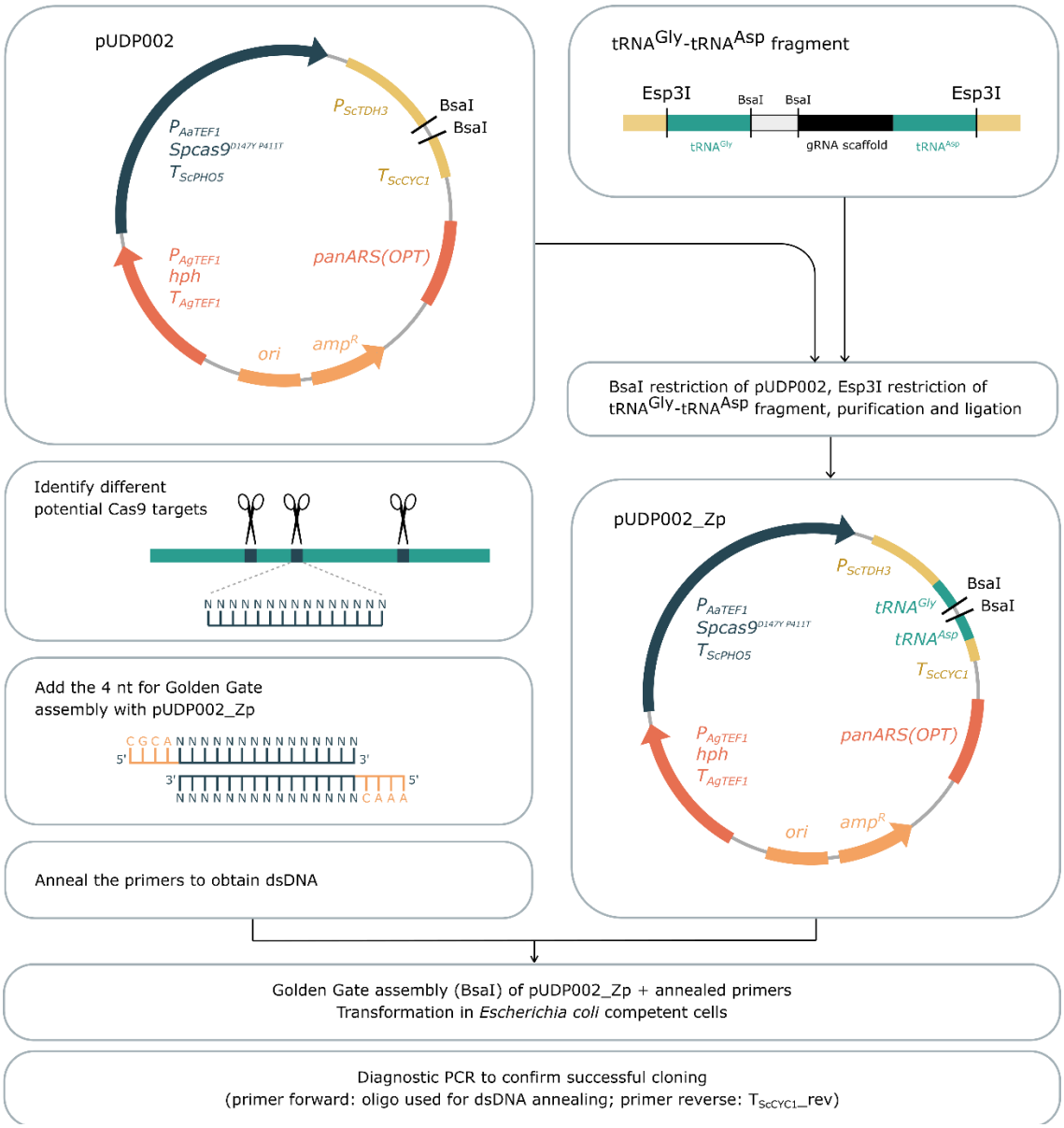


Figure S1: Steps involved in the construction of pUDPZb_U1 plasmid. The gRNA cassette was digested with Esp3I and inserted using the BsaI highlighted restriction sites under the control of *Sc*TDH3 promoter and *Sc*CYC1 terminator by ligation; Cas9 targets were identified and constructed by oligonucleotides hybridization and inserted in the universal plasmid in the newly constructed BsaI restriction site by Golden Gate Assembly.

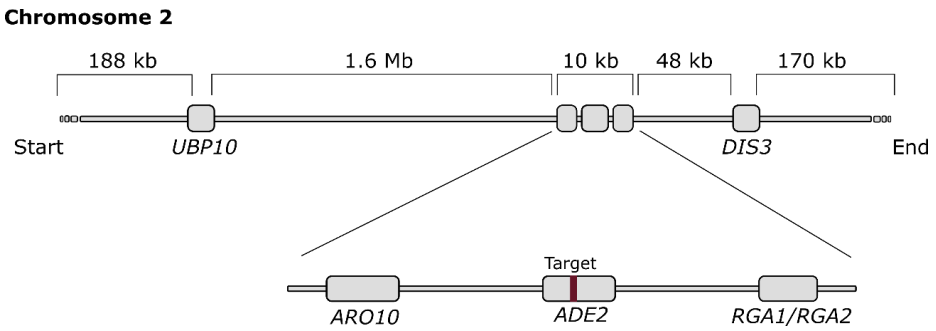


Figure S2. Position of amplified genes in chromosome 2 (not in scale) and in red the target DNA sequence of *ADE2chr2*

Table S1: Total number of transformants per replicate and ratio of red colonies of the overall analysed transformants

	Number of transformants				Red colonies			
replicates	1	2	3	Average	1	2	3	Average
tRNA-tRNA	129	102	57	96	43/52	35/52	39/52	39/52
tRNA-HDV	5	25	56	29	4/5	14/25	36/52	18/27
HH-tRNA	63	108	469	213	0/52	0/52	0/52	0/52
HH-HDV	47	180	278	168	4/47	0/52	0/52	1/50
pUDPZb	59	64	212	112	0/52	0/52	0/52	0/52

Table S2: Total number of transformants per replicate and frequency of confirmed insertion of repair fragments *kanMX* and *natMX* into the targeted allele region of *ADE2*.

	Number of transformants			Confirmed integration of *kanMX*				Confirmed integration of *natMX*			
replicates	1	2	3	1	2	3	Average %	1	2	3	Average %
tRNA-tRNA + R1 + R2	78	77	164	1 (1.28 %)	0	0	0.43 ± 0.74	1 (1.28 %)	2 (2.60 %)	3 (1.83 %)	1.90 ± 0.66
tRNA-HDV + R1 + R2	213	189	93	1 (0.47 %)	4 (2.12 %)	1 (1.08 %)	1.22 ± 0.83	7 (3.29 %)	2 (1.06 %)	0	1.45 ± 1.68
HH-tRNA + R1 + R2	109	106	698	0	0	0	0.00	0	0	0	0.00
HH-HDV + R1 + R2	82	7	481	0	0	1	0.07 ± 0.12	0	0	0	0.00
pUDPZb + R1 + R2	109	491	663	0	0	0	0.00	0	0	0	0.00

Table S3: Total number of transformants per replicate, and red/white colonies ratio for *Z. parabailii* Wt and *Z. parabailii Δdnl4.*

	Number of transformants				Red colonies					
	Zp Wt		Zp_ *Δdnl4*		Zp Wt			Zp_*Δdnl4*		
replicates	1	2	1	2	1	2	Efficiency (%)	1	2	Efficiency (%)
pUDPZb	75	200-250	236	350-400	0/52	-	0	0/52	-	0
pUDPZb:Ade2	15	56	10	17	9/15	14/51	60, 27.5	0/10	0/17	0
pUDPZb:Ade2+R1+R2	52	75	27	89	31/52	25/51	59.6, 49.0	17/27	39/47	63.0, 83.0
R1+R2	G418-66, NAT-*	-	G418-27, NAT-*	-	0/27	-	0	0/27	-	0
pUDPZb:Ade2+R1	-	210	-	47	-	23/51	45.1	-	20/47	42.6

Table S4: Total number of colonies growing on G418 and clonNAT and double integration efficiency confirmed by molecular analysis

	Growing on G418				Growing on clonNAT				Confirmed double integration			
	Zp Wt		Zp Δdnl4		Zp Wt		Zp Δdnl4		Zp Wt		Zp Δdnl4	
replicates	1	2	1	2	1	2	1	2	1	2	1	2
pUDPZb:Ade2+R1+R2	3/52	10/51	11/27	33/47	-	6/51	-	23/47	0/52	-	4/27 (15%)	-
pUDPZb:Ade2+R1	-	3/51	-	20/47	-	-	-	-	-	-	-	-

References

Allen, F., Crepaldi, L., Alsinet, C. et al. (2019) Predicting the mutations generated by repair of Cas9-induced double-strand breaks. Nat Biotechnol. https://doi.org/10.1038/nbt.4317

Alsammar H, Delneri D. (2020) An update on the diversity, ecology and biogeography of the *Saccharomyces* genus. FEMS Yeast Res. https://doi: 10.1093/femsyr/foaa013.

Arras, S. D. M. et al. (2016) 'Targeted genome editing via CRISPR in the pathogen *Cryptococcus neoformans*', PLoS ONE. https://doi: 10.1371/journal.pone.0164322.

Branduardi, P. et al. (2004) 'The yeast *Zygosaccharomyces bailii*: a new host for heterologous protein production, secretion and for metabolic engineering applications', FEMS Yeast Research. https://doi:10.1016/S1567-1356(03)00200-9.

Branduardi, P., Dato, L. and Porro, D. (2014) 'Molecular Tools and Protocols for Engineering the Acid-Tolerant Yeast *Zygosaccharomyces bailii* as a Potential Cell Factory', in Mapelli, V. (ed.) Yeast Metabolic Engineering. New York, NY: Springer New York (Methods in Molecular Biology). https://doi:10.1007/978-1-4939-0563-8_4.

Cernak, P., Estrela, R. and Poddar, S. (2018) 'Engineering *Kluyveromyces marxianus* as a Robust Synthetic', American Society For Microbiology. https://doi:10.1128/mBio.01410-18

Christine, L. J. et al. (2007) 'Characterization of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces bayanus* var. uvarum', FEMS Yeast Research. https://doi: 10.1111/j.1567-1364.2007.00207.x.

Dato, L. et al. (2010) 'Advances in molecular tools for the use of *Zygosaccharomyces bailii* as host for biotechnological productions and construction of the first auxotrophic mutant', FEMS Yeast Research. https://doi: 10.1111/j.1567-

1364.2010.00668.x.

Gao, Y. and Zhao, Y. (2014) 'Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing', Journal of Integrative Plant Biology. https://doi:10.1111/jipb.12152.

González, S. S., Barrio, E. and Querol, A. (2008) 'Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and *S. kudriavzevii* in brewing', Applied and Environmental Microbiology. https://doi: 10.1128/AEM.01867-07.

Gorter de Vries AR, Pronk JT, Daran JG. (2017b) Industrial Relevance of Chromosomal Copy Number Variation in *Saccharomyces* Yeasts. Appl Environ Microbiol. https://doi: 10.1128/AEM.03206-16.

Gorter de Vries, A.R. et al. (2017) 'CRISPR-Cas9 mediated gene deletions in lager yeast *Saccharomyces pastorianus'*, Microbial Cell Factories, 16(1), p. 222. doi:10.1186/s12934-017-0835-1.

Hoffman, C.S. and Winston, F. (1987) 'A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli'*, Gene. https://doi:10.1016/0378-1119(87)90131-4.

Horwitz, A. A. et al. (2015) 'Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas', Cell Systems. https://doi: 10.1016/j.cels.2015.02.001.

Hulin M, Wheals A. (2014) 'Rapid identification of *Zygosaccharomyces* with genus-specific primers'. Int J Food Microbiol. 3;173:9-13. https://doi: 10.1016/j.ijfoodmicro.2013.12.009.

Jacobs, J. Z. et al. (2014) 'Implementation of the CRISPR-Cas9 system in fission yeast', Nature Communications. Nature Publishing Group. https://doi: 10.1038/ncomms6344.

Jiao, X. et al. (2019) 'Developing a CRISPR/Cas9 System for Genome Editing in the Basidiomycetous Yeast *Rhodosporidium toruloides*', Biotechnology Journal.

https://doi: 10.1002/biot.201900036.

Juergens, H. et al. (2018) 'Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid', FEMS Yeast Research, https://doi: 10.1093/femsyr/foy012.

Karaalioğlu O, Yüceer YK. (2021). Nonconventional yeasts to produce aroma compounds by using agri-food waste materials. FEMS Yeast Res. https://doi: 10.1093/femsyr/foab063.

Krogerus, K. et al. (2017) 'Novel brewing yeast hybrids: creation and application', Applied Microbiology and Biotechnology. Applied Microbiology and Biotechnology. https://doi: 10.1007/s00253-016-8007-5.

Kuanyshev, N. et al. (2016) 'Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation', FEMS Yeast Research. https://doi: 10.1093/femsyr/fow058.

Kuanyshev, N. et al. (2017) 'The spoilage yeast *Zygosaccharomyces bailii*: Foe or friend?', Yeast. https://doi: 10.1002/yea.3238.

Kuanyshev, N. et al. (2020) 'Domesticating a food spoilage yeast into an organic acid-tolerant metabolic engineering host: Lactic acid production by engineered *Zygosaccharomyces bailii*', Biotechnology and Bioengineering. https://doi:10.1002/bit.27576.

Löbs AK, Schwartz C, Wheeldon I (2017). Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. Synth Syst Biotechnol. https://doi: 10.1016/j.synbio.2017.08.002.

Löbs, A. K. et al. (2017) 'CRISPR-Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*', Biotechnology for Biofuels. BioMed Central https://doi:10.1186/s13068-017-0854-5.

Lowe, T.M. and Eddy, S.R. (1997) 'tRNAscan-SE: a program for improved detection of

transfer RNA genes in genomic sequence', Nucleic Acids Res. https://doi:10.1093/nar/25.5.955

Mingfeng Cao , Meirong Gao , Carmen Lorena Lopez-Garcia , Yutong Wu, Arun Somwarpet Seetharam, Andrew Josef Severin, Z. S. (2017) 'Centromeric DNA facilitates Nonconventional Yeast Genetic Engineering'. ACS Synth Biol. https://doi:10.1021/acssynbio.7b00046

Mixão, V. and Gabaldón, T. (2020) 'Genomic evidence for a hybrid origin of the yeast opportunistic pathogen *Candida albicans*', BMC Biology. https://doi:10.1186/s12915-020-00776-6.

Morse, N. J. et al. (2018) 'T7 Polymerase Expression of Guide RNAs in vivo Allows Exportable CRISPR-Cas9 Editing in Multiple Yeast Hosts', ACS Synthetic Biology. https://doi: 10.1021/acssynbio.7b00461.

Nambu-Nishida, Y. et al. (2017) 'Development of a comprehensive set of tools for genome engineering in a cold- A nd thermo-tolerant *Kluyveromyces marxianus* yeast strain', Scientific Reports. Springer US. https://doi: 10.1038/s41598-017-08356-5.

Numamoto, M., Maekawa, H. and Kaneko, Y. (2017) 'Efficient genome editing by CRISPR/Cas9 with a tRNA-sgRNA fusion in the methylotrophic yeast *Ogataea polymorpha'*, Journal of Bioscience and Bioengineering. https://doi: 10.1016/j.jbiosc.2017.06.001.

O'Brien CE, Oliveira-Pacheco J, Ó Cinnéide E, Haase MAB, Hittinger CT, Rogers TR, Zaragoza O, Bond U, Butler G. Population genomics of the pathogenic yeast *Candida tropicalis* identifies hybrid isolates in environmental samples. PLoS Pathog. https://doi: 10.1371/journal.ppat.1009138.

Ortiz-Merino, R. A. et al. (2017) 'Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch', PLoS Biology. https://doi: 10.1371/journal.pbio.2002128.

Ortiz-Merino, R. A., Kuanyshev, N., Byrne, K. P., Varela, J. A., Morrissey, J. P., Porro, D., Branduardi, P. (2018). Transcriptional Response to Lactic Acid Stress in the Hybrid Yeast *Zygosaccharomyces parabailii*. Appl Environ Microbiol. https://doi:10.1128/AEM.02294-17

Qi, W., Zhu, T., Tian, Z. et al. (2016) High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. BMC Biotechnol. https://doi.org/10.1186/s12896-016-0289-2

Schultz, J. C., Cao, M. and Zhao, H. (2019) 'Development of a CRISPR/Cas9 system for high efficiency multiplexed gene deletion in *Rhodosporidium toruloides'*, Biotechnology and Bioengineering. https://doi: 10.1002/bit.27001.

Schwartz, C. M. et al. (2016) 'Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in *Yarrowia lipolytica*', ACS Synthetic Biology. https://doi: 10.1021/acssynbio.5b00162.

Shan, L., Dai, Z. and Wang, Q. (2021) 'Advances and Opportunities of CRISPR/Cas Technology in Bioengineering Non-conventional Yeasts', Frontiers in Bioengineering and Biotechnology. https://doi: 10.3389/fbioe.2021.765396.

Singh, J. et al. (2019) 'RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning', Nature Communications. https://doi:10.1038/s41467-019-13395-9.

Stovicek, V., Holkenbrink, C. and Borodina, I. (2017) 'CRISPR/Cas system for yeast genome engineering: advances and applications', FEMS yeast research. https://doi: 10.1093/femsyr/fox030.

Suh SO, et al. (2013) Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., two new species closely related to *Zygosaccharomyces bailii*. Int. J. Syst. Evol. Microbiol. https://10.1099/ijs.0.048058-0

Vyas, V. K., Barrasa, M. I. and Fink, G. R. (2015) 'A Candida albicans CRISPR system

permits genetic engineering of essential genes and gene families', Science Advances. . https://doi: 10.1126/sciadv.1500248.

Xie, S. et al. (2014) 'sgRNAcas9: A Software Package for Designing CRISPR sgRNA and Evaluating Potential Off-Target Cleavage Sites', PLoS ONE. https://doi:10.1371/journal.pone.0100448.

General Conclusion and future perspectives

The ongoing need to substitute the majority of fossil derived chemicals and fuels with biobased products requires a biorefinery to be efficient and economically viable. Therefore, it is important to develop effective and robust microbial cell factories, which can valorize lignocellulosic or other residual biomass by fermentations and produce economically viable biobased products. This can be achieved either by natural metabolism of the microbial host to transform biomass components into various compounds or by applying genetic engineering and synthetic biology approaches to improve their ability, thereby having to decide whether to follow or to force their nature respectively. Since high development costs, such as those associated with genetic modifications, can only be supported if the final product has a high value, the strategy to improve bioprocesses must also be consistent with cascading principles. The use of additional biologically derived tools such as enzymatic cocktails, which have the ability to liberate sugars from leftover biomasses for subsequent microbial fermentation, thereby maximizing their potential can be taken into consideration. Therefore, biodiversity must be a crucial cornerstone for bioprocesses to foster a transition to bioeconomy. Based on this outlook, in **chapter 1** of this thesis we reviewed the link between LCBs composition, choice of enzymatic cocktail and selection of yeast species and strains that need to be considered in an integrated fashion to enable the development of an efficient bio-based process. We discussed the pivotal role of enzymatic cocktail optimization to unlock the potential of non-Conventional yeasts, which, thanks to broader substrate utilization, inhibitor resistance and peculiar metabolism, can widen the array of feedstock and products of biorefineries.

The overall objective of this thesis was to expand the industrial potential of two non-conventional yeasts, *Zygosaccharomyces parabailii* and *Kluyveromyces marxianus,* by applying direct and indirect strain engineering approaches. **Chapter 2** of this thesis was focused on *K. marxianus.* Taking advantage of its ability to utilize a wide range of sugar substrates and thermotolerance traits, we aimed to improve robustness of this yeast for growth in SBP at low pH using an ALE approach. Preliminary growth kinetics suggested a temperature trade-off when *K. marxianus* parental strain was exposed to

SBP pH 4.3, as the growth of this strain was faster with shorter lag phase at 30 °C. Preadaptation of cells in SBP pH 4.3 facilitated the growth of *K. marxianus* parental strain in the stressed condition of SBP pH 3.0 at 30 °C and allowed the growth at 40 °C. By measuring mRNA levels of two molecular determinants involved in ROS scavenging we could demonstrate that preadaptation triggers a response involving at least Sod enzymes and glutathione. Furthermore, by implementing ALE approach we obtained *K. marxianus* variants capable of growth in 8% SBP pH 3.0 without the need for preadaptation in SBP at pH 4.3, at least at 30 °C. Among these isolates some showed a significantly reduced lag phase, or a higher specific growth rate if compared with the initial strain. Further characterizations of these isolates are required for understanding phenotype-genotype association. For this purpose, the best performing isolate with consistent lowest lag phase in the stressed condition can be selected for genome re-sequencing. Furthermore, analysis of mutations in the evolved isolate followed by reverse engineering in the parental strain can aid in determining the phenotype-genotype correlation. There is a limited knowledge on molecular characterisation of genes involved in improving the robustness of *K. marxianus* for growth in LCB. The findings from this study are important in terms of the strategy used to obtain a robust microbial host tolerant to multiple stressors in LCB, the significance of preadaptation as well as investigating the possible mechanims involved. This knowledge can be further used to ameliorate other yeast species with promising characteristics but lacking certain phenotypes. *K. marxianus* with its natural capability to utilize a wide range of sugars, associated with an improved growth in sugar beet pulp hydrolysate gained through ALE, may contribute to render it as a reliable microbial host for sustainable production of industrially relevant compounds. Also, integrating adaptive laboratory evolution into metabolic engineering of microbial cells can offer tuning possibilities at multiple levels of the engineering process. Significant improvements in microbial cell factory design may be anticipated by combining two worlds, one where we depend on the creative power of the human mind and the other depends on nature's inherent capacity to optimize existing building blocks in a non-directional way under selective pressure.

Chapter 3 and chapter 4 of this thesis were focused on *Z. parabailii.* Taking advantage of its WOA tolerance trait, we attempted to understand the mechanism involved in

this robust phenotype in chapter 3. For this, we investigated the role of *PDR12* by characterisation of a set of single and double *Z. parabailii pdr12* mutants. The phenotypic characterization determined that in *Z. parabailii* Pdr12p is not involved in tolerance towards longer chain organic acids stress such as sorbic and benzoic, while it is involved in mediating tolerance to butyric and acetic acids, though individual mutants displayed phenotypes consistent with differential roles for each parental allele. The *Z. parabailii PDR12* allele on chromosome 8 contributes to tolerance towards butyric acid rather than the allele on chromosome 2. The above mentioned phenotype can be further confirmed by performing growth kinetics of wild type strain and *pdr12* mutants in the presence of weak organic acids particularly acetic and butyric acid in which *pd12* mutants displayed sensitivity to growth. Moreover, detailed time-resolved studies to check the expression levels of *PDR12* genes at various time points after challenging the wild type cells with weak organic acid stress should allow for a thorough analysis of their exact role in resistance development in *Z. parabailii*. The structural analysis of the two Pdr12p isoforms shows 93% identity and 97% similarity: further investigations will concentrate on the amino acid differences that could justify the different contribution of the two isoforms to butyric acid tolerance. Understanding the mechanism involved in tolerance to organic acids can be used for avoiding the growth of this yeast in food production pipelines to promote its use as a cell factory for the production of organic acids and other bio-products. This work is a significant contribution in understanding the phenotype-genotype correlation involved in the weak acid tolerance trait of *Z. parabailii*.

Despite the potential of *Z. parabailii* as a cell factory because of its robustness in withstanding stressors, genetic engineering is hindered due to the complex genome of the hybrid and lack of optimized gene editing tools such as CRISPR-Cas9. Therefore, the aim of the study in chapter 4 was to develop a CRISPR-Cas9 gene editing system in *Z. parabailii* which allows simultaneous disruption or deletion of two alleles of a gene. For this purpose, we evaluated four different gRNA expression systems consisting of combinations of tRNAs, tRNA and ribozyme or ribozymes as self-cleaving flanking. The functionality of the gRNA systems was tested by analysing the inactivation of *ADE2* in the wild type *Z. parabailii.* Disruption efficiency of 70% by non-homologous end joining was achieved using the gRNA system flanked by two tRNA. Moreover, this gRNA

system was successfully used in the construction of a *Z. parabailii dnl4* mutant. This mutant exhibited homology directed repair efficiency of 15% in the deletion of both *ADE2* alleles compared to the lack of double deletion by HDR observed in the wild type strain. Analysis of mutations in the gRNA target regions of both *ADE2* and *DNL4* genes suggested inter-allelic rearrangements between the two gene loci, as well as absence of large regions of chromosomes. As *Z. parabailii* is resilient to genetic modification, the developments in this study significantly improve its genetic accessibility and facilitate future research into the complex genome of this hybrid. An interesting finding of this study based on use of a gRNA system consisting of two tRNA that portrayed highest efficiency can be further extended to improve gene editing efficiencies in other yeasts. Although CRISPR-Cas9 has ample advantages in the field of metabolic engineering compared to conventional gene recombination techniques, development of a novel CRISPR-Cas9 approach in a yeast species such as *Z. parabailii* can lead to unexpected problems such as absence of large regions of the chromosome as seen in this study.To enable the wide adaptation of this CRISPR-Cas9 approach, the current limitations need to be addressed. These include i) enabling large-scale multiplexing, ii) avoiding the deletion of large regions of chromosomes iii) enabling efficient HR in wild type strain rather than in *dnl4* mutants, as the impairment of DNA damage repair might have a negative effect on fitness during industrial process conditions. On the whole, the findings and methods described in this study enable avenues of research both in genome engineering and in exploration of *Z. parabailii* for industrial uses.

Both *Z. parabailii* and *K. marxianus* are promising species due to their industrial relevance and potential for exploitation. The studies from this work, collectively, represent an important progress in the field. Nonetheless, more efforts are required to expand the industrial potential of these yeasts. Further, this work emphasized the importance of yeast biodiversity exploration, implying that there are still yeasts with unique traits that should be studied and characterized. As a result, we should not limit ourselves merely to the organism we know the best for industrial application, but also benefit from nature's diversity. The knowledge obtained from these studies can be used as future tools and guidelines for exploiting other non conventional yeasts. This is

paving the way for further application of microbial bioprocesses in the worldwide scenario of bioeconomy.

Acknowledgments

My PhD journey within YEASTDOC program was an amazing adventure both from personal and scientific perspective. I am deeply grateful to having gotten this opportunity to be a part of this challenging and interesting program, funded by the **EU**. The research performed in this thesis was funded by the European Union Horizon 2020 research and innovation programme via the **Marie Skłodowska-Curie Actions Grant** Agreement No. 764927.

First, I would like to express my deepest gratitude to my supervisors **Prof. Paola Branduardi** and **Prof. John P. Morrissey** for giving me opportunity to study and career out research at your labs. You helped me to build the foundations for my scientific career, through encouragement and support, mentoring, patience and constructive criticism. You helped me to become more independent in the research field and also to develop the skills needed to pursue a career in research. Also, I greatly appreciate the support and assistance from **Dr. Francesca Doonan and Lucy Taylor,** who did a great job in organizing YEASTDOC meetings, answering inquiries and always being available when needed.

This PhD journey wouldn't have been the same if I had not met my YEASTDOC fellow, great friend and supporter **Luis Ferraz**. We started off on this journey together and it has been a great pleasure knowing and working with you. You made this journey a memorable and an enjoyable one. Also, I deeply acknowledge being a part of the multi-national **YEASTDOC** group along with the other amazing **fellow colleagues**, great friends with who I cherished a lot of wonderful memories and also learnt a lot in the field of science and also from their personalities.

Special thanks to the members of **Brandulab** for being great colleagues. You helped me a lot to go through the cultural and language barrier that I faced during my stay in Italy. Also, the advices and guidance during experimental procedures was crucial for my PhD course. Also, special thanks to my **colleagues** at **UCC** lab, for helping me to successfully get through this journey with great help in terms of scientific knowledge, experiments as well as with great memorable times.

At last, I would like to express my deepest gratitude to my **mom** and **brother.** I greatly appreciate your love and never-ending support and the belief you had which helped me to go through this challenging but amazing journey. Also, my gratitude to my **extended family** who have been a constant source of encouragement and support. And to **my friends**, who have stood with me through thick and thin and have been a great support. I love you with all my heart.