SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Dipartimento di / Department of

Dipartimento di Biotecnologie e Bioscienze/School of Microbiology

Dottorato di Ricerca in/PhD program **SCIENZE DELLA VITA/Microbiology**

Ciclo/Cycle **XXIX**

Elucidation of weak organic acid resistance mechanisms in non-Saccharomyces yeast

A case study of *Zygosaccharomyces parabailii* and *Kluyveromyces marxianus*

Cognome / Surname **Kuanyshev** Nome / Name **Nurzhan**

Matricola / Registration number **788751**

Tutore / Tutor: **Prof. Paola Branduardi**

Cotutore / Co-tutor: **Dr. John P. Morrissey**
(se presente / if there is one)

Coordinatore / Coordinator: **Prof. Marco Vanoni**

ANNO ACCADEMICO / ACADEMIC YEAR 2016-2017

“WORD THIRTY-TWO

… You should love learning for its own sake and strive for it. If you value knowledge as a supreme blessing, each new truth you uncover will bring peace and satisfaction to your soul. Memorize well what is new to you, and you will feel the desire-for new quests, and a love of knowledge will be born in your heart. Then your memory will absorb whatever you have seen and heard…”

From The Book of Words (Kazakh: қара сөздері, Qara sözderi) by Abai Qunanbaiuly, a Kazakh poet, composer and philosopher

Abstract

The efficient implementation of biorefineries to produce bio-based chemicals and fuels requires sustainable source of feedstock and robust microbial factories. Among others, lignocellulose and whey, which are residual wastes deriving from wood/agriculture and dairy industries, represent cheap, sugar-enriched feedstocks. The conversion of lignocellulose and whey into the desired products using microbial cell factories is a promising option to replace the fossil based petrochemical refinery. Different bacteria, algae and yeasts are currently used as microbial hosts, and their number is predicted to increase over next years. 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 production of natural and recombinant products, including bulk chemicals such as organic acids. However, efficient and economically viable production of organic acids has to face problems related to low productivity/titer and toxicity of the final product. Therefore, the exploration of yeast biodiversity to exploit unique native features and the understanding of mechanisms to endure harsh conditions are essential to develop ultra-efficient and robust industrial yeast with novel properties.

The aim of the research thesis is to evaluate the mechanism of weak acid stress response in the non-Saccharomyces yeasts *Zygosaccharomyces parabailii* and *Kluyveromyces marxianus*. To better understand the weak acid stress response of *Z. parabailii*, we summarized recent finding on the species. Knowing the relevant scientific reports, the next study was focused on the effect of lactic acid stress on *Z. parabailii*. This organic acid can be used as monomer for the production of biodegradable bioplastic polymers, such as poly lactic acid (PLA). The study revealed that cells are able to tolerate 40g/l of lactic acid without inducing a lag phase of growth and exhibit a negligible percentage of dead cells. More importantly, during lactic acid exposure, we observed structural modifications at the level of cell wall and membrane. These findings confirmed the peculiar ability of *Z. parabailii* to adapt to weak organic acids via remodeling of cellular components.

The lack of a complete genome assembly and annotation encouraged us to perform a genome sequencing and genome study of our *Z. parabailii* strain. The results revealed that *Z. parabailii* is undergoing fertility restoration after interspecies hybridization event, which may shed a light to the process of whole genome duplication. The availability of *Z. parabailii*

complete genome information allowed us to perform the first RNA-sequencing analysis on the species exposed to lactic acid stress. The results showed upregulation of mitochondrial and oxidative stress genes, and downregulation of a subset of cell wall genes, in addition to other specific regulation related to redox balance and ion homeostasis. Remarkably, several differentially regulated genes differ significantly from the *S. cerevisiae* counterpart or, in some cases, even seem not to have a homologue.

Increased interest of *K. marxianus* application in industrial biotechnology led us to study its multidrug resistance transporters during acetic and lactic acid stress, the first being a contaminant related to the use o lignocellulose as feedstocks, while the second as final product of interest, as mentioned above. The results showed a strain-specific response to weak organic acid stress, and a possible involvement of Km*PDR12* in acetic and lactic acid stress resistance, opening potential for future discoveries and novel studies.

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.

Keywords: *Zygosaccharomyces parabailii*, *Kluyveromyces marxianus*, FTIR, RNA-seq, biorefinery, acetic acid, lactic acid, hybrid, fertility restoration, *PDR12*

Riassunto

Al fine di rendere efficiente la produzione di biomolecole e biocarburanti tramite le bioraffinerie, è necessario avere a disposizione materie prime rinnovabili e *cell factory* microbiche robuste. Fra le diverse opzioni, la lignocellulosa e il siero di latte, sottoprodotti delle industrie forestali, agricole e casearie, rappresentano materie prime economiche e ricche di zuccheri. La conversione della lignocellulosa e del siero di latte nei prodotti desiderati tramite *cell factory* microbiche è un'opzione interessante per la sostituzione delle raffinerie petrolchimiche. Diversi batteri, alghe e lieviti sono già utilizzati come *cell factory*, ma è previsto che il loro numero aumenti nei prossimi anni. Scarse richieste nutrizionali e robustezza sono caratteristiche che hanno reso i lieviti una classe di microrganismi largamente utilizzata nelle biotecnologie industriali. Ciò è stato possibile grazie allo sfruttamento delle loro qualità naturali e all'ingegnerizzazione di *pathway* metabolici per la produzione di prodotti naturali o ricombinanti, tra i quali molecole come gli acidi organici. Per essere competitiva e sostenibile la produzione di acidi organici deve affrontare problematiche connesse alla bassa produttività/produzione e alla tossicità del prodotto finale. Di conseguenza, lo studio della biodiversità dei lieviti al fine di far emergere particolari caratteristiche naturali e l'analisi dei meccanismi di resistenza a condizioni estreme sono essenziali per lo sviluppo di lieviti industriali ultra-efficienti e robusti e identificabili per proprietà innovative.

Lo scopo della tesi di ricerca è stato quello di valutare i meccanismi di risposta allo stress indotto da acidi deboli nei lieviti non-Saccaromiceti *Zygosaccharomyces parabailii* e *Kluyveromyces marxianus*. Per capire al meglio la risposta nei confronti dello stress indotto da acidi deboli di *Z. parabailii*, abbiamo innanzi tutto ricapitolato tutti le recenti scoperte riguardo questa specie. Una volta venuti a conoscenza delle scoperte scientifiche più rilevanti, ci siamo focalizzati sull'effetto indotto da acido lattico nei confronti del ceppo di *Z. parabailii* utilizzato nel nostro laboratorio. Lo studio ha rivelato che le cellule sono in grado di sopportare fino a 40g/L di acido lattico senza mostrare una fase lag nelle cinetiche di crescita, ed una percentuale irrisoria di cellule morte. Ma ancor più importante è da sottolineare il fatto che durante l'esposizione all'acido lattico abbiamo osservato modificazioni strutturali a livello della parete e della membrana cellulare. Questi risultati hanno confermato la peculiare abilità di *Z. parabailii* di adattarsi agli acidi deboli tramite il rimodellamento di alcune componenti cellulari.

La mancanza di un genoma di riferimento completo ci ha spinto a compiere il lavoro di sequenziamento, assemblaggio ed annotazione: questo lavoro, oltre a permetterci di evidenziare la natura ibrida del ceppo di *Z. parabailii* considerato, ha aperto la possibilità di ulteriori studi. I risultati hanno rivelato che *Z. parabailii* sta subendo un ripristino della fertilità, a seguito dell'evento di ibridazione interspecie, cosa che potrebbe chiarire il processo di duplicazione dell'intero genoma avvenuta in *S. cerevisiae* ed altri lieviti appartenenti al medesimo clade. Avere a disposizione le informazioni riguardo il genoma completo di *Z. parabailii* ci ha permesso di portare a termine la prima analisi di sequenziamento dell'RNA sulla specie, quando esposta allo stress da acido lattico. I risultati hanno mostrato l'up-regolazione di geni mitocondriali e connessi allo stress ossidativo, e la down-regolazione di una serie di geni codificanti per determinanti della parete cellulare, in aggiunta alle regolazioni specifiche riguardanti il bilanciamento redox e l'omeostasi di ioni, tra cui il Ferro. È degno di nota il fatto che molti geni sono regolati differentemente rispetto alla controparte di *S. cerevisiae*, o addirittura non sembrano possedere un omologo nel lievito di riferimento.

L'interesse sempre crescente nei confronti delle applicazioni di *K. marxianus* nelle biotecnologie industriali ci ha portato allo studio dei suoi trasportatori *multidrug resistance* durante lo stress indotto da acido acetico e acido lattico. I risultati mostrano una risposta specifica agli acidi organici e un putativo coinvolgimento di *KmPDR12* nella resistenza all'acido acetico e all'acido lattico, aprendo le porte a future scoperte e studi innovativi.

Parole chiave: *Zygosaccharomyces parabailii, Kluyveromyces marxianus,* FTIR, RNA-seq, bioraffineria, acido acetico, acido lattico, ibrido, ripristino della fertilità, *PDR12*

List of Publications

This thesis is based on following publications:

Paper 1: Kuanyshev, N., Adamo, G. M., Porro, D. and Branduardi, P. (2017). **The spoilage yeast *Zygosaccharomyces bailii*: foe or friend?** *Yeast, 34: 359–370.* doi: 10.1002/yea.3238

Paper 2: Kuanyshev, N., Ami, D., Signori, L., Porro, D., Morrissey, J. P. and Branduardi, P. (2016). **Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation**. *FEMS Yeast Res. 2016 Aug;16(5). pii: fow058.* doi: 10.1093/femsyr/fow058

Paper 3: Ortiz-Merino, R. A., Kuanyshev, N., Braun-Galleani, S., Byrne, K. P., Porro, D., Branduardi, P. and 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 May 16;15(5):e2002128.* doi: 10.1371/journal.pbio.2002128

Paper 4: Ortiz-Merino, R. A. *****, Kuanyshev, N. *****, Byrne, K. P., Valera J.A., Morrissey J.P., Porro, D., Wolfe, K. H. and Branduardi, P. (2017). **Transcriptional response to lactic acid stress in the hybrid yeast *Zygosaccharomyces parabailii***. *Appl Environ Microbiol (submitted)*

***** Contributed equally

Contents

INTRODUCTION

Environmental context of microbial cell factories

Environmental concerns and the economic landscape around the world are driving society and industry to renewable, ecologically friendly and economically viable sources of fuel and chemicals. Science and technology are bringing new solutions to address the demands of the modern world to develop alternative sources of energy. These solutions include renewable energy such as solar and wind, as well as the replacement of petroleum as a source of fuel. This also fits into a larger goal of developing a more sustainable economy that is not reliant on fossil resources as a source of fuel energy, or of precursors for the chemical industry. A key goal is the development of an economy based on sustainable feedstocks and zero carbon emissions.

This goal underpins the concept of the biorefinery as utilization of renewable biomass, waste products, recycling of secondary products and valorization of co-products to produce fuel, power and value-added chemicals (Cherubini, 2010). A biorefinery is similar in concept to a petroleum refinery, which converts crude oil to different value-added chemicals and fuels. The main drawbacks of petroleum refineries are non-sustainability of the resource (crude oil) and contribution to global increase of greenhouse gases emission, as a result perturbing the Earth's climate and environment (Solomon *et al.*, 2007). In this context, biorefineries are seen as a potential alternative to completely substitute petroleum-based refineries (Clark and Deswarte, 2014) (Fig. 1).

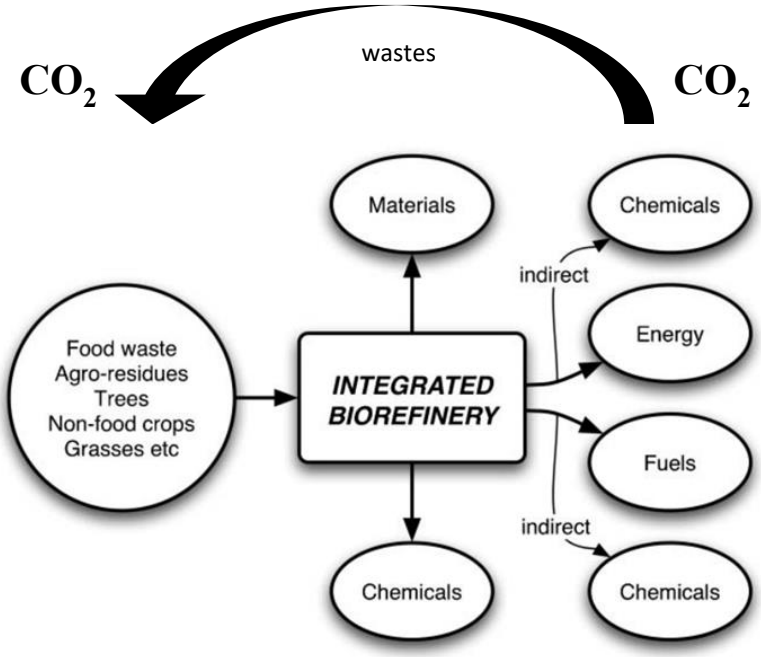


Figure 1 Biorefinery concept diagram. Adopted from Clark *et al.*, 2006

The biorefinery concept embraces a wide range of technologies ranging from biomass pretreatment to final product transportation. The technologies that are jointly applied to efficiently convert biomass into valuable products can be divided into four groups:

1. Thermochemical processes: The aim of thermochemical process is to convert biomass into energy and chemical products by gasification and pyrolysis. Gasification is biomass treatment at high temperature (above 700°C) with low oxygen levels to produce syngas. Pyrolysis is biomass treatment at intermediate temperature (300-600°C) without oxygen to produce pyrolytic oil (or bio-oil), solid charcoal and light gases like syngas. Both products can be directly used as a stationary biofuel or as precursors to produce other fuels (Spath and Dayton, 2003, Bridgwater and Peacocke, 2000).
2. Mechanical/physical processes: The aim of a mechanical process is to reduce the size and separate the components of the biomass, without changing the state and composition of it. This process is usually performed before biomass utilization to reduce the size of biomass within specific ranges for an efficient subsequent processing (Sun and Cheng, 2002).
3. Chemical processes: The aim of chemical processes is to change the chemical composition and structure of biomass by hydrolysis. Hydrolysis uses acids, bases and enzymes

to depolymerize polysaccharides into monomers (cellulose into glucose) for subsequent processing of the monomers into valuable products (Sun and Cheng, 2002).

4. Biochemical processes: Biochemical processes involve application of fermentation via microbial cell factories to convert fermentable substrate into desired product. The process occurs at low temperature and slower rate in comparison to thermochemical processing (Hamelinck *et al.*, 2005).

One of the key instruments of the biochemical processes is development and usage of efficient, robust, versatile microbial cell factories with innate or engineered metabolism, capable of converting bio-based substrates into desirable products. Thus, the production microorganism should have minimum nutritional requirements, fast growth, high yield and productivity, and be safe and amenable for genetic manipulation. In addition, the microorganism should have high tolerance to inhibitors present in raw material and to the final product (Gustavsson and Lee, 2016).

Being a model organism for fundamental research, it is not surprising that the bacterium *Escherichia coli* has been extensively used as a platform organism in various industrial productions (Kung *et al.*, 2012, Du *et al.*, 2011). The knowledge gathered in decades of research allowed the industry to engineer and exploit *E. coli* as a microbial cell factory for production of various chemicals including pharmaceuticals, biofuels and amino acids (Chen *et al.*, 2013). Nonetheless, this bacterium possesses innate disadvantages like low tolerance to organic acids, low pH, furan derivatives and phenolic compounds. In addition, low osmotic tolerance in high sugar medium may hinder overall fermentation ability of the bacteria (Liu and Khosla, 2010). Although, these limitations are not a serious problem when growing on synthetic medium to produce high value products for biopharma, they greatly restrict the applications of *E. coli* in industrial biotechnology. This creates the need to develop alternative hosts to serve as cell factories. In this regard, yeasts, especially the yeast *Saccharomyces cerevisiae*, have provided a useful option in conditions where their natural resilience to low pH and a variety of toxic compounds can be exploited.

S. cerevisiae is considered as the first domesticated microorganism, which have been used for thousands of years in wine making, bakery and brewing. The most prominent feature of *S. cerevisiae* is an ability to ferment glucose into ethanol and carbon dioxide, even in the presence

of oxygen. This phenomenon is called the Crabtree effect, which works by repressing respiration and prioritizing fermentative utilization of fermentative carbon source (glucose, sucrose and etc) (Piskur *et al.*, 2006). Since *S. cerevisiae* have been used for a long time in food applications, it has become a eukaryotic model organism for basic research and have been extensively studied and characterized. Moreover, *S. cerevisiae* is recognized as generally regarded as safe (GRAS) by the American Food and Drug Administration (FDA). *S. cerevisiae* meets all the requirements for efficient, robust and versatile microbial cell factories. Therefore, the yeast has become the main cell factory for production of chemicals from biomass. The advances in molecular biology, led to extensive investigation to expand the application of *S. cerevisiae*. In the last decades, many works were published reporting engineered yeast strains capable of producing a wide range of chemicals starting from bioethanol (Lu *et al.*, 2012, Wallace-Salinas *et al.*, 2014), organic acids (Ito *et al.*, 2014, Raab *et al.*, 2010, Porro *et al.*, 1995) to pharmaceutical compounds such as antimalarial precursors and opioids (Ro *et al.*, 2006, Galanie *et al.*, 2015). Moreover, *S. cerevisiae* was engineered to ferment pentose and arabinose, which it naturally unable to consume (Cai *et al.*, 2012, Matsushika *et al.*, 2009). The development of advanced molecular tools together with robust fermentation techniques will further expand the microbial cell factory capabilities of *S. cerevisiae*.

Indeed, *S. cerevisiae* is a popular organism used as cell factory, however yeast biodiversity can be a copious source for cell factory applications and inspirations. The use of yeast hosts, other than *S. cerevisiae,* that have natural traits such as utilization of new carbon sources and improved stress tolerance are being explored to be used as novel cell factories.

Kluyveromyces marxianus belongs to Saccharomycotina subphylum and is therefore phylogenetically related to *S. cerevisiae* (Llorente *et al.*, 2000). Its respiro-fermentative metabolism, along with some unique traits, led to many studies to exploit *K. marxianus* for bioethanol production from lignocellulosic (Goshima *et al.*, 2013) and whey biomass (Guimaraes *et al.*, 2010). The yeast has several important features that are highly desirable for industrial applications, such as an ability to utilize various substrates including C5-C6 sugars, thermotolerance (up to 52 C^{o}), fast growth in comparison to *S. cerevisiae*, and ability to grow at pH below 3 (Lane and Morrissey, 2010). Moreover, several strains of *K. marxianus* have obtained GRAS status similarly to *S. cerevisiae*, which makes it particularly interesting from an industrial point of view. The yeast already has been proposed for production of various value-added chemicals including bioethanol, flavor and fragrance molecules, solvent ethyl acetate

(Morrissey *et al.*, 2015, Radecka *et al.*, 2015, Loser *et al.*, 2015). However, the wide genetic diversity of *K. marxianus* species, together with limited research on its physiology, biochemistry and genetics, create difficulties to fully exploit its potential as microbial cell factory (Lane *et al.*, 2011, Rocha *et al.*, 2011, Fonseca *et al.*, 2008).

Zygosaccharomyces bailii is notoriously known as one the most aggressive food spoilage yeasts, with an ability to thrive in acidic environment (Stratford *et al.*, 2013). *Z. bailii* has a remarkable tolerance to weak organic acids (WOA), in particular to acetic, sorbic, benzoic and propionic acids, in addition to high osmotic and ethanol tolerance (Martorell *et al.*, 2007). These traits, including fermentative behavior in aerobic and anaerobic conditions (with specific nutritional requirements), make *Z. bailii* an attractive candidate for microbial cell factories. Several studies in the last decades were dedicated to engineer the yeast for industrially relevant chemical production. *Z. bailii* has been studied for production of L-ascorbic acid (vitamin C) (Sauer *et al.*, 2004, Branduardi *et al.*, 2004). Recently, the potential of *Z. bailii* to produce bioethanol has been reported (Paixao *et al.*, 2013). Despite the importance of the yeast for industrial and fundamental microbiology, accurate identification of *Z. bailii* and related strains is problematic. The phylogenetic relationships of many industrial isolates formerly known as *Z. bailii* have been re-evaluated, and significant differences in rRNA gene sequences were found. These led to the proposal that there are two novel species closely related to *Z. bailii*, namely *Zygosaccharomyces parabailii* and *Zygosaccharomyces pseudobailii* (Suh *et al.*, 2013). Indeed, further genome sequencing projects of commonly-used strains confirmed that in fact ISA1307 and ATCC60483 are interspecies hybrids of *Z. bailii* and closely related species rather than pure strains (Mira *et al.*, 2014, Ortiz-Merino *et al.*, 2017). Regardless of molecular differences between *Z. bailii sensu lato* species, physiological traits are undistinguishable.

Yeast cell factories for production of organic acids

Global climate change and non-sustainability of oil reserves are the major motivations to develop non-petrochemical alternative for bulk chemicals or bio-based products. These products are wholly or partially derived from renewable materials of biological origin. Most of the bio-based products are derived from plant material, which helps to reduce CO_2 emission and offers other advantages such as lower toxicity and recyclability, thus contributing to the bioeconomy. Among them the organic acids represent a class of interesting chemicals that can be produced

via microbial fermentation to fulfill the role as a platform chemical for production of more complex products (Sauer *et al.*, 2008, Becker *et al.*, 2015, Becker and Wittmann, 2015, Chen and Nielsen, 2016). The US department of Energy has identified top 10 organic acids with multiple functional groups that can be produced from plant biomass (Werpy and Petersen, 2004). Currently, the organic acids market is small and associated with food applications. However, global endorsement of green technologies, including biorefineries, is growing, therefore one can envision the future market potential of microbial produced organic acids, with transition to it from food applications to building blocks for wide range of chemical commodities (i.e. plastics, polymers, resins) (Chen and Nielsen, 2016). For example, organic acids with hydroxyl or carboxyl groups can be used as building blocks for polyesters, while dicarboxylic acids can be used for polyamide production (Becker *et al.*, 2015). As the field that investigates microbial acid production is moving fast, the industrial scale production of succinic and lactic acid has been reached by different companies (Table 1).

Product	Annual production (t/a)	Main microbial host	Main feedstock	Main established business
Lactic acid	~400 000	LAB Yeast	Beet sugar, corn sugar, wheat, cane, carbohydrates	NatureWorks (USA), Galactic (Belgium), Purac (The Netherlands)
3-hydroxy propionic acid	~3 700	*E. coli*	NA	OPXbio (USA), Dow chemicals (USA), Perstorp (France)
Succinic acid	~37 000	*E. coli* Yeast (*S. cerevisiae C. krusei*) *Basfia succiniciprodcens*	Glycerol, sugar, sorghum, corn, starch	Succinity (Germany), Myriant (USA), BioAmber (Canada), Reverdia (The Netherlands)
Glycolic acid	Pilot scale	*E. coli*	Glycerol, carbohydrates	Metabolic Explore/ Roquette (France)
Itaconic acid	~40 000	*Aspergillus terreus*	Molasses, starch, glucose, glycerol, xylose	~30 suppliers

Table 1. Industrial production of organic acids from biomass feedstocks (Becker *et al.*, 2015, Wittmann and Becker 2015)

The increased interest for succinic acid as a building block for various products such as butanediol, maleic anhydrate and nylon type polymers, has driven the development of succinic fermentation process for large scale production. Since succinic acid is an intermediate of tricarboxylic acid cycle, the rational strategy to develop a producing strain relies on elimination of by-product formation and amplification of anaplerotic flux toward reductive TCA cycle (Cheng *et al.*, 2013, Raab *et al.*, 2010). Several bacterial hosts have been developed for succinic acid production including *Basfia succiniciproducens* and *E. coli*, which are currently in the process of commercialization (Table 1). Having advantages of growth in low pH and stress tolerance, yeasts production hosts have received considerable interest as they can be more cost effective when dealing with downstream processing and purity of the product. DSM/Roquette constructed an efficient *S. cerevisiae* strain that capable to produce 100 gL^{-1} of succinic acid at pH 3. Moreover, Bioamber has changed from the *E. coli* platform to the yeast *Candida krusei* for low pH fermentation (Jansen and van Gulik, 2014).

Lactic acid has been traditionally used in the food, pharmaceutical and cosmetic industries. Lactic acid is also a desirable microbial cell factory product of high importance for sustainable production of bioplastics. In addition lactic acid can be an inhibitory compound present in whey biomass (Christensen *et al.*, 2011). The main source of lactic acid remains microbial fermentation, and the global annual production is estimated at 400 000 tons (Choi *et al.*, 2015). Lactic acid is typically produced by lactic acid bacterial (LAB) fermentation at neutral pH buffered with chemicals, usually $CaCO_3$ (Datta and Henry, 2006). The main desired final product of the fermentation is optically pure protonated lactic acid, whereas LAB fermentation generates considerable amounts of lactate salts and gypsum, as well as other impurities derived from complex medium required for the bacteria. This increases a purification cost of the final product, in addition to raw substrate costs (Fitzpatrick *et al.*, 2003).

S. cerevisiae is the most attractive candidate for production of lactic acid in its free form, resulting in less operational costs related to fermentation and downstream purification. Although *S. cerevisiae* is not a natural producer of lactic acid, it can be engineered to this aim by the heterologous expression of lactate dehydrogenase (*LDH*), which can efficiently convert pyruvate to lactic acid using NADH as a cofactor. In *S. cerevisiae*, pyruvate is generally channeled into ethanol production via pyruvate decarboxylation (*PDC* gene), creating a competing reaction for *LDH* (Fig. 2). Dequin and Barre pioneered in mixed ethanol/lactic acid fermentation in *S. cerevisiae* expressing *L-LDH* gene from *Lactobacillus casei* in 1994 (Dequin

and Barre, 1994). Yet, the final titer and yield of the product was low due to ethanol fermentation and not optimized production process. Therefore, the next intuitive strategy to increase the final yield and purity of the product was to rewire the metabolism towards lactic acid production by reducing the competing reactions like ethanol and later glycerol formation and by optimizing the production process. The strategy was implemented by many groups throughout the world with good rate of success (Sauer *et al.*, 2010). Further improvement via manipulation of intracellular redox enabled *S. cerevisiae* to produce 117 gL^{-1} of L-lactic acid in fed batch mode with pH controlled at 3.5 (Lee *et al.*, 2015). Several studies showed that *K. marxianus* expressing *LDH* either via genome integration or plasmid could produce lactic acid up to 25gL^{-1} (Lee *et al.*, 2017, Pecota *et al.*, 2007). Although, the final titer is still low in comparison to *S. cerevisiae* and conditions are not optimized for industrial purposes, these examples clearly demonstrate a potential of the yeast. The first proof on concept study on *Z. bailii* lactic acid production was implemented by heterologous expression of bovine lactate dehydrogenase (Branduardi *et al.*, 2004). Although the final titer was very low, it shows that in combination with exceptional low pH and weak organic acid (WOA) tolerance, *Z. bailii* should not be overlooked as a potential WOA production host. The production of lactic acid has been already commercialized by different companies using different strategies and fermentation processes, however there is always a room for further production improvement. Therefore, it is important to explore novel microbial hosts with innate advantageous traits, which can be further studied to either improve the trait or transfer it to existing hosts.

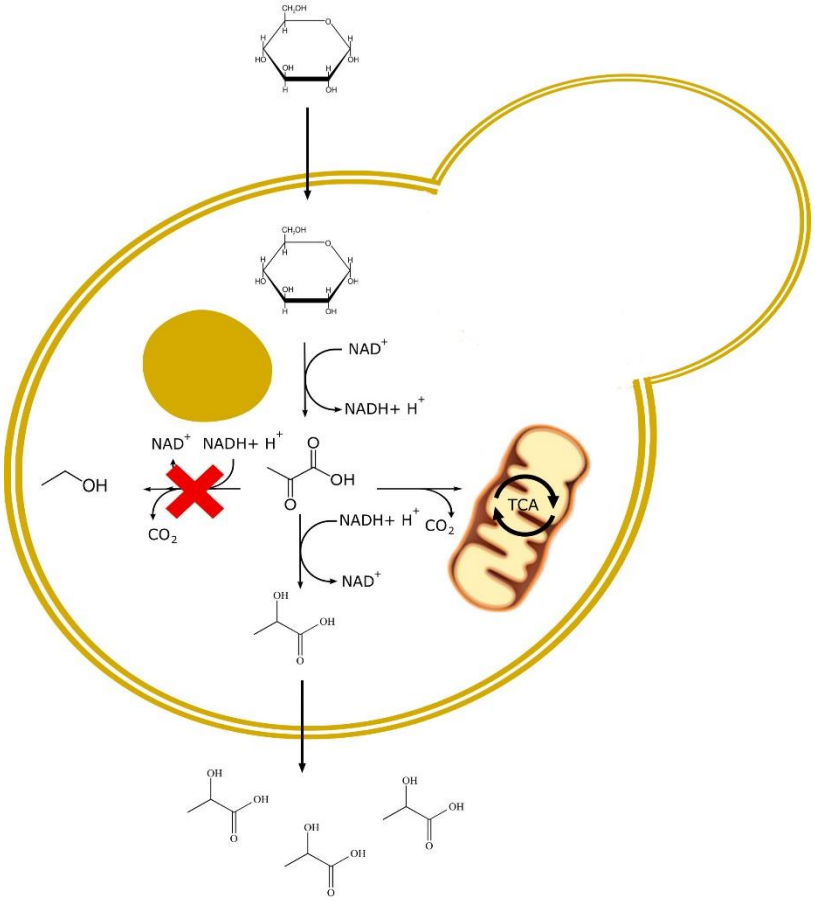


Figure 2 Schematic overview of lactic acid production from glucose in yeast. The glucose is metabolized via glycolysis into pyruvate. The pyruvate is channeled to *PDC* reaction leading to ethanol or to *LDH* leading to lactic acid. Only small portion of pyruvate goes to TCA cycle due to the Crabtree effect. In order to increase production of lactic acid, competitive ethanol production should be abolished.

Weak acid stress response in yeast

Although some organic acids are successfully produced in yeast, the intrinsic toxicity of weak organic acids (WOA) is a factor limiting development of ultra-efficient, high-yielding strains. Indeed, the toxicity of WOA is exploited as a preservation strategy, with the addition of WOA such as benzoate, citrate and sorbate to certain foods. Initial work on how yeasts responded with WOA was centered on the food industry, but latterly there has been a move towards industrial biotechnology because of cell factory applications. This has also led to a shift in balance from food-preservation acids such as citrate, towards more industrially-relevant ones like lactate and acetate. Whereas, lactic acid tolerance in *S. cerevisiae* is mainly considered in the context of

developing high yield production strains, acetic acid has wider importance because it is present at inhibitory concentrations in lignocellulosic hydrosylates.

Lignocellulosic biomass for biochemical production

Complete replacement of petroleum-based chemical production requires usage of considerable portions of renewable biomass. Currently, most of the biobased chemical production relies on sources like starch and sugar, arising ethical concerns about the food used as chemical raw material and the cost effectiveness of such substrates. Therefore, the key aspect of biorefineries is to explore the potential of cheap non-edible feedstocks such as lignocellulosic biomass. Lignocellulosic material is cheap, ubiquitous and rich in sugars. Moreover, as it does not compete with food production, its use contributes to environmental sustainability (Sun and Cheng, 2002). However, lignocellulosic biomass is a sturdy material, consisting of cellulose, hemicellulose and lignin and their relative percentage depends on the plant material used. To release the main components, lignocellulose needs to be pretreated and hydrolyzed. The pretreatment procedure decreases cellulose crystallinity and increase porosity, which enables easier hydrolysis of the sugar polymers. Steam explosion is the most commonly used way of lignocellulosic biomass pretreatment. This method relies on high pressure steam treatment with subsequent sudden reduction of the pressure, which makes the material undergo explosive decomposition (Agbor *et al.*, 2011). After pretreatment, the released polymers are hydrolyzed into free monomer molecules via enzymatic hydrolysis or acid hydrolysis. Depending on the pretreatment and hydrolysis methods used, sugars can be further degraded into inhibitory furans, hydroxymethylfurfural, furfural and weak organic acids (Limayem and Ricke, 2012). Acetic acid is the most abundant inhibitory weak acid, which mainly released from acetyl groups on the hemicellulose polymer (Almeida *et al.*, 2007, Jonsson and Martin, 2016). The elevated concentration of acetic acid can be detrimental to overall fermentation, in addition to the remaining inhibitory compounds. Therefore, understanding the weak acid resistance mechanism is important to develop a robust microbial host, which can use lignocellulosic biomass to produce bulk chemicals like WOAs.

Effects of weak organic acids on yeast

At pH below pKa, WOA will predominantly exist in the protonated form (RCOOH) (Fig. 3). The protonated form of the acid readily diffuses across the plasma membrane. At near neutral cytosolic pH, WOA dissociates to a proton and respective anion (H^+ and $RCOO^-$) (Fig. 3). Being charged molecules, the proton and WOA anion cannot leave the cell interior, causing cytosol acidification and anion dependent toxicity (Piper *et al.*, 2001). The acidification results in general decline of metabolism and enzyme activity, while anion accumulation may exert toxicity depending on the structure (Piper, 1999, Holyoak *et al.*, 1999). For example, acetic acid stress generates reactive oxygen species (ROS), damaging cellular functions and resulting in ROS-mediated apoptosis (Giannattasio *et al.*, 2005). Lactic acid affects iron availability in the medium and cell membrane composition (Narendranath *et al.*, 2001, Stratford and Eklund, 2003).

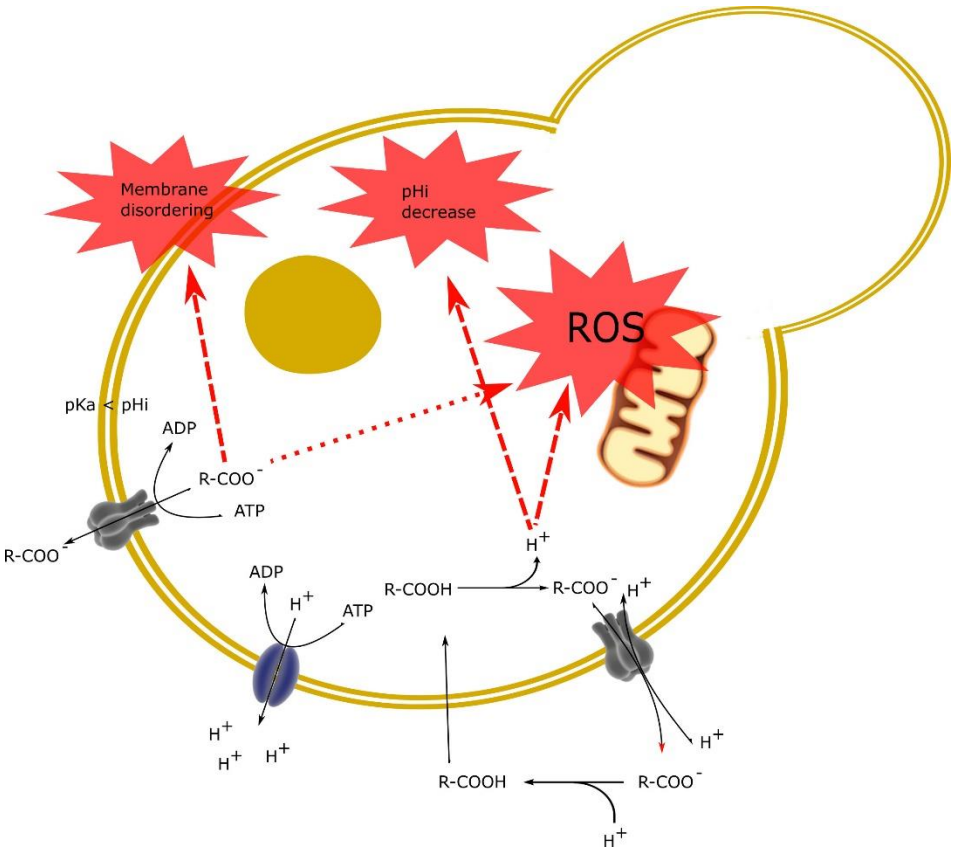


Figure 3 Simplified overview of weak organic acid toxicity mechanism. The illustration shows the major toxic effect exerted by the weak organic acids. The loop "in and out" describes the ability of weak acid to re-diffuse into the cell after it was extruded via proton and counteranion pumps.

Yeast response to weak organic acids

Most of the knowledge on weak organic acid response comes from *S. cerevisiae* studies. The yeast implements a multi-faceted response to WOAs that includes expression of transporters to extrude protons and the toxic counterions, remodeling of cell membrane and wall to restrict WOA diffusion, and activation of anti-oxidant systems to counter intracellular damage.

The immediate response of the yeast cell to the WOA stress is a maintenance of intracellular pH homeostasis by proton extrusion out of cytosol (Piper *et al.*, 2001). The H^+-ATPase and the V-ATPase are the main ATP dependent proton pumps of the cell which are activated during intracellular acidification (Martinez-Munoz and Kane, 2008). The proton extrusion is energetically expensive and consumes most of the ATP generated in the cell (Serrano, 1983, Serrano, 1991), resulting in growth rate reduction (Holyoak *et al.*, 1996).

To avoid toxic counterion accumulation, yeast activates multidrug resistance transporters. The expression of the transporters is WOA-specific and regulated by the War1p and Haa1p transcription factors. The War1p regulon is essential for induction of *PDR12*, a member of the ATP binding cassette transporter family (Kren *et al.*, 2003). War1p acts through a cis-acting weak acid responsive element (WARE) present in *PDR12* promoter region. The induction of *PDR12* following sorbic and benzoic acids stress is rapid, resulting in high levels of Pdr12p in the cell membrane comparable to Pma1p, the most abundant plasma membrane protein (Piper *et al.*, 2001, Kren *et al.*, 2003). However, *PDR12* is not responsive to acetic and formic acids (Hatzixanthis *et al.*, 2003), but in this case its deletion increase strain robustness (Nygard *et al.*, 2014). In line with these results, other studies demonstrated that Pdr12p is responsible for active extrusion of longer chain organic acids such as propionate, sorbate and benzoate (Holyoak *et al.*, 1999, Piper *et al.*, 1998), and not shorter organic acids like acetate and formate (Nygard *et al.*, 2014).

Haa1p, known to be a major gene responsible for regulation of acetic acid responsive genes, is implicated also in lactic acid response (Mira *et al.*, 2010a, Abbott *et al.*, 2008). Among the regulated genes, Haa1p is directly responsible for the induction of *TPO2* and *TPO3* genes, both

encoding for H^+ antiporters belonging to the major facilitator superfamily (Fernandes *et al.*, 2005). Previously thought to be involved in polyamine transport, Tpo2p and Tpo3p are responsible for weak acid anion extrusion, mainly acetic, propionic and benzoic acids (Mira *et al.*, 2010b). Haa1p has been considered as an interesting target for the rational engineering of *S. cerevisiae* to improve certain stress responses. For example, Tanake *et al*., showed that constitutive expression of *HAA1* in *S. cerevisiae* resulted in enhanced acetic acid tolerance and decrease of intracellular acetic acid concentration (Tanaka *et al.*, 2012).

Although the efficiency of weak acid efflux system in yeast is adequate for transient stress, long term weak acid stress may lead to futile "in and out loop" scenarios as shown in figure 3. Therefore, yeast modifies the cell membrane and/or cell wall and morphology to reduce overall diffusion of WOAs. The studies conducted on *S. cerevisiae* demonstrated that modulation of a membrane fluidity by changing phospholipids and ergosterol content to a certain extent may lead to a decrease of cell membrane permeability (Lindberg *et al.*, 2013, Berterame *et al.*, 2016, Xia and Yuan, 2009). Along with cell membrane, *S. cerevisiae* also remodels its cell wall; Simoes *et al.*, showed an important role of the Spi1p GPI-anchored cell wall protein in weak acid resistance. In particular, the authors suggest that expression of *SPI1* decreases the cell wall porosity, thus preventing weak acid access to the cell membrane (Simoes *et al.*, 2006). Yet, another way for long term adaptation to weak organic stress can be related to a morphology change. For example, formation of multicellular clusters can decrease weak acid diffusion rate into individual cells (Fletcher *et al.*, 2017, Oud *et al.*, 2013, Lei *et al.*, 2007). In general, the modification of cell wall and/or membrane and morphology provides stress adaptation in a weak acid specific way.

While WOA efflux and diffusion reduction are essential part of the defense, yeast has to deal with damage done by permeated WOAs. The yeast exposed to the acids, has to increase the pool of ATP via mitochondrial respiration (Mira *et al.*, 2010c) to empower the efflux system, which increases oxidative stress caused by ROS accumulation. In addition, certain counteranions can contribute to ROS accumulation. For example, hydroxyl radicals are formed via the Fenton reaction when cells exposed to lactic acid (Ali *et al.*, 2000). In order to decrease ROS levels, *S. cerevisiae* employs several antioxidant mechanisms starting from catalase and superoxide dismutase activation as a natural vanguard defense, ending with increasing in ROS scavenger

compounds and precursors like glutathione (an abundant non-enzymatic antioxidant in yeast) as reviewed in (Farrugia and Balzan, 2012). Knowing the importance of ROS detoxification during WOA stress, Branduardi *et al.*, demonstrated that introduction of the L-ascorbic acid (a natural antioxidant compound) biosynthesis pathway into *S. cerevisiae* improved resistance of cells to oxidative and WOA stress (Branduardi *et al.*, 2007).

Yeast biodiversity for lactic acid production

Giving the wide diversity of yeasts, traditional baker's yeast's position as the main work-horse for biotechnology can be challenged with emerging non-Saccharomyces yeasts. Non-saccharomyces yeasts are evolved to flourish in specific natural habitats, possessing beneficial native traits which baker's yeast lack. In addition, understanding the unique feature of non-saccharomyces yeasts can be used as a source of inspiration for rational engineering of *S. cerevisiae*.

Summary of research aims and outcomes

1. Assess the relevance and potential of *Z. bailii* for biotechnology.

Z. bailii has been studied for decades by food scientists and microbiologists since the yeast is involved in food spoilage problems. Only in last decades it started to get attention from biotechnologists as an interesting species for industrial applications. Therefore, **Chapter 1** of my thesis provides a summary of *Z. bailii* research from both food science and biotechnology perspective. It aims to give a reader clear overview of recent findings related to the *Z. bailii*.

2. Determine the physiological effects of lactic acid on the yeast *Z. bailii*

Since our long-term plan is to exploit *Z. bailii* for lactic acid production, we first need to understand how the acid can influence fermentation performance, and more importantly stress response. Hence, **Chapter 2** examines *Z. bailii* performance and viability in the presence of sub-inhibitory levels of lactic acid. Moreover, using FTIR (Fourier transform infrared spectroscopy), we studied the impact of lactic acid on cells at a macromolecular level. The information gathered is important to understand lactic acid-induced changes in *Z. bailii*.

3. Analysis of the genome sequence and structure of *Z. bailii*

Despite the relevance of the species, the science community still lack a complete genome assembly of *Z. bailii.* Currently, the public database only has little genome sequence data of *Z.*

bailii (Galeote *et al.*, 2013, Mira *et al.*, 2014, Palma *et al.*, 2017), and none of them can provide full genomic information about the genome organization and possible evolution. Thus, a complete genome sequencing of our strain was imperative to get further insight into the species. **Chapter 3** is dedicated to the genome and evolution of the strain. The study revealed that our strain is in fact a hybrid, thus to be named *Z. parabailii* as indicated for a different strain in (Suh *et al.*, 2013). Moreover, the study demonstrated how hybridization affected the mating switch, possibly explaining the full genome duplication event occurred in the lineage leading to *S. cerevisiae* (Wolfe, 2015).

4. Assess the genome-wide response to lactic acid stress in *Z. bailii*

Knowing both physio-morphological response and complete genome assembly, we proceeded with a transcriptomics study of lactic acid-treated *Z. parabailii* cells isolated from the previous study. We performed RNA sequencing analysis to get an insight to the molecular responses of lactic acid stress, and to possibly correlate the observation to the physio-morphological data. The study discussed in **Chapter 4** represents the first RNA sequencing of *Z. parabailii*. We combined different RNA-seq analysis methods to optimize an entire analysis and got interesting results regarding the transcriptome. The RNA-seq study can provide a basis for future research on *Z. bailii sensu lato* as a microbial cell factory.

5. Explore the potential synergies in organic acid response pathways in the alternative cell factory yeast *K. marxianus*

Commercially competitive properties of *K. marxianus* have resulted in a wide range of industrial applications. Whether being utilized by the food or the pharmaceutical industry, *K. marxianus* performance, similar to other yeasts, is often dependent on process environment conditions. Little information is available on the environmental stress resistance pathways in *K. marxianus* and what mechanism changes account for strain-to-strain variations in response to stress. Therefore, **Chapter 5** is dedicated to study the *K. marxianus* response to weak organic acid stress (acetic acid and lactic acid) which is commonly encountered during industrial fermentations for diverse applications. We focused our attention on the response mediated by multidrug resistance transporters during the weak organic acid stress and its relatedness to the model yeast *S. cerevisiae*.

References

ABBOTT, DA, SUIR, E, VAN MARIS, AJ & PRONK, JT 2008. Physiological and transcriptional responses to high concentrations of lactic acid in anaerobic chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol,* 74**,** 5759-5768.

AGBOR, VB, CICEK, N, SPARLING, R*, et al.* 2011. Biomass pretreatment: fundamentals toward application. *Biotechnol Adv,* 29**,** 675-685.

ALI, MA, YASUI, F, MATSUGO, S & KONISHI, T 2000. The lactate-dependent enhancement of hydroxyl radical generation by the Fenton reaction. *Free Radic Res,* 32**,** 429-438.

ALMEIDA, JRM, MODIG, T, PETERSSON, A*, et al.* 2007. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *Journal of Chemical Technology & Biotechnology,* 82**,** 340-349.

BECKER, J, LANGE, A, FABARIUS, J & WITTMANN, C 2015. Top value platform chemicals: bio-based production of organic acids. *Curr Opin Biotechnol,* 36**,** 168-175.

BECKER, J & WITTMANN, C 2015. Advanced biotechnology: metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products. *Angew Chem Int Ed Engl,* 54**,** 3328-3350.

BERTERAME, NM, 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,* 15**,** 39.

BRANDUARDI, P, FOSSATI, T, SAUER, M*, et al.* 2007. Biosynthesis of vitamin C by yeast leads to increased stress resistance. *PLoS One,* 2**,** e1092.

BRANDUARDI, P, VALLI, M, BRAMBILLA, L*, et al.* 2004. The yeast *Zygosaccharomyces bailii*: a new host for heterologous protein production, secretion and for metabolic engineering applications. *FEMS Yeast Res,* 4**,** 493-504.

BRIDGWATER, AV & PEACOCKE, GVC 2000. Fast pyrolysis processes for biomass. *Renewable and Sustainable Energy Reviews,* 4**,** 1-73.

CAI, Z, ZHANG, B & LI, Y 2012. Engineering *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: reflections and perspectives. *Biotechnol J,* 7**,** 34-46.

CHEN, X, ZHOU, L, TIAN, K*, et al.* 2013. Metabolic engineering of *Escherichia coli*: a sustainable industrial platform for bio-based chemical production. *Biotechnol Adv,* 31**,** 1200-1223.

CHEN, Y & NIELSEN, J 2016. Biobased organic acids production by metabolically engineered microorganisms. *Curr Opin Biotechnol,* 37**,** 165-172.

CHENG, KK, WANG, GY, ZENG, J & ZHANG, JA 2013. Improved succinate production by metabolic engineering. *Biomed Res Int,* 2013**,** 538790.

CHERUBINI, F 2010. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Conversion and Management,* 51**,** 1412-1421.

CHOI, S, SONG, CW, SHIN, JH & LEE, SY 2015. Biorefineries for the production of top building block chemicals and their derivatives. *Metab Eng,* 28**,** 223-239.

CHRISTENSEN, AD, KADAR, Z, OLESKOWICZ-POPIEL, P & THOMSEN, MH 2011. Production of bioethanol from organic whey using *Kluyveromyces marxianus*. *J Ind Microbiol Biotechnol,* 38**,** 283-289.

CLARK, J & DESWARTE, F 2014. *Introduction to Chemicals from Biomass,* Published Online, John Wiley & Sons, Ltd.

DATTA, R & HENRY, M 2006. Lactic acid: recent advances in products, processes and technologies — a review. *Journal of Chemical Technology & Biotechnology,* 81**,** 1119-1129.

DEQUIN, S & BARRE, P 1994. Mixed lactic acid-alcoholic fermentation by *Saccharomyces cerevisiae* expressing the *Lactobacillus casei* L(+)-LDH. *Biotechnology (N Y),* 12**,** 173-177.

DU, J, SHAO, Z & ZHAO, H 2011. Engineering microbial factories for synthesis of value-added products. *J Ind Microbiol Biotechnol,* 38**,** 873-890.

FARRUGIA, G & BALZAN, R 2012. Oxidative stress and programmed cell death in yeast. *Front Oncol,* 2**,** 64.

FERNANDES, AR, MIRA, NP, VARGAS, RC*, et al.* 2005. *Saccharomyces cerevisiae* adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. *Biochem Biophys Res Commun,* 337**,** 95-103.

FITZPATRICK, JJ, MURPHY, C, MOTA, FM & PAULI, T 2003. Impurity and cost considerations for nutrient supplementation of whey permeate fermentations to produce lactic acid for biodegradable plastics. *International Dairy Journal,* 13**,** 575-580.

FLETCHER, E, FEIZI, A, BISSCHOPS, MM*, et al.* 2017. Evolutionary engineering reveals divergent paths when yeast is adapted to different acidic environments. *Metab Eng,* 39**,** 19-28.

FONSECA, GG, HEINZLE, E, WITTMANN, C & GOMBERT, AK 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl Microbiol Biotechnol,* 79**,** 339-354.

GALANIE, S, THODEY, K, TRENCHARD, IJ*, et al.* 2015. Complete biosynthesis of opioids in yeast. *Science,* 349**,** 1095-1100.

GALEOTE, V, BIGEY, F, DEVILLERS, H*, et al.* 2013. Genome Sequence of the Food Spoilage Yeast *Zygosaccharomyces bailii* CLIB 213T. *Genome Announc,* 1.

GIANNATTASIO, S, GUARAGNELLA, N, CORTE-REAL, M*, et al.* 2005. Acid stress adaptation protects *Saccharomyces cerevisiae* from acetic acid-induced programmed cell death. *Gene,* 354**,** 93-98.

GOSHIMA, T, TSUJI, M, INOUE, H*, et al.* 2013. Bioethanol production from Lignocellulosic biomass by a novel *Kluyveromyces marxianus* strain. *Biosci Biotechnol Biochem,* 77**,** 1505-1510.

GUIMARAES, PM, TEIXEIRA, JA & DOMINGUES, L 2010. Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. *Biotechnol Adv,* 28**,** 375-384.

GUSTAVSSON, M & LEE, SY 2016. Prospects of microbial cell factories developed through systems metabolic engineering. *Microb Biotechnol,* 9**,** 610-617.

HAMELINCK, CN, HOOIJDONK, GV & FAAIJ, APC 2005. Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. *Biomass and Bioenergy,* 28**,** 384-410.

HATZIXANTHIS, K, MOLLAPOUR, M, SEYMOUR, I*, et al.* 2003. Moderately lipophilic carboxylate compounds are the selective inducers of the *Saccharomyces cerevisiae* Pdr12p ATP-binding cassette transporter. *Yeast,* 20**,** 575-585.

HOLYOAK, CD, BRACEY, D, PIPER, PW*, et al.* 1999. The *Saccharomyces cerevisiae* weak-acid-inducible ABC transporter Pdr12 transports fluorescein and preservative anions from the cytosol by an energy-dependent mechanism. *J Bacteriol,* 181**,** 4644-4652.

HOLYOAK, CD, STRATFORD, M, MCMULLIN, Z*, et al.* 1996. Activity of the plasma membrane H(+)-ATPase and optimal glycolytic flux are required for rapid adaptation and growth of *Saccharomyces cerevisiae* in the presence of the weak-acid preservative sorbic acid. *Appl Environ Microbiol,* 62**,** 3158-3164.

ITO, Y, HIRASAWA, T & SHIMIZU, H 2014. Metabolic engineering of *Saccharomyces cerevisiae* to improve succinic acid production based on metabolic profiling. *Biosci Biotechnol Biochem,* 78**,** 151-159.

JANSEN, ML & VAN GULIK, WM 2014. Towards large scale fermentative production of succinic acid. *Curr Opin Biotechnol,* 30**,** 190-197.

JONSSON, LJ & MARTIN, C 2016. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour Technol,* 199**,** 103-112.

KREN, A, MAMNUN, YM, BAUER, BE*, et al.* 2003. War1p, a novel transcription factor controlling weak acid stress response in yeast. *Mol Cell Biol,* 23**,** 1775-1785.

KUNG, Y, RUNGUPHAN, W & KEASLING, JD 2012. From fields to fuels: recent advances in the microbial production of biofuels. *ACS Synth Biol,* 1**,** 498-513.

LANE, MM, BURKE, N, KARREMAN, R*, et al.* 2011. Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. *Antonie Van Leeuwenhoek,* 100**,** 507-519.
LANE, MM & MORRISSEY, JP 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. *Fungal Biology Reviews,* 24**,** 17-26.
LEE, JW, IN, JH, PARK, JB*, et al.* 2017. Co-expression of two heterologous lactate dehydrogenases genes in *Kluyveromyces marxianus* for l-lactic acid production. *J Biotechnol,* 241**,** 81-86.
LEE, JY, KANG, CD, LEE, SH*, et al.* 2015. Engineering cellular redox balance in *Saccharomyces cerevisiae* for improved production of L-lactic acid. *Biotechnol Bioeng,* 112**,** 751-758.
LEI, J, ZHAO, X, GE, X & BAI, F 2007. Ethanol tolerance and the variation of plasma membrane composition of yeast floc populations with different size distribution. *J Biotechnol,* 131**,** 270-275.
LIMAYEM, A & RICKE, SC 2012. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Progress in Energy and Combustion Science,* 38**,** 449-467.
LINDBERG, L, SANTOS, AX, RIEZMAN, H*, et al.* 2013. Lipidomic profiling of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* reveals critical changes in lipid composition in response to acetic acid stress. *PLoS One,* 8**,** e73936.
LIU, T & KHOSLA, C 2010. Genetic engineering of *Escherichia coli* for biofuel production. *Annu Rev Genet,* 44**,** 53-69.
LLORENTE, B, MALPERTUY, A, BLANDIN, G*, et al.* 2000. Genomic exploration of the hemiascomycetous yeasts: 12. *Kluyveromyces marxianus* var. marxianus. *FEBS Lett,* 487**,** 71-75.
LOSER, C, URIT, T, KEIL, P & BLEY, T 2015. Studies on the mechanism of synthesis of ethyl acetate in *Kluyveromyces marxianus* DSM 5422. *Appl Microbiol Biotechnol,* 99**,** 1131-1144.
LU, Y, CHENG, YF, HE, XP*, et al.* 2012. Improvement of robustness and ethanol production of ethanologenic *Saccharomyces cerevisiae* under co-stress of heat and inhibitors. *J Ind Microbiol Biotechnol,* 39**,** 73-80.
MARTINEZ-MUNOZ, GA & KANE, P 2008. Vacuolar and plasma membrane proton pumps collaborate to achieve cytosolic pH homeostasis in yeast. *J Biol Chem,* 283**,** 20309-20319.
MARTORELL, P, STRATFORD, M, STEELS, H*, et al.* 2007. Physiological characterization of spoilage strains of *Zygosaccharomyces bailii* and Zygosaccharomyces rouxii isolated from high sugar environments. *Int J Food Microbiol,* 114**,** 234-242.
MATSUSHIKA, A, INOUE, H, KODAKI, T & SAWAYAMA, S 2009. Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Appl Microbiol Biotechnol,* 84**,** 37-53.
MIRA, NP, BECKER, JD & SA-CORREIA, I 2010a. Genomic expression program involving the Haa1p-regulon in *Saccharomyces cerevisiae* response to acetic acid. *OMICS,* 14**,** 587-601.
MIRA, NP, MUNSTERKOTTER, M, DIAS-VALADA, F*, et al.* 2014. The genome sequence of the highly acetic acid-tolerant *Zygosaccharomyces bailii*-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. *DNA Res,* 21**,** 299-313.
MIRA, NP, PALMA, M, GUERREIRO, JF & SA-CORREIA, I 2010b. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. *Microb Cell Fact,* 9**,** 79.
MIRA, NP, TEIXEIRA, MC & SA-CORREIA, I 2010c. Adaptive response and tolerance to weak acids in *Saccharomyces cerevisiae*: a genome-wide view. *OMICS,* 14**,** 525-540.
MORRISSEY, JP, ETSCHMANN, MM, SCHRADER, J & DE BILLERBECK, GM 2015. Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. *Yeast,* 32**,** 3-16.
NARENDRANATH, NV, THOMAS, KC & INGLEDEW, WM 2001. Effects of acetic acid and lactic acid on the growth of *Saccharomyces cerevisiae* in a minimal medium. *J Ind Microbiol Biotechnol,* 26**,** 171-177.

NYGARD, Y, MOJZITA, D, TOIVARI, M*, et al.* 2014. The diverse role of Pdr12 in resistance to weak organic acids. *Yeast,* 31**,** 219-232.
ORTIZ-MERINO, RA, KUANYSHEV, N, BRAUN-GALLEANI, S*, et al.* 2017. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol,* 15**,** e2002128.
OUD, B, GUADALUPE-MEDINA, V, NIJKAMP, JF*, et al.* 2013. Genome duplication and mutations in ACE2 cause multicellular, fast-sedimenting phenotypes in evolved *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A,* 110**,** E4223-4231.
PAIXAO, SM, TEIXEIRA, PD, SILVA, TP*, et al.* 2013. Screening of novel yeast inulinases and further application to bioprocesses. *N Biotechnol,* 30**,** 598-606.
PALMA, M, DIAS, PJ, ROQUE, FC*, et al.* 2017. The *Zygosaccharomyces bailii* transcription factor Haa1 is required for acetic acid and copper stress responses suggesting subfunctionalization of the ancestral bifunctional protein Haa1/Cup2. *BMC Genomics,* 18**,** 75.
PECOTA, DC, RAJGARHIA, V & DA SILVA, NA 2007. Sequential gene integration for the engineering of *Kluyveromyces marxianus*. *J Biotechnol,* 127**,** 408-416.
PIPER, P, CALDERON, CO, HATZIXANTHIS, K & MOLLAPOUR, M 2001. Weak acid adaptation: the stress response that confers yeasts with resistance to organic acid food preservatives. *Microbiology,* 147**,** 2635-2642.
PIPER, P, MAHE, Y, THOMPSON, S*, et al.* 1998. The pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast. *EMBO J,* 17**,** 4257-4265.
PIPER, PW 1999. Yeast superoxide dismutase mutants reveal a pro-oxidant action of weak organic acid food preservatives. *Free Radic Biol Med,* 27**,** 1219-1227.
PISKUR, J, ROZPEDOWSKA, E, POLAKOVA, S*, et al.* 2006. How did Saccharomyces evolve to become a good brewer? *Trends Genet,* 22**,** 183-186.
PORRO, D, BRAMBILLA, L, RANZI, BM*, et al.* 1995. Development of metabolically engineered *Saccharomyces cerevisiae* cells for the production of lactic acid. *Biotechnol Prog,* 11**,** 294-298.
RAAB, AM, GEBHARDT, G, BOLOTINA, N*, et al.* 2010. Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. *Metab Eng,* 12**,** 518-525.
RADECKA, D, MUKHERJEE, V, MATEO, RQ*, et al.* 2015. Looking beyond Saccharomyces: the potential of non-conventional yeast species for desirable traits in bioethanol fermentation. *FEMS Yeast Res,* 15.
RO, DK, PARADISE, EM, OUELLET, M*, et al.* 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature,* 440**,** 940-943.
ROCHA, SN, ABRAHAO-NETO, J & GOMBERT, AK 2011. Physiological diversity within the *Kluyveromyces marxianus* species [corrected]. *Antonie Van Leeuwenhoek,* 100**,** 619-630.
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,* 70**,** 6086-6091.
SAUER, M, PORRO, D, MATTANOVICH, D & BRANDUARDI, P 2008. Microbial production of organic acids: expanding the markets. *Trends Biotechnol,* 26**,** 100-108.
SAUER, M, PORRO, D, MATTANOVICH, D & BRANDUARDI, P 2010. 16 years research on lactic acid production with yeast – ready for the market? *Biotechnology and Genetic Engineering Reviews,* 27**,** 229-256.
SERRANO, R 1983. In vivo glucose activation of the yeast plasma membrane ATPase. *FEBS Lett,* 156**,** 11-14.
SERRANO, R 1991. *9 Transport across Yeast Vacuolar and Plasma Membranes*.
SIMOES, T, MIRA, NP, FERNANDES, AR & SA-CORREIA, I 2006. The SPI1 gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. *Appl Environ Microbiol,* 72**,** 7168-7175.

SOLOMON, S, QIN, D, MANNING, M*, et al.* 2007. *Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC*, Cambridge University Press.

SPATH, PL & DAYTON, DC 2003. Preliminary Screening - Technical and Economic Assessment of Synthesis Gas to Fuels and Chemicals With Emphasis on the Potential for Biomass-Derived Syngas. NATIONAL RENEWABLE ENERGY LAB GOLDEN CO.

STRATFORD, M & EKLUND, T 2003. Organic acids and esters. *In:* NICHOLAS J. RUSSELL , GWG (ed.) *Food Preservatives.* 2 ed.: Springer US.

STRATFORD, M, STEELS, H, NEBE-VON-CARON, G*, et al.* 2013. Extreme resistance to weak-acid preservatives in the spoilage yeast *Zygosaccharomyces bailii*. *Int J Food Microbiol,* 166**,** 126-134.

SUH, SO, GUJJARI, P, BERES, C*, et al.* 2013. Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. *Int J Syst Evol Microbiol,* 63**,** 1922-1929.

SUN, Y & CHENG, J 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour Technol,* 83**,** 1-11.

TANAKA, K, ISHII, Y, OGAWA, J & SHIMA, J 2012. Enhancement of acetic acid tolerance in *Saccharomyces cerevisiae* by overexpression of the HAA1 gene, encoding a transcriptional activator. *Appl Environ Microbiol,* 78**,** 8161-8163.

WALLACE-SALINAS, V, SIGNORI, L, LI, YY*, et al.* 2014. Re-assessment of YAP1 and MCR1 contributions to inhibitor tolerance in robust engineered *Saccharomyces cerevisiae* fermenting undetoxified lignocellulosic hydrolysate. *AMB Express,* 4**,** 56.

WERPY, T & PETERSEN, G 2004. Top Value Added Chemicals from Biomass: Volume I – Results of Screening for Potential Candidates from Sugars and Synthesis Gas. Springfield: National Renewable Energy Laboratory a DOE national Laboratory.

WOLFE, KH 2015. Origin of the Yeast Whole-Genome Duplication. *PLoS Biol,* 13**,** e1002221.

XIA, JM & YUAN, YJ 2009. Comparative lipidomics of four strains of *Saccharomyces cerevisiae* reveals different responses to furfural, phenol, and acetic acid. *J Agric Food Chem,* 57**,** 99-108.

Chapter 1

The spoilage yeast *Zygosaccharomyces bailii*: foe or friend?

Kuanyshev, N., Adamo, G. M., Porro, D. and Branduardi, P. (2017). **The spoilage yeast *Zygosaccharomyces bailii*: foe or friend?** *Yeast.* DOI: 10.1002/yea.3238

The spoilage yeast *Zygosaccharomyces bailii*: foe or friend?

Nurzhan Kuanyshev, Giusy M. Adamo, Danilo Porro and Paola Branduardi*
Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, Milano 20126 (Italy).

Running title: The spoilage yeast *Zygosaccharomyces bailii*: foe or friend?

*Corresponding authors
Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy. Tel: 39 02 6448 3418;
E-mail: paola.branduardi@unimib.it

Abstract

Zygosaccharomyces bailii is a non-Saccharomyces budding yeast known as one of the most aggressive food spoilage microorganism, often isolated as contaminant during wine fermentation, as well as from many acidic, high sugar and canned foods. The spoilage ability relies on the yeast's unique feature to tolerate the most common preservatives such as sulfite, dimethyl dicarbonate, acetic acid, and sorbic acid. Therefore, many studies focused on the description of this peculiar tolerance with the aim of developing preventative measures against *Z. bailii* food spoilage. These studies demonstrated the involvement of diverse molecular and physiological mechanisms in the yeast resistance, comprising detoxification of preservatives, adaptation of the cytoplasmic pH and modulation of the cell wall/membrane composition. At the same time, the described traits unveiled *Z. bailii* as a novel potential workhorse for industrial bioprocesses. Here we present the yeast *Z. bailii* starting from important aspects of its robustness, and concluding with the exploitation of its potential in biotechnology. Overall, the article describes *Z. bailii* from different perspectives, converging in presenting it as one of the interesting species of the *Saccharomycotina* subphylum.

Keywords *Zygosaccharomyces bailii*, Weak Organic Acids, genetic engineering, food spoilage yeast, industrial bioprocesses

Introduction

Food scientists and biotechnologists have extensively studied *Z. bailii* over the last decades. This yeast has been described on the one hand as a foe to be eradicated due to its spoilage ability and on the other hand as a prospective friend boasting unique traits to be exploited. Regardless what was the final aim of the study scientists admired *Z. bailii* for its capability to thrive in a number of harsh human-made environments (*i.e.* canned food, wine, mustard sauce, to cite few). For this reason, additional studies comprising both perspectives can be effective at unraveling its secrets, which can be used not only for developing novel preventative measures against food contaminations, but also to inspire novel tailoring of cell factories engaged in microbial-based biotransformations, as industrial processes very often impose several constraints that limit cell potential. In this review, we recapitulate the major findings related to *Z. bailii* in respect to food microbiology and biotechnology.

The spoilage yeast *Z. bailii*: a foe for food industry

The process of food quality deterioration, leading to changes in nutritional and organoleptic properties, is called food spoilage (Doyle, 2007). Growth and metabolic activities of microorganisms, including bacteria, molds and yeasts, are among the main factors responsible for food spoilage. The yeast group named *Zygosaccharomyces*, introduced as a genus at the beginning of the 20th century and belonging to the phylum Hemiascomycetes, comprises several of the most challenging and notorious spoilage yeasts. The name of the genus derives from the dumbbell-shaped ascus that contains smooth ascospores (Van der Walt and Johanssen, 1975). The genus comprises twelve yeast species: *Z. bisporus*, *Z. kombuchaensis*, *Z. lentus*, *Z. gambellarensis*, *Z. machadoi*, *Z. rouxii*, *Z. sapae*, *Z. siamensis*, *Z. mellis*, *Z. parabailii*, *Z. pseudobailii*, and *Z. bailii* (Kurtzman and Robnett, 2003, James and Stratford, 2011); (Rosa and Lachance, 2005, Saksinchai *et al.*, 2012, Torriani *et al.*, 2011, Solieri *et al.*, 2008, Suh *et al.*, 2013).

Z. bailii can colonize acidic foods such as soft drinks, fruit juices, dairy products and dressing, causing significant economic losses in food industry (Martorell *et al.*, 2007, Stratford, 2006, Fleet, 2007). In addition to acidic food, *Z. bailii* can contaminate wine must during the fermentation, although its activity on grape-juice has been also reported to be beneficial for vinification (Domizio *et al.*, 2011, Romano *et al.*, 2003, Romano and Suzzi, 1993). Indeed, the yeast is extremely tolerant to many conditions that are usually detrimental for cell growth such

as high osmotic pressure, high ethanol concentration, low pH values, and the presence of weak organic acids and/or various food preservatives (sulfite - SO_2, dimethyl dicarbonate - DMDC) (Martorell *et al.*, 2007, Stratford *et al.*, 2013) (Table 1). For example, *Z. bailii*'s weak organic acid tolerance varies between 375 to 550 mM for acetic acid and between 4.55 to 9.45 mM for sorbic acid, depending on the strain (Stratford *et al.*, 2013): these concentrations exceed the legally-permitted for use as preservatives (Anon, 1995). *Z. bailii* osmotolerance is well exemplified by different experiments: for example, Martorell and co-authors (Martorell *et al.*, 2007) reported that the yeast is able to grow in media containing up to 72% of glucose (w/v). The marked osmotolerance and the high fermentation capacity worsen the effects of spoilage, since the CO_2 generated during alcoholic fermentation was reported to be responsible of the explosion of canned and bottled foods (Stratford, 2006). Furthermore, despite the fact that it is widely accepted that *Z. bailii* is more sensitive to ethanol than *Saccharomyces cerevisiae (Kalathenos et al., 1995)*, there are examples of growth in media containing 20% of ethanol (Thomas and Davenport, 1985), indicating that the tolerance varies significantly among strains. In respect to wine contamination, it has been also reported that *Z. bailii* is able to counteract important SO_2 concentrations (> 3mg/L) through broad cellular response including sulfur extrusion, acetaldehyde production and limitation of cellular wall/membrane permeability to harmful substances (Stratford *et al.*, 1987, Park and Bakalinsky, 2000, Pilkington and Rose, 1988, Thomas and Davenport, 1985). In addition to strain-dependent tolerance, an inoculum effect has been also described: concentrations around 200 mg/L of DMDC usually do not significantly affect *Z. bailii* growth, but if the size of the inoculum is small, cells can be sensitive to a concentration of the preservative as low as 25 mg/L (Martorell *et al.*, 2007, Costa *et al.*, 2008).

Table 1. Physiology and different stress condition/compounds resistance threshold values (MIC – maximum inhibition concentration).

Yeast	Optimum growth temperature range (C°)	MIC of glucose (M)	MIC of Sulfur Dioxide (mg/L)	MIC of Dimethyl decarbonate (mg/L)	MIC of Ethanol (%, w/v)	MIC of Acetic acid (mM)	MIC of Sorbic acid (mM)
Z. bailii	25-30[4]	4.25[2]	32[2,5]	25-200[2, 3]	18-20[5]	375-550[2,5]	4.55-9.45[2,5]

[1]Stratford *et al.*, 2013
[2]Martorell *et al*., 2007
[3]Costa *et al.*, 2008
[4]James and Stratford 2011
[5]Thomas and Davenport, 1985

The management of production procedures including processing technologies, formulation of processed foods, the use of antimicrobial compounds (reviewed in (Doyle, 2007)) represents the main weapon to counteract spoilage. Microbial spoilage modeling can be effectively used for spoilage prevention (Doyle, 2007) and examples of modeling of spoilage involving *Z. bailii* were described in literature (Braun and Sutherland, 2004, Quintas *et al.*, 2005). However, microorganisms have the extraordinary ability to adapt to challenging environments, a quality that is generally difficult to include in the current *in silico* description. In addition, in the case of spoilage microorganisms, the description of the model has to take into account that these microbes can become effective even starting from extremely low concentrations. *Z. bailii* is not an exception. Indeed, it has been shown that a single viable cell in a 10 liters volume of beverages was enough to trigger an undesired spoilage event (James and Stratford, 2003). For this reason, the comprehension and the description of the molecular mechanisms that account for spoilage activity become compelling for the development of novel preservation strategies and *Z. bailii* is not an exception.

As mentioned above, high fermentation performance is one of the main traits that allow *Z. bailii* to thrive in high sugar substrate causing spoilage problems during food production processes.

Sugar metabolism in *Z. bailii*

The main characteristics of *Z. bailii* that pose severe contamination problems in high sugar concentration food are the fructophilic metabolism and the tolerance for high osmotic environments. Fructophilic yeasts are largely diffused in several ecological niches – for example in honeycomb, grape and grape juice – and *Z. bailii* is one of the few that have been characterized (Zott *et al.*, 2008).

Z. bailii transports fructose by a high-capacity and low-affinity facilitator, whereas a low capacity and high-affinity nonspecific hexose sugar transporter drives glucose uptake (Sousa-Dias *et al.*, 1996, Pina *et al.*, 2004). Furthermore, fructose can effect a moderate inactivation of the glucose transporter and can compete with glucose for the same transport system (Sousa-Dias *et al.*, 1996). The gene encoding for *Z. bailii* fructose transporter, *FFZ1* (Fructose Facilitator *Zygosaccharomyces*), was characterized by the functional complementation when expressed in *S. cerevisiae hxt* null mutant, where it conferred the ability to grow on fructose (Pina *et al.*, 2004). This and earlier works (Fuhrmann *et al.*, 1992) support the hypothesis that different molecular mechanisms are accountable for fructose and other hexose transport. Ffz1 protein has a predicted size of 616 amino acids arranged in twelve membrane spanning domains characteristic of membrane transporters. The predicted amino acid sequence of the Ffz1 protein shares low identity with the sequences of *S. cerevisiae* glucose transporters. Ffz1 is phylogenetically distant also from other sugar transporters, such as Fsy1 and Frt1 that have been described as two H^+-fructose symporter of *S. pastorianus* and *K. lactis*, respectively (Pina *et al.*, 2004). Yet, it is phylogenetically close to the Ffz1 fructose transporters described in *Z. rouxii* (Leandro *et al.*, 2011), therefore constituting a novel sugar transporter family mediating hexose transport via facilitated diffusion. Overall, these observations suggest that *Zygosaccharomyces* evolved as a clade with *FFZ* genes constituting one of the specific traits (Cabral *et al.*, 2015) that contribute to the peculiar physiology, metabolism and spoilage activity of this group of yeasts.

A clear example of the peculiar spoilage nature of *Z. bailii* linked to fructose transport can be observed considering wine fermentation. Galeote *et al.* described in different *S. cerevisiae* wine yeast strains the presence of a 17kb genetic cluster that originated from *Z. bailii* (Galeote *et al.*, 2013). Interestingly, the cluster comprises *FFZ1* and *FSY1* genes and several transporters with high similarity to the *S. cerevisiae HXT* gene family, indicating the occurrence of at least one

event of horizontal gene transfer, possibly captured by niche competition between *Z. bailii* and wine strains.

Besides utilizing common hexose sugar monomers Arez, *et al.*, (2014) demonstrated that *Z. bailii* is able to use sucrose as carbon source thanks to the production of an invertase, the enzyme catalyzing the hydrolysis of sucrose into glucose and fructose. The authors further demonstrated that glucose and sucrose were weak inducers of the gene encoding for the invertase, while the Jerusalem artichoke pulp triggered the highest invertase activity (Arez *et al.*, 2014), opening the possibility of describing the typical cascade of a two-component regulated system. It is important to mention that sucrose fermentation, which is described as delayed in the absence or in the case of poor invertase activity (Thomas and Davenport, 1985), can be favored at low pH, since this condition promotes the hydrolysis of the disaccharide.

Like *S. cerevisiae, Z. bailii* is a Crabtree positive yeast (Leyva *et al.*, 1999): at high sugar concentration it redirects part of the carbon metabolism toward ethanol production in aerobic condition. In *Z. bailii* the Crabtree effect is closely related to the carbon source provided: in the presence of fructose, aerobic ethanol production is more pronounced than in glucose, possibly sustained by the higher fructose-phosphorylating activity (Merico *et al.*, 2003). This well correlates with the different capacity of fructose and glucose transport systems. Nevertheless, Merico *et al.*, 2003 observed a lower fructose fermentation capacity of *Z. bailii* (0.83 mol^{-1}) in comparison to glucose fermenting *S. cerevisiae* (1.6 mol^{-1}) cells (Gombert *et al.*, 2001). The authors concluded that in *Z. bailii* the pyruvate dehydrogenase bypass redirects more efficiently pyruvate towards the oxidative than the fermentative pathway.

Finally, depending on the composition of the growth substrate, *Z. bailii* can proliferate under oxygen-restrictive condition. Specifically, in synthetic media supplemented with ergosterol and Tween 80, *Z. bailii* exhibits extremely slow growth, while *S. cerevisiae* proliferates efficiently. However, in complex media *Z. bailii* grows rapidly, suggesting specific nutrient requirement for supporting the anaerobic growth (Rodrigues *et al.*, 2001). This observation is consistent with the spoilage ability of *Z. bailii* detected in hermetically sealed products such as canned fruits, juices, etc. There are not significant differences with *S. cerevisiae* sugar catabolism, except for the distinct fructophilic ability due to *FFZ* gene products and to efficient pyruvate dehydrogenase bypass. Yet, this does not fully explain how *Z. bailii* readily colonizes acidic food products otherwise hostile to other microorganism including *S. cerevisiae*.

Weak Organic Acids (WOAs) tolerance of *Z. bailii* from a *S. cerevisiae* perspective

Z. bailii displays the remarkable ability to tolerate high concentrations of organic acids, among which acetic acid in the range of 375-550 mM (Stratford *et al.*, 2013), whereas for the less tolerant *S. cerevisiae* the minimum inhibitory concentration is 80-150 mM (Martorell *et al.*, 2007, Mira *et al.*, 2010a, Mira *et al.*, 2010b). However, most of the information concerning WOAs toxicity and tolerance/adaptation are investigated in the model yeast *S. cerevisiae* (Mira *et al.*, 2010a, Mira *et al.*, 2010b).

Here we start recapitulating some general concepts, together with the main findings and the current vision of the mechanisms involved in WOAs response in the bakers' yeast, as the basis to compare what was revealed in *Z. bailii*.

The common growth inhibition effect of WOAs derives from their structure and chemical properties. In aqueous solution, a WOA exists in a pH-dependent equilibrium between the uncharged acidic and charged anionic form; when the external pH is below the pKa of the WOA, the undissociated form of the acid (RCOOH) predominates, and permeate the cellular plasma membrane mainly by simple diffusion. Once inside the cell, the near-neutral cytosolic pH (Valli *et al.*, 2005) leads to the dissociation of the acid in protons (H^+) and the counteranion ($RCOO^-$). The two charged species, being unable to diffuse back across the membrane bilayer, accumulate inside the cell (Brul and Coote, 1999, Lambert and Stratford, 1999). Acidification due to protons release can influence different metabolic functions by perturbing the ionization state of amino acid side chains: this affects protein activity (Krebs *et al.*, 1983, Bracey *et al.*, 1998, Orij *et al.*, 2011) and the plasma membrane electrochemical gradient. The counteranion accumulation, besides the generation of high turgor pressure, can lead to free radical production triggering oxidative stress, protein aggregation, lipid peroxidation and the inhibition of membrane trafficking (reviewed by (Piper *et al.*, 2001, Teixeira *et al.*, 2007)) (Figure 1).

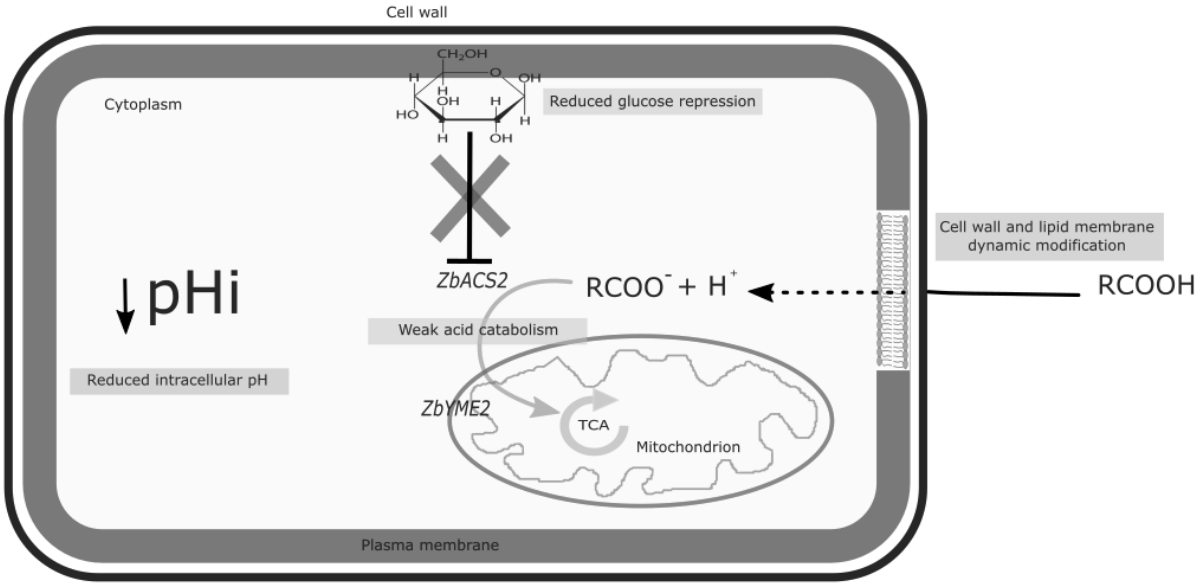


Figure 1 Schematic illustration of the major adaptive responses to weak organic acid stress in *Z. bailii*, as derived by the current state of the art. Cell wall and membrane dynamic remodeling act to decrease weak acid diffusion into the cell. Detoxification of weak acid by catabolic pathways insensitive to catabolite repression, as via TCA cycle. Reduction of pH_i can contribute to the limitation of weak organic acid accumulation.

S. cerevisiae cells respond to WOAs exposure by the activation of protection systems that are committed to restore the cellular homeostasis. The plasma membrane H^+-ATPase Pma1 and vacuolar H^+-ATPase (V-ATPase) contributes in maintaining the optimum intracellular pH by proton extrusion and sequestering (de Kok *et al.*, 2012, Kane, 2006). Nevertheless, those proteins alone would be ineffective in the absence of movement charge that compensates the ion homeostasis. The extrusion of the counteranion ($RCOO^-$) is mediated by the plasma membrane Multi Drug Resistance (MDR) transporters (Holyoak *et al.*, 1999, Piper *et al.*, 1998). The combined action of proton and counteranion extrusion is energy demanding, resulting in the reduction of biomass yield during acidic stress (reviewed by (Piper *et al.*, 2001)). In the last decade, many studies focused on understanding how cells can avoid or reduce the futile cycle of diffusional entry and active extrusion of WOAs. Indeed, once extruded from the cell, in the acidic environment protons and anions associate again and diffuse back across the membrane, resulting in a pointless and energy consuming process (Piper *et al.*, 2001). Therefore, the

limitation of passive diffusion by the modulation of cell wall composition and membrane fluidity is recognized as fundamental processes for the adaptation to WOA stress (Mira *et al.*, 2010b, Klose *et al.*, 2012, Simoes *et al.*, 2006) (Figure 1).

In the last two decades, the extreme WOAs tolerance of *Z. bailii* was also intensively explored and characterized at molecular level, giving emphasis to metabolic rewiring, proteomic response and lipid surface remodeling (Rodrigues *et al.*, 2012, Guerreiro *et al.*, 2012, Lindberg *et al.*, 2013). Although some initial studies suggested that *S. cerevisiae* and *Z. bailii* share similar preservative tolerance mechanisms, for example via cellular export system (Warth, 1989), it was then clear that superimposing the *S. cerevisiae* model might lead to underestimation of the unique evolution of the *Z. bailii*. As an example, no Pdr12-like activity was detected in *Z. bailii* during acidic stress (Papadimitriou *et al.*, 2007): this is in contrast with the widely accepted mechanism described in *S. cerevisiae*, where the Pdr12 extrusion pump is strongly induced upon WOAa exposure (reviewed in (Piper *et al.*, 2001)).

The first *Z. bailii* gene appointed for WOAs resistance was *ZbYME2*, which has been proposed to provide benzoic and sorbic acid resistance through acid degradation (Mollapour and Piper, 2001b). The protein encoded by this gene has a high degree of similarity with the N-terminal domain of the *S. cerevisiae* larger mitochondrial protein Yme2p/Rna12p. By the heterologous expression of *ZbYME2* gene, *S. cerevisiae* tolerance to benzoate and sorbate becomes comparable to *Z. bailii*, yet the counteranion degradation remains dubious. Interestingly, the *S. cerevisiae* counterpart *Sc*Yme2 do not provide the ability to degrade sorbate and benzoate, but is involved in free radicals detoxification produced by the respiratory chain (Mollapour and Piper, 2001b). This function was also hypothesized for *Zb*Yme2 and opened the pending question whether the enzyme lost the catabolic activity in *S. cerevisiae*, determining its lower tolerance to acidic environments, or gained a novel function in *Z. bailii* for adaptation to harsh conditions.

Another feature that distinguishes *Z. bailii* from *S. cerevisiae* and other yeasts is the absence of glucose repression, mainly named as catabolite repression (Ronne, 1995). The absence of such regulation allows *Z. bailii* to co-metabolize glucose and acetate (Sousa *et al.*, 1998). The involvement of a specific membrane transporter for acetic acid (Sousa *et al.*, 1996) and the presence of a high metabolic flux through the acetyl-CoA-synthetase, encoded by the *ZbACS2* gene, allow a fast acetic acid consumption even in the presence of glucose, thus maintaining the cytosolic concentration of this acid below the toxic levels, and generating additional energy for

cellular maintenance (Rodrigues *et al.*, 2012). Moreover, the effectiveness of *ZbACS2* is demonstrated by the increase of *S. cerevisiae* tolerance to acid when this gene is heterologously expressed (Rodrigues *et al.*, 2004). Noticeably, despite *ScACS2* has a high degree of similarity with *ZbACS2* (Rodrigues *et al.*, 2004), the *ScACS2* is not involved in acetate catabolism in *S. cerevisiae* (Van den Berg and Steensman, 1995). A metabolic study conducted on *Z. bailii* cells cultivated with labeled substrates confirmed the co-consumption of acetate and glucose (Rodrigues *et al.*, 2012): glucose is mainly channeled into the glycolytic pathway, while acetate provides acetyl-CoA as precursor for protein (through the tricarboxylic acid cycle) and lipid synthesis. This metabolic reshaping is in agreement with proteomic data obtained by Guerreiro and collaborators (Guerreiro *et al.*, 2012). Although acetyl-CoA synthase mediates acetic acid metabolism in *Z. bailii*, this protein was not over-represented during growth on glucose-acetic acid medium. Nevertheless, several proteins involved in the TCA cycle (mitochondrial malate dehydrogenase – Mdh1, aconitase – Aco1, mitochondrial isocitrate dehydogenase – Idh2, citrate synthase – Cit1, dihydolipoamide dehydrogenase – Lpd1) were found over-represented in this condition. This response is consistent with the necessity to oxidize the acetyl-CoA, and might also indicate that *ZbACS2* is constitutively expressed. The higher levels of the *Z. bailii* mitochondrial ATP synthase subunits, Atp1 and Atp2, in glucose-acetic acid growing cells suggests an increase of oxidative phosphorylation due to the increase of the TCA flux and to the concomitant production of reduced cofactors.

The growth of *Z. bailii* in a non-fermentable carbon source like acetic acid (in the absence of glucose) is similar to the growth on ethanol. Up-regulation of enolase – Eno1, triose phosphate isomerase – Tpi1, 3-phosphoglycerate kinase – Pgk1 and fructose1,2 biphosphate aldolase – Fba1 (a key regulator of gluconeogenesis), are consistent with the increase of gluconeogenic flux for the biosynthesis of anabolic precursors (Guerreiro *et al.*, 2012). Moreover, an increase of the transaldolase Tal1 signal, an enzyme of the pentose phosphate pathway, is coherent with a requirement for redox balance maintenance and the production of anabolic metabolites for nucleotides, proteins and fatty acid biosynthesis. Finally, *Z. bailii* growth on acetic acid seems dependent on the glutamate node, as suggested by the increase of the enzymes glutamate dehydrogenase – Gdh1, and of the mitochondrial and cytosolic isocitrate dehydrogenases - Idh2 and Idp2, respectively (Guerreiro *et al.*, 2012).

The same study showed higher levels of proteins involved in stress response and adaptation when *Z. bailii* cells were exposed to acetic acid. In particular, they found the mitochondrial

manganese superoxide dismutase – Sod2 and the mitochondrial porin – Omp2, which are known for their protection against oxidative stress (Pereira *et al.*, 2007), but also the dihydrossiacetone kinase – Dak2 and the phosphomannomutase – Sec53. Dak2 up-regulation might be involved in protection mechanisms evoked to reduce the toxicity of dihydroxyacetone during oxidative stress and in the maintenance of the ATP balance, since a similar mechanism was proposed in *S. cerevisiae* (Gasch *et al.*, 2000, Norberck and Blomberg, 1997, Boy-Marcotte *et al.*, 1999, Molin *et al.*, 2003).

As mentioned before, in *S. cerevisiae* the cell wall reorganization in response to WOAs is an important protection mechanism to counteract acidic injury by limiting the permeability of undissociated acid within the cells (Simoes *et al.*, 2006, Mira *et al.*, 2010b). The same defense mechanism might occur in *Z. bailii*. Recent studies on *Z. bailii* using Fourier Transformed Infrared micro-spectroscopy (FTIR) showed lactic acid dependent changes in the signals of bands corresponding to glucans and mannans, an indication of cell wall reorganization (Kuanyshev *et al.*, 2016).

Yeast plasma membrane fluidity plays a crucial role in the adaptation to environmental stresses (Klose *et al.*, 2012), including WOAs. It was demonstrated that in *S. cerevisiae* the presence of ethanol or butanol – perturbing cell surface organization - increase passive diffusion of acetic acid across the membrane, exacerbating the toxicity. Ethanol and other alkanols have also an inhibitory effect on active membrane facilitators, reducing their transport capacity (Casal *et al.*, 1998). This inhibitory effect was also reported for *Z. bailii*, but in this yeast ethanol up to 2.43 M seems to play a protective role leading to the inhibition of active uptake of some organic acids and thus reducing its intracellular concentration (Sousa *et al.*, 1996). This physiological mechanism can explain the occurrence of *Z. bailii* during wine fermentation, characterized by the abundant concentration of both ethanol and acetic acid.

However, this observation might sound counterintuitive if we consider that the membrane integrity and the activity of the embedded/associated proteins are strictly correlated. It is unlikely that membrane damages do not influence protein structure/activity and *vice versa*, unless we do not imply some differences in *Z. bailii* that can justify the major robustness and its adaptation in harsh environments.

As expected, *Z. bailii* strains exposed to various stressors show differences in the fatty acid composition. The first evidence reported in literature described that the long chain fatty acids (stearic, oleic and linoleic acids) increased in response to high ethanol concentration, whereas

the fatty acids with a shorted acyl chain (palmitic and palmitoleic acids) increased mainly at low temperatures (Baileras Couto and Huis in't Veld, 1995). Additional investigations generically suggested that differences in the cell surface structure might also account for higher acidic tolerance of *Z. bailii* compared to *S. cerevisiae* (Prudencio *et al.*, 1998). Only recently, an accurate comparative study of the lipidome of *S. cerevisiae* and *Z. bailii* revealed the pronounced diversity occurring in the plasma membrane composition between the two yeasts (Lindberg *et al.*, 2013). Specifically, *Z. bailii* adaptation to acetic acid correlated with quali- and quantitative significant changes in fatty acid composition: the authors described enrichment in the sphingolipids fraction at the expense of glycerophospholipids, leading to a reduced permeability of the membrane to the acid and reasonably to an improved support for the functionality of different membrane proteins (*i.e.* Pma1). On the contrary, only minor changes were observed in *S. cerevisiae*, suggesting a limited ability for lipid membrane adaptation toward the stress compared to *Z. bailii* (Lindberg *et al.*, 2013). The results were supported by an *in silico* molecular dynamics simulation, further confirmed by *in vivo* experiments, of *Z. bailii* plasma membrane, suggesting that the difference in sphingolipid content can account for thicker and more dense membranes, limiting the diffusion of acetic acid (Lindahl *et al.*, 2016).

Other WOAs could also evoke additional and/or different changes: the above mentioned FTIR-based study showed that the phosphatidylcholine content decreased under lactic acid treatment, once more suggesting a mechanism of adaptation through changes in membrane fluidity (Kuanyshev *et al.*, 2016), and strengthening the active role of the *Z. bailii* lipidome dynamic adaptation in providing stress tolerance.

Together with the preservative degradation ability and the diminished permeability, a reduced intracellular pH (pH_i) could also contribute to the tolerance of *Z. bailii* to weak organic acids (Figure 1). It was recently demonstrated a recurrent heterogeneity in *Z. bailii* cellular population: a sub-population characterized by a lower pH_i (0.4-0.8 pH units) is extremely resistant to a number of organic acids (Stratford *et al.*, 2013), a behavior that was also observed by others (Arneborg *et al.*, 2000, Dang *et al.*, 2012). Stratford and coworkers (Stratford *et al.*, 2013) calculated the increase in intracellular accumulation of sorbic acid (accumulation index) at different pH_i (from 4.0 to 6.6) while fixing the extracellular pH (4.0 units). According with this calculation, cells with lower pH_i accumulated less acid, with beneficial effects on metabolism and growth, and generating the observed resistant sub-population, which may survive hostile conditions.

Undeniably, *Z. bailii* represents industrially attractive yeast, whose robust phenotype can be exploited and/or transferred to known industrial microorganisms.

Exploitation of *Z. bailii*: a potential friend for industrial biotechnology

The ability to endure low pH and high WOAs concentration makes *Z. bailii* a promising platform to be applied in industrial bioprocesses (Branduardi *et al.*, 2004, Sauer *et al.*, 2004, Vigentini *et al.*, 2005, Camattari *et al.*, 2007, Piper *et al.*, 2001).

To investigate the potential of *Z. bailii* as cell factory, traditional yeast molecular tools, including a set of episomal plasmids based on *Sc*ARS and *Sc*CEN sequences, have been developed, leading to the selection of transformants producing the desired products. A stable integration and multiple integration vectors, both based on *LEU2* marker selection, were also successfully adapted for *Z. bailii* (Branduardi *et al.*, 2004, Dato *et al.*, 2010).

Heterologous protein production was successfully carried out in *Z. bailii* obtaining better yields compared to *S. cerevisiae* (Branduardi *et al.*, 2004, Vigentini *et al.*, 2005). Proteins of different origins, dimensions and biochemical functions were expressed both at cytoplasmic and extracellular level. In comparison to *S. cerevisiae*, *Z. bailii* displays a better secretory capacity in defined media, which can be further improved by *GAS1* gene deletion (Passolunghi *et al.*, 2010). From an industrial perspective, this is an advantageous feature, since protein secretion (that can simplify recovery and purification procedures and costs) is among the principal bottlenecks of heterologous productions.

As in other non-*Saccharomyces* yeasts, the poor efficiency of homologous recombination (or the prevalence of the non-homologous end-joining recombination) makes the genome manipulations troublesome and limited (MacKenzie *et al.*, 1987, Mollapour and Piper, 2001a). Indeed so far most of the investigations on *Z. bailii* were performed at the single gene level, and with limited availability of auxotrophic strains (Mollapour and Piper, 2001b, Dato *et al.*, 2008, Passolunghi *et al.*, 2010). The optimization of novel genome editing tools, such as the CRISPR-Cas9 system, is essential to expand the potential of *Z. bailii's* industrial application.

Yeast-based fermentations for the production of organic acids starting from cheap carbon sources are well documented in literature (Ilmen *et al.*, 2013, Koivistoinen *et al.*, 2013, Valli *et al.*, 2006, Xu *et al.*, 2012, Zelle *et al.*, 2008), often taking advantage from the ability of these cells to sustain adequate growth at stressful but desired conditions. Organic acids represent building-blocks for several products (polymers, cosmetics, and pharmaceutics) of industrial and

commercial interests. Their production by microbial fermentation is willing to represent a sustainable alternative to fossil oil-based processes (Sauer *et al.*, 2008), but the production costs need to be maintained as low as possible to be competitive with the non-sustainable counterpart. If fermentation processes are run at low pH, the organic acid of interest is produced in the desired undissociated form, with a significant reduction of the overall costs. However, this generates a stressful environment for the microorganism. In this context, *Z. bailii* endowed with extreme tolerance to acidic environment represents the ideal candidate for organic acids production. The first attempts were done to produce lactic acid (Dato *et al.*, 2010) and L-ascorbic acid (Sauer *et al.*, 2004) in *Z. bailii*. Despite yield and productivity were low, the proof of concept was clearly obtained, and opened the room for further developments and improvements.

As mentioned before, the viability of a bioprocess relies on the reduction of the overall costs, that for low-added value compounds almost equally distributes between downstream and substrate costs. Therefore, the use of residual or wasted biomass can play an important role for the viability of microbial bioconversion, named second-generation productions.

Among abundant polysaccharides that can be a cheap source of sugars for bioethanol production there is inulin, which is present in a variety of plants (Chi *et al.*, 2011). Inulin is hydrolyzed by microbial inulinase producing inulo-oligosaccharides, glucose and fructose, which can be further fermented to bioethanol (Kango and Jain, 2011). Paixao *et al.*, isolated inulinase producing *Z. bailii* strains, which have an indubitable potential for industrial applications (Paixao *et al.*, 2013). Second-generation bioethanol obtained from lignocellulose fermentation is by far the most sustainable source of bioenergy, due to the abundance of this substrate, and has the potential to displace fossil raw materials for energy and chemical feedstock (Laluce *et al.*, 2012). Several yeasts (*Saccharomyces* and non-*Saccharomyces*) are employed for commercial production of bioethanol using sugars released from lignocellulosic hydrolysate such as glucose, xylose, arabinose (Margeot *et al.*, 2009). However, acetic acid, formic acid, furan aldehydes and phenolic compounds (released from lignocellulosic hydrolysate) impose important limitations to microbial growth and therefore to ethanol yield (Almeida *et al.*, 2007) that still need to be resolved. In addition, high ethanol concentrations are also harmful for yeast (Laluce *et al.*, 2012). *Z. bailii* can represent a solution due to its superior ability to endure weak acid toxicity and harsh conditions, which is important for biofuels and bio-based chemicals production.

Conclusion

The knowledge gathered about the yeast *Z. bailii* is extremely intriguing and deserves to be further developed and investigated. Understanding the peculiar physiology of *Z. bailii* may help to improve not only spoilage prevention measures, but also the knowledge about stress resistance capability in yeasts. Being a problem for food industry, *Z. bailii* may indicate how to develop new robust strains for biotechnology exploitation. More accurate genome annotations and genomics analysis will be helpful to improve well-known cell factories and further develop *Z. bailii* as a novel one. Interestingly, in the last years it appeared that among the best performing *Z. bailii* strains there are some hybrid strains, named *Z. parabailii*. Some of those genomes have been partially (Galeote *et al.*, 2013, Mira *et al.*, 2014) or completely (Ortiz-Merino *et al* in press) assembled and annotated, promising to leverage similar potential as for the hybrids of *S. cerevisiae* industrial strains (Krogerus *et al.*, 2017, Steensels *et al.*, 2014).

Overall, the extensive researches devoted to unveil the basis of *Z. bailii* tolerance, as well as of other robust microbial strains, will contribute to improving the sustainability and viability of industrial bioprocesses.

Acknowledgments

This work was supported by European Union FP7 Marie Curie Programme [YEASTCELL - 7PQ MARIE CURIE (12-4-2001100-40)] P.B. and D.P. also acknowledge the support by FAR (Fondo di Ateneo per la Ricerca) of the University of Milano-Bicocca. We thank John Morrissey and Francesca Doonan for encouragement and support.

The authors declare that there is no conflict of interest

References

ALMEIDA, JRM, MODIG, T, PETERSSON, A*, et al.* 2007. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *Journal of Chemical Technology and Biotechnology,* 82**,** 340-349.

ANON 1995. Directive No 95/2/EC of 20 February 1995 on food additives other than colours and sweeteners. *EC Official Journal (L61)***,** 1-40.

AREZ, BF, ALVES, L & PAIXAO, SM 2014. Production and characterization of a novel yeast extracellular invertase activity towards improved dibenzothiophene biodesulfurization. *Appl Biochem Biotechnol,* 174**,** 2048-2057.

ARNEBORG, N, JESPERSEN, L & JAKOBSEN, M 2000. Individual cells of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* exhibit different short-term intracellular pH responses to acetic acid. *Arch Microbiol,* 174**,** 125-128.

BAILERAS COUTO, MM & HUIS IN'T VELD, J 1995. Influence of ethanol and temperature on the cellular fatty acid composition of *Zygosaccharomyces bailii* spoilage yeast. *J Appl Bacteriol,* 78**,** 327-334.

BOY-MARCOTTE, E, LAGNIEL, G, PERROT, M*, et al.* 1999. The heat shock response in yeast: differential regulations and contributions of the Msn2p/Msn4p and Hsf1p regulons. *Mol Microbiol,* 33**,** 274-283.

BRACEY, D, HOLYOAK, CD & COOTE, PJ 1998. Comparison of the inhibitory effect of sorbic acid and amphotericin B on *Saccharomyces cerevisiae*: is growth inhibition dependent on reduced intracellular pH? *J Appl Microbiol,* 85**,** 1056-1066.

BRANDUARDI, P, VALLI, M, BRAMBILLA, L*, et al.* 2004. The yeast *Zygosaccharomyces bailii*: a new host for heterologous protein production, secretion and for metabolic engineering applications. *FEMS Yeast Res,* 4**,** 493-504.

BRAUN, P & SUTHERLAND, JP 2004. Predictive modeling of growth and enzymatic synthesis and activity by a cocktail of *Yarrowia lipolytica*, *Zygosaccharomyces bailii* and *Pichia anomala*. *Food Microbiol,* 21**,** 459-467.

BRUL, S & COOTE, P 1999. Preservative agents in foods. Mode of action and microbial resistance mechanisms. *Int J Food Microbiol,* 50**,** 1-17.

CABRAL, S, PRISTA, C, LOUREIRO-DIAS, MC & LEANDRO, MJ 2015. Occurrence of FFZ genes in yeasts and correlation with fructophilic behaviour. *Microbiology,* 161**,** 2008-2018.

CAMATTARI, A, BIANCHI, MM, BRANDUARDI, P*, et al.* 2007. Induction by hypoxia of heterologous-protein production with the KlPDC1 promoter in yeasts. *Appl Environ Microbiol,* 73**,** 922-929.

CASAL, M, CARDOSO, H & LEAO, C 1998. Effects of ethanol and other alkanols on transport of acetic acid in *Saccharomyces cerevisiae*. *Appl Environ Microbiol,* 64**,** 665-668.

CHI, ZM, ZHANG, T, CAO, TS*, et al.* 2011. Biotechnological potential of inulin for bioprocesses. *Bioresour Technol,* 102**,** 4295-4303.

COSTA, A, BARATA, A, MALFEITO-FERREIRA, M & LOUREIRO, V 2008. Evaluation of the inhibitory effect of dimethyl dicarbonate (DMDC) against wine microorganisms. *Food Microbiol,* 25**,** 422-427.

DANG, TD, DE MAESENEIRE, SL, ZHANG, BY*, et al.* 2012. Monitoring the intracellular pH of *Zygosaccharomyces bailii* by green fluorescent protein. *Int J Food Microbiol,* 156**,** 290-295.

DATO, L, BRANDUARDI, P, PASSOLUNGHI, S*, 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 Res,* 10**,** 894-908.

DATO, L, SAUER, M, PASSOLUNGHI, S*, et al.* 2008. Investigating the multibudded and binucleate phenotype of the yeast *Zygosaccharomyces bailii* growing on minimal medium. *FEMS Yeast Res,* 8**,** 906-915.

DE KOK, S, YILMAZ, D, DARAN, JM*, et al.* 2012. In vivo analysis of *Saccharomyces cerevisiae* plasma membrane ATPase Pma1p isoforms with increased in vitro H+/ATP stoichiometry. *Antonie Von Leuwenhoek,* 102**,** 401-406.

DOMIZIO, P, ROMANI, C, LENCIONI, L*, et al.* 2011. Outlining a future for non-Saccharomyces yeasts: selection of putative spoilage wine strains to be used in association with *Saccharomyces cerevisiae* for grape juice fermentation. *Int J Food Microbiol,* 147**,** 170-180.

DOYLE, ME 2007. Microbial food spoilage - losses and control strategy. *Fri Briefings*.

FLEET, GH 2007. Yeasts in foods and beverages: impact on product quality and safety. *Curr Opin Biotechnol,* 18**,** 170-175.

FUHRMANN, GF, STORCH, D, BODE, HP & VOLKER, B 1992. Inhibition of glucose transport in *Saccharomyces cerevisiae* by uranyl ions. *J Biotechnol,* 27**,** 75-84.

GALEOTE, V, BIGEY, F, DEVILLERS, H*, et al.* 2013. Genome Sequence of the Food Spoilage Yeast *Zygosaccharomyces bailii* CLIB 213T. *Genome Announc,* 1.

GASCH, AP, SPELLMAN, PT, KAO, CM*, et al.* 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell,* 11**,** 4241-4257.

GOMBERT, AK, MOREIRA DOS SANTOS, M, CHRISTENSEN, B & NIELSEN, J 2001. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *J Bacteriol,* 183**,** 1441-1451.

GUERREIRO, JF, MIRA, NP & SA-CORREIA, I 2012. Adaptative response to acetic acid in the highly resistant yeast species *Zygosaccharomyces bailii* revealed by quantitative proteomics. *Proteomics,* 12**,** 2303-2318.

HOLYOAK, CD, BRACEY, D, PIPER, PW*, et al.* 1999. The *Saccharomyces cerevisiae* weak-acid-inducible ABC transporter Pdr12 transports fluorescein and preservative anions from the cytosol by an energy-dependent mechanism. *J Bacteriol,* 181**,** 4644-4652.

ILMEN, M, KOIVURANTA, K, RUOHONEN, L*, et al.* 2013. Production of L-lactic acid by the yeast Candida sonorensis expressing heterologous bacterial and fungal lactate dehydrogenases. *Microb Cell Fact,* 12**,** 53.

JAMES, SA & STRATFORD, M 2011. Zygosaccharomyces Barker (1901), 5th edn. In: Kurtzman, C.P., Fell, J.W., Boekhout, T. (Eds.), The Yeasts, a Taxonomic Study, vol. 2. Elsevier, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo, pp. 937–947.

KALATHENOS, P, SUTHERLAND, JP & ROBERTS, TA 1995. Resistance of some wine spoilage yeasts to combinations of ethanol and acids present in wine. *Journal of Applied Bacteriology,* 78**,** 245-250.

KANE, PM 2006. The where, when, and how of organelle acidification by the yeast vacuolar H+-ATPase. *Microbiol Mol Biol Rev,* 70**,** 177-191.

KANGO, N & JAIN, SC 2011. Production and Properties of Microbial Inulinases: Recent Advances. *Food Biotechnology,* 25**,** 165-212.

KLOSE, C, SURMA, MA, GERL, MJ*, et al.* 2012. Flexibility of a eukaryotic lipidome--insights from yeast lipidomics. *PLoS One,* 7**,** e35063.

KOIVISTOINEN, OM, KUIVANEN, J, BARTH, D*, et al.* 2013. Glycolic acid production in the engineered yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. *Microb Cell Fact,* 12**,** 82.

KREBS, HA, WIGGINS, D, STUBBS, M*, et al.* 1983. Studies on the mechanism of the antifungal action of benzoate. *Biochem J,* 214**,** 657-663.

KROGERUS, K, MAGALHAES, F, VIDGREN, V & GIBSON, B 2017. Novel brewing yeast hybrids: creation and application. *Appl Microbiol Biotechnol,* 101**,** 65-78.

KUANYSHEV, N, AMI, D, SIGNORI, L*, et al.* 2016. Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation. *FEMS Yeast Res,* 16.

KURTZMAN, CP & ROBNETT, CJ 2003. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. *FEMS Yeast Res,* 3**,** 417-432.

LALUCE, C, SCHENBERG, AC, GALLARDO, JC*, et al.* 2012. Advances and developments in strategies to improve strains of *Saccharomyces cerevisiae* and processes to obtain the lignocellulosic ethanol--a review. *Appl Biochem Biotechnol,* 166**,** 1908-1926.

LAMBERT, RJ & STRATFORD, M 1999. Weak-acid preservatives: modelling microbial inhibition and response. *J Appl Microbiol,* 86**,** 157-164.

LEANDRO, MJ, SYCHROVA, H, PRISTA, C & LOUREIRO-DIAS, MC 2011. The osmotolerant fructophilic yeast Zygosaccharomyces rouxii employs two plasma-membrane fructose uptake systems belonging to a new family of yeast sugar transporters. *Microbiology,* 157**,** 601-608.

LEYVA, JS, MANRIQUE, M, PRATS, L*, et al.* 1999. Regulation of fermentative CO2 production by the food spoilage yeast *Zygosaccharomyces bailii Enzyme Microbiol Technol,* 24**,** 270-275.

LINDAHL, L, GENHEDEN, S, ERIKSSON, LA*, et al.* 2016. Sphingolipids contribute to acetic acid resistance in *Zygosaccharomyces bailii*. *Biotechnol Bioeng,* 113**,** 744-753.

LINDBERG, L, SANTOS, AX, RIEZMAN, H*, et al.* 2013. Lipidomic profiling of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* reveals critical changes in lipid composition in response to acetic acid stress. *PLoS One,* 8**,** e73936.

MACKENZIE, DA, PEARSON, BM, FULLER, LJ & KEENMAN, MH 1987. Stability of chimaeric vectors based on plasmids from the yeast *Zygosaccharomyces bailii*. *Micorbios Letters,* 36**,** 55-63.

MARGEOT, A, HAHN-HAGERDAL, B, EDLUND, M*, et al.* 2009. New improvements for lignocellulosic ethanol. *Curr Opin Biotechnol,* 20**,** 372-380.

MARTORELL, P, STRATFORD, M, STEELS, H*, et al.* 2007. Physiological characterization of spoilage strains of *Zygosaccharomyces bailii* and *Zygosaccharomyces rouxii* isolated from high sugar environments. *Int J Food Microbiol,* 114**,** 234-242.

MERICO, A, CAPITANIO, D, VIGENTINI, I*, et al.* 2003. Aerobic sugar metabolism in the spoilage yeast *Zygosaccharomyces bailii*. *FEMS Yeast Res,* 4**,** 277-283.

MIRA, NP, MUNSTERKOTTER, M, DIAS-VALADA, F*, et al.* 2014. The genome sequence of the highly acetic acid-tolerant *Zygosaccharomyces bailii*-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. *DNA Res,* 21**,** 299-313.

MIRA, NP, PALMA, M, GUERREIRO, JF & SA-CORREIA, I 2010a. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. *Microb Cell Fact,* 9**,** 79.

MIRA, NP, TEIXEIRA, MC & SA-CORREIA, I 2010b. Adaptive response and tolerance to weak acids in *Saccharomyces cerevisiae*: a genome-wide view. *OMICS,* 14**,** 525-540.

MOLIN, M, NORBERCK, J & BLOMBERG, A 2003. Dihydroxyacetone kinases in *Saccharomyces cerevisiae* are involved in detoxification of dihydroxiacetone. *J Biol Chem,* 278**,** 1415-1423.

MOLLAPOUR, M & PIPER, P 2001a. Targeted gene deletion in *Zygosaccharomyces bailii*. *Yeast,* 18**,** 173-186.

MOLLAPOUR, M & PIPER, PW 2001b. The ZbYME2 gene from the food spoilage yeast *Zygosaccharomyces bailii* confers not only YME2 functions in *Saccharomyces cerevisiae*, but also the capacity for catabolism of sorbate and benzoate, two major weak organic acid preservatives. *Mol Microbiol,* 42**,** 919-930.

NORBERCK, J & BLOMBERG, A 1997. Metabolic and regulatory changes associated with growth of *Saccharomyces cerevisiae* in 1.4 M NaCl. Evidences for osmotic induction of glycerol dissimilation via the gihydroxyacetone pathway. *J Biol Chem,* 272**,** 5544-5554.

ORIJ, R, BRUL, S & SMITS, GJ 2011. Intracellular pH is a tightly controlled signal in yeast. *Biochim Biophys Acta,* 1810**,** 933-944.

PAIXAO, SM, TEIXEIRA, PD, SILVA, TP*, et al.* 2013. Screening of novel yeast inulinases and further application to bioprocesses. *N Biotechnol,* 30**,** 598-606.

PAPADIMITRIOU, MN, 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,* 113**,** 173-179.

PARK, H & BAKALINSKY, AT 2000. SSU1 mediates sulphite efflux in *Saccharomyces cerevisiae*. *Yeast,* 16**,** 881-888.

PASSOLUNGHI, S, RIBOLDI, L, DATO, L*, et al.* 2010. Cloning of the *Zygosaccharomyces bailii* GAS1 homologue and effect of cell wall engineering on protein secretory phenotype. *Microb Cell Fact,* 9**,** 7.

PEREIRA, C, CAMOUGRAND, N, MANON, S*, et al.* 2007. ADP/ATP carrier is required for mitochondrial outer membrane permeabilization and cytochrome c release in yeast apoptosis. *Mol Microbiol,* 66**,** 571-582.

PILKINGTON, BJ & ROSE, AH 1988. Reactions of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* to sulphite. *J Gen Microbiol,* 134**,** 2823-2830.

PINA, C, GONCALVES, P, PRISTA, C & LOUREIRO-DIAS, MC 2004. Ffz1, a new transporter specific for fructose from *Zygosaccharomyces bailii*. *Microbiology,* 150**,** 2429-2433.

PIPER, P, CALDERON, CO, HATZIXANTHIS, K & MOLLAPOUR, M 2001. Weak acid adaptation: the stress response that confers yeasts with resistance to organic acid food preservatives. *Microbiology,* 147**,** 2635-2642.

PIPER, P, MAHE, Y, THOMPSON, S*, et al.* 1998. The pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast. *EMBO J,* 17**,** 4257-4265.

PRUDENCIO, C, SANSONETTY, F & CORTE-REAL, M 1998. Flow cytometric assessment of cell structural and functional changes induced by acetic acid in the yeasts *Zygosaccharomyces bailii* and *Saccharomyces cerevisiae*. *Cytometry,* 31**,** 307-313.

QUINTAS, C, LEYVA, JS, SOTOCA, R*, et al.* 2005. A model of the specific growth rate inhibition by weak acids in yeast based on enery requirements. *Int J Food Microbiol,* 100**,** 125-130.

RODRIGUES, F, CORTE-REAL, M, LEAO, C*, et al.* 2001. Oxygen requirements of the food spoilage yeast *Zygosaccharomyces bailii* in synthetic and complex media. *Appl Environ Microbiol,* 67**,** 2123-2128.

RODRIGUES, F, SOUSA, MJ, LUDOVICO, P*, et al.* 2012. The fate of acetic acid during glucose co-metabolism by the spoilage yeast *Zygosaccharomyces bailii*. *PLoS One,* 7**,** e52402.

RODRIGUES, F, ZEEMAN, AM, CARDOSO, H*, et al.* 2004. Isolation of an acetyl-CoA synthetase gene (ZbACS2) from *Zygosaccharomyces bailii*. *Yeast,* 21**,** 325-331.

ROMANO, P, FIORE, C, PARAGGIO, M*, et al.* 2003. Function of yeast species and strains in wine flavour. *Int J Food Microbiol,* 86**,** 169-180.

ROMANO, P & SUZZI, G 1993. Higher alcohol and acetoin production by Zygosaccharomyces wine yeast. *J Appl Bacteriol,* 75**,** 541-545.

RONNE, H 1995. Glucose repression in fungi. *Trends Genet,* 11**,** 12-17.

ROSA, CA & LACHANCE, M-A 2005. *Zygosaccharomyces machadoi* sp. n., a yeast species isolated from a nest of the stingless bee Tetragonisca angustula. *Lundiana,* 6**,** 27-29.

SAKSINCHAI, S, SUZUKI, M, CHANTAWANNAKUL, P*, et al.* 2012. A novel ascosporpgenous yeast species, *Zygosaccharomyces siamensis*, and the sugar tolerant yeast associated with raw honey collected in Thailand. *Fungal Divers,* 52**,** 123-139.

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,* 70**,** 6086-6091.

SAUER, M, PORRO, D, MATTANOVICH, D & BRANDUARDI, P 2008. Microbial production of organic acids: expanding the markets. *Trends Biotechnol,* 26**,** 100-108.

SIMOES, T, MIRA, NP, FERNANDES, AR & SA-CORREIA, I 2006. The SPI1 gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. *Appl Environ Microbiol,* 72**,** 7168-7175.

SOLIERI, L, CASSANELLI, S, CROCE, MA & GIUDICI, P 2008. Genome size and ploidy level: new insights for elucidating relationships in Zygosaccharomyces species. *Fungal Genet Biol,* 45**,** 1582-1590.

SOUSA-DIAS, S, GONCALVES, T, LEYVA, JS*, et al.* 1996. Kinetics and regulation of fructose and glucose transport systems are responsible for fructophily in *Zygosaccharomyces bailii*. *Microbiology,* 142**,** 1733-1738.

SOUSA, MJ, MIRANDA, L, CORTE-REAL, M & LEAO, C 1996. Transport of acetic acid in *Zygosaccharomyces bailii*: effects of ethanol and their implications on the resistance of the yeast to acidic environments. *Appl Environ Microbiol,* 62**,** 3152-3157.

SOUSA, MJ, RODRIGUES, F, CORTE-REAL, M & LEAO, C 1998. Mechanisms underlying the transport and intracellular metabolism of acetic acid in the presence of glucose in the yeast *Zygosaccharomyces bailii*. *Microbiology,* 144 (Pt 3)**,** 665-670.

STEENSELS, J, SNOEK, T, MEERSMAN, E*, et al.* 2014. Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol Rev,* 38**,** 947-995.

STRATFORD, M 2006. Food and beverage spoilage yeast. In: Querol, A. and Fleet, G. H., eds, Yeats in fodd and beverages. Springer: Heidelberg.

STRATFORD, M, MORGAN, P & ROSE, AH 1987. Sulphur Dioxide Resistance in *Saccharomyces cerevisiae* and Saccharomycodes ludwigii. *Microbiology,* 133**,** 2173-2179.

STRATFORD, M, STEELS, H, NEBE-VON-CARON, G*, et al.* 2013. Extreme resistance to weak-acid preservatives in the spoilage yeast *Zygosaccharomyces bailii*. *Int J Food Microbiol,* 166**,** 126-134.

SUH, SO, GUJJARI, P, BERES, C*, et al.* 2013. Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. *Int J Syst Evol Microbiol,* 63**,** 1922-1929.

TEIXEIRA, MC, DUQUE, P & SA-CORREIA, I 2007. Environmental genomics: mechanistic insights into toxicity of and resistance to the herbicide 2,4-D. *Trends Biotechnol,* 25**,** 363-370.

THOMAS, DS & DAVENPORT, RR 1985. *Zygosaccharomyces bailii* — a profile of characteristics and spoilage activities. *Food Microbiology,* 2**,** 157-169.

TORRIANI, S, LORENZINI, M, SALVETTI, E & FELIS, GE 2011. *Zygosaccharomyces gambellarensis* sp. nov., an ascosporogenous yeast isolated from an Italian 'passito' style wine. *Int J Syst Evol Microbiol,* 61**,** 3084-3088.

VALLI, M, SAUER, M, BRANDUARDI, P*, et al.* 2005. Intracellular pH distribution in *Saccharomyces cerevisiae* cell populations, analyzed by flow cytometry. *Appl Environ Microbiol,* 71**,** 1515-1521.

VALLI, M, SAUER, M, BRANDUARDI, P*, et al.* 2006. Improvement of lactic acid production in *Saccharomyces cerevisiae* by cell sorting for high intracellular pH. *Appl Environ Microbiol,* 72**,** 5492-5499.

VAN DEN BERG, MA & STEENSMAN, HY 1995. ACS2, a *Saccharomyces cerevisiae* gene encoding acetyl-coenzyme A synthase, essential for growth on glucose. *Eur J Biochem,* 231**,** 704-713.

VAN DER WALT, JP & JOHANSSEN, E 1975. The genus Torulaspora. *Lindner. Counc. Sci. Ind. Res. Res. Rep. Pretoria* 325**,** 1-23.

VIGENTINI, I, BRAMBILLA, L, BRANDUARDI, P*, et al.* 2005. Heterologous protein production in *Zygosaccharomyces bailii*: physiological effects and fermentative strategies. *FEMS Yeast Res,* 5**,** 647-652.

WARTH, AD 1989. Transport of benzoic and propanoic acids by *Zygosaccharomyces bailii*. *J Gen Microbiol,* 135**,** 1383-1390.

XU, G, ZOU, W, CHEN, X*, et al.* 2012. Fumaric acid production in *Saccharomyces cerevisiae* by in silico aided metabolic engineering. *PLoS One,* 7**,** e52086.

ZELLE, RM, DE HULSTER, E, VAN WINDEN, WA*, et al.* 2008. Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. *Appl Environ Microbiol,* 74**,** 2766-2777.

ZOTT, K, MIOT-SERTIER, C, CLAISSE, O*, et al.* 2008. Dynamics and diversity of non-Saccharomyces yeasts during the early stages in winemaking. *Int J Food Microbiol,* 125**,** 197-203.

Chapter 2

Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation

Kuanyshev, N., Ami, D., Signori, L., Porro, D., Morrissey, J. P. and Branduardi, P. (2016). **Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation**. *FEMS Yeast Res* 16. DOI: 10.1093/femsyr/fow058

Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation

Nurzhan Kuanyshev[1], Diletta Ami[1,2,3*], Lorenzo Signori[1], Danilo Porro[1], John P. Morrissey[4] and Paola Branduardi[1*]

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, Milano 20126 (Italy).
[2]Department of Physics, University of Milano-Bicocca, Piazza della Scienza 3, Milano 20126 (Italy).
[3]Consorzio Nazionale Interuniversitario per le Scienze fisiche della Materia (CNISM) UdR Milano-Bicocca, Via R. Cozzi 53, Milano 20126 (Italy).
[4]School of Microbiology, University College Cork, Cork, Ireland

*Corresponding authors
Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di
Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy. Tel: 39 02 6448 3418;
E-mail: paola.branduardi@unimib.it
E-mail: diletta.ami@unimib.it (for FTIR analysis)

Abstract

The ability of *Zygosacchromyces bailii* to grow at low pH and in the presence of considerable amounts of weak organic acids, lethal condition for *Saccharomyces cerevisiae*, increased the interest in the biotechnological potential of the yeast. To understand the mechanism of tolerance and growth effect of weak acids on *Z. bailii*, we evaluated the physiological and macromolecular changes of the yeast exposed to sub lethal concentrations of lactic acid. Lactic acid represents one of the important commodity chemical which can be produced by microbial fermentation. We assessed physiological effect of lactic acid by bioreactor fermentation using synthetic media at low pH in the presence of lactic acid. Samples collected from bioreactors were stained with propidium iodide (PI) which revealed that, despite lactic acid negatively influence the growth rate, the number of PI positive cells is similar to that of the control. Moreover, we have performed Fourier Transform Infra-Red (FTIR) microspectroscopy analysis on intact cells of the the same samples. This technique has been never applied before to study *Z. bailii* under this condition. The analyses revealed lactic acid induced macromolecular changes in the overall cellular protein secondary structures, and alterations of cell wall and membrane physico-chemical properties.

Keywords: *Zygosaccharomyces bailii*, lactic acid, FTIR, bioreactor, fermentation

Introduction

The yeast *Zygosaccharomyces bailii* is well-known to be responsible for major spoilage losses in the food and beverage industry (Fleet, 2007). This spoilage nature is explained by its remarkable ability to grow in harsh environments such as high osmotic pressure, low pH, low water activity, and high concentration of weak organic acids (Thomas and Davenport, 1985, James and Stradford, 2011). These same traits have drawn attention to *Z. bailii* as a potential cell factory for production of biomolecules (Branduardi *et al.*, 2004, Sauer *et al.*, 2004, Vigentini *et al.*, 2005). Although initial studies focused on understanding tolerance to weak organic acids with a view to controlling the yeast in food spoilage, more recent work aims to dissect the mechanisms of tolerance to acids, such as acetic and lactic acids, to facilitate exploitation for biotechnological applications. Lactic acid is of particular relevance as it represents an important commodity chemical that can be produced by microbial fermentation as reviewed by (Sauer *et al.*, 2010, Becker *et al.*, 2015).

Most of our understanding of the toxicity of weak organic acids to yeast cells, and the cellular response, comes from studies on the model yeast *Saccharomyces cerevisiae*. The lipophilic nature of most undissociated organic weak acids (Lambert and Stratford, 1999) allows them to diffuse across the semi-permeable cell membrane. The ratio of undissociated:dissociated acid, and hence degree of diffusion into the cell, increases with decrease of external pH. Once inside of the cell where the cytoplasmic pH is close to neutral, the weak acid dissociates to a proton and respective anion. The presence of proton acidifies the internal pH causing inhibition of most metabolic processes, and the released anions may have additional toxicity causing programmed cell death (Ullah *et al.*, 2012, Russell, 1992). Various multiple responses are detected in *S. cerevisiae* to tolerate weak acid toxicity. Most pronounced among them is activation of H^+-ATPases present in the plasma and vacuolar membranes to stabilize internal pH, cell wall and membrane remodeling to decrease fluidity, thus preventing/decreasing undissociated weak acid diffusion, alteration in central carbon metabolism to increase ATP pool, anion extrusion through MDR-MFS transporters. The mechanism of acetic acid toxicity, model organic acid for stress response studies, has been extensively studied in the yeast *S. cerevisiae* (Paiva *et al.*, 2004, Mira *et al.*, 2010, Nygard *et al.*, 2014). Remarkably, different weak acids may have different physiological and morphological responses (Stratford *et al.*, 2013), differing also among diverse yeasts.

There have been several studies in *Z. bailii*, focused on acetic acid tolerance due to its common use as a weak acid preservative and ubiquitous inhibitor in industrial fermentation (Lindberg *et al.*, 2013, Sousa *et al.*, 1998, Guerreiro *et al.*, 2012, Rodrigues *et al.*, 2012). The reports suggest diverse weak acid toxicity responses of *Z. bailii*, during acetic acid exposure. *Z. bailii* is able to modulate cell membrane to decrease weak acid diffusion and acid uptake through active transport, allowing dose dependent uptake of acetic acid, thus preventing oversaturation of acetic acid in the cell and its high toxicity (Sousa *et al.*, 1998, Lindberg *et al.*, 2013, Rodrigues *et al.*, 2012). In addition, *Z. bailii* can utilize acetic acid as an extra carbon source in the presence of glucose, suggesting that under certain conditions *Z. bailii* may even benefit from the presence of the acid. This simultaneous co-consumption of glucose and acetic acid is possible because *Z. bailii* acetyl-CoA synthetase is not subject to glucose repression (Guerreiro *et al.*, 2012, Rodrigues *et al.*, 2012).

Being the product of interest in industrial fermentation and food acidulant, lactic acid represents great interest. Lactic acid has generally a weak inhibitory effect in yeast, affecting internal cell pH, acidification, and ROS accumulation. However, low amount of lactic acid has beneficial effect due to its buffering capacity, but with increase of concentration the effect disappears (Dang *et al.*, 2009, Nugroho *et al.*, 2015).

In this study we investigated the physiological and macromolecular responses of *Z. bailii* exposed to lactic acid stress at low aeration and pH. Bioreactor fermentation profile, propidium iodide staining and Fourier Transform Infrared (FTIR) microspectroscopy analysis were chosen to study the growth performance and the cellular response of *Z. bailii* to lactic acid at various time points. The study revealed substantial phenotypic response of *Z. bailii* to lactic acid, in particular the study shows that exposure to lactic acid inhibits the growth without affecting cell viability and induced macromolecular changes, which are mostly augmented by time.

Materials and Methods

Cell cultivation

The *Z. bailii* strains ATCC36947, ATCC60483, ATCC8766 and ATCC for convenience named Zb1, Zb2, Zb3 and Zb4 were used. The *S. cerevisiae* laboratory CEN.PK113-7D (obtained from Kötter P, Institut fur Mikrobiologie der Johann Wolfgang Goethe Universitat, Frankfurt, Germany) and commercial alcohol yeast (Dry Ethanol Red®; Fermentis, Marcq-en-Baroeul, France) strains were used. The cells were stored at -80°C in YPD glycerol stock. For liquid

cultivation, the cells were pre-grown on YPD plates (20 gL^{-1} peptone, 10 gL^{-1} yeast extract, 20 gL^{-1} glucose, 20 gL^{-1} agar). Phosphate – citrate buffer at pH 3 was used for buffering the flask fermentation. All liquid cultures were grown on at 30°C on synthetic minimal Verduyn medium (Verduyn *et al.*, 1992). Pre-culture for microaerobic bioreactor batch cultivation was prepared by transferring one full loop of cells from YPD plate to 20 mL Verduyn medium in 125 mL flask. The flasks were incubated at 160 rpm at 30°C overnight. The culture from first flask pre-culture was re-inoculated into 50mL Verduyn medium in 250 mL flask at OD_{660nm} 1.5 and let it grow for 4-5 hours until OD_{660nm} 4. The inoculum for the bioreactor cultivation was harvested at 3000g for 3 min at 20°C and resuspended in 20 mL of sterile ddH_2O, and aseptically added to the bioreactor.

Microaerobic bioreactor cultivation

Bioreactor experiment were performed in 2L volume bioreactors (BIOSTAT B, Sartorius AG, Germany) with operative volume of 1.5L. Zb2 was cultivated in 2x Verduyn medium containing 40 gL^{-1} glucose with 40 gL^{-1} lactic acid or no lactic acid. The cells were grown to mid exponential phase and inoculated to the bioreactor, to final absorbance of OD_{660} 0.1. The temperature was maintained at 30°C, pH at 3 by addition of 4M NaOH and the stirrer speed was setup to 400 rpm. The inlet gas flow was adjusted by two mass flow controllers (Bronkhornst®High Tech- EL-FLOW®Select). The mass flow was setup to mixture N_2 and air with final concentration of inlet oxygen 5%. The mixture was sparged at 0.75 vvm. Antifoam (Antifoam 204, Sigma Aldrich) was used for foaming control. A minimum of 3 independent cultivations were performed per each condition (Table 1).

The concentration of produced CO_2 was monitored by on-line gas analyzer (Omnitec). The gas analyzer was always calibrated 24 h before starting the cultivation using synthetic air containing defined concentration of CO_2.

Samples (20 mL) were collected regularly from the bioreactor in vials; 1 mL was used for OD_{660nm} measurement, after appropriate dilution; 1 mL was centrifuged at 4°C, 14.000 rpm for 5 min and supernatants were collected and stored at −20°C for later determination of extracellular metabolites concentrations.

Table 1 Bioreactor settings used in this study. The settings were modified to closely resemble industrial organic acid batch fermentation, where low pH and aeration is essential.

	Control	Lactic Acid
Medium	Verdyun, 40 gL^{-1} Glucose	Verdyun, 40 gL^{-1} Glucose, 40 gL^{-1} Lactic acid
pH	3, controlled with NaOH	3, controlled with NaOH
Air flow rate, (vvm)	0,75	0,75
Stirring speed, (rpm)	400	400
Inlet oxygen, (%)	Set to 5%, not controlled	Set to 5%, not controlled
Initial OD_{660nm}	0,1	0,1
Temperature, (°C)	30	30

Dry cell weight

The dry weight of the cell mass was measured per each sampling by washing samples in ddH_2O and pelleting the cells. The cell pellets were dried in vacuum concentrator using default mode (Concentrator 5301, Eppendorf, Germany) before measuring.

Extracellular metabolite quantification

Residual glucose, ethanol and lactic acid were determined via high-performance liquid chromatography (HPLC, Model 1100, Agilent Technologies) using Aminex HPX-87H ion exchange column 300 mm ×7.8 mm (Bio-Rad) thermostated at 45°C. The mobile phase was 5 mM sulphuric acid with a flow of 0.5 $mlmin^{-1}$. Lactic acid was detected with an UV-detector at 210 nm. Glucose and ethanol were detected with a RI detector.

Propidium iodide staining and flow cytometry

For identification of dead/severely compromised cells, cells were washed three times (Tris-HCl 50 mM, $MgCl_2$ 15 mM, pH 7.7) and resuspended in propidium iodide (PI, Sigma-Aldrich CO., St. Louis, MO, USA) solution 0.23 mM, incubated on ice for 20 min. Positive and negative controls were also prepared. In particular, positive control was prepared by killing cells by incubating in 70% ice cold ethanol for 20 mins. Samples were then analyzed using a CYTOMICS FC 500 flow cytometer (Beckman Coulter, Fullerton, CA, USA) equipped with an Argon laser (excitation wavelength 488 nm, laser power 20 mW). The fluorescence emission was measured through a 670 nm long pass filter (FL4 parameter) for PI signal. The sample flow

rate during analysis did not exceed 600–700 cells/s. Threshold settings were adjusted so that the cell debris was excluded from the data acquisition; 25000 cells were measured for every sample. Data analysis was performed afterwards with Cyflogic 1.2.1 software (©Perttu Terho & ©CyFlo Ltd).

FTIR microspectroscopy analysis

The bioreactor cultivated intact cells of *Z. bailii* were collected at 18, 24 and 42 hours. The cells were washed three times in distilled water to eliminate medium contamination. Approximately 3 μl of the cell suspensions were then deposited onto an IR transparent BaF_2 support, and dried at room temperature for at least 30 minutes to eliminate the excess water.

FTIR absorption spectra were acquired in transmission mode, between 4000 and 700 cm^{-1}, by means of a Varian 610-IR infrared microscope coupled to the Varian 670-IR FTIR spectrometer (both from Varian Australia Pty Ltd), equipped with a mercury cadmium telluride (MCT) nitrogen-cooled detector. The variable microscope aperture was adjusted to ~ 100 μm × 100 μm. Measurements were performed at 2 cm^{-1} spectral resolution; 25 KHz scan speed, triangular apodization, and by the accumulation of 512 scan co-additions.

Second-derivatives spectra were obtained following the Savitsky-Golay method (third-grade polynomial, 9 smoothing points), after a binomial 13 smoothing points of the measured spectra (Susi and Byler, 1986), using the GRAMS/32 software (Galactic Industries Corporation, USA). To verify the reproducibility and reliability of the spectral results, three independent preparations were analyzed and for each preparation at least ten spectra for sample were measured.

Moreover, to better illustrate the discussed spectral variations, for each of the selected IR absorption bands we calculated the difference of the intensity (ΔI) between the lactic acid treated and untreated cells. To evaluate their statistical significance, we reported in Supplementary material, Figure S3, the average ΔIs from the performed independent experiments with their standard deviation.

Results and Discussion

Establishing the suitable lactic acid concentration to investigate the effects on *Z. bailii* cells

Four *Z. bailii* strains (see Materials and Methods) obtained from ATCC collection were tested by plate assay experiment using different combinations of weak acids at pH 3 on Verdyun minimal medium to identify the best strain to be further studied for lactic acid tolerance. In addition, laboratory and industrial *S. cerevisiae* strains were included for comparison. The cells were pre-grown on YPD media until exponential phase and drop plated at 10, 10^{-1}, 10^{-2}, 10^{-3} dilution. No differences in yeast growth were observed in control plates without weak acid addition (data not shown). When evaluated for lactic acid tolerance (Figure 1A and 1B), other than Zb4 all strains were only minimal affected by 40 gL^{-1} of lactic acid. At 80 gL^{-1} lactic acid only Zb2 and Zb3 showed growth, and even those strains were strongly inhibited by this condition. Next, the effects of acetic acid and a combination of acetic acid and lactic acid were assessed (Figure 1C and 1D). At 5 gL^{-1} of acetic acid, growth of *Z. bailii* Zb4 and the two *S. cerevisiae* strains was completely inhibited, whereas *Z. bailii* strains Zb1, Zb2 and Zb3 were unaffected. No synergistic or additive inhibitory effects of 40 gL^{-1} lactic acid and 3 gL^{-1} acetic acid on any of the strain were observed (Figure 1, D). Being strains of the same species, *Z. bailii* show inconsistent phenotype toward various stresses as shown in Fig 1. Moreover, genetic heterogeneity of *Z. bailii sensu lato* strains may contribute to the phenotypic differences observed early (Suh *et al.*, 2013). The full genome sequence of the strains used in this study, which may provide detailed information about the phenotype divergence of *Z. bailii* strains, is not yet available. Overall, these results highlight the superior weak acid tolerance relative to *S. cerevisiae* of *Z. bailii* strains to weak organic acids. Strain Zb2 showed high levels of tolerance and this particular strain was previously reported to be amenable to genetic manipulation (Dato *et al.*, 2010, Passolunghi *et al.*, 2010), Zb2 was therefore chosen for subsequent detailed tests on the effects of lactic acid. Preliminary shake flask fermentations of Zb2 using Verduyn medium at pH 3 with different concentrations of lactic acid were carried out to assess sensitivity to the stressor in liquid medium (Supplementary material, Figure S1). Growth was not significantly impaired at 40 gL^{-1} lactic acid but was strongly reduced at 60 or 80 gL^{-1} lactic acid. Interestingly, growth was stimulated by 20 gL^{-1} lactic acid. This may be attributed to a buffering effect of the weak organic acid, which may mitigate acidification of the growth medium caused by yeast growth. At higher concentrations, the toxic effect of the acid would dominate over this mild buffering. In addition, we performed propidium iodide staining analysis to evaluate the

percentage of damaged cells under these test conditions (Supplementary material, Figure S2). In comparison to the control condition, there was only little difference in the percentage of PI positive cells at 18 and 22 hours (5-15%) for cultures grown with 20 gL^{-1} and 40 gL^{-1} lactic acid, indicating little cellular damage, whereas cells treated with 60 gL^{-1} and 80 gL^{-1} lactic acid show a high percentage of PI positive cells, which is a clear evidence these or higher concentrations significantly affect the cell viability. Based on these data, 40 gL^{-1} lactic acid was selected as the optimum sub-lethal lactic acid concentration for further tests. It should be noted that as the pKa of lactic acid is 3.86, according to Henderson-Hasselbalch equation the total concentration of undissociated lactic acid in the medium was approximately 34.5 gL^{-1}.

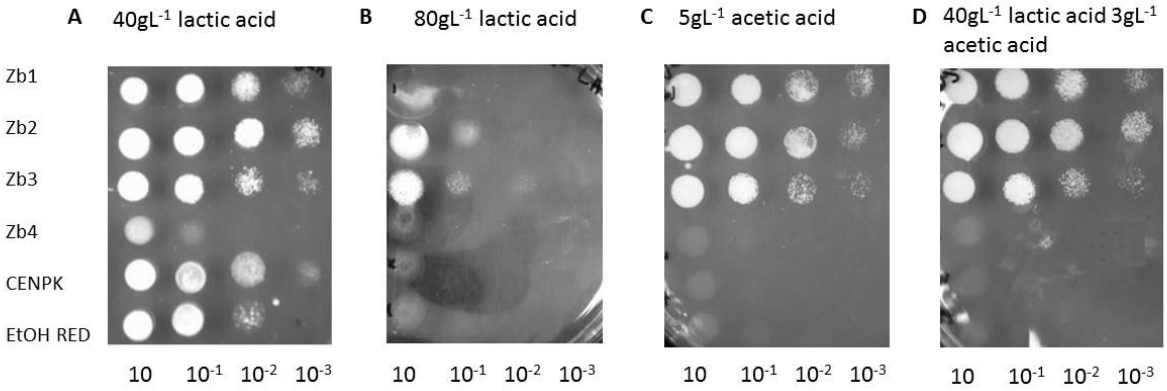


Figure 1 Spotting growth assay for weak acid tolerance screening of various *Z. bailii* and *S. cerevisiae* strains. Cells of the indicated *Z. bailii* and *S. cerevisiae* strains were cultivated until mid-exponential phase and spotted to Verduyn minimum medium plate (2% glucose) at pH 3, added with different concentration of lactic and acetic acid. Cells were 10-fold serially diluted and incubated at 30°C for 2 days.

Table 2 Physiological data obtained from microaerobic batch fermentation. Zb2 cells were cultured in minimal medium using bioreactors. The results were calculated from at least three biological replicates, and are given as the means with corresponding standard deviation.

	Control	Lactic Acid
Specific growth rate μ, h^{-1}	0,19 ±0.012	0,14 ±0.017
Biomass yield x/s (g g^{-1})	0,13 ±0,01	0,11 ±0,007
Ethanol yield EtOH/s (g g^{-1})	0,41 ±0,05	0,4 ±0,07
Specific glucose consumption rate (g g $DCW^{-1}h^{-1}$)	1,38 ±0.05	1,15 ±0.07

Effects of lactic acid on the growth and metabolic profiles of *Z. bailii* during microaerobic bioreactor fermentation

Bioreactor cultivation, in addition of allowing a precise monitoring and control of fermentation parameters, better represents industrial conditions. Therefore, experiments to assess the effect of 40 gL^{-1} lactic acid on strain Zb2 were performed in a 2L bioreactor (see Material and Methods for growth conditions). Cells were inoculated from overnight cultures at OD 0.1 and growth, as monitored by the production of CO_2, commenced after 5-8 hours of adaptation to the new environment (Figure 2). The similar lag phase, regardless of the presence/absence of lactic acid, suggests that a specific pre-adaption to lactic acid is not required. Nevertheless, it was noted that there was an effect on growth rate and yield (Figure 2), detailed with biomass and metabolite profiles in Figure 3. Cells grown in presence of lactic acid (40 gL^{-1}) exhibited a 25% reduction in growth rate (0.19h^{-1} ±0.012 vs 0.14h^{-1}±0.017) and a 15% reduction in final biomass titer (5.833gL^{-1}±0.16 vs 5.54gL^{-1}±0.21). Also the specific glucose consumption rate decreased of 13% reduction in lactic acid treated cells in comparison to control (1.38g $gDCW^{-1}h^{-1}$±0.05 vs 1.15g $gDCW^{-1}h^{-1}$±0.07). The reduction in growth rate and yield under bioreactor conditions is consistent with the described effects of other organic acids in yeast and in part may be attributable to the energetic cost of maintaining pH homeostasis by pumping H^+ ions from the cytoplasm using the plasma membrane ATPase, which requires energy for its activity in one proton per ATP rate (Stratford and Anslow, 1996, van der Rest *et al.*, 1995). However, we should not exclude other intracellular effects of lactic acid, which may contribute to overall growth inhibition. Indeed, the reduction in the glucose consumption and growth rate can be else

ascribed to a general conservative response, including a decrease in protein synthesis and inhibition of glycolytic enzymes activity in yeast (Pearce *et al.*, 2001). Despite the difference in biomass yield, the presence of lactic acid under our experimental setting did not influence ethanol fermentation. In other words, weak acid stressed cells prefer fermentative carbon utilization, which may be due to higher energy demand at O_2 limitation and/or lactate induced oxidative stress which affect mitochondria (Table 2) (Sousa *et al.*, 2012). After glucose depletion, the residual ethanol was slowly consumed, contributing to slight increase of biomass observed starting from 28 hours on, in both conditions.

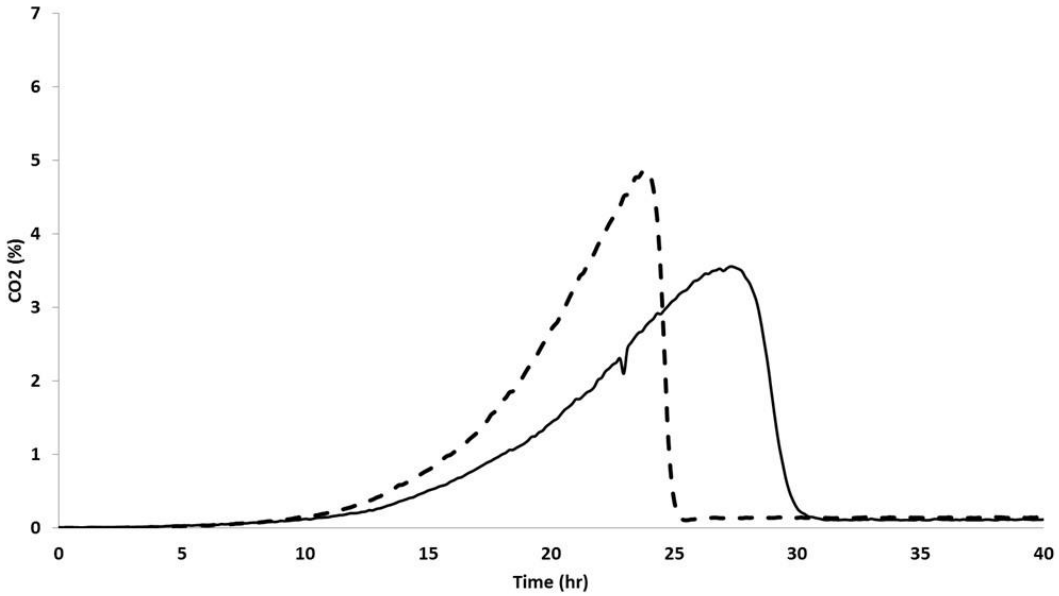


Figure 2 CO_2 profile of the *Z. bailii* bioreactor batch fermentation. Cells were cultivated in bioreactor. Gas samples were taken every 10 minutes. Dash line (control): Zb2 without lactic acid, solid line: Zb2 with 40 gL^{-1} lactic acid. Results are average values of three replicates.

S. cerevisiae is able to consume lactic acid through mitochondrial L-lactate ferricytochrome c oxidoreductase (L-LCR) and D-lactate ferricytochrome c oxidoreductase (D-LCR) activities (Lodi and Ferrero, 1993), encoded by genes that are targets for glucose repression and highly depended on aeration. Comparable studies are not present in literature for *Z. bailii*. In our experimental setting, only a slight decrease of lactic acid concentration was detected (Figure 3),

possibly because of lactic acid influx into the cell, even when glucose was exhausted in the medium. At high aeration condition in bioreactor fermentation, Zb2 consumed lactic acid after glucose depletion (data not shown). In addition, we observed increase of OD, during flask fermentation at 80 and 160rpm, which may be stimulated by consumption of 20 gL^{-1} and 40 gL^{-1} lactic acid after glucose depletion (Supplementary material, Figure S1). The consumption of lactic acid only under conditions of high aeration and glucose depletion suggests that comparable mechanisms are involved in *Z. bailii* and *S. cerevisiae*. During fermentation, control cultures used around 30 mL of NaOH base for pH maintenance, while cells grown with lactic acid barely used 1 mL NaOH. This confirms the indication deriving from the shake flask experiments (Supplementary material, Figure S1) of the lactic acid buffering effect. As with the shake flask experiment, growth is not promoted in media with 40 gL^{-1} lactic acid because of the more dominant inhibitory effects. Interestingly, no acetic acid accumulation was observed in both conditions at low aeration, which is very likely in agreement with previous reports describing *Z. bailii* for its ability to consume acetic acid even at presence of glucose and microaerobic condition (Sousa *et al.*, 1998, Rodrigues *et al.*, 2012).

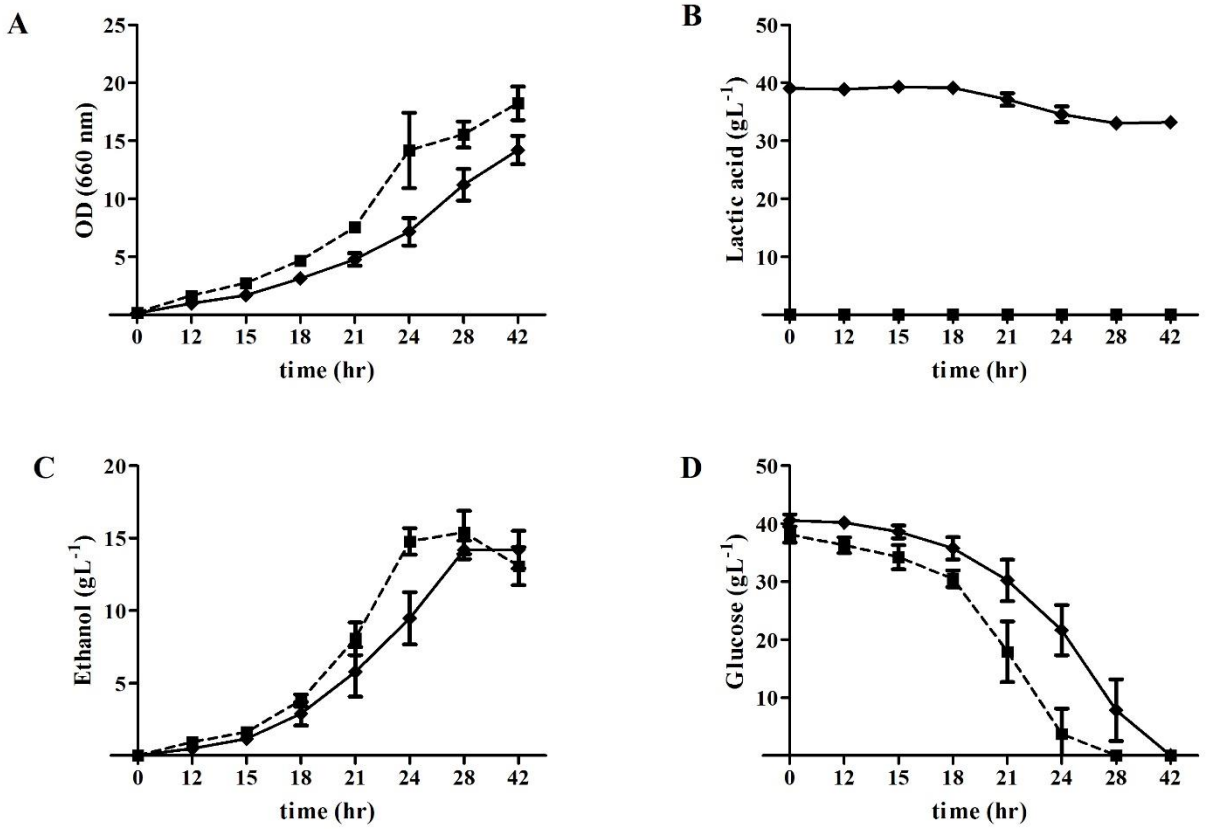


Figure 3 Fermentation performance of Zb2 with and without lactic acid. Cells were cultivated in bioreactor using Verduyn minimum medium (40 gL^{-1}glucose) under controlled condition (5% inlet oxygen, pH 3). Dash line (control): Zb2 without lactic acid. Solid line: Zb2 with 40 gL^{-1} lactic acid. **A.** Optical density at 660nm **B.** Lactic acid concentration **C.** Ethanol production rate **D.** Glucose consumption rate. Results are average values of three replicates. Error bars represent standard deviation from three independent fermentations.

The viability/integrity of cells recovered from bioreactors with and without lactic acid treatment was assessed using propidium iodide (PI) staining (Figure 4). There were no major differences in the percentage of PI positive cells between control and treatment conditions, confirming the data obtained in shake flasks (Supplementary material, Figure S1) and supporting the explanation that an increased energy burden rather than a toxic effect is responsible for the slight reduction in growth rate and yield seen in treated cultures (Figure 3). The minor differences in PI positivity measured at 12h were no

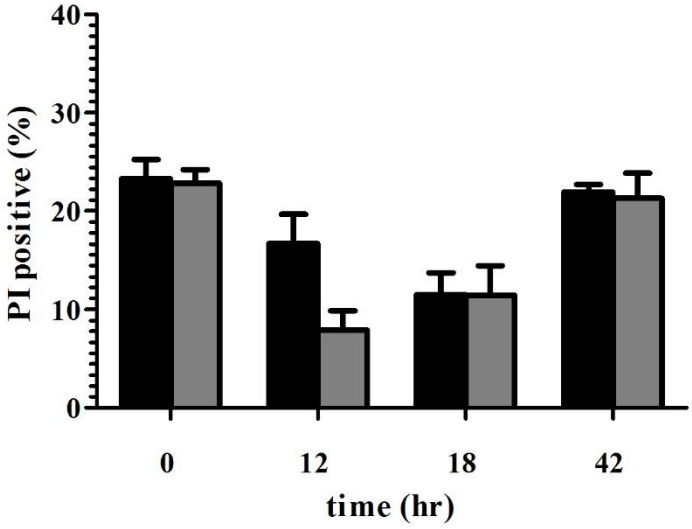


Figure 4 Propidium Iodide (PI) staining of Zb2 cells during course of bioreactor fermentation. Cells were cultivated in bioreactor. Samples were taken at 0, 12, 18 and 42 hours and stained with PI. Damaged/dead cells, positive for the staining, were detected using flow cytometry. The columns represent percentage of damaged/dead cells measured by fluorescence emission at 670nm (FL3). Black columns (control): Zb2 without lactic acid. Grey columns: Zb2 with 40 gL^{-1} lactic acid. Error bars represent standard deviation from at least three independent experiments.

longer apparent at 18 h. There was an increase in the percentage of damaged cells from ~12% to ~20% between 18 h and 42 h and this is likely to be due to glucose starvation and ethanol exposure.

FTIR microspectroscopy analysis of *Z. bailii* reveals gradual macromolecular changes during exposure to lactic acid

The results obtained from PI staining indicate that 40gL^{-1} lactic acid has no obvious cell damaging property on this strain, which may be a consequence of some cellular adaptation.

To investigate possible effects of lactic acid on the composition and structure of the main cell molecules, we used Fourier transform infrared (FTIR) microspectroscopy, a non-invasive and label free technique that enables to obtain a unique molecular fingerprint of the sample under investigation within a single experiment (Ami *et al.*, 2012). Intact *Z. bailii* cells were collected at 18, 24 and 42 hours after inoculation, respectively, corresponding to the early, late exponential, and the stationary phases of growth, and analyzed by FTIR microspectroscopy. As an example, in Figure 5 we reported the measured absorption spectrum of *Z. bailii* cells, grown in the absence of lactic acid, at 24 hours after the inoculation. As illustrated, the spectrum is due to the overlapping absorption of multiple components representing the different specific cellular macromolecules. Therefore, to better resolve the absorption bands, an essential prerequisite for the identification of peak positions and for their assignment to the different biomolecules, we analyzed the second derivative spectra (Susi and Byler, 1986). The amide I band (Figure 6a), between 1700 and 1600 cm^{-1}, principally gives information on the whole cell protein secondary structures and aggregation (Tamm and Tatulian, 1997, Barth, 2007). In early exponential phase, there are no significant differences observed between the treated and untreated cells (see also Supplementary material, Figure S3). The spectra are dominated by a band at ~1657 cm^{-1}, mainly due to alpha-helix and random-coil structures, and by a band at ~1638 cm^{-1}, due to intramolecular native beta-sheets. Moreover, two minor absorptions at ~1692 cm^{-1} and ~1685 cm^{-1} were present, respectively due to beta-sheet and beta-turn structures (Tamm and Tatulian, 1997, Barth, 2007). As the cells entered later exponential and stationary phase there were only minor changes to the whole cell protein structure seen in the untreated cells. Indeed, in the late exponential phase we observed only the appearance of two well resolved absorptions at ~1690 cm^{-1} and ~1680 cm^{-1}, respectively assigned to beta-sheets and beta-turns. More pronounced changes were instead evident in the treated cells. In late exponential phase, we detected in

particular a minor but significant reduction in the intensity of the alpha helix/random coil and of the native beta-sheet absorptions, accompanied by the appearance of a weak shoulder around 1627 cm^{-1}, mostly due to intermolecular beta-sheets, typical of protein aggregates (Seshadri *et al.*, 1999, Ami *et al.*, 2003). Moreover, the upshift of the ~1690 cm^{-1} beta-sheet absorption to ~1694 cm^{-1} again indicates the enrichment in proteins with intermolecular beta-sheets. These changes were even more evident in the stationary phase consistent with a progressive effect of the exposure to lactic (see also Supplementary material, Figure S3), and possibly ethanoloic stress on protein structure and folding.

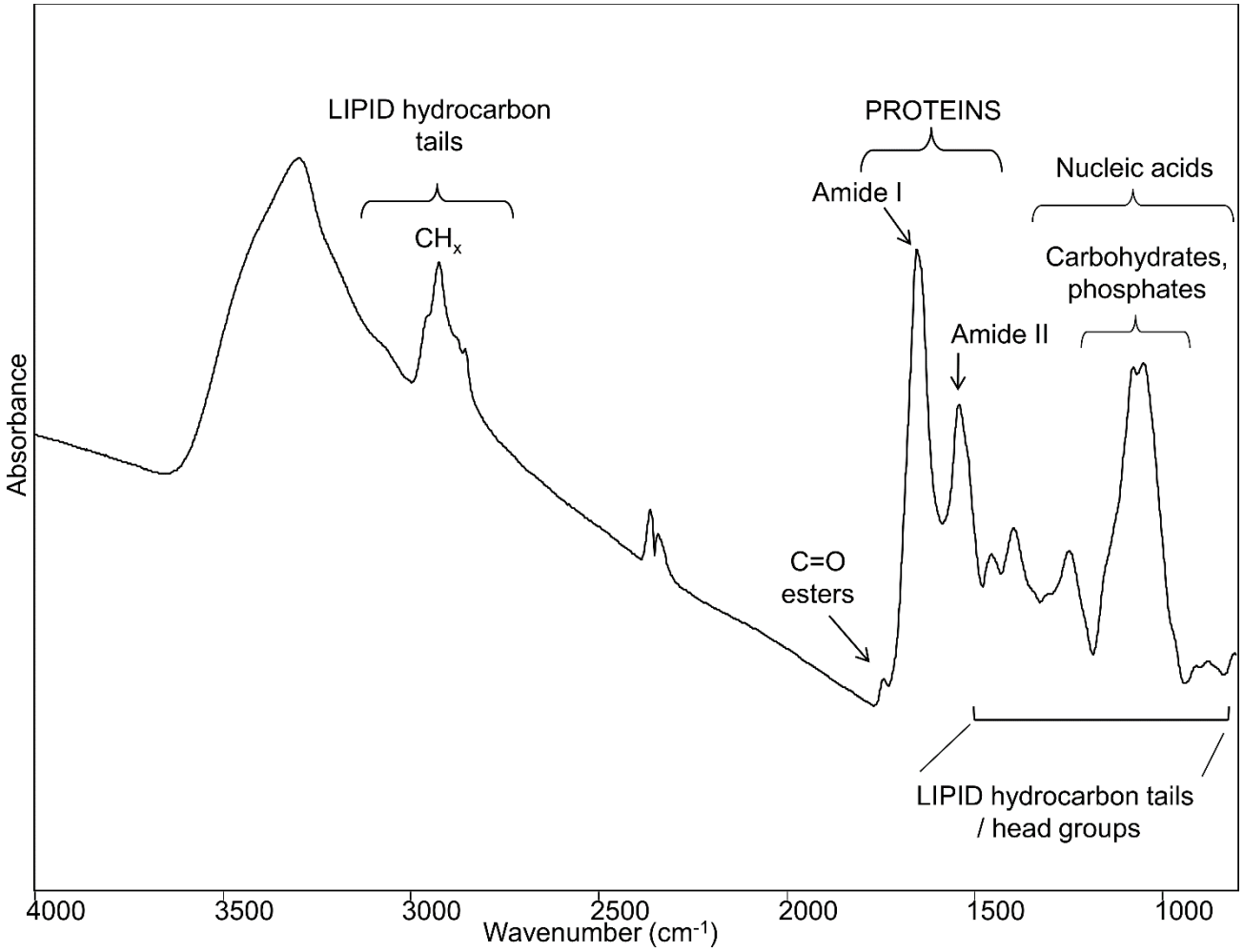


Figure 5 FTIR spectrum of *Z. bailii* intact cells. FTIR absorption spectrum of *Z. bailii* cells, grown in Verduyn minimum medium in the absence and in the presence of 40 g/L of lactic acid (LA). FTIR analysis was performed at 24 hours after the inoculation, corresponding to the late exponential phase of growth. The assignment of selected bands to the main biomolecules is reported.

Next we explored the IR response between 1500-1200 cm^{-1} (Figure 6b), mainly due to the deformation modes of the lipid hydrocarbon tails and head groups. In particular, the second derivative spectrum of the unchallenged cells is characterized by a component at ~ 1467 cm^{-1}, due to the overlapping absorption of CH_2 and CH_3; moreover, the bands at ~ 1455 cm^{-1}, 1438 cm^{-1} and 1368 cm^{-1} are due to CH_3, and the ~ 1416 cm^{-1} absorption to CH_2 (Casal and Mantsch, 1984, Arrondo and Goni, 1998, Natalello *et al.*, 2013). In addition, two spectral components were present at ~ 1397 cm^{-1} and ~ 1387 cm^{-1}, respectively mainly due to the CH_3 bending vibration of the $N(CH_3)_3$ head group of phosphatidylcholine (PC) and to the CH_3 deformation arising from ergosterol (Casal and Mantsch, 1984, Berterame *et al.*, 2016). Finally, a broad band at ~1248 cm^{-1} was also observed, due to the PO_2- stretching mode mainly of phospholipids and nucleic acids (Casal and Mantsch, 1984, Banyay *et al.*, 2003). There were no major alterations of the lipid components in lactic acid – treated cells in the early exponential phase, but in the later exponential and stationary phases lactic acid induced in particular significant reduction in the intensity of the ~1400 cm^{-1} band (Supplementary material, Figure S3), marker of PC (Berterame *et al.*, 2016). Notably, phosphatidylcholine - one of the most abundant membrane phospholipids - affects membrane fluidity (Nagle and Tristram-Nagle, 2000, Fajardo *et al.*, 2011). Therefore, the PC reduction observed in cells challenged with lactic acid might contribute to make the membrane more compact and, consequently, to counteract the lactic acid influx (Berterame *et al.*, 2016). Moreover, in later growth, the PC reduction was accompanied by a slight increase of the ergosterol component at ~ 1387 cm^{-1}, though this change was less evident in stationary phase.

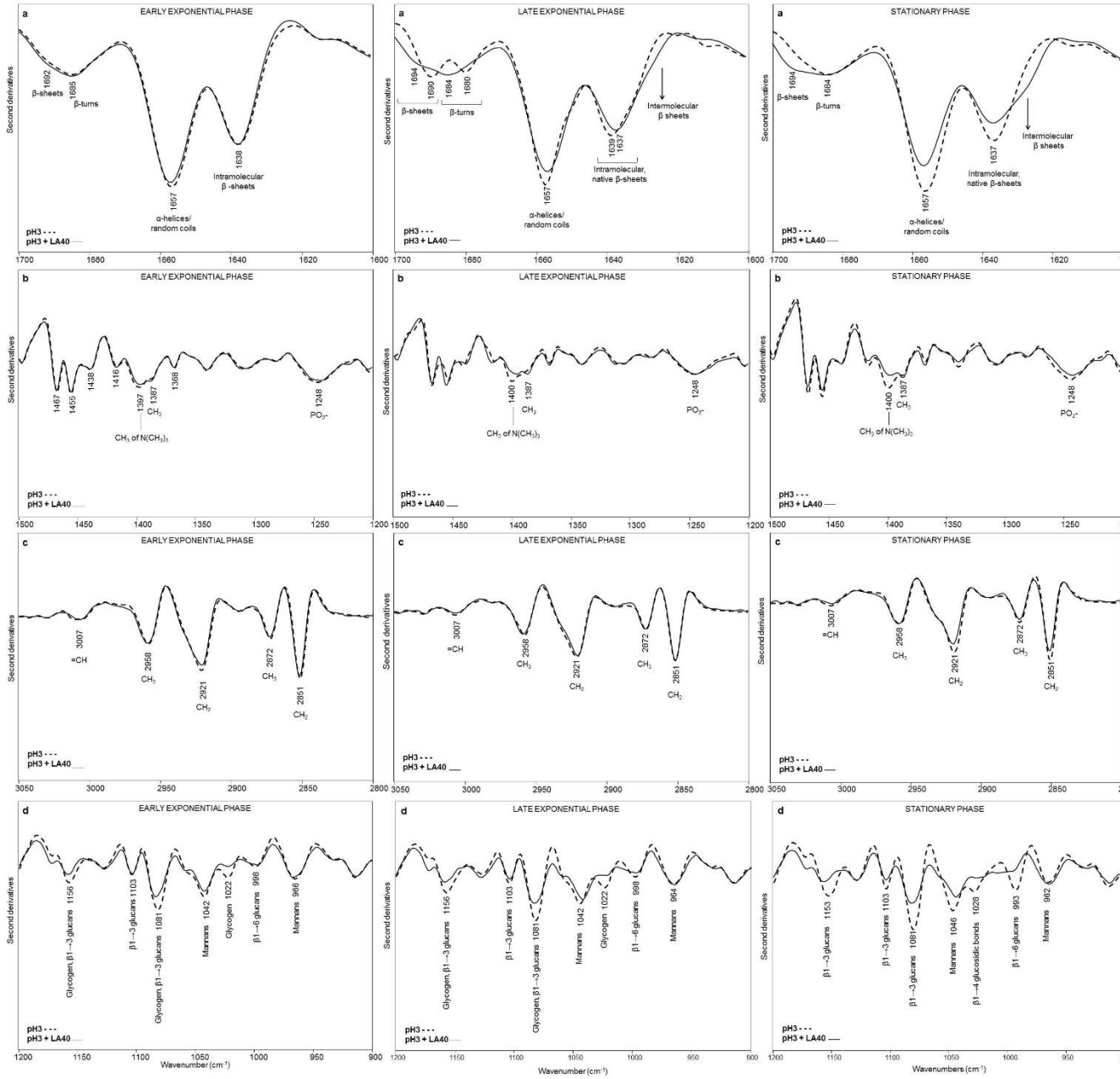


Figure 6 Second derivatives of the FTIR absorption spectra of *Z. bailii* cells, in the absence and in the presence of lactic acid. Cells were grown in Verduyn minimum medium in the absence (control) and in the presence of 40 g/L of lactic acid (LA). FTIR analysis was performed at 18 hours, 24 hours and 42 hours after the inoculation, corresponding to the i: early exponential phase; ii: mid exponential phase; iii: stationary phase of growth. **a**: amide I band; **b**: vibrational modes mainly due to lipid hydrocarbon tails and head groups, as well as to phosphate groups; **c**: stretching modes from lipid hydrocarbon tails; **d**: spectral range dominated by the absorption of the cell wall carbohydrates. In **a**, **b**, and **d** derivative spectra have been normalized to the tyrosine band at ~ 1516 cm^{-1}, while in **c** spectra have been normalized at the CH_3 band at ~ 2958 cm^{-1}.

Next, we analyzed the spectral range between 3050-2800 cm^{-1} (Figure 6c) that is mainly due to the stretching vibrations of the lipid hydrocarbon tails (Casal and Mantsch, 1984, Arrondo and Goni, 1998). In particular, the spectrum of cells grown in absence of lactic acid is characterized by four well resolved bands due to the CH_2 (at ~2921 cm^{-1} and 2851 cm^{-1}) and CH_3 (at ~2958 cm^{-1} and 2872 cm^{-1}) absorption. In addition, a low intensity band was detected at ~ 3007 cm^{-1}, due to the olefinic =CH groups in acyl chains (Casal and Mantsch, 1984). Lactic acid treated cells displayed very similar spectral features to those detected in not exposed cells, suggesting that lipid acyl chain length and or/saturation degree were not significantly affected by the exposure to the stressing agent in the exponential phase of growth. In stationary phase cells exposed to lactic acid displayed a slightly lower intensity of the lipid hydrocarbon tail CH_2 bands (~ 2921 cm^{-1} and ~2851 cm^{-1}) compared to cells grown in the absence of lactic acid, which could reflect a decrease of the acyl chain length ~~on~~ of lipids (Supplementary material, Figure S3) likely affecting membrane fluidity.

We finally analyzed the complex range between 1200-900 cm^{-1} (Figure 6d), dominated by the absorption of carbohydrates, with additional overlapping contributions of phosphate groups mainly from phospholipids and nucleic acids (Casal and Mantsch, 1984, Kacurakova and Mathlouthi, 1996). The analysis of this spectral range can provide information on cell wall properties that involve in particular the yeast envelope carbohydrate composition (Galichet *et al.*, 2001, Zimkus *et al.*, 2013).The second derivative spectrum of *Z. bailii* cells not exposed to lactic acid is characterized in particular by the simultaneous presence of three absorptions at ~1156 cm^{-1}, ~1081 cm^{-1} and ~1022 cm^{-1}, altogether marker of glycogen (Naumann, 2000). We should note that β(1-3) glucans, as well as absorbing at ~1103 cm^{-1}, can have overlapping contributions with glycogen at ~1156 cm^{-1} and at ~1081 cm^{-1} (Galichet *et al.*, 2001, Zimkus *et al.*, 2013). The two bands at ~1042 cm^{-1} and ~966 cm^{-1} are mainly assigned to mannans (Galichet *et al.*, 2001, Zimkus *et al.*, 2013). Even in early exponential phase, these spectral features were partly found to change in cells challenged with lactic acid (Supplementary material, Figure S3), and in particular an important reduction of glycogen occurred. This result suggests that lactic acid treated cells have faster glycogen turnover, possibly due to energy demand required to maintain cell homeostasis (Francois and Parrou, 2001). By late exponential phase, in lactic acid-treated cells there was a reduction in the intensity of the absorption mainly due to glycogen (~1156 cm^{-1}, 1081 cm^{-1}, 1022 cm^{-1}). In addition, a slight reduction of the β 1→3 glucan absorption (~1156 cm^{-1}, ~1103 cm^{-1}, ~1081 cm^{-1}) and of the β1→6 glucans at ~998 cm^{-1} was

also detected, suggesting that the exposure to the stressing agent induced a reorganization of the cell wall components. The changes were even more profound in the stationary phase (Supplementary material, Figure S3). Interestingly, glycogen was not detectable in either untreated or treated cells, but dramatic changes in the carbohydrate components of the cell wall were observed. Particularly, the ~ 998 cm^{-1} absorption due to β1→6 glucans downshifted to ~ 993 cm^{-1}, likely reflecting a modification of the carbohydrate interactions with the surrounding molecules of the cell wall. Furthermore, lactic acid-treated cells displayed a significant reduction in the intensity of the β1→3 and β1→6 glucan bands, and of the mannan band at 1046 cm^{-1} (Zimkus *et al.*, 2013) compared to cells in absence of lactic acid. The spectrum of cells grown in the absence of lactic acid was then characterized by a new band, not observed in the other phases of growth, at ~ 1028 cm^{-1} that can be assigned to β1→4 glucosidic bonds (Naumann, 2000). Interestingly, this component almost disappeared in cells challenged with lactic acid. These results indicate that lactic acid led to a dramatic rearrangement of the cell wall properties - mainly involving the carbohydrate components - that started in the early exponential phase and continued progressively to the stationary phase.

Significant modification in cell membrane and wall composition of *Z. bailii* induced by lactic acid inform us about adaptive response of the cells toward the stress. Similar studies were also implemented in *S. cerevisiae* using different approaches. The results, however, displaying similar pattern still showed diverse outcome. In particular, lipodomics analysis of *S. cerevisiae* showed the modulation of cell membrane composition toward acetic acid stress, yet the process and degree of modulation was different in *Z. bailii* (Lindberg *et al.*, 2013). The study of transcriptional changes in *S. cerevisiae* show a significant induction of *SED1* gene, responsible for cell wall architecture modulation, during lactic acid stress (Kawahata *et al.*, 2006). The reports suggest that SED1 encodes for cell wall protein which being overexpressed confers Zymulase resistance to the cells (Shimoi *et al.*, 1998), which might reflect that the gene involved either in cell wall repair or thickening. Recently, FTIR study of *S. cerevisiae* exposed to lactic acid demostrated a lactic acid induced cell wall modifications, specifically slight decrease in signals of glucans and mannans (Berterame *et al.*, 2016). In the case of current study, *Z. bailii* FTIR analysis clearly reveals a dramatic decrease in glucans and mannans. Being closely related to *S. cerevisiae, Z. bailii* growth niche increased and improved the weak acid stress resistance by enhancement not only catabolism of weak acid, but also the degree of cell and wall modifications which may play crucial role in extreme resistance of *Z. bailii*.

Summarizing the insights obtained from the FTIR analysis, the lactic acid treatment results in four major types of cellular changes. (1) An increase in the level of protein aggregation caused by lactic and ethanol stress, suggesting that it likely induces protein misfolding; (2) a reduction in glycogen, possibly caused by energy requirements for homeostasis; (3) modification of lipids, affecting in particular membrane fluidity; and (4) carbohydrate cell wall remodeling. Some of these alterations are consistent with other studies on organic/lactic acid stresses in *Z. bailii* or other yeasts. The reduction of PC and the simultaneous slight increase in ergosterol could account for a reduction of membrane fluidity (Fajardo *et al.*, 2011) that in turn could lead to the observed increased resistance to lactic acid. Lipidomic profiling of *Z. bailii* under acetic acid stress reported a high basal level of sphingolipids (Lindberg *et al.*, 2013). In our system, however, we didn't detect by FTIR analysis important variations of the sphingolipid content induced by lactic acid treatment. In addition, the slight but significant decrease detected in lipid acyl chain length in the stationary phase of growth could also contribute to lower membrane fluidity.

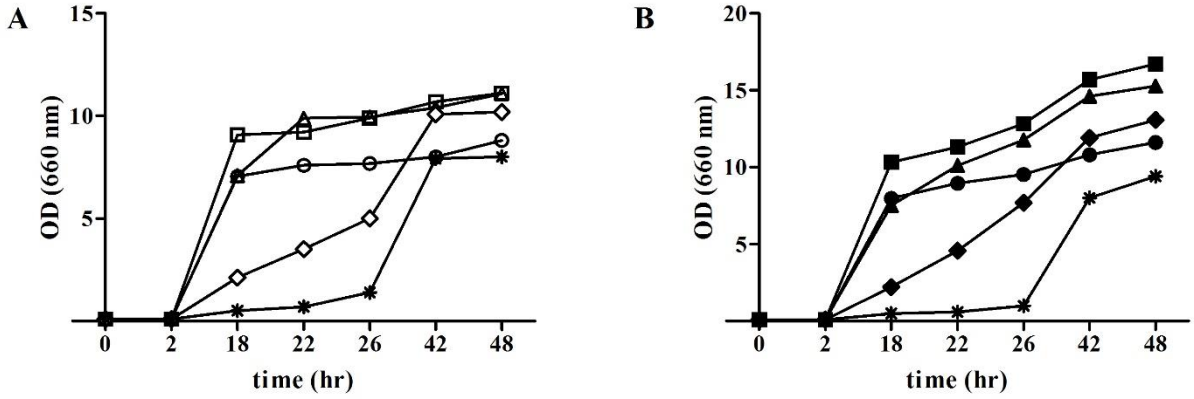
Interestingly, changes in cell wall properties induced by weak acids have previously been reported in *S. cerevisiae*, where exposure to weak acids induced the formation of a more rigid cell wall resistant to zymolyase digestion (Simoes *et al.*, 2006). Although the mechanism of response may be different in *Z. bailii*, the common response of cell wall modifications suggests that this is an important tolerance mechanism in yeasts.

In conclusion, the knowledge gathered during the study will help to better understand the weak acid tolerance of *Z. bailii* in the view of further ameliorating its biotechnological potential.

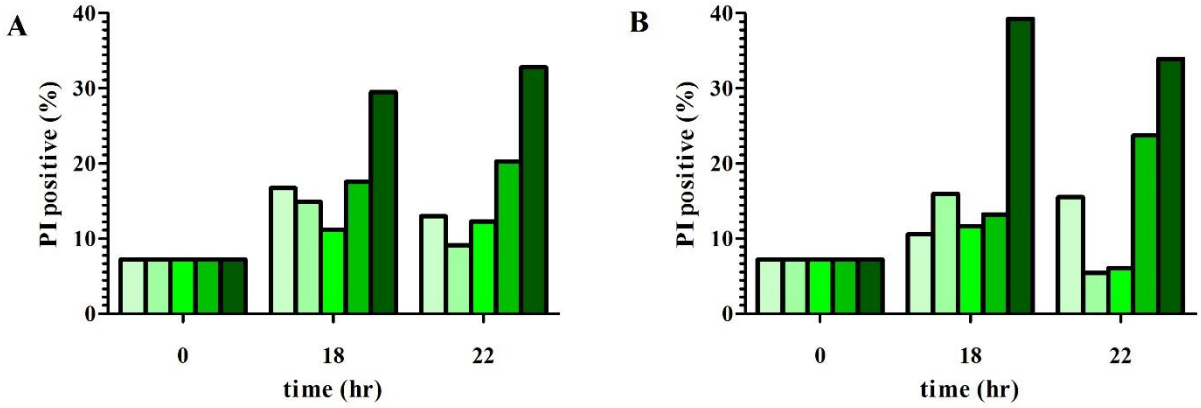
Acknowledgments

This work was supported by European Union FP7 Marie Curie Programme [YEASTCELL - 7PQ MARIE CURIE (12-4-2001100-40)] and partially by the SYSBIO - Centre of Systems Biology (SysBioNet, Italian Roadmap for ESFRI Research Infrastructure). D.A. acknowledges the University of Milano Bicocca (Fondo Grandi Apparecchiature) for the acquisition of the FTIR spectrometer Varian 670-IR. The authors are grateful to Prof. Silvia Maria Doglia (University of Milano-Bicocca) for helpful discussions.

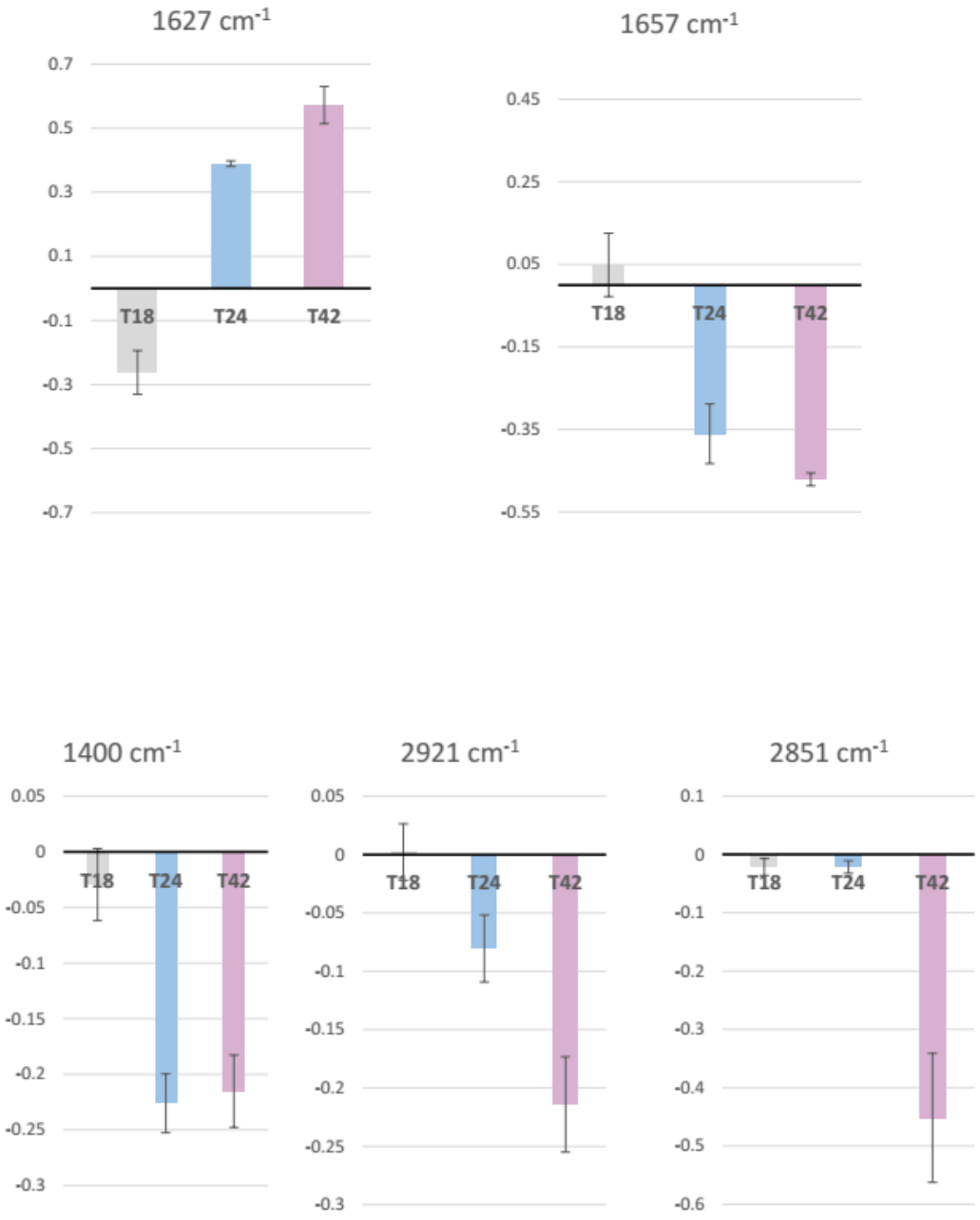
Supplementary materials



Supplementary Figure 1 (S1) Flask fermentations using different lactic acid concentration and rpm. Cells were shake-flasks cultivated at 30°C using Verduyn minimum medium (40 gL^{-1}glucose) with different lactic acid concentration, as indicated. **A.** *Z. bailii* flask fermentation under 80 rpm agitation. Open circle 0 gL^{-1} (control), open square 20 gL^{-1}, open triangle 40 gL^{-1}, open diamond 60 gL^{-1}, asterisk 80 gL^{-1}. **B.** *Z. bailii* flask fermentation under 160 rpm agitation. Solid circle 0 gL^{-1} (control), solid square 20 gL^{-1}, solid triangle 40 gL^{-1}, solid diamond 60 gL^{-1}, asterisk 80 gL^{-1}.



Supplementary Figure 2 (S2) Propidium Iodide (PI) staining of Zb2 cells during preliminary flask fermentation. Cells were cultivated in Verduyn minimal medium. Samples were taken at 0, 18 and 22 hours and stained with PI. Damaged/dead cells were detected using flow cytometry. The columns represent percentage of damaged/dead cells measured by fluorescence emission at 670nm (FL3). **A.** *Z. bailii* flask fermentation under 80 rpm of agitation. **B.** *Z. bailii* flask fermentation under 160 rpm of agitation. Different green-tone scale represents the increasing concentrations of lactic acid in the medium, gL^{-1} (0, 20, 40, 60, 80).



Supplementary Figure 3 (S3) Statistical analysis of FTIR data. For each of the selected IR absorption bands we calculated the difference of the intensity (ΔI) between the lactic acid treated and untreated cells. The average ΔIs from the performed independent experiments are reported with their standard deviation. For the β1→4 glucosidic band at 1028 cm^{-1}, the ΔI calculated for the T18 and T24 hours - where the 1028 cm^{-1} is absent - has to be ascribed to the contribution of the tail of the the glycogen band at 1022 cm^{-1}. Analogously, the ΔI calculated for the 1022 cm^{-1} band at T42. Intensities are taken from the second derivative spectra.

References

AMI, D, BONECCHI, L, CALI, S*, et al.* 2003. FT-IR study of heterologous protein expression in recombinant Escherichia coli strains. *Biochim Biophys Acta,* 1624**,** 6-10.

AMI, D, NATALELLO, A & DOGLIA, SM 2012. Fourier transform infrared microspectroscopy of complex biological systems: from intact cells to whole organisms. *Methods Mol Biol,* 895**,** 85-100.

ARRONDO, JL & GONI, FM 1998. Infrared studies of protein-induced perturbation of lipids in lipoproteins and membranes. *Chem Phys Lipids,* 96**,** 53-68.

BANYAY, M, SARKAR, M & GRASLUND, A 2003. A library of IR bands of nucleic acids in solution. *Biophys Chem,* 104**,** 477-488.

BARTH, A 2007. Infrared spectroscopy of proteins. *Biochim Biophys Acta,* 1767**,** 1073-1101.

BECKER, J, LANGE, A, FABARIUS, J & WITTMANN, C 2015. Top value platform chemicals: bio-based production of organic acids. *Curr Opin Biotechnol,* 36**,** 168-175.

BERTERAME, NM, 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,* 15**,** 39.

BRANDUARDI, P, VALLI, M, BRAMBILLA, L*, et al.* 2004. The yeast *Zygosaccharomyces bailii*: a new host for heterologous protein production, secretion and for metabolic engineering applications. *FEMS Yeast Res,* 4**,** 493-504.

CASAL, HL & MANTSCH, HH 1984. Polymorphic phase behaviour of phospholipid membranes studied by infrared spectroscopy. *Biochim Biophys Acta,* 779**,** 381-401.

DANG, TD, VERMEULEN, A, RAGAERT, P & DEVLIEGHERE, F 2009. A peculiar stimulatory effect of acetic and lactic acid on growth and fermentative metabolism of *Zygosaccharomyces bailii*. *Food Microbiol,* 26**,** 320-327.

DATO, L, BRANDUARDI, P, PASSOLUNGHI, S*, 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 Res,* 10**,** 894-908.

FAJARDO, VA, MCMEEKIN, L & LEBLANC, PJ 2011. Influence of phospholipid species on membrane fluidity: a meta-analysis for a novel phospholipid fluidity index. *J Membr Biol,* 244**,** 97-103.

FLEET, GH 2007. Yeasts in foods and beverages: impact on product quality and safety. *Curr Opin Biotechnol,* 18**,** 170-175.

FRANCOIS, J & PARROU, JL 2001. Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol Rev,* 25**,** 125-145.

GALICHET, A, SOCKALINGUM, GD, BELARBI, A & MANFAIT, M 2001. FTIR spectroscopic analysis of *Saccharomyces cerevisiae* cell walls: study of an anomalous strain exhibiting a pink-colored cell phenotype. *FEMS Microbiol Lett,* 197**,** 179-186.

GUERREIRO, JF, MIRA, NP & SA-CORREIA, I 2012. Adaptive response to acetic acid in the highly resistant yeast species *Zygosaccharomyces bailii* revealed by quantitative proteomics. *Proteomics,* 12**,** 2303-2318.

JAMES, SA & STRADFORD, M 2011. *The Yeast*, Elsevier.

KACURAKOVA, M & MATHLOUTHI, M 1996. FTIR and laser-Raman spectra of oligosaccharides in water: characterization of the glycosidic bond. *Carbohydr Res,* 284**,** 145-157.

KAWAHATA, M, MASAKI, K, FUJII, T & IEFUJI, H 2006. Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in *Saccharomyces cerevisiae* cultures induce expression of intracellular metal metabolism genes regulated by Aft1p. *FEMS Yeast Res,* 6**,** 924-936.

LAMBERT, RJ & STRATFORD, M 1999. Weak-acid preservatives: modelling microbial inhibition and response. *J Appl Microbiol,* 86**,** 157-164.

LINDBERG, L, SANTOS, AX, RIEZMAN, H*, et al.* 2013. Lipidomic profiling of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* reveals critical changes in lipid composition in response to acetic acid stress. *PLoS One,* 8**,** e73936.

LODI, T & FERRERO, I 1993. Isolation of the DLD gene of *Saccharomyces cerevisiae* encoding the mitochondrial enzyme D-lactate ferricytochrome c oxidoreductase. *Mol Gen Genet,* 238**,** 315-324.

MIRA, NP, TEIXEIRA, MC & SA-CORREIA, I 2010. Adaptive response and tolerance to weak acids in Saccharomyces cerevisiae: a genome-wide view. *OMICS,* 14**,** 525-540.

NAGLE, JF & TRISTRAM-NAGLE, S 2000. Structure of lipid bilayers. *Biochim Biophys Acta,* 1469**,** 159-195.

NATALELLO, A, SASSO, F & SECUNDO, F 2013. Enzymatic transesterification monitored by an easy-to-use Fourier transform infrared spectroscopy method. *Biotechnol J,* 8**,** 133-138.

NAUMANN, D 2000. Infrared spectroscopy in microbiology. *In:* MEYERS, RA (ed.) *Encyclopedia of Analytical Chemistry.* Chichester: John Wiley & Sons.

NUGROHO, RH, YOSHIKAWA, K & SHIMIZU, H 2015. Metabolomic analysis of acid stress response in Saccharomyces cerevisiae. *J Biosci Bioeng,* 120**,** 396-404.

NYGARD, Y, MOJZITA, D, TOIVARI, M*, et al.* 2014. The diverse role of Pdr12 in resistance to weak organic acids. *Yeast,* 31**,** 219-232.

PAIVA, S, DEVAUX, F, BARBOSA, S*, et al.* 2004. Ady2p is essential for the acetate permease activity in the yeast Saccharomyces cerevisiae. *Yeast,* 21**,** 201-210.

PASSOLUNGHI, S, RIBOLDI, L, DATO, L*, et al.* 2010. Cloning of the *Zygosaccharomyces bailii GAS1* homologue and effect of cell wall engineering on protein secretory phenotype. *Microb Cell Fact,* 9**,** 7.

PEARCE, AK, BOOTH, IR & BROWN, AJ 2001. Genetic manipulation of 6-phosphofructo-1-kinase and fructose 2,6-bisphosphate levels affects the extent to which benzoic acid inhibits the growth of Saccharomyces cerevisiae. *Microbiology,* 147**,** 403-410.

RODRIGUES, F, SOUSA, MJ, LUDOVICO, P*, et al.* 2012. The fate of acetic acid during glucose co-metabolism by the spoilage yeast *Zygosaccharomyces bailii*. *PLoS One,* 7**,** e52402.

RUSSELL, JB 1992. Another explanation for the toxicity of fermentation acids at low pH: anion accumulation versus uncoupling. *Journal of Applied Bacteriology,* 73**,** 363-370.

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,* 70**,** 6086-6091.

SAUER, M, PORRO, D, MATTANOVICH, D & BRANDUARDI, P 2010. 16 years research on lactic acid production with yeast - ready for the market? *Biotechnol Genet Eng Rev,* 27**,** 229-256.

SESHADRI, S, KHURANA, R & FINK, AL 1999. Fourier transform infrared spectroscopy in analysis of protein deposits. *Methods Enzymol,* 309**,** 559-576.

SHIMOI, H, KITAGAKI, H, OHMORI, H*, et al.* 1998. Sed1p is a major cell wall protein of *Saccharomyces cerevisiae* in the stationary phase and is involved in lytic enzyme resistance. *J Bacteriol,* 180**,** 3381-3387.

SIMOES, T, MIRA, NP, FERNANDES, AR & SA-CORREIA, I 2006. The *SPI1* gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. *Appl Environ Microbiol,* 72**,** 7168-7175.

SOUSA, MJ, LUDOVICO, P & RODRIGUES, F 2012. Stress and Cell Death in Yeast Induced by Acetic Acid. *In:* BUBULYA, P (ed.) *Cell Metabolism - Cell Homeostasis and Stress Response.* InTech.

SOUSA, MJ, RODRIGUES, F, CORTE-REAL, M & LEAO, C 1998. Mechanisms underlying the transport and intracellular metabolism of acetic acid in the presence of glucose in the yeast *Zygosaccharomyces bailii*. *Microbiology,* 144 (Pt 3)**,** 665-670.

STRATFORD, M & ANSLOW, PA 1996. Comparison of the inhibitory action on *Saccharomyces cerevisiae* of weak-acid preservatives, uncouplers, and medium-chain fatty acids. *FEMS Microbiol Lett,* 142**,** 53-58.

STRATFORD, M, STEELS, H, NEBE-VON-CARON, G*, et al.* 2013. Extreme resistance to weak-acid preservatives in the spoilage yeast *Zygosaccharomyces bailii*. *Int J Food Microbiol,* 166**,** 126-134.

SUH, SO, GUJJARI, P, BERES, C*, et al.* 2013. Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. *Int J Syst Evol Microbiol,* 63**,** 1922-1929.

SUSI, H & BYLER, DM 1986. Resolution-enhanced Fourier transform infrared spectroscopy of enzymes. *Methods Enzymol,* 130**,** 290-311.

TAMM, LK & TATULIAN, SA 1997. Infrared spectroscopy of proteins and peptides in lipid bilayers. *Q Rev Biophys,* 30**,** 365-429.

THOMAS, DS & DAVENPORT, RR 1985. *Zygosaccharomyces bailii* — a profile of characteristics and spoilage activities. *Food Microbiology,* 2**,** 157-169.

ULLAH, A, ORIJ, R, BRUL, S & SMITS, GJ 2012. Quantitative analysis of the modes of growth inhibition by weak organic acids in Saccharomyces cerevisiae. *Appl Environ Microbiol,* 78**,** 8377-8387.

VAN DER REST, ME, KAMMINGA, AH, NAKANO, A*, et al.* 1995. The plasma membrane of Saccharomyces cerevisiae: structure, function, and biogenesis. *Microbiol Rev,* 59**,** 304-322.

VERDUYN, C, POSTMA, E, SCHEFFERS, WA & VAN DIJKEN, JP 1992. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast,* 8**,** 501-517.

VIGENTINI, I, BRAMBILLA, L, BRANDUARDI, P*, et al.* 2005. Heterologous protein production in *Zygosaccharomyces bailii*: physiological effects and fermentative strategies. *FEMS Yeast Res,* 5**,** 647-652.

ZIMKUS, A, MISIŪNAS, A & CHAUSTOVA, L 2013. Li+ effect on the cell wall of the yeast *Saccharomyces cerevisiae* as probed by FT-IR spectroscopy. *Central European Journal of Biology,* 8**,** 724-729.

Chapter 3

Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch

Ortiz-Merino, R. A., Kuanyshev, N., Braun-Galleani, S., Byrne, K. P., Porro, D., Branduardi, P. and 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* 15, e2002128. DOI: 10.1371/journal.pbio.2002128

Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch

Raúl A. Ortiz-Merino[1], Nurzhan Kuanyshev[2], Stephanie Braun-Galleani[1], Kevin P. Byrne[1], Danilo Porro[2], Paola Branduardi[2], Kenneth H. Wolfe[1*]

[1]UCD Conway Institute, School of Medicine, University College Dublin, Dublin 4, Ireland, [2]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, Milano 20126, Italy

* kenneth.wolfe@ucd.ie

Short title: Mechanism of whole genome duplication in a fertile hybrid yeast

Abstract

Many interspecies hybrids have been discovered in yeasts, but most of these hybrids are asexual and can replicate only mitotically. Whole-genome duplication has been proposed as a mechanism by which interspecies hybrids can regain fertility, restoring their ability to perform meiosis and sporulate. Here, we show that this process occurred naturally during the evolution of *Zygosaccharomyces parabailii*, an interspecies hybrid that was formed by mating between two parents that differed by 7% in genome sequence and by many interchromosomal rearrangements. Surprisingly, *Z. parabailii* has a full sexual cycle and is genetically haploid. It goes through mating-type switching and auto-diploidization, followed by immediate sporulation. We identified the key evolutionary event that enabled *Z. parabailii* to regain fertility, which was breakage of one of the two homeologous copies of the mating-type (*MAT*) locus in the hybrid, resulting in a chromosomal rearrangement and irreparable damage to one *MAT* locus. This rearrangement was caused by HO endonuclease, which normally functions in mating-type switching. With one copy of *MAT* inactivated, the interspecies hybrid now behaves as a haploid. Our results provide the first demonstration that *MAT* locus damage is a naturally occurring evolutionary mechanism for whole-genome duplication and restoration of fertility to interspecies hybrids. The events that occurred in *Z. parabailii* strongly resemble those postulated to have occurred to cause ancient whole-genome duplication in an ancestor of *Saccharomyces cerevisiae*.

Author Summary

It has recently been proposed that the whole-genome duplication (WGD) event that occurred during evolution of an ancestor of the yeast *Saccharomyces cerevisiae* was the result of a hybridization between two parental yeast species that were significantly divergent in DNA sequence, followed by a doubling of the genome content to restore the hybrid's ability to make viable spores. However, the molecular details of how genome doubling could occur in a hybrid were unclear because most known interspecies hybrid yeasts have no sexual cycle. We show here that *Zygosaccharomyces parabailii* provides an almost exact precedent for the steps proposed to have occurred during the *S. cerevisiae* WGD. Two divergent haploid parental species, each with 8 chromosomes, mated to form a hybrid that was initially sterile but regained fertility when one copy of its mating-type locus became damaged by the mating-type switching apparatus. As a result of this damage, the *Z. parabailii* life cycle now consists of a 16-chromosome haploid phase and a transient 32-chromosome diploid phase. Each pair of homeologous genes behaves as two independent Mendelian loci during meiosis.

Introduction

A whole-genome duplication (WGD) occurred more than 100 million years ago in the common ancestor of six yeast genera in the ascomycete family Saccharomycetaceae, including *Saccharomyces* (Wolfe and Shields, 1997, Wolfe *et al.*, 2015). Recent phylogenomic analysis has shown that the WGD was an allopolyploidization – that is, a hybridization between two different parental lineages (Marcet-Houben and Gabaldon, 2015). One of these parental lineages was most closely related to a clade (ZT) containing *Zygosaccharomyces* and *Torulaspora*, whereas the other was closer to a clade (KLE) containing *Kluyveromyces*, *Lachancea* and *Eremothecium*. The ZT and KLE clades are the two major groups of non-WGD species in family Saccharomycetaceae. The WGD had a profound effect on the genome, proteome, physiology and cell biology of the yeasts that are descended from it, but the genomes of these yeasts have changed substantially in the time since the WGD occurred, with extensive chromosomal rearrangement, deletion of duplicate gene copies, and sequence divergence between ohnologs (pairs of paralogous genes produced by the WGD). These changes have made it difficult to ascertain the molecular details of how the WGD occurred. Ancient hybridizations are rare in fungi, or at least difficult to detect (Campbell *et al.*, 2016), but numerous relatively recent hybridizations have been identified using genomics, particularly in the ascomycete genera *Saccharomyces* (Hittinger, 2013, Wendland, 2014), *Zygosaccharomyces* (James *et al.*, 2005, Gordon and Wolfe, 2008, Solieri *et al.*, 2013b), *Candida* (Pryszcz *et al.*, 2014, Pryszcz *et al.*, 2015, Schroder *et al.*, 2016) and *Millerozyma* (Leh Louis *et al.*, 2012).

Marcet-Houben and Gabaldón (Marcet-Houben and Gabaldon, 2015) proposed two alternative hypotheses for the mechanism of interspecies hybridization that led to the ancient WGD in the *Saccharomyces* lineage. Hypothesis A was hybridization between diploid cells from the two parental species, perhaps by cell fusion. Hypothesis B was mating between haploid cells from the two parental species to produce an interspecies hybrid zygote, followed by genome doubling. Under both hypotheses, the product is a cell with two identical copies of each parental chromosome. These identical copies should be able to pair during meiosis, leading to viable spores. While there are no known examples of natural yeast hybrid species formed by diploid-diploid fusion (hypothesis A), three examples have been discovered where hybrid species were apparently formed simply by mating between haploids of opposite mating types from different species (hypothesis B). These are *Candida metapsilosis* (Pryszcz *et al.*, 2015), *C. orthopsilosis*

(Pryszcz *et al.*, 2014, Schroder *et al.*, 2016), and *Zygosaccharomyces* strain ATCC42981 (Gordon and Wolfe, 2008, Bizzarri *et al.*, 2016). These interspecies hybridizations occurred by mating between parents with 4-15% nucleotide sequence divergence between their genomes. However, none of these three hybrids can sporulate, which could be either because the homeologous chromosomes from the two parents are too divergent in sequence to pair up during meiosis, or because pairing occurs but evolutionary rearrangements (such as translocations) between the parental karyotypes result in DNA duplications or deficiencies after meiosis (Hunter *et al.*, 1996, Delneri *et al.*, 2003, Liti *et al.*, 2006, Morales and Dujon, 2012). None of these three hybrids has undergone the genome doubling step envisaged in hypothesis B.

Several groups (Marcet-Houben and Gabaldon, 2015, Scannell *et al.*, 2006, Wolfe, 2015, Morales and Dujon, 2012) have proposed that genome doubling could occur quite simply by means of damage to one copy of the *MAT* locus in the interspecies hybrid, which could cause the hybrid cell to behave as a haploid, switch mating type, and hence auto-diploidize. This proposal mimics laboratory experiments carried out by Greig *et al*. (2002) in which hybrids between different species of *Saccharomyces* were constructed by mating. The hybrids were unable to segregate chromosomes properly and were sterile, but when one allele of the *MAT* locus was deleted they spontaneously auto-diploidized by mating-type switching and were then able to complete meiosis and produce spores with high viability. Each spore contained a full set of chromosomes from both parental species (2002). While genome doubling via *MAT* locus damage is an attractive hypothesis consistent with hypothesis B above (Marcet-Houben and Gabaldon, 2015), no examples of it occurring in nature have been described. We show here that *Z. parabailii* has gone through this process.

There are 12 formally described species in the genus *Zygosaccharomyces* (Hulin and Wheals, 2014). The most studied of these is *Z. rouxii*, originally found in soy sauce and miso paste (Ohnishi, 1963, Mori and Windisch, 1982). Others include *Z. mellis* frequently found in honey (James and Stratford, 2011), and *Z. sapae* from balsamic vinegar (Solieri *et al.*, 2013a, Solieri *et al.*, 2014). Species in the *Z. bailii sensu lato* clade (*Z. bailii*, *Z. parabailii* and *Z. pseudobailii*; (Suh *et al.*, 2013)) are of economic importance because they are exceptionally resistant to osmotic stress and low pH. Their resistance to the weak organic acids commonly used as food preservatives makes them the most frequent spoilage agent of packaged foods with high sugar

content such as fruit juices and jams, or low pH such as mayonnaise (Thomas and Davenport, 1985, Deak and Beuchat, 1996, Mollapour and Piper, 2001, Stratford *et al.*, 2013, Palma *et al.*, 2015). These same characteristics make *Zygosaccharomyces* relevant to biotechnology since high stress tolerance and rapid growth are often desirable traits in microorganisms to be used as cell factories. The strain we analyze here, *Z. parabailii* ATCC60483, has previously been used for production of vitamin C (Sauer *et al.*, 2004), lactic acid (Dato *et al.*, 2010), and heterologous proteins (Vigentini *et al.*, 2005).

Despite the diversity of the genus, genome sequences have been published for only two non-hybrid species of *Zygosaccharomyces*: the type strains of *Z. rouxii* (CBS732^{T}; (Souciet *et al.*, 2009)) and *Z. bailii* (CLIB213^{T}; (Galeote *et al.*, 2013)). The genus also includes many interspecies hybrids with approximately twice the DNA content of pure species (20 Mb instead of 10 Mb; (James *et al.*, 2005, Gordon and Wolfe, 2008, Mira *et al.*, 2014, Bizzarri *et al.*, 2016)). Mira *et al.* (Mira *et al.*, 2014) sequenced the genome of *Zygosaccharomyces* strain ISA1307 and found that it is a hybrid between *Z. bailii* and an unidentified *Zygosaccharomyces* species. In 2013, Suh *et al*. (Suh *et al.*, 2013) proposed that some strains that were historically classified as *Z. bailii* should be reclassified as two new species, *Z. parabailii* and *Z. pseudobailii,* based on phylogenetic analysis of a small number of genes. The sequences of the *RPB1* and *RPB2* genes that they obtained from *Z. parabailii* and *Z. pseudobailii* contained multiple ambiguous bases, consistent with a hybrid nature (Mira *et al.*, 2014). In the current study, we sequenced the genome of a second hybrid strain, ATCC60483. We show that ATCC60483 and ISA1307 are both *Z. parabailii* and are both descended from the same interspecies hybridization event. By sequencing ATCC60483 using Pacific Biosciences (PacBio) technology we obtained near-complete sequences of every *Z. parabailii* chromosome, which enabled us to study aspects of chromosome evolution in this species that were not evident from the Illumina assembly of ISA1307 (Mira *et al.*, 2014).

Results

Z. parabailii ATCC60483 genome assembly by PacBio

We first tried to sequence the *Z. parabailii* genome using Illumina technology, but even with high coverage we were unable to obtain long contigs. The data indicated that the genome was a hybrid, so instead we switched to PacBio technology which generates long sequence reads (6 kb on average in our data). Our initial assembly had 22 nuclear scaffolds, which we refined into 16 complete chromosome sequences with a cumulative size of 20.8 Mb by manually identifying overlaps between the ends of scaffolds, and by tracking centromere and telomere locations. We annotated genes using the YGAP pipeline, assisted by RNAseq data to identify introns. The nuclear genome has 10,087 protein-coding genes, almost twice as many as *Z. bailii* CLIB213^T (**Table 1**).

Table 1. Comparison of *Z. bailii* and *Z. parabailii* genome assemblies.

Strain	Species	Genome size (Mb)	Scaffolds	Scaffold N50 (Mb)	Reference	tRNA genes[a]	Protein-coding genes
CLIB213^T	*Z. bailii*	10.2	27	0.9	Galeote *et al.* (2013)	161	5084
ISA1307	*Z. parabailii*	21.2	154	0.2	Mira *et al.* (2014)	513	9925
ATCC60483	*Z. parabailii*	20.8	16	1.3	This study	499	10087

[a] We predicted the tRNA gene content of each genome assembly using tRNAscan-SE (Lowe and Eddy, 1997).

Most of the chromosome sequences extend into telomeric repeats at the ends. The consensus sequence of the telomeres is TGTGGGTGGGG, which matches exactly the sequence of the template region of the two homeologous *TLC1* genes for the RNA component of telomerase that are present in the genome. Chromosome sequences that do not extend into telomeres instead terminate at gene families that are amplified in subtelomeric regions, or contain genes that are at chromosome ends in the inferred Ancestral (pre-WGD) gene order for yeasts (Gordon *et al.*,

2009) indicating that they are almost full-length, except for three chromosome ends that appear to have undergone break-induced replication (BIR) and homogenization with other chromosome ends.

We identified one scaffold as the mitochondrial genome, which maps as a 30 kb circle containing orthologs of all *S. cerevisiae* mitochondrial genes. We also found a plasmid in the 2-micron family (5427 bp), with 99% sequence identity to pSB2 which was first isolated (1987) from the type strain of *Z. parabailii* (NBRC1047 / ATCC56075).

Z. parabailii ATCC60483 is an interspecies hybrid, with *Z. bailii* as one parent

Visualization of the genome using a Circos plot (Zhang *et al.*, 2013) shows that most of the genome is duplicated, indicating a polyploid origin (**Fig. 1**). However, although most genes have a homeolog, the chromosomes do not form simple collinear pairs. Instead, sections of each chromosome are collinear with sections of other chromosomes.

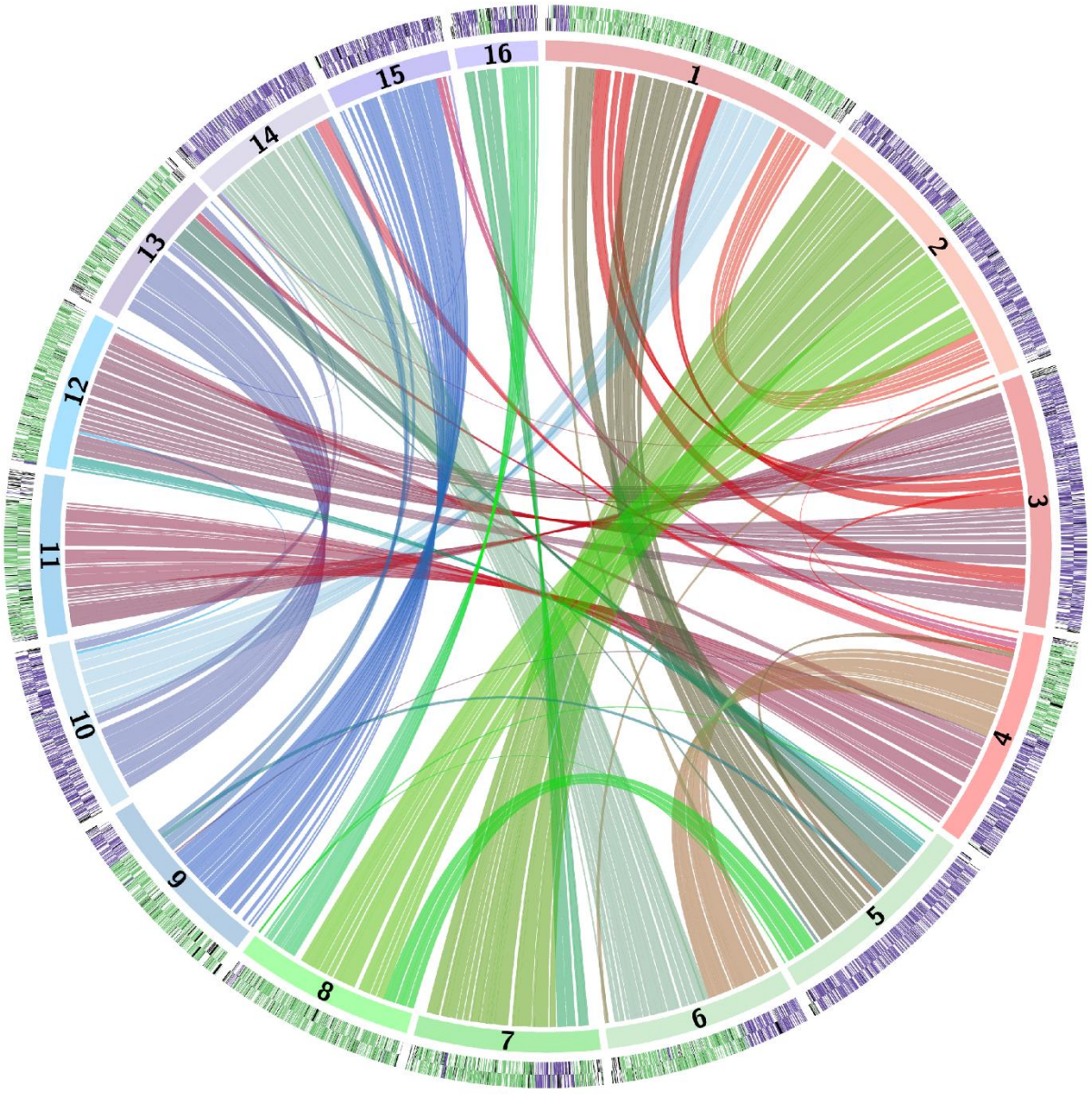


Figure 1 Circos plot of relationships among the *Z. parabailii* ATCC60483 chromosomes. In the outer arcs, purple and green coloring indicates A- and B-genes on the Watson and Crick strands of each chromosome. Arcs in the center of the diagram link homeologous (A:B) gene pairs.

Comparison to *Z. bailii* CLIB213^T shows that, for each region of the *Z. bailii* genome there are two corresponding regions of the *Z. parabailii* genome: one almost identical in sequence, and one with approximately 93% sequence identity, which demonstrates a hybrid (allopolyploid) origin of *Z. parabailii* and suggests that *Z. bailii* was one of its parents. To analyze this relationship in detail, we estimated the parental origin of every *Z. parabailii* ATCC60483 gene based on the number of synonymous substitutions per synonymous site (K_S) when compared to its closest *Z. bailii* homolog (**Fig. 2A**). This analysis revealed a bimodal distribution of K_S values where 47.1% of the ATCC60483 genes are almost identical to CLIB213^T genes ($K_S \leq 0.05$), and a further 42.5% are more divergent ($0.05 < K_S \leq 0.25$).

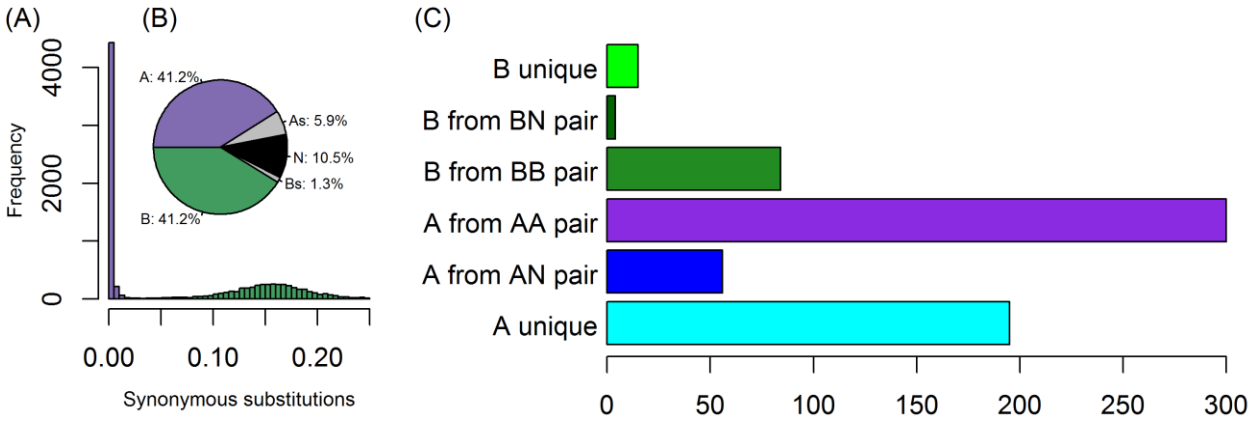


Figure 2 (A) Histogram of the distribution of synonymous site divergence (KS) values for 10,087 *Z. parabailii* ATCC60483 genes compared to their closest *Z. bailii* CLIB213T homologs. (B) Pie chart showing the proportions of genes classified into each category. The two largest categories refer to A-genes and B-genes that are in A:B pairs. N means genes for which no *Z. bailii* homolog was found or KS to Z. bailii exceeded 0.25. 'As' and 'Bs' indicate other A-genes and B-genes, as analyzed in panel C. (C) Breakdown of the numbers of genes assigned to the A- or B-subgenomes, that are not in A:B pairs. See S1 Data for category counts and KS values for each gene.

From this relationship, we infer that *Z. parabailii* ATCC60483 is an interspecies hybrid formed by a fusion of two parental cells, which we refer to as Parent A (purple) and Parent B (green). Parent A was a cell with a genome essentially identical to *Z. bailii* CLIB213^T. Parent B was a cell of an unidentified *Zygosaccharomyces* species with approximately 93% overall genome sequence identity to *Z. bailii*, corresponding to a synonymous site divergence peak of $K_S = 0.16$

(Fig. 2A). We refer to the two sets of DNA in *Z. parabailii* that were derived from Parents A and B as the A-subgenome and the B-subgenome respectively. We refer to the A- and B-copies of a gene as homeologs, and we use a suffix ('_A' or '_B') in gene names to indicate which subgenome they come from.

The genome contains ribosomal DNA (rDNA) loci inherited from each of its parents. Our assembly includes two complete rDNA units with 26S, 5.8S, 18S and 5S genes. Phylogenetic analysis of their internal transcribed spacer (ITS) sequences shows that the rDNA on chromosome 11 is derived from *Z. bailii* (Parent A), whereas the rDNA on chromosome 4 is derived from Parent B and contains an ITS variant seen only in other *Z. parabailii* strains (**S1 Figure**). A third rDNA locus in our assembly (at one telomere of chromosome 15) is incomplete and does not extend into the ITS region. The rDNA unit on chromosome 4 is also telomeric, whereas the unit on chromosome 11 is located at an internal site 165 kb from the right end. None of the genes in the interval between this rDNA and the right telomere of chromosome 11 have orthologs in *Z. bailii* CLIB213^{T}.

Z. parabailii has 16 chromosomes. We identified its 16 centromeres bioinformatically, which correspond to two copies (A and B) of each of the 8 centromeres in the Ancestral pre-WGD yeast genome (**Table 2**) (Gordon *et al.*, 2009, Gordon *et al.*, 2011b). In contrast, *Z. rouxii* has only 7 chromosomes due to a telomere-to-telomere fusion between two chromosomes followed by loss of a centromere (Gordon *et al.*, 2011b). The missing centromere in *Z. rouxii* is Ancestral centromere *Anc_CEN2*, which maps to *Z. parabailii* centromeres *CEN4* and *CEN11*, located between the genes *MET14* and *VPS1*. The *Z. rouxii* centromere must have been lost after it diverged from the *Z. bailii*/*Z. parabailii* lineage. Alignment of the *Z. rouxii MET14-VPS1* intergenic region with the *Z. parabailii CEN4* and *CEN11* regions shows that the CDE III motif of the point centromere has been deleted in *Z. rouxii* (**S2 Figure**).

Table 2 *Z. parabailii* ATCC60483 chromosomes and centromeres.

	bp	Protein-coding genes	tRNA genes	Ancestral centromere[a]	*Z. rouxii* centromere
1	2,110,500	1010	62	Anc_CEN5 (B)	Zr_CEN2
2	2,005,801	1009	55	Anc_CEN6 (A)	Zr_CEN3
3	1,751,495	868	31	Anc_CEN4 (A)	Zr_CEN7
4	1,516,135	718	29	Anc_CEN2 (A)	absent[b]
5	1,443,312	709	44	Anc_CEN5 (A)	Zr_CEN2
6	1,315,104	614	22	Anc_CEN7 (B)	Zr_CEN4
7	1,283,838	638	29	Anc_CEN6 (B)	Zr_CEN3
8	1,249,162	634	28	Anc_CEN8 (B)	Zr_CEN6
9	1,240,939	579	43	Anc_CEN1 (B)	Zr_CEN5
10	1,189,704	576	28	Anc_CEN3 (A)	Zr_CEN1
11	1,115,933	522	16	Anc_CEN2 (B)	absent[b]
12	1,091,360	520	24	Anc_CEN4 (B)	Zr_CEN7
13	1,077,716	517	25	Anc_CEN3 (B)	Zr_CEN1
14	1,007,293	494	16	Anc_CEN7 (A)	Zr_CEN4
15	858,772	406	37	Anc_CEN1 (A)	Zr_CEN5
16	571,967	273	10	Anc_CEN8 (A)	Zr_CEN6
mtDNA	29,945	13	20		
Total (nuclear)	20,829,031	10087	499		

[a] Synteny correspondence between *Z. parabailii* centromeres and yeast Ancestral (pre-WGD) centromere locations (Gordon *et al.*, 2011b). A and B indicate the subgenome assignments of the *Z. parabailii* centromeres.

[b] *Z. rouxii* lost Anc_CEN2 in an evolutionary fusion of two chromosomes (Gordon *et al.*, 2011b).

Z. parabailii inherited the mitochondrial genome of its *Z. bailii* parent. A complete mitochondrial genome sequence for *Z. bailii* is not available, but we identified 55 small mtDNA contigs in the CLIB213^T assembly, which together account for most of the genome, and calculated an average of 96% sequence identity between these and ATCC60483 mtDNA. CLIB213^T lacks two of the five mitochondrial introns that are present in ATCC60483: the omega intron of the large subunit mitochondrial rDNA, and intron 2 of *COX1*. Intra-species polymorphism for intron presence/absence, and comparable levels of intra-species mtDNA sequence diversity, have been reported in other yeast species (Wolters *et al.*, 2015, Wu *et al.*, 2015).

Pre-hybridization chromosomal rearrangements in *Z. parabailii*'s parents relative to *Z. bailii*

When genes in the Circos plot are colored according to their parent of origin, it is striking that many *Z. parabailii* chromosomes are either almost completely 'A' (purple) or almost completely 'B' (green) (outer ring in Fig 1), even though the chromosomes do not form collinear pairs. This pattern can be seen in more detail in a dot matrix plot between *Z. bailii* and *Z. parabailii* **(Fig 3).** From this plot, it is evident that most of the A-subgenome is collinear with *Z. bailii* scaffolds, whereas the B-subgenome contains many rearrangements relative to *Z. bailii*. For example, *Z. parabailii* chromosome 1 is derived almost entirely from the B-subgenome but maps to about 12 different regions on the *Z. bailii* scaffolds. In contrast, *Z. parabailii* chromosome 3 is derived from the A-subgenome and is collinear with a single *Z. bailii* scaffold.

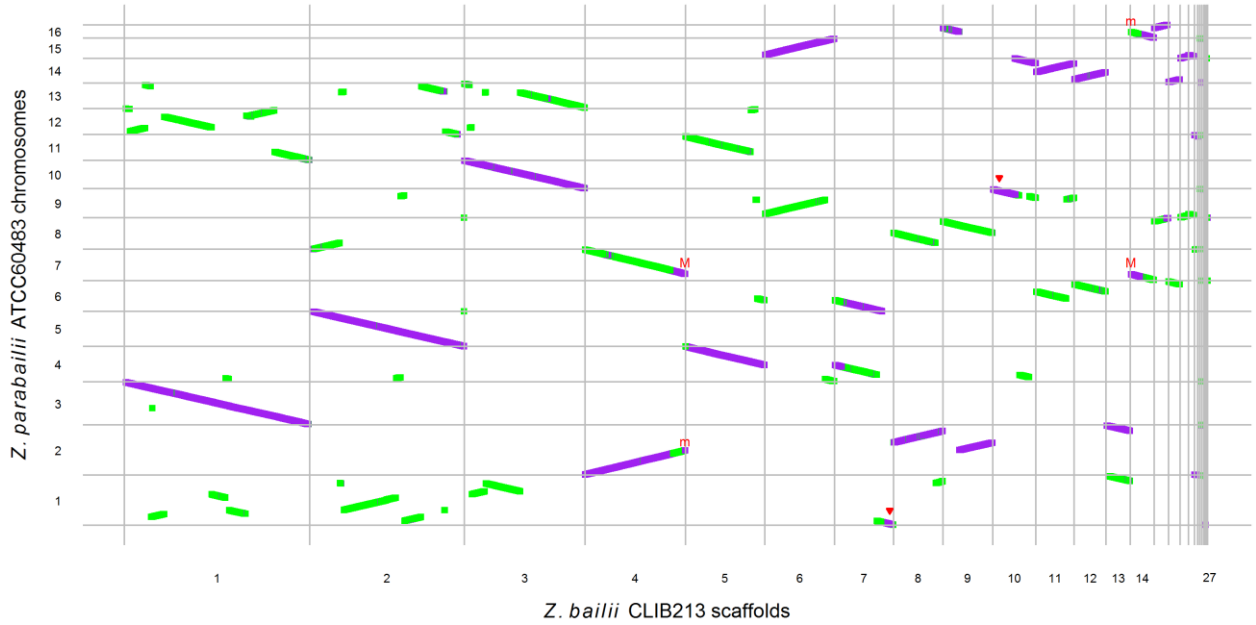


Figure 3 Dot-matrix plot between *Z. bailii* CLIB213T scaffolds (Galeote, *et al.*, 2013) and *Z. parabailii* ATCC60483 chromosomes. Each dot is a protein-coding gene (purple: A-genes; green, B-genes). Red triangles indicate chromosome ends that appear unpaired due to BIR. M and m indicate the active and broken MAT loci of *Z. parabailii*, respectively.

In total, from Fig 3 we estimate that there are approximately 34 breakpoints in synteny between the *Z. parabailii* B-subgenome and *Z. bailii*, but no breakpoints between the A-subgenome and *Z. bailii*, when post-hybridization rearrangement events (described below) are excluded. This difference in the levels of rearrangement in the A- and B-subgenomes relative to *Z. bailii* indicates that the two subgenomes were not collinear at the time the hybrid was formed. Therefore, most of the rearrangements between the two subgenomes are rearrangements that existed between the two parental species prior to hybridization. The two parents both had 8 chromosomes, but their karyotypes were quite different. Because each event of reciprocal translocation or inversion creates two synteny breakpoints (Sankoff, 1993), we estimate that about 17 events of chromosomal translocation or inversion occurred between the two parents in the time interval between when they last shared a common ancestor and when they hybridized. The situation in *Z. parabailii* (hybridization between parents differing by 17 rearrangements and 7% sequence divergence) contrasts with that in the hybrid *Millerozyma sorbitophila* (only 1 detectable rearrangement between the parents, despite 15% sequence divergence (Leh Louis *et al.*, 2012)).

Post-hybridization recombination, loss of heterozygosity (LOH) and break-induced replication (BIR)

Although the *Z. parabailii* genome largely contains unrearranged parental chromosomes, there have been two major types of rearrangement after hybridization. First, post-hybridization recombination between the subgenomes at homeologous sites has formed some chromosomes that are partly 'A' and partly 'B'. Second, a process of homogenization has occurred at some places in which one subgenome overwrote the other, resulting in gene pairs that are A:A or B:B. This process is commonly called loss of heterozygosity (LOH) or gene conversion. Based on their K_S distances from *Z. bailii*, the genome contains 4153 simple A:B homeologous gene pairs, 300 A:A pairs and 84 B:B pairs.

To examine the genomic locations of LOH and rearrangement events in more detail, we further classified genes using a scheme that takes account of their pairing status as well as their divergence from *Z. bailii*. Genes were defined as 'A' or 'B' as before, or 'N' if a K_S distance from *Z. bailii* could not be calculated (**Fig 2B,C**). We then assigned each gene to one of 7 categories such as "B-gene in an A:B pair" or "A-gene, unpaired", and plotted the locations of genes in each category. The resulting map of the genome (**Fig 4**) allows LOH and recombination events to be visualized. N-genes (black in Fig 4) are seen to be mostly located near telomeres. Several points of recombination between the A- and B-subgenomes are apparent, such as in the middle of chromosome 4. LOH tends to occur in stretches that span multiple genes. For example, on chromosome 13, LOH has formed eight runs of consecutive A-genes in a chromosome that is otherwise 'B'; these A-genes are members of A:A pairs. They were probably formed by homogenization (gene conversion without crossover), although they could also be the result of double crossovers followed by meiotic segregation of chromosomes. Patches of LOH are frequently seen adjacent to sites of recombination between the two subgenomes (Fig 4). Three large regions of apparently unpaired A-genes near the ends of chromosomes (1L, 5L and 9R; light blue in Fig 4) are probably artefacts caused by break-induced replication (BIR), which is a process that can make the ends of two chromosomes completely identical from an initiation point out to the telomere (Bosco and Haber, 1998). These regions have 2x sequence coverage

in our Illumina data, and we can identify the probable locations of an identical second copy of each of them at other chromosome ends (Fig 4).

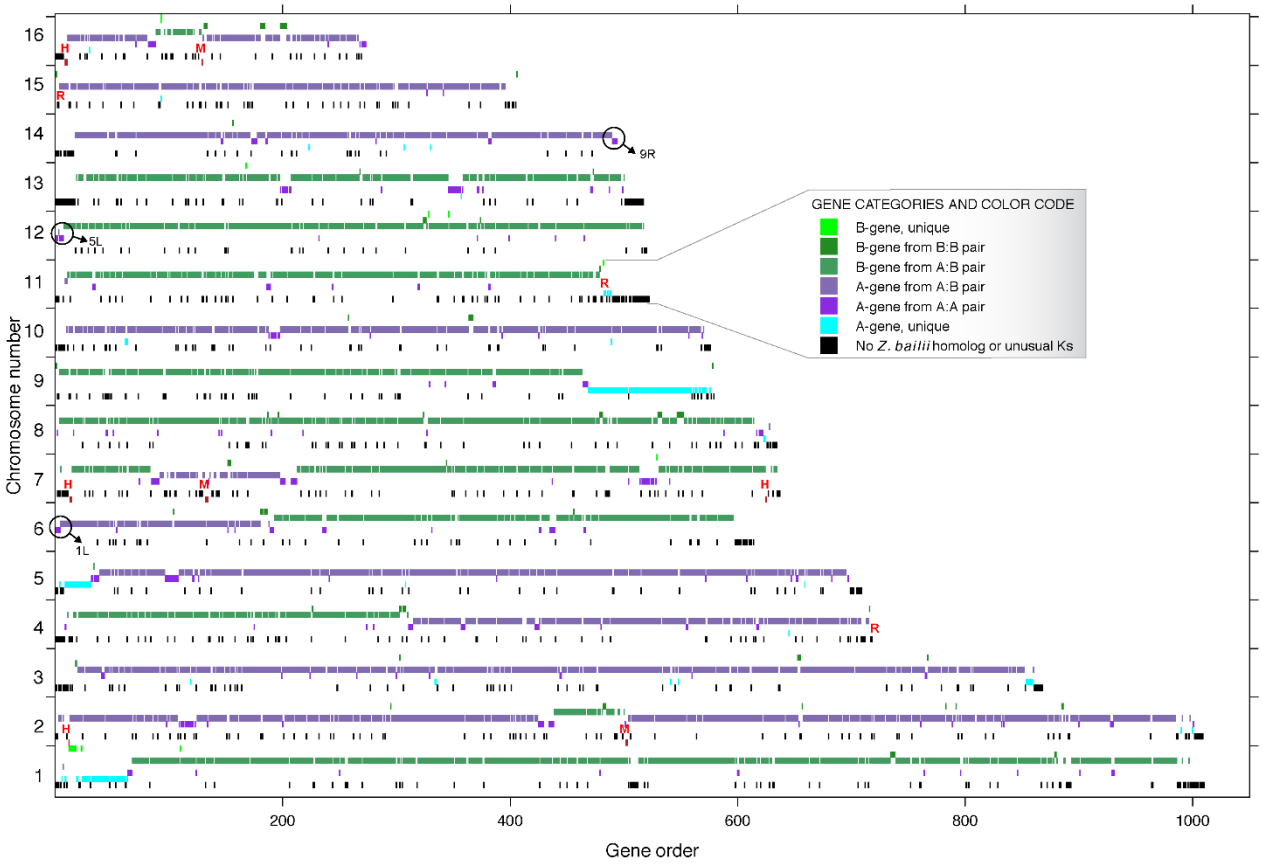


Figure 4 Subgenome and duplication status of each *Z. parabailii* gene. Each gene was classified into one of 7 categories and color-coded as shown in the legend. For each chromosome, seven rows were then drawn, showing the locations of genes in each category (the 7 rows appear in the same order from top to bottom as in the legend). R shows the locations of rDNA clusters. M and H indicate the locations of MAT and HML/HMR loci. Circles with arrows mark the three chromosome ends where our sequence is incomplete due to BIR; in each case the missing sequence is apparently identical to the end of another chromosome as shown. For example, we infer that at the right end of chromosome 14, our assembly artefactually lacks a second copy of the genes that are labeled as 'A unique' on the right end of chromosome 9. The high sequence identity of the chromosome 9 and 14 copies of this region caused them to co-assemble, and the co-assembled contig was arbitrarily assigned to chromosome 9.

Rearrangement catalyzed by HO endonuclease, and degeneration of the 'B' *MAT* locus

The *Z. parabailii* genome contains two *MAT* loci (one of which is broken) and four *HML*/*HMR* silent loci **(Fig 5)**. In *S. cerevisiae*, mating-type switching is a DNA rearrangement process that occurs in haploid cells to change the genotype of the *MAT* locus (Haber, 2012). During switching, the active *MAT* locus is first cleaved by an endonuclease called HO, and its **a**- or α-

specific DNA is removed by an exonuclease. The resulting double-strand DNA break at *MAT* is then repaired by copying the sequence of either the *HML*α or *HMR*a locus. This process converts a *MAT*a genotype to *MAT*α, or *vice versa*. Repeated sequences, called Z and X, located beside *MAT* and the *HM* loci act as guides for the DNA strand exchanges that occur during this repair process. The *HM* loci are 'silent' storage sites for the **a** and α sequence information because genes at these loci are not transcribed due to chromatin modification; only *MAT* is transcribed (Haber, 2012).

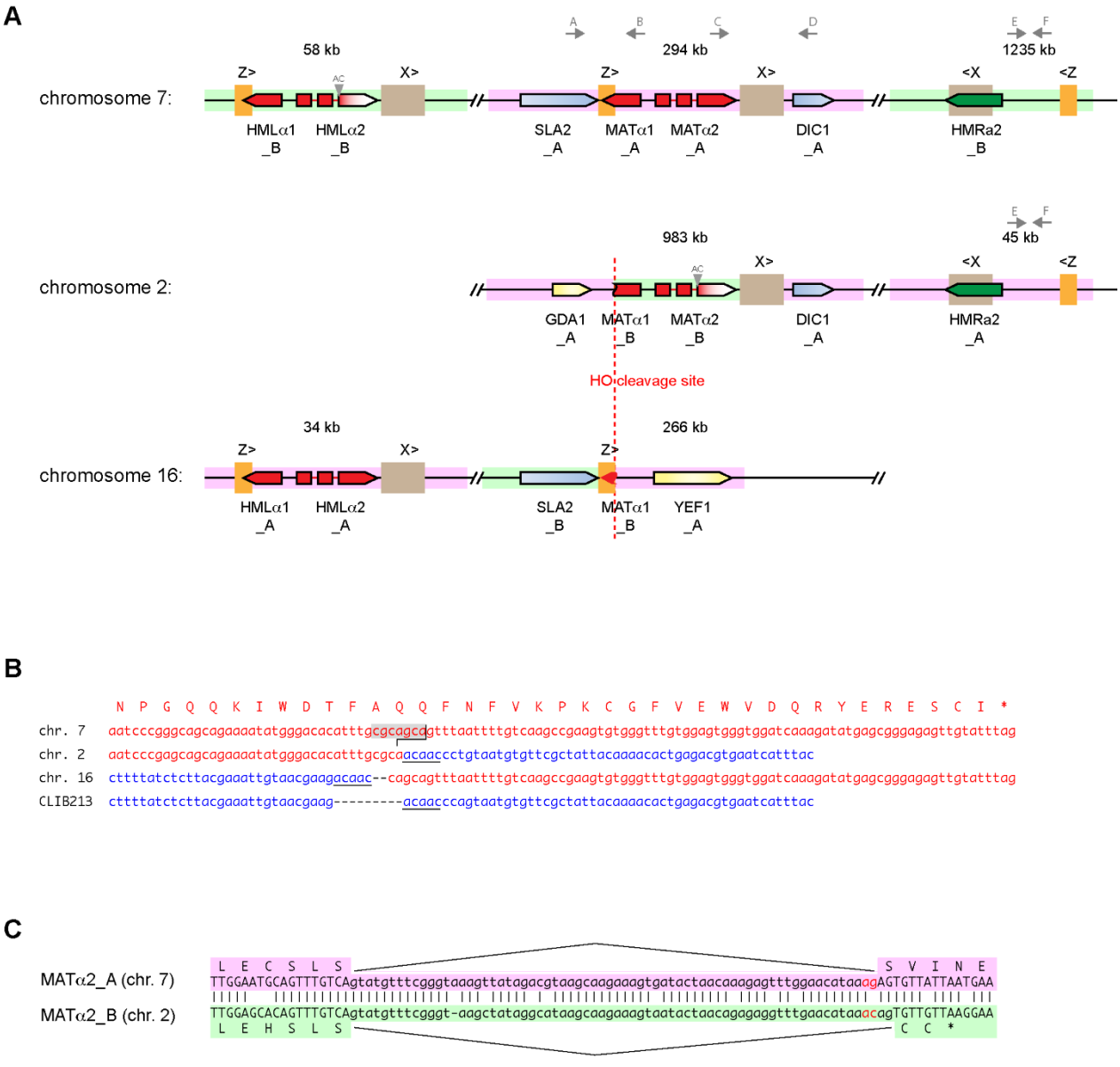


Figure 5 (A) Organization of MAT, HML and HMR loci in *Z. parabailii* ATCC60483. The genome contains six MAT-related regions, with one MAT, one HML and one HMR locus derived from each of

the A and B parents. Pink and green backgrounds indicate sequences from the A- and B-subgenomes, respectively. The MAT locus in the A-subgenome (position 294 kb on chromosome 7) is intact and expressed. The MAT locus of the B-subgenome has been broken into two parts by cleavage by HO endonuclease. All six copies of the X repeat region (654 bp) are identical in sequence, as are all six copies of the Z repeat region (266 bp). Gray triangles indicate the disruption of splicing of intron 2 in MATα2 and HMLα2 of the B-subgenome. Binding sites for primers A-F used for PCR amplification are indicated by gray arrows. (B) Sequences at the MAT locus breakpoint. Red, MATα1-derived sequences. The HO cleavage site (CGCAGCA, giving a 4-nucleotide 3' overhang) is highlighted in gray. Blue, the GDA1-YEF1 intergenic region from the equivalent region of *Z. bailii* CLIB213T, and homologous sequences from the A-subgenome on *Z. parabailii* chromosomes 2 and 16. A 5-bp sequence (ACAAC) that became duplicated during the rearrangement is underlined. (C) Sequences of MATα2 intron 2 (lowercase) from the A- and B-subgenomes. An AG-to-AC mutation (red) at the 3' end of the intron moved the splice site by 2 bp in the B-subgenome, causing a frameshift and premature translation termination. The splice sites in both genes were identified from RNAseq data.

We infer that the parents of *Z. parabailii* each contained a *MAT* locus and two silent loci (*HML*α and *HMR*a), similar to *S. cerevisiae* and *Z. rouxii* haploids (Watanabe *et al.*, 2013). Fig 5A shows that *Z. parabailii* has a *MAT* locus on chromosome 7, flanked by Z and X repeats and full-length copies of the genes *SLA2* and *DIC1*, similar to the *MAT* loci of many other species (Gordon *et al.*, 2011a, Watanabe *et al.*, 2013). This *MAT* locus is derived from Parent A. Chromosome 7 also contains *HML*α and *HMR*a loci (derived from Parent B) near its telomeres. However, the B-subgenome's *MAT* locus is broken into two pieces. Most of it is on chromosome 2, but its left part (the 3' end of *MAT*α1, the *Z* repeat and the neighboring gene *SLA2*) is on chromosome 16 (Fig 5A). Chromosomes 2 and 16 also each contain an *HML*α or *HMR*a locus from the A-subgenome.

Examination of the breakpoint in the B-subgenome's *MAT* locus shows that the break was catalyzed by HO endonuclease, because it occurs precisely at the cleavage site for this enzyme (Fig 5B). In *S. cerevisiae*, HO has a long (~18 bp) recognition sequence that is unique in the genome, and it cleaves DNA at a site within this sequence leaving a 4-nucleotide 3' overhang (Nickoloff *et al.*, 1990). Although the recognition and cleavage sites of HO endonucleases in other species have not been investigated biochemically, they can be deduced because the core of the HO cleavage site (CGCAGCA) invariably forms the first nucleotides of the Z region in each species (Gordon *et al.*, 2011a). Moreover, the HO cleavage site corresponds to an amino acid sequence motif (FAQQ) in the *MAT*α1 protein that is strongly conserved among species.

The two parts of the broken *MAT* locus are located beside the genes *GDA1* and *YEF1* (Fig 5A), which are neighbors in *Z. bailii* CLIB213^{T} and in the Ancestral yeast genome (Galeote *et al.*, 2013, Gordon *et al.*, 2009). Therefore, after HO endonuclease cleaved the 'B' *MAT* locus, the broken ends of the chromosome apparently interacted with the *GDA1-YEF1* intergenic region

of the A-subgenome, causing a reciprocal translocation. This site is the only synteny breakpoint between the A-subgenome of *Z. parabailii* and the genome of *Z. bailii* (scaffold 9; Fig 3). Comparison of the DNA sequences at the site (Fig 5B) shows no microhomology between the two interacting sequences, and that DNA repair led to duplications of a 5-bp sequence (ACAAC) from the *GDA1-YEF1* intergenic region and a 2-bp sequence (CA) from *MAT*α1, suggestive of nonhomologous end-joining (NHEJ) as the repair mechanism. We hypothesize that this genomic rearrangement occurred during a failed attempt to switch mating types, which resulted in a reciprocal translocation instead of normal repair of *MAT* by *HML* or *HMR*.

While the B-subgenome's *MAT*α1 gene is clearly broken, its *MAT*α2 gene also appears to be nonfunctional. *MAT*α2 has two introns, and our RNAseq data shows how both homeologs of this gene (*ZPAR0G01480_A* and *ZPAR0B05090_B*) are spliced. A point mutation at the 3' end of intron 2 of the B-gene changed its AG splice acceptor site to AC, with the result that splicing now uses another AG site two nucleotides downstream (Fig 5C). This change results in a frameshift, truncating the B-copy of the α2 protein to 57 amino acid residues instead of 211 and presumably inactivating it.

Surprisingly, the *Z. parabailii* genome does not contain any *MAT***a**1 (or *HMR***a**1) gene. This gene codes for the **a**1 protein, which is one subunit of the heterodimeric **a**1-α2 transcriptional repressor that is formed in diploid (**a**/α) cells and which acts as a sensor of diploidy by repressing transcription of haploid functions such as mating, while permitting diploid functions such as meiosis (Johnson, 1995). The **a**1 gene is present in *Z. rouxii* and *Z. sapae* (Souciet *et al.*, 2009, Watanabe *et al.*, 2013, Solieri *et al.*, 2014), but it is also absent from *Z. bailii* CLIB213^T and must have been absent from Parent B. The *Z. bailii* CLIB213^T *MAT* organization is not fully resolved (2013) but it contains a *MAT* locus with α1 and α2 genes on scaffold 14, and an *HMR* locus with only an **a**2 gene on scaffold 19. Evolutionary losses of *MAT***a**1 have previously been seen in some *Candida* species (Logue *et al.*, 2005, Butler, 2010), but not in any species of family Saccharomycetaceae. In contrast, the gene for the other subunit of the heterodimer, *MAT*α2, is present in all *Zygosaccharomyces* species and is probably maintained because it has a second role in repressing **a**-specific genes in this genus (Baker *et al.*, 2012). Solieri and colleagues have reported evidence that **a**1-α2 is nonfunctional in a *Zygosaccharomyces rouxii/pseudorouxii* hybrid where its two subunits are derived from different species (Bizzarri *et al.*, 2016).

Z. parabailii strains ATCC60483 and ISA1307 are descendants of the same interspecies hybridization event

The two subgenomes apparent in the Illumina scaffolds of the *Zygosaccharomyces* hybrid strain ISA1307, previously sequenced by Mira *et al.* (2014), are both 99-100% identical in sequence to the A- or B-subgenomes of ATCC60483. Therefore ISA1307 is also a strain of *Z. parabailii.* Importantly, the ISA1307 genome sequence contains the same HO-catalyzed reciprocal translocation between *MAT*α1 of the B-subgenome and the *GDA1-YEF1* intergenic region of the A-subgenome (Fig 5A). Because this rearrangement is so unusual, and because it did not involve recombination between repeated sequences, it is highly unlikely to have occurred twice in parallel. The rearrangement is much more likely to have occurred only once, in a common ancestor of the two *Z. parabailii* strains after the hybrid was formed. It cannot pre-date the hybridization because it formed junctions between the A- and B-subgenomes, which originated from different parents.

ATCC60483 and ISA1307 are independent isolates of *Z. parabailii,* both from industrial sources. ATCC60483 was isolated from citrus concentrate used for soft drinks manufacturing in the Netherlands (Put *et al.,* 1976, Put and De Jong, 1982), and ISA1307 was a contaminant in a sparkling wine factory in Portugal (Malfeito-Ferreira *et al.*, 1990, Malfeito-Ferreira *et al.*, 1997a, Malfeito-Ferreira *et al.,* 1997b, Mira *et al.,* 2014). We found several examples where the two strains differ in their patterns of LOH, which confirms that they have had some extent of independent evolution. All three large regions of BIR (on chromosomes 1, 5 and 9; Fig 4) are unique to ATCC60483. ISA1307 contains A:B homeolog pairs throughout these regions whereas ATCC60483 has only A-genes which we infer to be in A:A pairs. Other examples of differential LOH include a 4-kb region around homologs of the *S. cerevisiae* gene *YLR049C* that exists as B:B pairs in ATCC60483 but A:B pairs in ISA1307, and the gene *KAR4* which is an A:B pair in ATCC60483 but only a B-gene (single contig) in ISA1307. Notably, the section of the *RPB1* gene (also called *RPO21*) that Suh *et al.* (2013) used for taxonomic identification of *Z. parabailii* and *Z. pseudobailii* exists as an A:B pair in ATCC60483, but only as an A-gene in the ISA1307 genome. The absence of the B-copy of *RPB1* made Mira *et al.* (2014) hesitant to conclude that ISA1307 is *Z. parabailii*.

Z. parabailii ATCC60483 is fertile and haploid

Both ATCC60483 and the type strain of *Z. parabailii* ATCC56075^T have previously been reported to be capable of forming ascospores (Put *et al.*, 1976, Put and De Jong, 1982, Suh *et al.*, 2013). We confirmed that our stock of ATCC60483 is able to sporulate (**Fig 6A,B**). On malt extract agar plates, we observed that sporulation occurs directly in zygotes formed by conjugation between two cells, resulting in asci in which the two former parental cell bodies typically contain two ascospores each. Such dumbbell-shaped (conjugated) asci, indicative of sporulation immediately after mating, are characteristic of the genus *Zygosaccharomyces* (James and Stratford, 2011) and have previously been described in other *Z. bailii* (*sensu lato*) strains (Kudrjawzew, 1960, Phaff *et al.*, 1966, Barnett *et al.*, 1983, Barnett *et al.*, 2000, Kurtzman and James, 2006, James and Stratford, 2011). The presence of conjugating cells in a culture grown from a single strain indicates that ATCC60483 is functionally haploid (capable of mating), and that it is homothallic (capable of mating-type switching). Since the zygote proceeds immediately into sporulation without further vegetative cell divisions, the diploid state of *Z. parabailii* appears to be unstable. Although Suh *et al.* (2013) reported that asci of the type strain of *Z. parabailii* contain two spores, we consistently observed that asci occur in pairs of mated cells connected by a conjugation tube (Fig 6A,B), indicating that four spores are formed per meiosis.

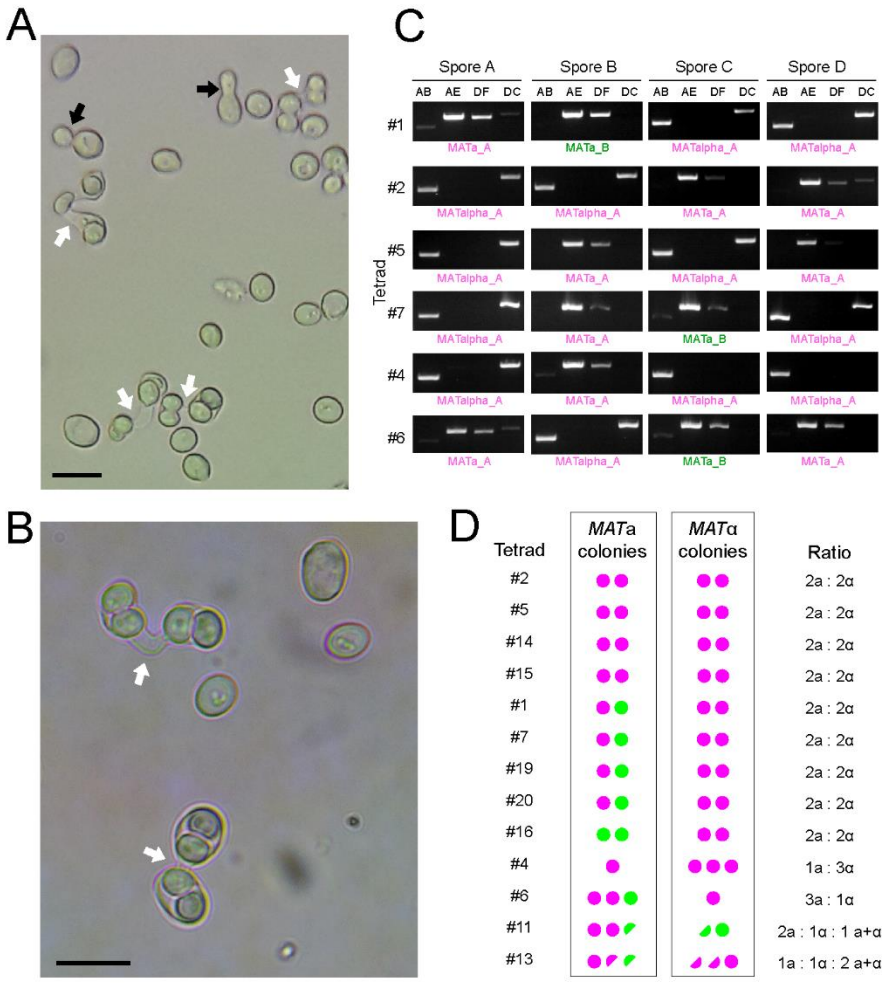


Figure 6 (A,B) Ascospore formation in *Z. parabailii* ATCC60483. White arrows show conjugation tubes in dumbbell-shaped asci. Black arrows show budding vegetative cells. Scale bars, 10 µm. Cultures were grown on 5% malt extract agar for 6-10 days at 25°C. (C) Examples of PCR determination of MAT locus genotypes in tetrads. Pairs of PCR primers as shown in Fig 5A were used to amplify the MAT locus in colonies grown from spores after dissection of conjugated asci. PCR primer pairs AB and AE amplify the left side of the MAT locus including the Z region (AB, 1485 bp product from MATα; AE, 2103 bp product from MATa). Primer pairs DC and DF amplify the right side of the MAT locus including the X region (DC, 2027 bp product from MATα; DF, 1882 bp product from MATa). PCR products were

sequenced to determine whether they originated from the A- or B-subgenome. (D) Summary of MAT genotypes in colonies grown from spores from 13 dissected tetrads. Magenta circles denote colonies with A-subgenome alleles (MATa_A or MATα_A), and green circles denote colonies with B-subgenome alleles (MATa_B or MATα_B). Half-circles represent colonies that gave both MATa and MATα PCR products.

We dissected tetrad asci from ATCC60483, grew colonies from the spores, and then used colony PCR to determine their genotype at the intact *MAT* locus on chromosome 7. Among 13 tetrads analyzed, nine showed a ratio of 2 *MAT***a** : 2 *MAT*α colonies (**Fig 6C,D**). Two tetrads showed 1:3 or 3:1 ratios, and the other two yielded both *MAT***a** and *MAT*α PCR products from some single-spore colonies. The genotype of the ATCC60483 starting strain is *MAT*α from the A-subgenome (designated *MAT*α_A), so the presence of *MAT***a** genotypes in colonies derived from spores made by this strain confirms that mating-type switching occurred at some point. We sequenced the PCR products and found that the A- and B-subgenome *HMR***a** loci were both used as donors for mating-type switching: among the pure *MAT***a** colonies, 18 were *MAT***a**_A and 7 were *MAT***a**_B (Fig 6D). Quite surprisingly, four tetrads with 2**a**:2α segregation had one *MAT***a**_A and one *MAT***a**_B spore colony, which is inconsistent with simple meiotic segregation from an **a**/α diploid. Because all the spores contain a functional *HO* gene, the genotypes of these four tetrads (#1, #7, #19, #20) probably result from additional switches during the early growth of some colonies. Similarly, switching during early colony growth may explain the presence of *MAT*α_B genotypes in tetrad #11, and the colonies with mixed **a**+α genotypes (in tetrads #11 and #13), as well as the presence of faint PCR products corresponding to the alternative *MAT* genotype in some other colonies (Fig 6C). In *S. cerevisiae*, homothallic diploid (*HO/HO MAT***a**/*MAT*α) strains show 2:2 segregation of *MAT* alleles in tetrads, but after spore germination the haploid cells can then switch mating types as often as once per cell division (Strathern and Herskowitz, 1979), leading to mating and colonies that contain mostly diploid cells (Herskowitz, 1988); by contrast, most (but not all) of the *Z. parabailii* spore-derived colonies contained a single mating type (Fig 6C,D).

We found that almost all the genes involved in mating and meiosis that Mira *et al.* (2014) reported to be missing from the *Z. parabailii* ISA1307 genome are in fact present in both ATCC60483 and ISA1307 (**S1 Table**). For example, we annotated A- and B-homeologs of *IME1*, *UME6*, *DON1*, *SPO21*, *SPO74*, *REC104* and *DIG1*/*DIG2* as well as *MAT***a**2, *MAT*α1 and *MAT*α2. We also identified genes for the α-factor and **a**-factor pheromones (*MF*α and *MF*a).

The *MFα* genes code for an unusually high number of copies (10–14) of a 13-residue peptide whose consensus sequence, AHLVRLSPGAAMF, is quite different from that of other yeasts including *Z. rouxii* (7/13 matches) and *S. cerevisiae* (4/13 matches) (Wolfe *et al.*, 2015). *Z. parabailii* and *Z. bailii* do lack most of the ZMM group of genes, involved in crossover interference during recombination (Shinohara *et al.*, 2008), even though these are present in *Z. rouxii* (S1 Table). Interestingly, identical sets of ZMM genes have been lost in *Z. bailii*/*Z. parabailii* relative to *Z. rouxii*, as were lost in most *Lachancea* species relative to *L. kluyveri* (Vakirlis *et al.*, 2016): *ZIP2*, *CST9 (ZIP3)*, *SPO22 (ZIP4)*, *MSH4*, *MSH5* and *SPO16* are absent, as well as *MLH2* which is not known to be a ZMM gene, whereas *ZIP1* is retained. A similar loss of ZMM genes has occurred in *Eremothecium gossypii* relative to *E. cymbalariae* (Wendland and Walther, 2011).

Post-hybridization gene inactivations

A small number of *Z. parabailii* ATCC60483 genes have 'disabling' mutations – frameshifts or premature stop codons that prevent translation of a normal protein product. The majority of these mutations are present in only one subgenome of ATCC60483 and are unique to this strain. For example, there is a 1-bp insert in the A-homeolog of the DNA repair gene *MLH1*, which is not present in the B-homeolog, nor in ISA1307 or CLIB213^{T}. In a systematic search we found a total of 10 A-genes and 9 B-genes that were inactivated only in strain ATCC60483 (**S2 Table**). In each case the other homeolog was intact and the mutations, discovered in the PacBio assembly, were confirmed by our Illumina contigs of the ATCC60483 genome.

We found a further eight disabling mutations that are shared between ATCC60483 and ISA1307. One of these is the AC to AG splice site mutation in the B-homeolog of *MATα2* described above (Fig 5C). Another is the *HO* endonuclease gene, whose A-homeolog contains an identical 1-bp deletion in both ATCC60483 and ISA1307, whereas the B-homeolog of *HO* is intact in both strains (S2 Table). It is perhaps surprising that the *HO* gene that degenerated is the A-homeolog, whereas the broken *MAT* locus is the B-homeolog, but the two endonucleases are likely to have had identical site specificities because the HO cleavage site is well conserved among species. The existence of these eight shared disabling mutations provides further support for the idea that the two strains of *Z. parabailii* are descended from the same hybrid ancestor, because these mutations may not be viable in the absence of the intact homeologous copies of these genes. Only one of them is present also in CLIB213^{T} (S2 Table).

In-frame introns and other features of the genome

We annotated 447 introns in the *Z. parabailii* ATCC60483 genome, most of which are confirmed by our RNAseq data. There are 428 intron-containing genes, including 19 that have two introns. We did not find any examples of intron presence/absence differences between homeologs. Interestingly, we found several genes with an in-frame intron – that is, an intron that is a multiple of 3 bp long and contains no stop codons, so that both the spliced and unspliced forms of the mRNA can be translated into proteins. Genes with in-frame introns are likely to undergo alternative splicing, making two forms of the protein with different functions. One of these loci is *PTC7* (*ZPAR0J04940_A* and *ZPAR0A06900_B*). Both of the *Z. parabailii* homeologs contain a 69 bp in-frame intron within the ORF of the gene. It has previously been shown that alternative splicing of a similar in-frame intron in *S. cerevisiae PTC7* leads to the translation of a mitochondrial protein isoform from the spliced mRNA, and a nuclear envelope protein isoform from the unspliced mRNA, and that the intronic region codes for a transmembrane domain of the protein (Juneau *et al.*, 2009). Thus, the alternative splicing mechanism in *PTC7* is conserved between *Saccharomyces* and *Zygosaccharomyces.* We also found in-frame introns in the *Z. parabailii* orthologs of *S. cerevisiae NUP100*, *NCB2* and *HEH2*, identically in their A- and B-homeologs. None of these genes is known to be alternatively spliced in *S. cerevisiae.* In each of these examples, there are typical splice donor, branch and acceptor sequences within the long form of the ORF.

Programmed '+1' ribosomal frameshifting, a process whereby the ribosome skips forward by one nucleotide when translating an mRNA, is known to occur in three genes in *S. cerevisiae*: *OAZ1*, *ABP140* and *EST3* (Farabaugh *et al.*, 2006), and we found that +1 frameshifting is also required to translate the *Z. parabailii* orthologs of these three genes, in both the A- and B-homeologs. We also found two new loci that apparently undergo +1 frameshifting. Translation of both homeologs of *BIR1* (*ZPAR0O02690_A, ZPAR0I02720_B*) requires a +1 frameshift at a sequence identical to the *EST3* frameshifting site: CTT-A-GTT where the A is the skipped nucleotide. Translation of both homeologs of *YJR112W-A* (*ZPAR0O02960_A, ZPAR0I02990_B*) requires a +1 frameshift at a sequence identical to the *ABP140* frameshifting site: CTT-A-GGC.

In *S. cerevisiae*, the *CUP1* locus confers resistance to copper toxicity by a gene amplification mechanism. *CUP1* codes for a metallothionein, a tiny cysteine-rich copper-binding protein. The

reference *S. cerevisiae* genome sequence contains two identical copies of *CUP1* duplicated in tandem, but under copper stress this locus can become amplified to contain up to 18 tandem copies of the gene (Karin *et al.*, 1984, Zhao *et al.*, 2014). There are at least five different types of *CUP1* repeats in different *S. cerevisiae* strains, which must have originated independently from progenitors with a single *CUP1* gene (Warringer *et al.*, 2011, Zhao *et al.*, 2014). In *Z. parabailii* we found a slightly different organization. At homeologous loci on chromosomes 2 and 7, ATCC60483 has multiple identical copies of a 1454-bp repeating unit. Each unit contains two metallothionein genes, *MT-58* and *MT-47*, coding for proteins of 58 and 47 residues respectively. There is only 56% amino acid sequence identity between MT-58 and MT-47 proteins. The chromosome 7 locus contains five copies of the repeating unit, and the chromosome 2 locus contains two copies, so ATCC60483 has 14 metallothionein genes in total. These loci are not syntenic with *S. cerevisiae CUP1*, but they are syntenic with metallothionein genes in *Candida glabrata* and *Z. rouxii* (Mehra *et al.*, 1989, Byrne and Wolfe, 2005).

Discussion

Our results show that *Z. parabailii* is a hybrid species that was formed by fusion between two 8-chromosome parental species, one of which was *Z. bailii*. The low sequence divergence of the ATCC60483 A-subgenome from the type strain of *Z. bailii* (the modal synonymous site divergence is less than 1%; Fig 2A), and the almost complete collinearity of these genomes (Fig 3), indicate that the A-parent of *Z. parabailii* should be regarded as *Z. bailii* itself, and not merely a species closely related to *Z. bailii*.

The unusual *MAT* locus structure of this hybrid raised questions about how it was formed and whether *Z. parabailii* currently has a full sexual cycle. At first glance the *MAT*α/*MAT*α hybrid genotype of ATCC60483 might suggest that *Z. parabailii* could not have been formed by mating. However, this genotype could also be the result of mating-type switching. We propose that the following steps occurred **(Fig 7)**. *Z. parabailii* was formed by mating between strains of parent A (*Z. bailii*) and parent B, of opposite mating types. These parental genomes already differed by about 34 chromosomal rearrangement breakpoints, so the hybrid was unable to produce viable spores by meiosis. The hybrid also had no *MAT***a**1 gene, so it could not form the **a**1-α2 heterodimer that stabilizes the diploid state in *S. cerevisiae* (Herskowitz, 1988). One of the roles of the **a**1-α2 dimer in *S. cerevisiae* is to repress transcription of *HO* endonuclease, which is only required in haploid cells. We suggest that in the newly-formed *Z. parabailii*

hybrid, transcription of *HO* was not repressed. Continued expression of this gene resulted in genotype switching at the *MAT* loci (perhaps several consecutive switches between **a** and α), and eventually breakage of the B-subgenome *MAT* locus due to an illegitimate recombination with the *GDA1-YEF1* intergenic region instead of *HML* or *HMR*. At some point after hybridization, the *HO* gene from the A-subgenome also degenerated by acquiring a frameshift mutation.

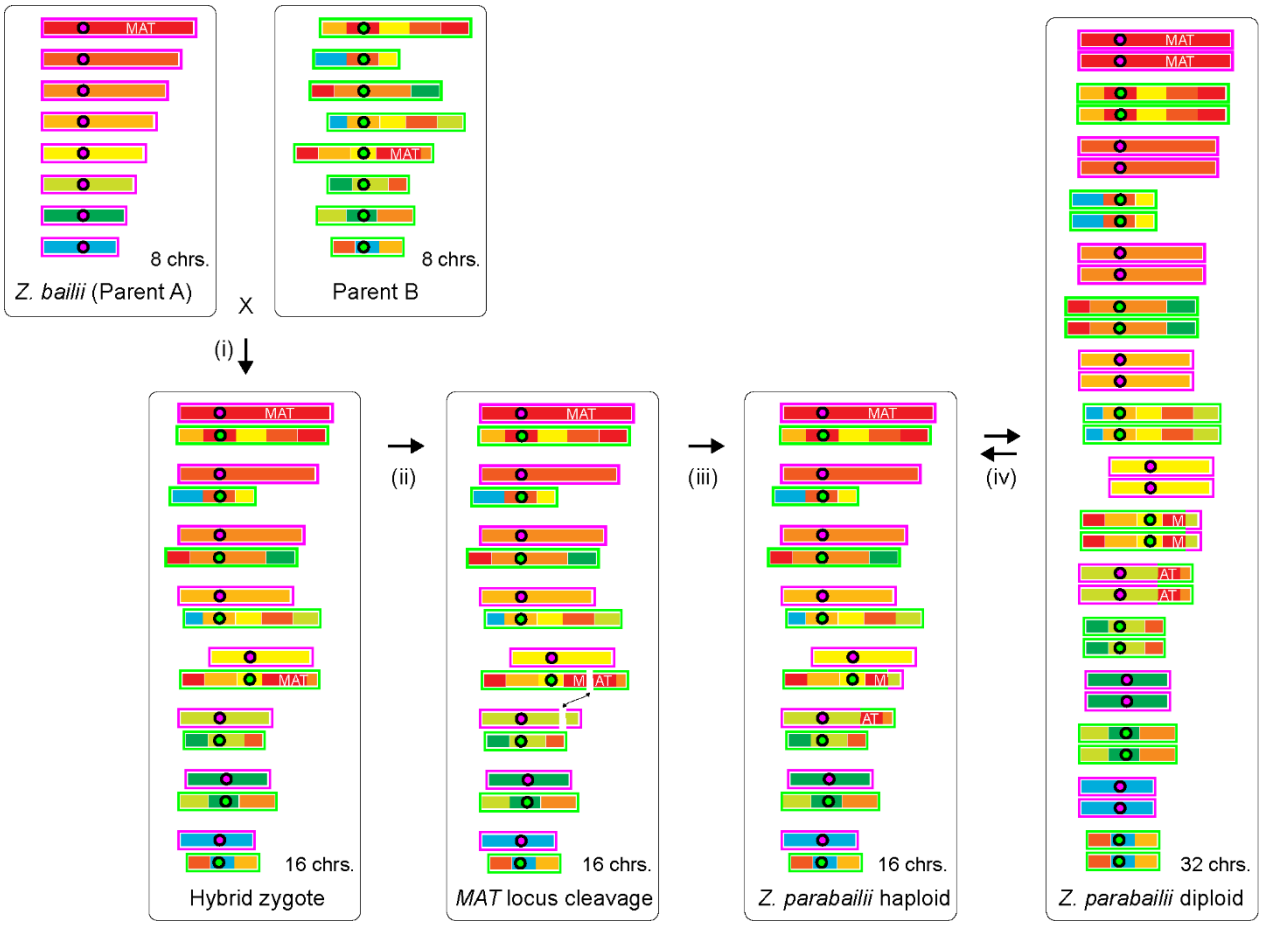


Figure 7 Cartoon of key steps in the origin of the *Z. parabailii* genome. Chromosome regions (thick bars) are colored according to their location in *Z. bailii* (magenta outlines). The corresponding homeologous regions are scrambled in Parent B (green outlines). Circles represent centromeres. (i) Interspecies mating occurred between Parent A (*Z. bailii*) and Parent B. The genomes differed by about 34 rearrangement breakpoints and 7% nucleotide sequence divergence. The resulting zygote was unable to form viable spores due to the non-collinearity of its chromosomes. (ii) Expression of HO endonuclease in the zygote, due to absence of a1-α2, resulted in cleavage of the B-copy of the MAT locus and ectopic recombination with the GDA1-YEF1 region of the A-subgenome, causing a reciprocal translocation. (iii) The resulting genome has only one functional MAT locus and behaves as a haploid. Recombinations and other exchanges between homeologous regions of the two subgenomes, such as those that exchanged the HML/HMR regions, occurred but are not shown here for simplicity. (iv) The current life cycle of *Z. parabailii* involves mating between 16-chromosome haploids to form 32-chromosome diploids, which immediately sporulate to regenerate 16-chromosome haploids. *Z. parabailii* is homothallic because it contains an intact HO gene, which allows interconversion between MATa and MATα haploids and hence auto-diploidization.

The breakage of the 'B' *MAT* locus can be inferred to have been one of the first rearrangement events that occurred after the hybridization, but also to have been recent. It must have been one of the first post-hybridization events, because the *GDA1-YEF1* breakage that occurred simultaneously with it is the only point of non-collinearity between the A-subgenome and the *Z. bailii* genome (apart from sites of inter-homeolog recombination or homogenization; Fig 4). It must have been recent because the pseudogene fragments of the broken *MAT* locus have not yet accumulated any other mutations. There are no nucleotide differences in 2298 bp between the broken *MAT*α_B locus on chromosome 2 and *HML*α_B on chromosome 7. Together, these two observations suggest that the interspecies mating that formed *Z. parabailii* occurred less than 10^5 generations or 1000 years ago (Rolland and Dujon, 2011). Such a recent origin is consistent with the very low numbers of gene inactivations that have occurred since hybridization, with the fact that most of these are not shared between the two sequenced *Z. parabailii* strains (only 8 of 27 inactivating mutations are shared; S2 Table), and with the retention of rDNAs from both parents. We expect that, if the *Z. parabailii* lineage survives, it will accumulate extensive inactivations and deletions of redundant duplicated genes over the next few million years as seen in older WGDs.

The net result of the evolutionary changes to the genome is that *Z. parabailii* now has 16 chromosomes (all different in structure but containing homeologous regions), one active *MAT* locus, one active *HO* gene, and four silent *HML/HMR* loci. A genome with this structure resembles haploid *S. cerevisiae* (Wolfe and Shields, 1997) and is potentially capable of both mating-type switching and mating. We confirmed that both of these processes occur in ATCC60483. *Z. parabailii* has a life cycle in which 16-chromosome haploids mate to produce 32-chromosome diploids (Fig 7), that sporulate immediately because the diploid state is unstable; there is no *MAT***a**1 gene and hence no **a**1-α2 heterodimer. Thus, *Z. parabailii* is an allopolyploid that regained fertility by genome doubling after interspecies mating, as a consequence of damage to one copy of its *MAT* locus.

Two previous reports that *Z. parabailii* strains produce only mitotic spores (Rodrigues *et al.*, 2003, Mollapour and Piper, 2001) can be re-interpreted in view of the hybrid nature of the genome. Their experimental data are fully compatible with the meiotic sexual cycle we propose for *Z. parabailii*. Rodrigues *et al.* (2003) made a derivative of ISA1307 in which one copy of *ACS2* was disrupted by the G418-resistance marker *APT1* and the other copy was not. After sporulation of this strain, all 80 spores they tested were G418-resistant, and all 16 spores from

4 tetrads contained both an intact copy of *ACS2* and an *acs2::APT1* disruption, which led Rodrigues *et al.* (2003) to conclude that the spores were made by mitosis. However, this inheritance pattern is exactly the pattern expected if the two copies of *ACS2* are homeologs (different Mendelian loci) rather than alleles of a single Mendelian locus, and if ISA1307 is a haploid that auto-diploidized before it sporulated. Thus their strain could be described as haploid *acs2_a::APT1 ACS2_B,* where the *ACS2_A* and *ACS2_B* loci have independent inheritance (they are on chromosomes 10 and 13 in our genome sequence). Similarly, Mollapour and Piper (Mollapour and Piper, 2001) disrupted one of the two copies of *YME2* in strain NCYC1427 with a *kanMX4* cassette, and found that all the spores produced by this strain retained both an intact *YME2* and *yme2::kanMX4*. They concluded that the spores were vegetative, but again the result is consistent with meiotic spore production if the two *YME2* loci have independent inheritance (they are on chromosomes 4 and 6), and if the disruption was made in a haploid strain that auto-diploidized before sporulating. The sequence data in (Mollapour and Piper, 2001) allows NCYC1427 to be identified as *Z. parabailii* and not *Z. bailii* as originally described. Furthermore, in both ISA1307 (Rodrigues *et al.*, 2003) and NCYC1427 (Mollapour and Piper, 2001, James and Stratford, 2011), spores are formed in pairs of conjugated cells, similar to Fig 6A. We conclude that ISA1307 and NCYC1427 have sexual cycles identical to the one we describe for ATCC60483.

The evolutionary steps that formed *Z. parabailii* by interspecies mating, and restored its fertility by damage to one of its *MAT* loci, are essentially identical to one of the mechanisms (hypothesis B) proposed for the origin of the ancient WGD in the *S. cerevisiae* lineage (Marcet-Houben and Gabaldon, 2015, Scannell *et al.*, 2006, Morales and Dujon, 2012, Wolfe, 2015). Our study therefore validates genome doubling after *MAT* locus damage as a real evolutionary process that occurs in natural interspecies hybrids, enabling them to resume mating and meiosis. The *Z. parabailii* hybridization was very recent, so any period of clonal reproduction that elapsed before fertility was restored must have been short, which is as expected because there is no selection to maintain meiosis genes during clonal growth (Morales and Dujon, 2012, Wolfe, 2015). The possible role of *MAT***a**1 in the ancient WGD remains unclear. In *Zygosaccharomyces*, the absence of this gene makes zygotes proceed into sporulation. In the ancient WGD it is likely that a *MAT***a**1 gene was present in the initial zygote, in which case the zygote would have been stable until it sustained *MAT* locus damage, but this is not certain because the ZT parent might have lacked *MAT***a**1. The specific cause of damage to the *MAT*

locus in *Z. parabailii* was incorrect DNA repair after cleavage by the mating-type switching endonuclease HO. The *HO* gene is present in the ZT clade, but not in the KLE clade (Gordon *et al.*, 2011a, Butler *et al.*, 2004), which were the two parental lineages of the interspecies hybridization that led to the ancient WGD (Marcet-Houben and Gabaldon, 2015). Species that contain *HO* show evolutionary evidence of repeated deletions of DNA from beside their *MAT* loci, caused by accidents during mating-type switching (Gordon *et al.*, 2011a). Indeed, the disappearance of the *MAT***a**2 gene from Saccharomycetaceae genomes, which occurred at approximately the same time as the WGD, must have been due to some sort of mutational damage to the *MAT* locus. Although HO-mediated damage can only occur in the small clade of yeasts that contain *HO*, other types of mutational damage to one copy of *MAT* are a plausible mechanism for fertility restoration in other fungal interspecies hybrids.

Materials and Methods

Strain and growth media

The strain analyzed here originally came from the collection of Thomassen & Drijver-Verblifa NV in the Netherlands (Put *et al.*, 1976, Put and De Jong, 1982) and was called '*Saccharomyces bailii* strain 242' in those studies. It was isolated from citrus concentrate being used as raw material for soft drinks. It was later deposited at the American Type Cultures Collection as ATCC60483. Suh *et al.* (2013) identified it as *Z. parabailii* by molecular methods.

PacBio DNA sequencing, assembly, and annotation

ATCC60483 genomic DNA was prepared using the Blood & Cell Culture DNA Mini Kit (Qiagen), according to the manufacturer's manual. To prevent fragmentation of the DNA, the sample was not vortexed. The final genomic DNA amount was 15 μg as determined by Qubit Fluorometer (Thermo Scientific). Pacific Biosciences sequencing was carried out by the Earlham Institute (Norwich, UK) using 8 SMRT cells, which generated 218x mean coverage for the nuclear scaffolds. We assembled the raw data using the computational facilities at the Irish Centre for High-End Computing (ICHEC), with the HGAP3 protocol of the SMRT Analysis suite version 2.3.0 (Chin *et al.*, 2013). We initially obtained 22 nuclear scaffolds, which we reduced to 16 chromosomes by manually identifying overlaps between scaffolds. In parallel, we also obtained 198x Illumina read coverage of the genome (Genome Analyzer IIx; University of Milano-Bicocca, Department of Clinical Medicine), which we assembled

separately into contigs that were used to verify the status of rearrangement points and pseudogenes discussed in the text.

The *Z. parabailii* chromosomes were annotated using an improved version of our YGAP automated pipeline (Proux-Wera *et al.*, 2012), which uses information in the Yeast Gene Order Browser (Byrne and Wolfe, 2005) and the Ancestral (pre-Whole Genome Duplication) gene order (2009) to generate a synteny-based annotation. The automated annotation was curated using transcriptome data from ATCC60483 cultures grown in a bioreactor; Illumina RNAseq was generated at Parco Tecnologico Padano (Italy). We made a *de novo* transcriptome assembly using Trinity (Grabherr *et al.*, 2011) and compared the transcripts against YGAP's gene models using PASA (Haas *et al.*, 2003) and by manual inspection of spliced mRNA reads.

Chromosomes were numbered 1-16 from largest to smallest. Genes were given systematic names by YGAP such as *ZPAR0D01210_B*, where *ZPAR* indicates the species; *0* indicates the genome sequence version; *D* indicates chromosome 4; *01210* is a sequential gene number counter that increments by 10 for each protein-coding gene (genes that were added manually have numbers that end in 5 or other digits); and the suffix *_B* indicates that this gene is assigned to the B-subgenome as described below. NCBI nucleotide sequence database accession numbers are CP019490 – CP019505 (nuclear chromosomes), CP019506 (mitochondrial genome), and CP019507 (2-micron plasmid).

The mitochondrial genome of *Z. bailii* CLIB213^T was not reported with the rest of this strain's genome (Galeote *et al.*, 2013), and is highly fragmented in the assembly. We identified mitochondrial contigs in the original CLIB213^T assembly by BLASTN using the ATCC60483 mtDNA as a query, assembled these contigs into 55 larger contigs using the CAP3 assembler and SSPACE3 (Huang and Madan, 1999, Boetzer *et al.*, 2011), and calculated a weighted average nucleotide identity of 96% from non-overlapping alignments totaling 23,197 bp.

Gene assignments to the A- and B-subgenomes

We assigned most genes in *Z. parabailii* ATCC60483 to either the A-subgenome (highly similar to the *Z. bailii* CLIB213^T genome) or the B-subgenome (derived from the other parent in the hybridization), using their levels of synonymous nucleotide sequence divergence from CLIB213^T genes. For this purpose, we used BLASTP (Altschul *et al.*, 1990) to compare every annotated protein from ATCC60483 to the CLIB213^T proteome and designated the best hit as a homolog. The corresponding ATCC60483 and CLIB213^T DNA sequence pairs were then

aligned using CLUSTALW (Larkin *et al.*, 2007) and their levels of sequence divergence were calculated using the yn00 program from the PAML suite (Yang, 2007). ATCC60483 genes were assigned to the A-subgenome if the level of synonymous divergence was $K_S \leq 0.05$, and to the B-subgenome if $0.05 < K_S \leq 0.25$, and given an *_A* or *_B* suffix on the gene name accordingly. Genes for which $K_S > 0.25$, or for which no *Z. bailii* homolog was identified, were given the suffix *_N*. To identify inactivated genes systematically, we searched the annotated A:B gene pairs for cases where one of the homeologs was less than 90% of the length of the other, and then examined these cases manually (S1 Table).

Note that our use of the labels 'A' and 'B' differs from the scheme used by Mira *et al.* (2014) for strain ISA1307. We designated each gene (homeolog) as either 'A' or 'B' based on its divergence from *Z. bailii* CLIB213^T, with 'A' always indicating the *Z. bailii*-like homeolog. Some chromosomes therefore contain mixtures of 'A' and 'B' genes due to post-hybridization recombination or homogenization between the two subgenomes. In contrast, Mira *et al.* (2014) identified homeologous pairs of scaffolds in their assembly and arbitrarily designated one scaffold as 'A' and the other as 'B', so that each scaffold is homogeneous but there is no consistent relationship between the 'A' and 'B' labels and the parent-of-origin of a homeolog in their scheme.

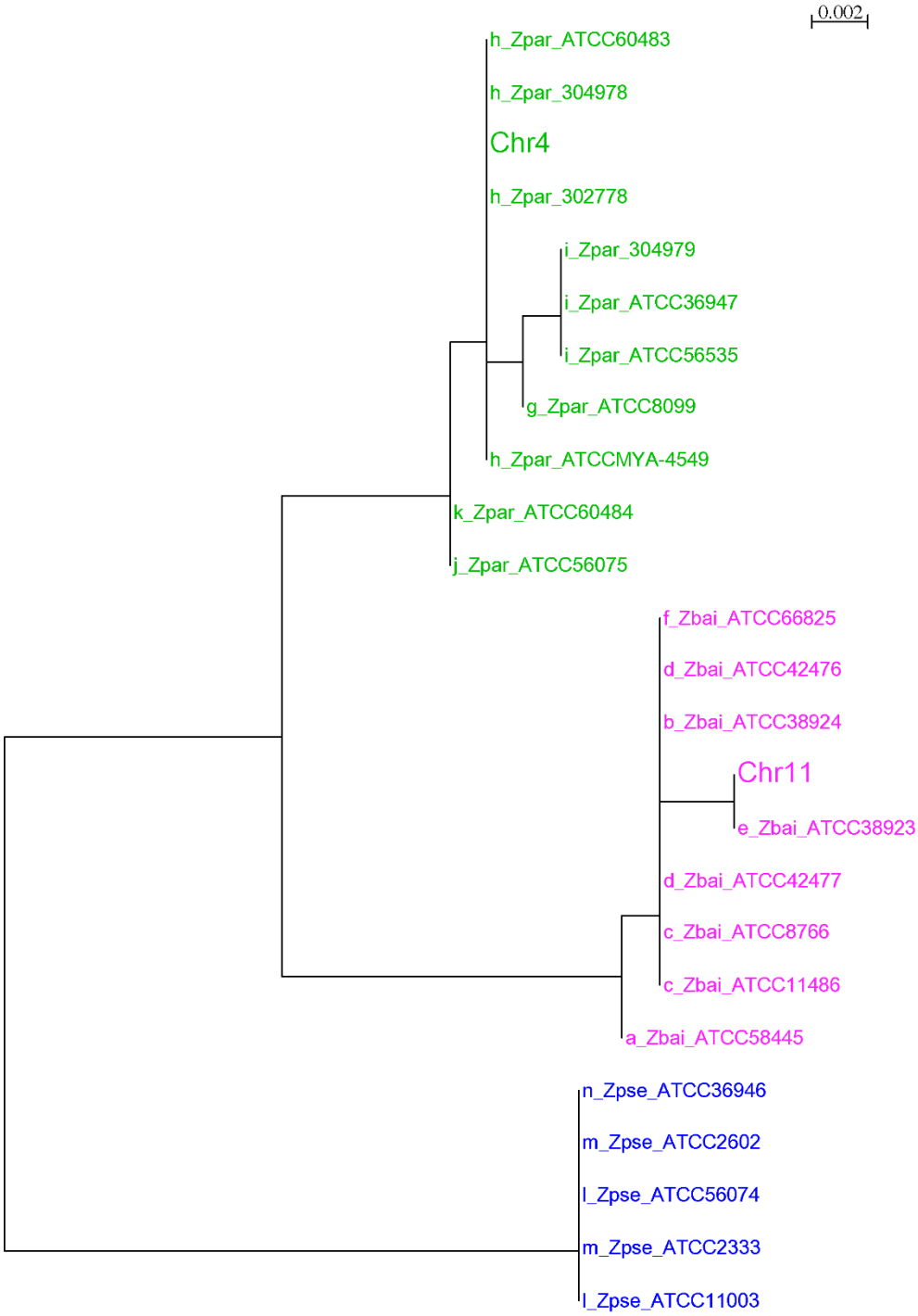
Tetrad dissection and *MAT* locus PCR amplification

Cells were left for sporulation on malt extract (5%) agar for 5 days. A small loop of cells was washed in sterile distilled water, resuspended in a 1:20 dilution of Zymolyase 100T and incubated for 10 min at 30°C. The Zymolyase solution was removed by centrifugation and the pellet resuspended in distilled water (500 µl). A 10 µl drop was placed in the middle of a YPD plate, and dumbbell-shaped asci were dissected using a Singer Sporeplay dissection microscope. The YPD plate was incubated for 2 days at 30°C. Individual spore-derived colonies were used for *MAT* locus genotyping by colony PCR using Q5 polymerase high-fidelity 2x master mix (NEB) and annealing temperature 55°C. Sequences of PCR primers A-F are given in S3 Table. Primers E and F were designed to bind equally to the *HMR* regions of the A- and B-subgenomes. Primers A-D are specific for the A-subgenome.

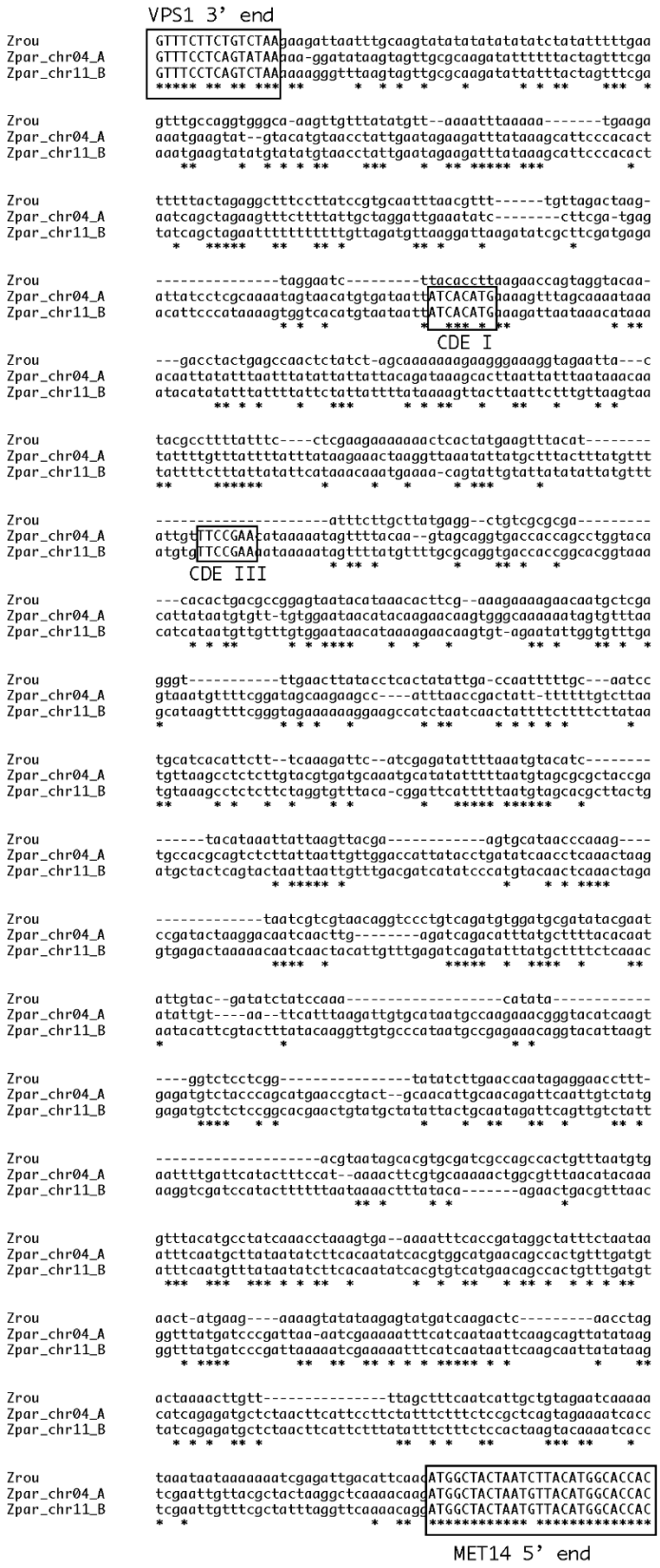
Acknowledgments

We thank John Morrissey and Francesca Doonan for encouragement and support, Simon Wong at ICHEC for help with genome assembly, Laura Dato for initial work on Illumina genome sequencing, Isabel Sá-Correia for strain ISA1307, Virginie Galeote for CLIB213 data, and Geraldine Butler for comments on the manuscript.

Supplementary Material



Suplementary Figure 1 (S1) Phylogenetic tree of internal transcribed spacer (ITS) regions of rDNA. Chr4 and Chr11 are the ITS sequences from the chromosome 4 and 11 rDNA units in the *Z. parabailii* ATCC60483 genome. All other sequences are from Suh *et al.* (Suh *et al.*, 2013) for strains of *Z. parabailii* (Zpar), *Z. bailii* (Zbai), and *Z. pseudobailii* (Zpse). Letters a-n are ITS variant designations (Suh *et al.*, 2013). The tree was constructed by PhyML in the Seaview package using default parameters.



Supplementary Figure 2 (S2) Sequence alignment of the VPS1-MET14 intergenic regions from *Z. rouxii* and *Z. parabailii*. The *Z. parabailii* regions contain CEN4 and CEN11 whereas the *Z. rouxii* region is not a centromere. Putative CDE I and CDE III motifs are boxed.

Supplementary Table 1 *Z. parabailii* ATCC60483 orthologs of *S. cerevisiae* genes involved in mating or meiosis. https://doi.org/10.1371/journal.pbio.2002128.s003

Supplementary Table 2 Homeolog pairs in which one gene is damaged and the other is intact, in ATCC60483. https://doi.org/10.1371/journal.pbio.2002128.s004

Supplementary Table 3 PCR primer sequences used for MAT locus genotyping. https://doi.org/10.1371/journal.pbio.2002128.s005

Data S1. Excel spreadsheet containing, in separate sheets, the underlying numerical data for Figure panels 2A, 2B and 2C. https://doi.org/10.1371/journal.pbio.2002128.s006

References

ALTSCHUL, SF, GISH, W, MILLER, W*, et al.* 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-410.

BAKER, CR, BOOTH, LN, SORRELLS, TR & JOHNSON, AD 2012. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell,* 151**,** 80-95.

BARNETT, JA, PAYNE, RW & YARROW, D 1983. *Yeasts: Characteristics and identification*, Cambridge University Press.

BARNETT, JA, PAYNE, RW & YARROW, D 2000. *Yeasts: Characteristics and identification*, Cambridge University Press.

BIZZARRI, M, GIUDICI, P, CASSANELLI, S & SOLIERI, L 2016. Chimeric Sex-Determining Chromosomal Regions and Dysregulation of Cell-Type Identity in a Sterile *Zygosaccharomyces* Allodiploid Yeast. *PLoS One,* 11**,** e0152558.

BOETZER, M, HENKEL, CV, JANSEN, HJ*, et al.* 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics,* 27**,** 578-579.

BOSCO, G & HABER, JE 1998. Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics,* 150**,** 1037-1047.

BUTLER, G 2010. Fungal sex and pathogenesis. *Clin Microbiol Rev,* 23**,** 140-159.

BUTLER, G, KENNY, C, FAGAN, A*, et al.* 2004. Evolution of the *MAT* locus and its Ho endonuclease in yeast species. *Proc Natl Acad Sci U S A,* 101**,** 1632-1637.

BYRNE, KP & WOLFE, KH 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res,* 15**,** 1456-1461.

CAMPBELL, MA, GANLEY, AR, GABALDON, T & COX, MP 2016. The Case of the Missing Ancient Fungal Polyploids. *Am Nat,* 188**,** 602-614.

CHIN, CS, ALEXANDER, DH, MARKS, P*, et al.* 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods,* 10**,** 563-569.

DATO, L, BRANDUARDI, P, PASSOLUNGHI, S*, 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,* 10**,** 894-908.

DEAK, T & BEUCHAT, LR 1996. *Handbook of Food Spoilage Yeasts,* Boca Raton, Florida, CRC Press.

DELNERI, D, COLSON, I, GRAMMENOUDI, S*, et al.* 2003. Engineering evolution to study speciation in yeasts. *Nature,* 422**,** 68-72.

FARABAUGH, PJ, KRAMER, E, VALLABHANENI, H & RAMAN, A 2006. Evolution of +1 programmed frameshifting signals and frameshift-regulating tRNAs in the order Saccharomycetales. *J Mol Evol,* 63**,** 545-561.

GALEOTE, V, BIGEY, F, DEVILLERS, H*, et al.* 2013. Genome sequence of the food spoilage yeast *Zygosaccharomyces bailii* CLIB 213T. *Genome Announc,* 1**,** e00606-00613.

GORDON, JL, ARMISEN, D, PROUX-WERA, E*, et al.* 2011a. Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proc Natl Acad Sci U S A,* 108**,** 20024-20029.

GORDON, JL, BYRNE, KP & WOLFE, KH 2009. Additions, losses and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet,* 5**,** e1000485.

GORDON, JL, BYRNE, KP & WOLFE, KH 2011b. Mechanisms of chromosome number evolution in yeast. *PLoS Genet,* 7**,** e1002190.

GORDON, JL & WOLFE, KH 2008. Recent allopolyploid origin of *Zygosaccharomyces rouxii* strain ATCC 42981. *Yeast,* 25**,** 449-456.

GRABHERR, MG, HAAS, BJ, YASSOUR, M*, et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology,* 29**,** 644-U130.

GREIG, D, BORTS, RH, LOUIS, EJ & TRAVISANO, M 2002. Epistasis and hybrid sterility in *Saccharomyces*. *Proc Biol Sci,* 269**,** 1167-1171.
HAAS, BJ, DELCHER, AL, MOUNT, SM*, et al.* 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research,* 31**,** 5654-5666.
HABER, JE 2012. Mating-type genes and *MAT* switching in *Saccharomyces cerevisiae*. *Genetics,* 191**,** 33-64.
HERSKOWITZ, I 1988. Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol Rev,* 52**,** 536-553.
HITTINGER, CT 2013. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet,* 29**,** 309-317.
HUANG, X & MADAN, A 1999. CAP3: A DNA sequence assembly program. *Genome Res,* 9**,** 868-877.
HULIN, M & WHEALS, A 2014. Rapid identification of *Zygosaccharomyces* with genus-specific primers. *Int J Food Microbiol,* 173**,** 9-13.
HUNTER, N, CHAMBERS, SR, LOUIS, EJ & BORTS, RH 1996. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J,* 15**,** 1726-1733.
JAMES, SA, BOND, CJ, STRATFORD, M & ROBERTS, IN 2005. Molecular evidence for the existence of natural hybrids in the genus *Zygosaccharomyces*. *FEMS Yeast Res,* 5**,** 747-755.
JAMES, SA & STRATFORD, M 2011. *Zygosaccharomyces* Barker (1901). *In:* KURTZMAN, CP, FELL, JW & BOEKHOUT, T (eds.) *The Yeasts, a Taxonomic Study (5th edition).* Amsterdam: Elsevier.
JOHNSON, AD 1995. Molecular mechanisms of cell-type determination in budding yeast. *Curr Opin Genet Dev,* 5**,** 552-558.
JUNEAU, K, NISLOW, C & DAVIS, RW 2009. Alternative splicing of *PTC7* in *Saccharomyces cerevisiae* determines protein localization. *Genetics,* 183**,** 185-194.
KARIN, M, NAJARIAN, R, HASLINGER, A*, et al.* 1984. Primary structure and transcription of an amplified genetic locus: the *CUP1* locus of yeast. *Proc Natl Acad Sci U S A,* 81**,** 337-341.
KUDRJAWZEW, WI 1960. *Die Systematik der Hefen,* Berlin, Akademie Verlag.
KURTZMAN, CP & JAMES, SA 2006. *Zygosaccharomyces* and related genera. *In:* BLACKBURN, CDW (ed.) *Food spoilage microorganisms.* Boca Raton: CRC Press.
LARKIN, MA, BLACKSHIELDS, G, BROWN, NP*, et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.
LEH LOUIS, V, DESPONS, L, FRIEDRICH, A*, et al.* 2012. *Pichia sorbitophila*, an interspecies yeast hybrid, reveals early steps of genome resolution after polyploidization. *G3,* 2**,** 299-311.
LITI, G, BARTON, DB & LOUIS, EJ 2006. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. *Genetics,* 174**,** 839-850.
LOGUE, ME, WONG, S, WOLFE, KH & BUTLER, G 2005. A genome sequence survey shows that the pathogenic yeast *Candida parapsilosis* has a defective *MTL*a1 at its mating type locus. *Eukaryot Cell,* 4**,** 1009-1017.
LOWE, TM & EDDY, SR 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res,* 25**,** 955-964.
MALFEITO-FERREIRA, M, LOUREIRO-DIAS, MC & LOUREIRO, V 1997a. Weak acid inhibition of fermentation by *Zygosaccharomyces bailii* and *Saccharomyces cerevisiae*. *Int J Food Microbiol,* 36**,** 145-153.
MALFEITO-FERREIRA, M, LOUREIRO, V, WIUM, H & AUBYN, S 1990. A rapid characterization of yeast contaminants associated with sparkling wine production. *Industrie dell Bevande,* 19**,** 504-506.
MALFEITO-FERREIRA, M, TARECO, M & LOUREIRO, V 1997b. Fatty acid profiling: a feasible typing system to trace yeast contamination in wine bottling plants. *Int J Food Microbiol,* 38**,** 143-155.
MARCET-HOUBEN, M & GABALDON, T 2015. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biol,* 13**,** e1002220.

MEHRA, RK, GAREY, JR, BUTT, TR*, et al.* 1989. *Candida glabrata* metallothioneins. Cloning and sequence of the genes and characterization of proteins. *J Biol Chem,* 264**,** 19747-19753.
MIRA, NP, MUNSTERKOTTER, M, DIAS-VALADA, F*, et al.* 2014. The genome sequence of the highly acetic acid-tolerant *Zygosaccharomyces bailii*-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. *DNA Res,* 21**,** 299-313.
MOLLAPOUR, M & PIPER, P 2001. Targeted gene deletion in *Zygosaccharomyces bailii*. *Yeast,* 18**,** 173-186.
MORALES, L & DUJON, B 2012. Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiol Mol Biol Rev,* 76**,** 721-739.
MORI, H & WINDISCH, S 1982. Homothallism in sugar-tolerant *Saccharomyces rouxii*. *J Ferment Technol,* 60**,** 157-161.
NICKOLOFF, JA, SINGER, JD & HEFFRON, F 1990. In vivo analysis of the *Saccharomyces cerevisiae* HO nuclease recognition site by site-directed mutagenesis. *Mol Cell Biol,* 10**,** 1174-1179.
OHNISHI, H 1963. Osmophilic yeasts. *Adv Food Res,* 12**,** 53-94.
PALMA, M, ROQUE FDE, C, GUERREIRO, JF*, et al.* 2015. Search for genes responsible for the remarkably high acetic acid tolerance of a *Zygosaccharomyces bailii*-derived interspecies hybrid strain. *BMC Genomics,* 16**,** 1070.
PHAFF, HJ, MILLER, MW & MRAK, EM 1966. *The life of yeasts*, Harvard University Press.
PROUX-WERA, E, ARMISEN, D, BYRNE, KP & WOLFE, KH 2012. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics,* 13**,** 237.
PRYSZCZ, LP, NEMETH, T, GACSER, A & GABALDON, T 2014. Genome comparison of *Candida orthopsilosis* clinical strains reveals the existence of hybrids between two distinct subspecies. *Genome Biol Evol,* 6**,** 1069-1078.
PRYSZCZ, LP, NEMETH, T, SAUS, E*, et al.* 2015. The genomic aftermath of hybridization in the opportunistic pathogen *Candida metapsilosis*. *PLoS Genet,* 11**,** e1005626.
PUT, HMC & DE JONG, J 1982. Heat resistance studies of yeasts; vegetative cells versus ascospores: erythromycin inhibition of sporulation in *Kluyveromyces* and *Saccharomyces* species. *J Appl Microbiol,* 53**,** 73-79.
PUT, HMC, DE JONG, J, SAND, FE & VAN GRINSVEN, AM 1976. Heat resistance studies on yeast spp. causing spoilage in soft drinks. *J Appl Bacteriol,* 40**,** 135-152.
RODRIGUES, F, LUDOVICO, P, SOUSA, MJ*, et al.* 2003. The spoilage yeast *Zygosaccharomyces bailii* forms mitotic spores: a screening method for haploidization. *Appl Environ Microbiol,* 69**,** 649-653.
ROLLAND, T & DUJON, B 2011. Yeasty clocks: dating genomic changes in yeasts. *C R Biol,* 334**,** 620-628.
SANKOFF, D 1993. Analytical approaches to genomic evolution. *Biochimie,* 75**,** 409-413.
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,* 70**,** 6086-6091.
SCANNELL, DR, BYRNE, KP, GORDON, JL*, et al.* 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature,* 440**,** 341-345.
SCHRODER, MS, MARTINEZ DE SAN VICENTE, K, PRANDINI, TH*, et al.* 2016. Multiple origins of the pathogenic yeast *Candida orthopsilosis* by separate hybridizations between two parental species. *PLoS Genet,* 12**,** e1006404.
SHINOHARA, M, OH, SD, HUNTER, N & SHINOHARA, A 2008. Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. *Nat Genet,* 40**,** 299-309.
SOLIERI, L, CHAND DAKAL, T & GIUDICI, P 2013a. *Zygosaccharomyces sapae sp. nov.*, isolated from Italian traditional balsamic vinegar. *Int J Syst Evol Microbiol,* 63**,** 364-371.

SOLIERI, L, DAKAL, TC, CROCE, MA & GIUDICI, P 2013b. Unravelling genomic diversity of *Zygosaccharomyces rouxii* complex with a link to its life cycle. *FEMS Yeast Res,* 13**,** 245-258.
SOLIERI, L, DAKAL, TC, GIUDICI, P & CASSANELLI, S 2014. Sex-determination system in the diploid yeast *Zygosaccharomyces sapae*. *G3 (Bethesda),* 4**,** 1011-1025.
SOUCIET, JL, DUJON, B, GAILLARDIN, C*, et al.* 2009. Comparative genomics of protoploid Saccharomycetaceae. *Genome Res,* 19**,** 1696-1709.
STRATFORD, M, STEELS, H, NEBE-VON-CARON, G*, et al.* 2013. Extreme resistance to weak-acid preservatives in the spoilage yeast *Zygosaccharomyces bailii*. *Int J Food Microbiol,* 166**,** 126-134.
STRATHERN, JN & HERSKOWITZ, I 1979. Asymmetry and directionality in production of new cell types during clonal growth: the switching pattern of homothallic yeast. *Cell,* 17**,** 371-381.
SUH, SO, GUJJARI, P, BERES, C*, et al.* 2013. Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. *Int J Syst Evol Microbiol,* 63**,** 1922-1929.
THOMAS, DS & DAVENPORT, RR 1985. *Zygosaccharomyces bailii* -- a profile of characteristics and spoilage activities. *Food Microbiology,* 2**,** 157-169.
UTATSU, I, SAKAMOTO, S, IMURA, T & TOH-E, A 1987. Yeast plasmids resembling 2 micron DNA: regional similarities and diversities at the molecular level. *J Bacteriol,* 169**,** 5537-5545.
VAKIRLIS, N, SARILAR, V, DRILLON, G*, et al.* 2016. Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Res,* 26**,** 918-932.
VIGENTINI, I, BRAMBILLA, L, BRANDUARDI, P*, et al.* 2005. Heterologous protein production in *Zygosaccharomyces bailii*: physiological effects and fermentative strategies. *FEMS Yeast Res,* 5**,** 647-652.
WARRINGER, J, ZORGO, E, CUBILLOS, FA*, et al.* 2011. Trait variation in yeast is defined by population history. *PLoS Genet,* 7**,** e1002111.
WATANABE, J, UEHARA, K & MOGI, Y 2013. Diversity of mating-type chromosome structures in the yeast *Zygosaccharomyces rouxii* caused by ectopic exchanges between *MAT*-like loci. *PLoS One,* 8**,** e62121.
WENDLAND, J 2014. Lager yeast comes of age. *Eukaryot Cell,* 13**,** 1256-1265.
WENDLAND, J & WALTHER, A 2011. Genome evolution in the *Eremothecium* clade of the *Saccharomyces* complex revealed by comparative genomics. *G3: Genes, Genomes, Genetics,* 1**,** 539-558.
WOLFE, KH 2015. Origin of the yeast whole-genome duplication. *PLoS Biol,* 13**,** e1002221.
WOLFE, KH, ARMISEN, D, PROUX-WERA, E*, et al.* 2015. Clade- and species-specific features of genome evolution in the Saccharomycetaceae. *FEMS Yeast Res,* 15**,** fov035.
WOLFE, KH & SHIELDS, DC 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature,* 387**,** 708-713.
WOLTERS, JF, CHIU, K & FIUMERA, HL 2015. Population structure of mitochondrial genomes in Saccharomyces cerevisiae. *BMC Genomics,* 16**,** 451.
WU, B, BULJIC, A & HAO, W 2015. Extensive Horizontal Transfer and Homologous Recombination Generate Highly Chimeric Mitochondrial Genomes in Yeast. *Mol Biol Evol,* 32**,** 2559-2570.
YANG, Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol,* 24**,** 1586-1591.
ZHANG, H, MELTZER, P & DAVIS, S 2013. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics,* 14**,** 244.
ZHAO, Y, STROPE, PK, KOZMIN, SG*, et al.* 2014. Structures of naturally evolved *CUP1* tandem arrays in yeast indicate that these arrays are generated by unequal nonhomologous recombination. *G3 (Bethesda),* 4**,** 2259-2269.

Chapter 4

Transcriptional response to lactic acid stress in the hybrid yeast *Zygosaccharomyces parabailii*

Ortiz-Merino, R. A., Kuanyshev, N., Byrne, K. P., Valera J.A., Morrissey J.P., Porro, D., Wolfe, K. H. and Branduardi, P. (2017). **Transcriptional response to lactic acid stress in the hybrid yeast *Zygosaccharomyces parabailii*.** *Appl Environ Microbiol (submitted)*

R.A.O.-M. and N.K. contributed equally to this work

Transcriptional response to lactic acid stress in the hybrid yeast *Zygosaccharomyces parabailii*

[1]Raúl A. Ortiz-Merino, [2,3]Nurzhan Kuanyshev, [1]Kevin P. Byrne, [3]Javier A. Varela, [3,4]John P. Morrissey, [2]Danilo Porro, [1]Kenneth H. Wolfe, [2]Paola Branduardi

[1]UCD Conway Institute, School of Medicine, University College Dublin, Dublin, Ireland
[2]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milano, Italy
[3]School of Microbiology, University College Cork, Cork, Ireland
[4]School of Microbiology/Centre for Synthetic Biology and Biotechnology/Environmental Research Institute/APC Microbiome Institute, University College Cork, Cork, Ireland

Running head: Hybrid yeast transcriptomics under lactic acid stress

#Address correspondence to Paola Branduardi, paola.branduardi@unimib.it

R.A.O.-M. and N.K. contributed equally to this work.

Abstract

Lactic acid has a wide range of applications starting from its undissociated form and its production using cell factories requires stress-tolerant microbial hosts. The interspecies hybrid yeast *Zygosaccharomyces parabailii* has a great potential to be exploited as a novel host for lactic acid production, due to high organic acid tolerance at low pH, a remarkable osmotolerance, and a fermentative metabolism with a fast growth rate. Here we used RNA-seq to analyze *Z. parabailii*'s transcriptional response to lactic acid exposure, and we explore the biological mechanisms involved in tolerance. *Z. parabailii* contains two homeologous copies of most genes. Under lactic acid stress, the two genes in each homeolog pair tend to diverge in expression to a significantly greater extent than in control conditions, indicating that stress tolerance is facilitated by interactions between the two gene sets in the hybrid. Lactic acid induces downregulation of genes related to cell wall and plasma membrane functions possibly altering the rate of diffusion of lactic acid into cells. Genes related to iron transport and redox processes were upregulated, suggesting an important role for respiratory functions and oxidative stress defense. We found differences in the expression profiles of genes putatively regulated by Haa1 and Aft1/2, previously described as lactic acid-responsive in *Saccharomyces cerevisiae.* Furthermore, formate dehydrogenase (*FDH*) genes form a lactic acid-responsive gene family that has been specifically amplified in *Z. parabailii* as compared to other closely related species. Our study provides a starting point for engineering *Z. parabailii* as a host for lactic acid production.

Importance

Hybrid yeasts are important in biotechnology because of their tolerance to harsh industrial conditions. The molecular mechanisms of tolerance can be studied by analyzing differential gene expression in conditions of interest, and relating gene expression patterns to protein functions. However, hybrid organisms present a challenge to the standard use of mRNA sequencing (RNA-seq) to study transcriptional responses to stress, because their genomes contain two similar copies of almost every gene. Here we used stringent mapping methods and a high-quality genome sequence to study the transcriptional response to lactic acid stress in *Zygosaccharomyces parabailii* ATCC60483, a natural interspecies hybrid yeast that contains two complete subgenomes that are approximately 7% divergent in sequence. Beyond the insights we gained into lactic acid tolerance in this study, the methods we developed will be broadly applicable to other yeast hybrid strains.

Introduction

Species belonging to the *Zygosaccharomyces bailii sensu lato* clade have a remarkable resilience against stress induced by weak acids, some of which are widely used as food preservatives and/or are versatile chemical platforms (Martorell *et al.*, 2007, Stratford *et al.*, 2013). Therefore, on the one hand these yeasts represent a challenging problem in the food industry because they are often found as contaminants in production pipelines for wine, high sugar products, and canned foods. On the other hand, they are promising cell factories for biotechnological applications involving organic acids that can be produced by microbial fermentation (Kuanyshev *et al.*, 2017, Becker *et al.*, 2015) or released by lignocellulosic pretreatment of biomass (Limayem and Ricke, 2012).

Lactic acid is one of the useful organic acids that can be produced by yeasts as a microbial factory. This compound has a wide range of industrial applications including food preservation, additives and pharmaceuticals (Castillo Martinez *et al.*, 2013), and potential to be used for bioplastic production from a renewable source (Sauer *et al.*, 2010). Natural fermentation by lactic acid bacteria has long been the main source of industrial lactic acid production (Datta and Henry, 2006). Nevertheless, the latter approach has low cost effectiveness due to complex nutritional requirements and low final product purity (Fitzpatrick *et al.*, 2003), together with the need to convert lactate to lactic acid, whereas engineered yeast

platforms cultivated at pH well below the pK_a of lactic acid (3.78) have already shown promising potential (Sauer *et al.*, 2010, Chen and Nielsen, 2016). The production of several weak organic acids, including lactic acid, has reached the industrial scale (Becker *et al.*, 2015) but there is still room for further production improvement by enhancing production host robustness and/or exploiting novel microbial hosts. Therefore, understanding the mechanism of weak acid tolerance in non-*Saccharomyces* yeasts such as *Zygosaccharomyces* is important for the future development of ultra-efficient production platforms.

The mechanisms of weak acid stress tolerance and response have been studied extensively in the model yeast *S. cerevisiae* (Giannattasio *et al.*, 2013, Mira *et al.*, 2010, Piper *et al.*, 2001, Berterame *et al.*, 2016, Martani *et al.*, 2015). However, this knowledge is far from complete and cannot be applied easily to non-*Saccharomyces* species. Previous research on tolerance to weak organic acids revealed the capability of *Z. bailii sensu lato* to catabolize acetic and benzoic acids even in the presence of glucose (Rodrigues *et al.*, 2012, Mollapour and Piper, 2001). In addition, different *Z. bailii* strains display specific adaptation traits such as the ability to modulate their cell wall and membrane composition in order to decrease the influx of weak acids (Lindberg *et al.*, 2013, Kuanyshev *et al.*, 2016).

Importantly, the *Z. bailii sensu lato* clade is characterized by substantial genetic diversity. Some strains that were previously considered to be *'Z. bailii*' were reclassified in 2013 into two new species called *Z. parabailii* and *Z. pseudobailii* (Suh *et al.*, 2013). The name '*Z. bailii sensu lato*' is used to refer to the species complex that includes these two new species as well as other strains that were not reclassified (*Z. bailii sensu stricto*). The widely studied strains CLIB213^T and IST302 are *Z. bailii sensu stricto* (Galeote *et al.*, 2013, Palma *et al.*, 2017b). The strains ATCC60483 (used in this study) and ISA1307 are *Z. parabailii*, which is a hybrid that was formed naturally by mating between *Z. bailii sensu stricto* and an unidentified *Zygosaccharomyces* species (Mira *et al.*, 2014, Ortiz-Merino *et al.*, 2017). *Z. parabailii* genomes contain two copies of almost every gene, differing by 7% in nucleotide sequence on average (Ortiz-Merino *et al.*, 2017). These genes are referred as homeologs given they are derived from different organisms and to distinguish them from paralogs which are originated from one particular organism (Glover *et al.*, 2016).

We are exploring the possibility of using *Z. bailii sensu lato* species as alternative yeast hosts for lactic acid production. We focused on *Z. parabailii* strain ATCC60483 because our previous work demonstrated its high tolerance to lactic acid at low pH, characterized by growth without any detectable lag phase or acid consumption (Kuanyshev *et al.*, 2016), under microaerobic conditions. These natural characteristics are promising in terms of possible exploitation for organic acid production only if the molecular basis of its unusual tolerance to low pH, high inhibitor concentrations, and other traits of interest are clarified. As a preliminary step towards metabolic engineering, in this study we sought to investigate the molecular mechanisms of lactic acid tolerance in ATCC60483 by means of RNA-seq. In general, we found that the *Z. parabailii* transcriptome responds to lactic acid stress by inducing genes related to oxidative stress response and iron homeostasis in a different way than *S. cerevisiae* does. In addition, *Z. parabailii* modulates the transcription of genes related to the cell wall, in agreement with our previous data.

Results

Transcriptional profile of *Z. parabailii* homeolog pairs and duplicated genes in lactic acid stress

Our previous study showed that lactic acid at a concentration of 40 g L^{-1} does not affect *Z. parabailii* ATCC60483 cell viability but exerts phenotypic and morphological changes (Kuanyshev *et al.*, 2016). Our aim here was to study this tolerance response, by comparing the transcriptomes of cultures grown in the presence or absence of lactic acid (40 g L^{-1}) at time points (18 h and 42 h) specifically chosen to ascertain comparable growth kinetics and exclude growth phase related bias (**Fig. 1**). After normalizing and filtering the raw RNA-seq read counts, we detected expression for >95% of the *Z. parabailii* genes in at least one condition, including 36 genes that were transcribed only in lactic acid and 31 that were transcribed only in control conditions (**Table 1**).

We used stringent mapping of RNA-seq reads to the genome (see Methods), in order to capture expression differences between homeologous gene pairs even when they are highly similar in sequence. About 82% of the 10,072 genes in the *Z. parabailii* nuclear genome show the pattern characteristic of hybrid genomes, forming pairs of ‘A’ and ‘B’ homeologs, where

the A-gene came from one parent in the hybridization and the B-gene came from the other (Ortiz-Merino *et al.*, 2017). Most of the remaining loci in the genome are also present in two copies, but are either A:A or B:B pairs due to loss of heterozygosity after hybridization (Ortiz-Merino *et al.*, 2017). We calculated the ratio of expression between each of 4136 A:B homeolog pairs as described in the methods section and the values are presented in **Fig. 2**. All but 21 pairs showed evidence of expression of both homeologs.

Strikingly, the distribution of expression ratios is broader in lactic acid than in control conditions, at both time points. In other words, in the stress condition one of the two genes in each homeolog pair tends to become predominantly expressed. If we define unbalanced expression as an expression ratio that lies outside the range 0.4-0.6 (**Fig. 2**), the proportion of homeolog pairs with unbalanced expression is 13.8-18.7% in the control conditions but increases to 31-33.4% in lactic acid conditions. The difference in variance of expression ratios is statistically significant (Fligner-Killeen test; P value = 1.5e-98 at 18 h, and P value = 4.5e-61 at 42 h).

The distribution of expression ratios is approximately symmetrical (**Fig. 2**) indicating that in some homeolog pairs the A-gene is more highly expressed than the B-gene (ratios > 0.5) whereas in others the B-gene is higher (ratios < 0.5). The A-genes were derived from the parental species *Z. bailii* in the hybridization, and the B-genes were derived from the other parent (an unidentified *Zygosaccharomyces* species) (Ortiz-Merino *et al.*, 2017). Thus, broadly speaking the cell responds to lactic acid stress by inducing greater divergence of expression between the genes in a homeolog pair, without a strong preference as to whether the A-gene or the B-gene is the higher-expressed one. However, statistical analysis indicates a weak bias towards A-genes, derived from the *Z. bailii* parent. All four distributions in **Fig. 2** have slight but consistent negative skew values, indicating a trend towards higher expression of the A-genes (**Table 2**). This difference is also illustrated by the larger numbers of loci for which the Reads Per Kilobase of transcript per Million mapped reads (RPKM) for the A-gene exceeds that of the B-gene, as opposed to the converse (**Table 2**, Ab and aB columns). A-genes showing higher RPKM values than B-genes are greater in lactic acid, this is significant at 18 h and borderline significant at 42 h (**Table 2**). In summary, *Z. parabailii* has a slight tendency to express its A-genes more highly than its B-genes, this tendency is maintained under lactic acid

stress, but the magnitude of this tendency is small compared to the grossly increased divergence of expression levels between homeologs that occurs in lactic acid stress.

Our method of high-stringency mapping of RNA-seq reads to a high-quality genome sequence detected transcriptional profiles of homeologous gene pairs even with highly similar genes. Nevertheless, we needed to modify it to determine read counts for genes that occur in identical pairs (See Methods). Specifically, among the 402 genes that have no evidence of expression in the "full" count dataset, 42% are genes that were affected by loss of heterozygosity (genes in A:A pairs or B:B pairs). The modified method enabled us to measure the combined expression of 230 duplicated genes in *Z. parabailii* including the two orthologs of the *S. cerevisiae* major mitochondrial D-lactate dehydrogenase *DLD1* (*I04780_A* and *N05010_A*) and the minor isoform *DLD2* (*B01190_N* and *G05430_N*). Although we cannot investigate the expression of these identical gene pairs individually, their RPKM values are low when compared with all duplicated genes on lactic acid conditions (*DLD1* combined=1317, *DLD2* combined=710.5; mean RPKM on lactic acid at 18 h = 4028, mean RPKM on lactic acid at 42 h = 9230). Furthermore, *DLD1* shows a statistically significant 2-fold expression decrease in lactic acid at both timepoints whereas *DLD2* shows no significant expression changes (**Data Set S4**). These observations are consistent to the previously reported lack of lactic acid consumption of *Z. parabailii* (Kuanyshev *et al.*, 2016).

Upregulated genes are related to oxidation-reduction processes and ion transport in the mitochondria.

The "full" set of counts (See Methods) for the 9,683 genes in the union set of expressed genes were then used for differential expression analysis, filtering the results for adjusted P value < 0.05 and |log-fold change| ≥ 1 (**Data Set S5**). This analysis is completely independent from that for the abovementioned duplicated genes and identified a total of 227 genes upregulated in lactic acid, of which 117 are specific to 18 h and 83 to 42 h (**Table 3**). Similarly, a total of 1019 downregulated genes were found, including 430 specific to 18 h and 431 to 42 h. We then performed a Gene Ontology (GO) term enrichment analysis to identify GO terms that were enriched at both time points in either the upregulated genes (**Fig. 3A**) or the downregulated genes (**Fig. 3B**).

When we refer to a *S. cerevisiae* gene in the following functional analysis, we refer to either one or both of its orthologs in a *Z. parabailii* homeolog pair. These genes are referred in our functional analysis if at least one of the members of a *Z. parabailii* homeolog pair was differentially expressed.

The enriched GO term associated with the highest number of upregulated genes in our dataset is GO:0055114 for "oxidation-reduction process" (**Fig. 3A**, **Data Set S6**). This term is associated with 33 genes including homologs of the *S. cerevisiae* genes *GOR1*, *AIM17*, *CCP1*, *MET13*, *SOD2*, *SOD1*, *GND1/2*, and *GRX1/2*, some of which are also related to enriched mitochondrial terms (GO:0005758 for example). We also observed enrichment for genes in the glyoxylate cycle (GO:0006097) and the glyoxysome (GO:0009514) including homologs of *ICL1* and *IDP2*. The upregulation of these genes along with *FBP1* could indicate activation of the anaplerotic reactions, probably caused by oxygen limitation. *ACH1* with CoA transferase activity, and *Z. parabailii* gene *L05300_N* predicted to have epoxide hydrolase activity, were upregulated at both time points and are presumably involved in enzymatic detoxification process.

The siderophore transmembrane transport term (GO:0044718) was also found enriched in upregulated genes. Genes in this category are members of the MFS_1 family of transporters, potentially involved in iron retention and/or transport (genes *A10040_B*, *B02380_A*, *G04250_B*, *I00120_N*, *I05800_A*, *O00120_N*), and upregulated at 18 h. These genes are all classified as integral components of the membrane. Other genes specifically upregulated at 42 h include *FIT2*, *STL1* and *K05040_N* which shows no sequence homology to *S. cerevisiae* genes but is predicted to be a transmembrane transporter (see Methods).

Downregulated genes are mainly related to components of the cell boundaries and protein translation.

The GO term enrichment analysis for downregulated genes showed 63 genes related to ribosomal functions (GO:0003735), and 38 to cytoplasmic translation (GO:0002181) (**Fig. 3B**, **Data Set S6**). Most of those genes were downregulated at 42 h, implying a general decrease in protein synthesis. This response seems to correspond to a general mechanism observed also in other yeasts used as cell factories, *e.g. S. cerevisiae* under stress conditions (Simpson and

Ashe, 2012) and *Komagataella phaffii* (*Pichia pastoris*) used for heterologous protein production induced by methanol (Sauer *et al.*, 2004) during stress, possibly related to resilience or energy maintenance. Some of these genes are also related to the enriched terms GO:0000932 and GO:0010494 for P-body and cytoplasmic stress granules known to be involved in mRNA translation and turnover during different stress conditions in *S. cerevisiae* (Decker and Parker, 2012). These categories are downregulated at 18 h in lactic acid treated cells. One of the components of stress granules that is also downregulated at 42 h encodes for a homolog of *S. cerevisiae* Pab1, the major polyA binding protein which has been demonstrated to promote the formation of stress granules (Swisher and Parker, 2010). A recent study conducted in *S. cerevisiae* reported that stress granules are not formed in lactic acid treated cells (Iwaki and Izawa, 2012) and a similar situation might be also true for *Z. parabailii.*

Among the downregulated genes, we also identified many with functions that we summarize as being related to the boundaries of the cell, *i.e.* the cell wall and the plasma membrane. The GO terms in this group include the actin cortical patch (GO:0030479), cell cortex (GO:0005938), extracellular region (GO:0005576), fungal-type cell wall (GO:0009277), and structural constituents of the cell wall (GO:0005199) (**Fig. 3B**). Consistent with this, we noticed enrichment of the GO terms for glucan endo-1,3-beta-D-glucosidase activity (GO:0042973) and chitin binding (GO:0008061). These observations indicate that the cell wall is modulated upon lactic acid stress, in agreement with our previous findings (Kuanyshev *et al.*, 2016). Other genes downregulated at 42 h predicted to be integral components of the membrane are *H01670_B* with unknown function, *OPT1*, and *HBT1*. We also found downregulation of *CWP1*, a cell wall protein homolog, and *LDS2*, which is involved in the assembly of the *S. cerevisiae* spore wall.

Involvement of Haa1 and Aft1/Aft2 regulated genes in lactic acid stress response

Previous studies on lactic acid stress response mechanisms in *S. cerevisiae* indicated an important role of the transcription factors Haa1, Aft1/Aft2 (Abbott *et al.*, 2008, Kawahata *et al.*, 2006). Therefore, we extracted all the *S. cerevisiae* genes reported to be targets of either Haa1 or Aft1/Aft2 in YEASTRACT (Teixeira *et al.*, 2006), in addition to those identified as lactic acid-responsive (Abbott *et al.*, 2008). We then tested whether the *Z. parabailii* orthologs

of these *S. cerevisiae* genes are differentially expressed in our dataset. In this case we ignored the log-fold change cut-off to enable detection of small but still significant changes. The results are shown in **Data Set S7**.

We found differential expression of 42 orthologs of *S. cerevisiae* genes putatively controlled by Haa1 (**Fig. 4A**). These include the membrane-bound and major weak acid response genes *YPC1/YDC1*, *TPO2/3*, *VPS62/TDA6*, *PDR16*, and *PDR12*. This set also includes the transcription factors *MSN4/2*, *COM2*, and the transcription factor itself (*HAA1/CUP2*). Interestingly, we observed the major weak acid stress response genes, *TPO2/TPO3* and *SPS100/YGP1* as downregulated, although they are upregulated during lactic acid stress in *S. cerevisiae* (Abbott *et al.*, 2008). Here we also found *PFK27* with a response changing from upregulated at 18 h to downregulated at 42 h, and *MTH1/STD1* going from downregulated at 18 h to upregulated at 42 h. These changes in glucose-responsive genes could possibly reflect the diauxic shift.

We performed a similar search strategy as above for genes putatively under the control of Aft1/Aft2 which in *S. cerevisiae* are related to iron utilization and homeostasis (Rutherford *et al.*, 2003). Results are shown in **Fig. 4B** and **Data Set S7**. This revealed orthologs of 27 *S. cerevisiae* genes where 6 are downregulated at both time points: *AFT1/AFT2* coding for the transcription factor itself, *AKR1/AKR2* an integral component of the membrane with palmitoyltransferase activity, *LEU2* involved in leucine biosynthesis, *MRS3/MRS4* iron transporters, *APE1* with metalloaminopeptidase activity, and *AHP1*, a thiol-specific peroxiredoxin. The rest of these genes are upregulated, at both timepoints: *TIS11/CTH1* involved in mRNA processing, *CCC2* a Cu^{++}-transporting P-type ATPase, *UBC8* which negatively regulates gluconeogenesis, *ECL1* that increases oxygen consumption and respiratory activity, *SMF3* a putative divalent metal ion transporter, *ARA2*, a NAD-dependent arabinose dehydrogenase, *FET3*, a Ferro-O_2-oxidoreductase, and *PEP4*, a vacuolar protease. Most of these upregulated genes are related to ion transport and redox functions, in agreement with our GO term enrichment analysis.

Multi-gene families significantly modulated upon lactic acid exposure

We identified an unusual regulatory pattern in a family of genes related to *S. cerevisiae FDH1*, which codes for formate dehydrogenase. The *Z. parabailii* genome contains six genes in this family (*I01900_B*, *O01850_A*, *P02220_N*, *H05680_N*, *N02280_N*, and *F04070_N*) although formate dehydrogenase activity has not been demonstrated for any of them. All six *FDH*-like genes were highly upregulated at 18 h of lactic acid exposure, and the last two were also significantly downregulated at 42 h (**Data Set S5**). Formate dehydrogenases perform the NAD^+ dependent oxidation of formate to carbon dioxide. The *S. cerevisiae* strain CEN.PK 113-7D contains two *FDH* genes (*FDH1* and *FDH2*), whereas only *FDH1* is intact in the laboratory strain BY4741 because *FDH2* is truncated (Overkamp *et al.*, 2002). The function of *FDH* genes in *S. cerevisiae* is not well characterized, but these enzymes have been better studied in methylotrophic yeasts such as *Komagataella phaffi* where they are involved in the last step of the methanol dissimilation pathway (Tishkov and Popov, 2004).

Interestingly, the phylogenetic distribution of *FDH* genes among sequenced yeast genomes is rather patchy (Tishkov and Popov, 2004) and indicative both of recent gene amplifications and of multiple gene losses. We searched for *FDH* homologs in the NCBI databases and constructed a phylogenetic tree (**Fig. 5**). Many yeast species lack *FDH* genes completely, containing only homologs of distantly related genes such as *GOR1* (glyoxylate reductase). Nevertheless, the phylogenetic relationship among the *FDH* genes of the few species that do retain this gene agrees well with the expected relationship among these species (**Fig. 5**). This observation suggests that the patchy distribution of the gene is due to numerous losses of an ancestral *FDH* gene (for example, in the genera *Torulaspora*, *Lachancea* and *Kluyveromyces*), and not the result of horizontal gene transfer. There is essentially no conservation of synteny among the existing *FDH* genes, which shows that multiple species-specific gene duplications and gene relocations have occurred. Of the six *Z. parabailii FDH*-like genes, four are closely related and form a phylogenetic cluster with *Saccharomyces* species (**Fig. 5**). The other two form a cluster with the only *FDH*-like gene we identified in the genome of $CLIB213^T$, a *Z. bailii sensu stricto* strain. The sister species *Z. rouxii* has four *FDH*-like genes that cluster together in the tree. Thus, amplifications of *FDH*-like genes by gene duplication have occurred separately in *Z. parabailii* and *Z. rouxii*, and in the former species they are highly induced by lactic acid. This difference in *FDH* gene copy number between *Z. parabailii* and

Z. bailii may be a contributory factor to the difference in their tolerance to lactic acid as our previous study showed that *Z. parabailii* ATCC60483 is more resilient to lactic acid than the *Z. bailii sensu stricto* strains ATCC8766 and ATCC58445^T (synonymous with CLIB213^T) (Kuanyshev *et al.*, 2016).

We searched systematically for other *Z. parabailii* genes assigned into multigene families, which have significant expression changes in lactic acid. This was done by searching for sets of 3 or more *Z. parabailii* genes that share the same *Z. bailii* ortholog. We examined a total of 123 *Z. parabailii* genes in multigene families of this type, of which 22 are differentially expressed in at least one timepoint (classified as category 9 in **Data Set S5**). These 22 genes belong to 12 different multigene families significantly modulated in lactic acid. For example, *F06230_N*, *N00190_A*, and *O04100_A* are homologs of the *FFZ2* transporters specific to *Zygosaccharomyces* species and able to transport both fructose and glucose when overexpressed in *S. cerevisiae* (Leandro *et al.*, 2011, Cabral *et al.*, 2015). In this family, *F06230_N* is upregulated in lactic acid at both time points whereas *N00190_A* is upregulated only at 18 h, and *O04100_A* did not show significant expression changes. Another interesting family is *A10020_N*, *G00240_N*, and *P00180_N* which are all significantly upregulated in lactic acid (when ignoring the log-fold change cut-off) and are homologous to the iron siderophore transporter *FIT2* putatively under control of Aft1/2 (**Fig. 4B, Data Set S7**). This family also includes *K00140_A* and *C00210_N* for which we did not observe any evidence of expression. Furthermore, given that *K00140_A* is identical to the only *FIT2* homolog annotated in *Z. bailii* strain CLIB213^T (BN860_19394g1_1) and it is not differentially regulated, there might be certain functional relevance only for genes specific to *Z. parabailii*.

Discussion

We found that lactic acid stress induces robust and statistically significant divergent expression responses between the two genes in homeologous gene pairs in *Z. parabailii*. These differences need to be further explored when considering differentially expressed genes as engineering targets, but the overall stress response we saw amongst them is striking. Homeologous gene pairs are present in all hybrid (allopolyploid) organisms (Glover *et al.*, 2016). Most previous transcriptomic analyses including homeolog pairs have been carried out in plant species (Rapp *et al.*, 2009, Yoo *et al.*, 2013, Combes *et al.*, 2015), although there are

examples with fungi (Cox *et al.*, 2014) and yeasts (Tirosh *et al.*, 2009, Wang *et al.*, 2015). We are not aware of any previous studies that found a similar genome-wide increase in homeolog expression divergence under stress conditions. Our study differs from the previous work on yeast hybrids as we examined gene expression in a natural hybrid isolate whereas preceding studies analyzed synthetic hybrids (Tirosh *et al.*, 2009, Wang *et al.*, 2015). Furthermore we compared expression between homeolog pairs under two different growth conditions while previous comparisons where done against the parental genes (Tirosh *et al.*, 2009, Wang *et al.*, 2015), even when using more than one condition (44).

We observed upregulation of genes related to reactive oxygen species (ROS) detoxification which could be linked to the respiratory chain upregulation. Lactic acid stress has been reported to imbalance the prooxidant/antioxidant ratio (Piper, 1999), and trigger the accumulation of ROS via the Fenton reaction (Ali *et al.*, 2000). Accordingly, overexpression of cytosolic catalase or introduction of the pathway for biosynthesis of L-ascorbic acid (a well-known antioxidant) into *S. cerevisiae* improved resistance to oxidative and lactic acid stress (Abbott *et al.*, 2009, Branduardi *et al.*, 2007). Therefore, an increase in ROS detoxification can help to alleviate lactic acid stress. The upregulation of the *FDH* multigene family in *Z. parabailii* could be related to the ROS detoxification. One hypothesis for this involves ROS detoxification by ketoacids leading to formate accumulation, which consequently catabolized by Fdh to NADH and CO_2 (Yokota *et al.*, 1983). This mechanism was described in the bacterium *Pseudomonas fluorescens* as an anti-oxidative defence mechanism (Thomas *et al.*, 2016, Alhasawi *et al.*, 2015) and we speculate that the multiple Fdh enzymes in *Z. parabailii* might serve a similar role.

There are significant differences between the response to lactic acid that we observed in *Z. parabailii* and the responses previously reported in *S. cerevisiae* (Kawahata *et al.*, 2006, Abbott *et al.*, 2008). While many of these differences may reflect differences in the physiology of the two species, there were also differences in the experimental setup used. We used microaerobic conditions, whereas previous studies used anaerobic chemostat conditions (Abbott *et al.*, 2008), and batch flask fermentation (Kawahata *et al.*, 2006). Nevertheless, we also identified some similarities between the lactic acid responses in *S. cerevisiae* and *Z. parabailii*, involving iron homeostasis genes such as siderophore transporters and iron

transporters. It has been shown that high concentration of lactate ions in the growth medium chelates free iron reducing its availability for cellular functions (Abbott *et al.*, 2008), and triggering a strong regulation of iron homeostasis in *S. cerevisiae* (Abbott *et al.*, 2008, Kawahata *et al.*, 2006).We observed a similar response to lactic acid stress in *Z. parabailii*.

Z. parabailii cell wall modulation shows to be a response towards lactic acid stress. The cell wall is generally considered as a barrier for large molecules (Aguilar-Uscanga and Francois, 2003, Lesage and Bussey, 2006). Nevertheless, studies on *S. cerevisiae* have reported regulation of genes coding for cell wall components (Simoes *et al.*, 2006), or related to cell wall integrity (Rego *et al.*, 2014) as a response to acetic acid or a low pH environment (Kapteyn *et al.*, 2001). In *Z. parabailii*, downregulation of cell wall related genes can be linked to the decrease of cell wall mannoproteins and β 1→3 glucan levels observed in our previous FTIR analysis (Kuanyshev *et al.*, 2016), and together with the peculiar plasma membrane composition (Lindberg *et al.*, 2013) this can contribute to the superior lactic acid tolerance of this yeasts compared to the baker's yeast.

The expression of Haa1 regulated genes during the stress in *Z, parabailii* is rather different from S. *cerevisiae*. Haa1 is a transcriptional activator of genes responsive to acetic and lactic acid in *S. cerevisiae* (Abbott *et al.*, 2008, Mira *et al.*, 2010, Keller *et al.*, 2001) and in *Z. bailii* (only in acetic acid) (Palma *et al.*, 2017a, Palma *et al.*, 2017b). It is intriguing to observe a different expression patterns for those genes in *Z. parabailii* during lactic acid stress and further studies are necessary to explain these observations.

Our study is a pioneering approach to examining a hybrid yeast response to lactic acid stress. This was possible by the availability if a high quality genome reference (Ortiz-Merino *et al.*, 2017) which is often not the case for other hybrid organisms. It also required highly-stringent and tailored methods to study the expression of highly similar genes and even identical copies. With this we showed that homeolog gene pairs have different expression patterns when subjected to acid stress: this could reflect or override transcriptional control mechanisms inherited from the parental strains of this hybrid. This hybrid nature is one of a few differences we observed in comparison with the lactic acid response reported for *S. cerevisiae* and *Z. bailii*. Our observations still need experimental validation given that changes in transcript

levels are not always reflected in protein activities *in vivo*. Nevertheless, our observations for the duplicated homologs of *DLD1* and *DLD2* being lowly expressed, and even repressed, in lactic acid are consistent to the previously reported lack of lactic acid consumption of *Z. parabailii* (Kuanyshev *et al.*, 2016) which is a key feature needed for a lactic acid producing host. Our study provides methods and data to facilitate the understanding of molecular responses during acid stress in this or other hybrid yeasts, which is important both for fundamental and applied science.

Materials and Methods

Cell growth, RNA extraction and sequencing

Z. parabailii strain ATCC60483 was used for bioreactor fermentation. Cell aliquots, stored at -80°C in YPD glycerol stock, were grown to mid exponential phase before being inoculated to the bioreactor at final absorbance of OD_{660} 0.1. We used 2x Verduyn growth medium (Verduyn *et al.*, 1992) at pH 3 containing 40 gL^{-1} glucose with 40 gL^{-1} lactic acid or no lactic acid (control condition) and the fermentations were performed in 2 L volume bioreactors (BIOSTAT B, Sartorius AG, Germany) with operative volume of 1.5 L. The temperature was maintained at 30°C, pH at 3 by the addition of 4 M NaOH and the stirrer speed was set to 400 rpm. The inlet gas flow was adjusted by two mass flow controllers (Bronkhornst®High Tech-EL-FLOW®Select). The mass flow was set to obtain a mixture of N_2 and air with final concentration of inlet oxygen of 5%. The mixture was sparged at 0.75 vvm. Antifoam (Antifoam 204, Sigma Aldrich) was used for foaming control.

The samples for RNA sequencing were taken in triplicate at 18 h and 42 h from the bioreactor fermentation, corresponding to log phase and post diauxic shift, respectively (Kuanyshev *et al.*, 2016). The total RNA was then extracted using Zymo Research Fungal/Bacterial RNA MiniPrep™ kit (Irvine, USA) and the quality of RNA samples were evaluated with Agilent Bioanalyzer. The RNA samples were sequenced using the Illumina HiSeq2000 platform with 100 nt-long paired-end reads at Parco Tecnologico Padano (Lodi, Italy).

RNA-seq analysis

We used our recently published *Z. parabailii* ATCC60483 genome annotation as a reference (Ortiz-Merino *et al.*, 2017). This consists of 10,072 nuclear and 13 mitochondrial protein-coding obtained using an improved version of the Yeast Genome Annotation Pipeline (Proux-

Wera *et al.*, 2012) and includes additional metadata as an aid for functional interpretation. Briefly, because of its hybrid nature, the *Z. parabailii* genome contains two homeologous copies of most genes. We use suffixes _A and _B in gene names to indicate the two copies, where _A indicates gene copies that are virtually identical to their *Z. bailii sensu stricto* orthologs, and _B indicates copies that are more divergent (5-25% synonymous sequence divergence). A few genes have the suffix _N because they could not be assigned to either of these two groups.

Some extra information was added to the original annotation and is contained in **Data Set S1**. This includes functional domains and protein family memberships obtained by aligning all the *Z. parabailii* ATCC60483 amino acid sequences against the PFAM database (Finn *et al.*, 2014) using HMMER v. 3.0 (Finn *et al.*, 2011). A genome-wide annotation of transmembrane proteins was also performed by comparing the *Z. parabailii* proteome against the TransportDB 2.0 (Elbourne *et al.*, 2017) database using BLAST v. 2.2.22(Altschul *et al.*, 1990). The sequences were then filtered based on identity (>35 %) and coverage (>80%) and submitted to the TMHMM server v. 2.0 (Krogh *et al.*, 2001) to determine a minimum of 2 potential transmembrane domains per sequence. Blast2GO (Gotz *et al.*, 2008) was then used to generate a custom Gene Ontology (GO) annotation for *Z. parabailii* available in **Data Set S2**.

The raw RNA-seq reads were mapped against the *Z. parabailii* ATCC60483 nuclear and mitochondrial genome (Ortiz-Merino *et al.*, 2017) using bowtie v1.1.2 (Langmead, 2010) with the parameters -v 0 -k 10 --best -M 1. The parameter -v 0 gives high stringency by allowing no mismatches in the alignments discriminating between highly similar regions in the genome, and discarding reads with sequencing artefacts. The parameters -k 10 --best -M 1 report only the best possible alignment out of up to 10 alternatives and, in case there are two equivalent best hits, only one is reported at random. This reports so-called multi-mapping reads with the tag “XM:i:2” and a mapping quality (MAPQ) equal to 0.

The mapped reads were subsequently counted using htseq-count v0.6.0 (Anders *et al.*, 2015) using two different settings. In the first case htseq-count was used over the full set of *Z. parabailii* genes with default parameters and referred as “full” counts. This setting discards the alignments for multi-mapping reads as their quality is artificially set to the lowest possible

value. Therefore, a different htseq-count run was performed with the parameter -a 0 allowing for MAPQ >= 0. To avoid spurious low-quality alignments, this second run used only the alignments with the "XM:i:2" tag and a subset of 232 duplicated genes defined by showing full-length hits and 100% blastn (Altschul *et al.*, 1990) nucleotide sequence identity with one or more different *Z. parabailii* genes. The counts from this second htseq-run over duplicated genes with multi-mapping reads and are referred as "duplicated" counts. The duplicated counts represent a homogeneized signal of two or more identical genes and potential different quality values which is not the case with the "full" counts. Therefore, the two sets of counts were analysed completely separate. All the analyses reported were done with the full counts unless stated otherwise.

The read counts were split in 4 groups according to condition and time point, each group containing 3 libraries. One of the libraries for the control condition at 18 h contained few reads (5.9 million in total compared to the average of 30.5 million from the other libraries) and was excluded from further analyses. We therefore used the TMM method (Robinson and Oshlack, 2010) implemented in edgeR v. 3.18.1 (McCarthy *et al.*, 2012) to normalize the read counts and provide better comparability across our different sized samples. Counts per million (CPM) were then calculated from the normalized counts using edgeR v. 3.18.1 (McCarthy *et al.*, 2012) and the genes with less than 1 CPM in at least 3 samples from the same condition were considered to have no evidence of expression. We also calculated Reads Per Kilobase of transcript per Million mapped reads (RPKM) using edgeR v. 3.18.1 (McCarthy *et al.*, 2012). This was done for the normalized and filtered sets of "full" counts for the 4139 homeolog pairs, and for the "duplicated" counts for the 232 duplicated genes.

An expression ratio index was calculated for the 4139 A and B homeolog pairs as: *Expression ratio = avg* $RPKM_A$ */ (avg* $RPKM_A$ *+ avg* $RPKM_B$*)* where the subscripts A and B indicate the parental origin of the corresponding gene on the homeolog pair. This index ranges from 0 to 1 where values of 0.5 mean there is no difference in the expression levels of the A gene as compared to its corresponding B homeolog. Averaged RPKM values for each homeolog on a pair are available in **Data Set S3**. We calculated descriptive statistics from the expression ratio for the different groups using the R package psych v. 1.7.5 (Revelle, 2017). Exact binomial

tests and Fligner-Killeen tests were performed using the R functions binom.test and fligner.test correspondingly.

The normalized and filtered datasets were then Voom-transformed (Law *et al.*, 2014) to consider the differences in count sizes (or sequencing depth) and the overall dataset variability. This was followed by differential expression analysis (DEA) with adjusted P value < 0.05 and $|\text{log-fold change}| \geq 1$ for statistical significance (**Data Set S4** for the "full" set; **Data Set S5** for the "duplicated" set). Both the Voom transformation and the differential expression analysis were done using Limma v. 3.32.2 (Ritchie *et al.*, 2015). The *Z. parabailii* GO annotation was then utilized for GO term enrichment analysis with the R package goseq v. 1.28.0 (Young *et al.*, 2010). This was performed for the 3 sets of differentially expressed genes found to be upregulated at 18 h, upregulated at 42 h and upregulated at both time points, in addition to the corresponding 3 sets of differentially expressed genes found as downregulated (**Data Set S6**). The output of goseq for the upregulated genes at both time points was visualized using UpsetR v. 1.3.3 (Conway *et al.*, 2017) in the same way as for the downregulated genes.

Nucleotide sequence accession numbers.

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) and are accessible through GEO Series accession number GSE104654 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104654). This includes the raw RNAseq fastq files, counts for the "full" and "duplicated" sets (both raw and normalized), and RPKMs for the duplicated genes.

Acknowledgments

This work was funded by the European Union FP7 Marie Curie Programme [YEASTCELL - 7PQ MARIE CURIE (12-4-2001100-40)]. P.B. and D.P. also acknowledge the support by FAR (Fondo di Ateneo per la Ricerca) of the University of Milano-Bicocca. R. A. O.-M. was partially supported by CONACyT, Mexico (fellowship number 440667).
The authors declare that there is no conflict of interest

Category	Control	Lactic Acid	Intersect	Union
No evidence of expression	438	433	402	469
Expressed	9647	9652	9616	9683
Specific	31	36	NA	NA

Table 1. General overview of the *Z. parabailii* transcriptional profile.

Group	Mean	Median	Standard deviation	Skew	Ab	aB	P value	Unbalanced (%)
C18	0.5	0.5	0.079	-0.045	2077	2059	0.396	13.8
LA18	0.503	0.505	0.113	-0.009	2151	1985	0.005	31
C42	0.501	0.501	0.089	-0.146	2088	2048	0.272	18.7
LA42	0.5	0.503	0.115	-0.085	2119	2017	0.058	33.4

Table 2. Expression ratio between homeolog gene pairs. P values obtained from one sided exact binomial test using confidence intervals of 95% for the gene pairs where the A member shows higher expression than the B member (Ab). C18: control at 18 h; LA18: lactic acid at 18 h; C42: control at 42 h; LA42: lactic acid at 42 h

Category	18hr specific	42hr specific	Intersect	Union
Upregulated	117	83	27	227
Downregulated	430	431	158	1019

Table 3. *Z. parabailii* differential expression analysis. The upregulated and downregulated rows show the numbers of genes with an adjusted P value < 0.05 and a log-fold change ≥ 1, or log-fold change ≤ 1 respectively. Those sets of genes were further classified into 18 h specific and 42 h specific, where the intersect shows expression in both time points and the union shows the total.

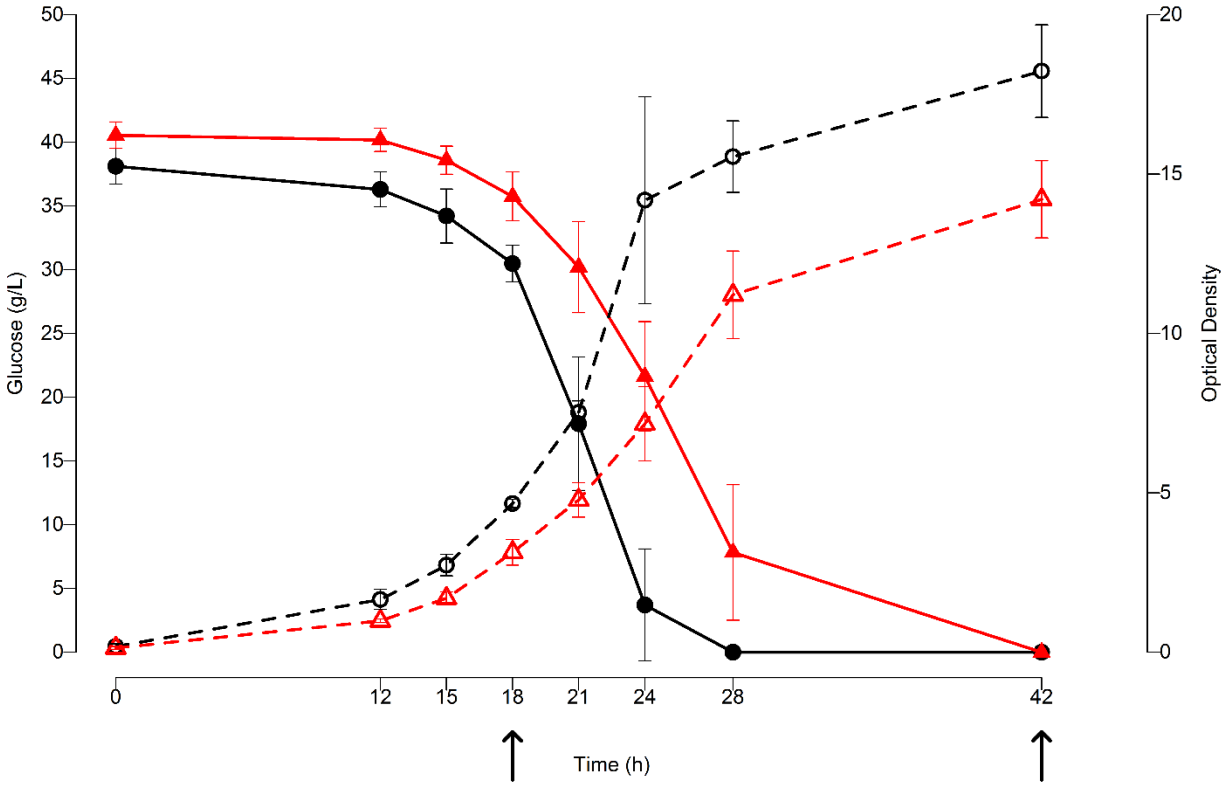


Figure 1. *Z. parabailii* fermentation profile. Batch bioreactor fermentation was performed in Verduyn medium at pH 3 with addition of 40 gL^{-1} lactic acid (red lines) or without lactic acid (black lines). The samples for RNA sequencing were taken at 18 h and 42 h (indicated by arrows), corresponding to exponential phase and post diauxic shift. Solid lines represent glucose consumption rate while dash lines corresponding optical density values at 660nm.

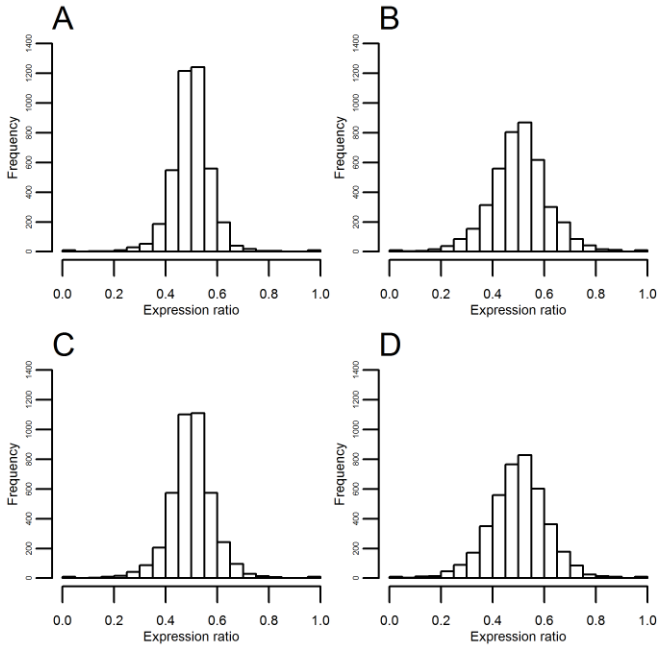


Figure 2. Expression ratios in 4136 homeologous gene pairs. Expression ratio is defined as A/(A+B) where A and B are the RPKM values (reads per kilobase of mRNA per million transcripts) of the A- and B- homeologous genes, respectively, averaged among replicates. Histograms show the distribution of expression ratio values in (A) control conditions at 18 h; (B) lactic acid at 18 h; (C) control conditions at 42 h; (D) lactic acid at 42 h.

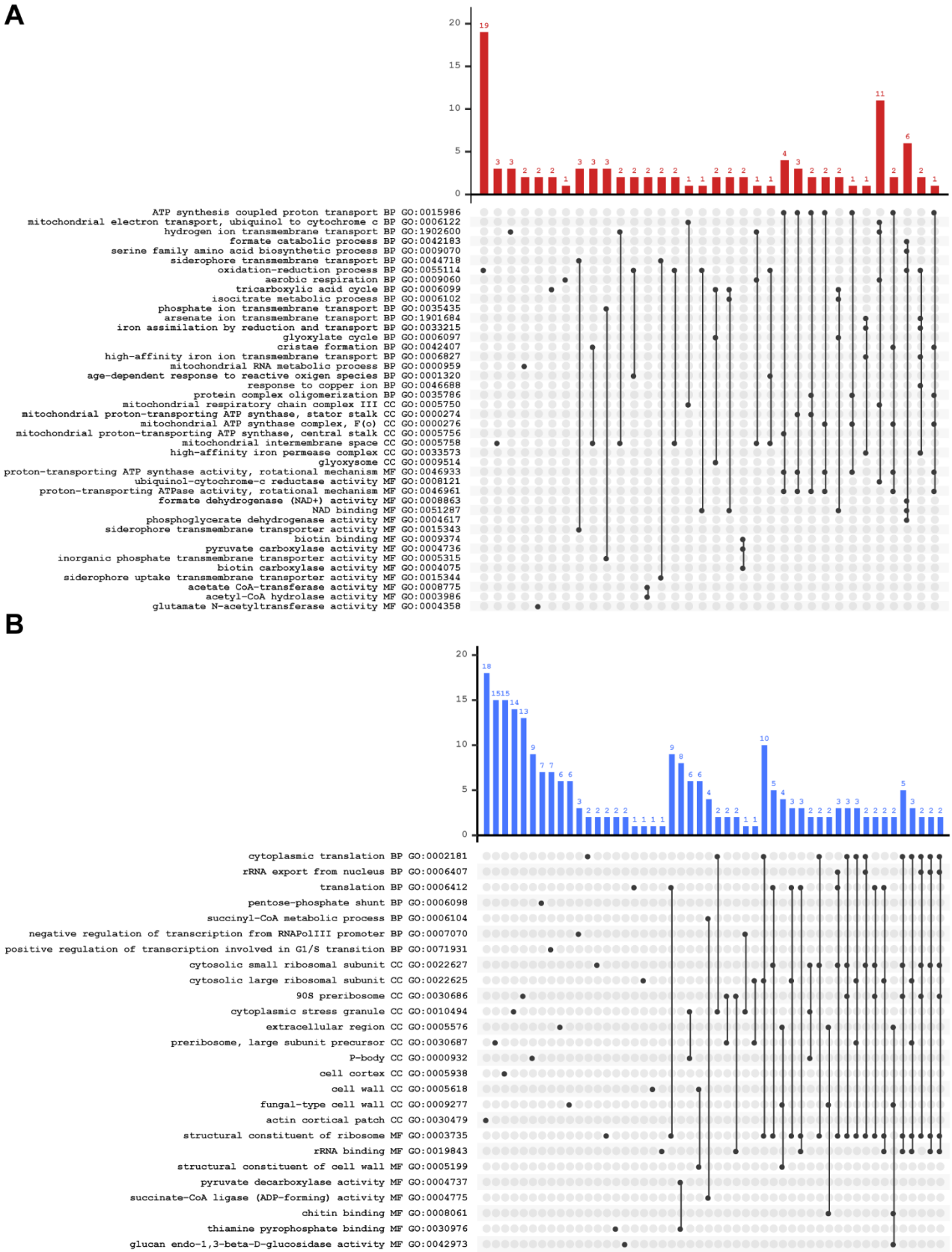

A
ATP synthesis coupled proton transport BP GO:0015986
mitochondrial electron transport, ubiquinol to cytochrome c BP GO:0006122
hydrogen ion transmembrane transport BP GO:1902600
formate catabolic process BP GO:0042183
serine family amino acid biosynthetic process BP GO:0009070
siderophore transmembrane transport BP GO:0044718
oxidation-reduction process BP GO:0055114
aerobic respiration BP GO:0009060
tricarboxylic acid cycle BP GO:0006099
isocitrate metabolic process BP GO:0006102
phosphate ion transmembrane transport BP GO:0035435
arsenate ion transmembrane transport BP GO:1901684
iron assimilation by reduction and transport BP GO:0033215
glyoxylate cycle BP GO:0006097
cristae formation BP GO:0042407
high-affinity iron ion transmembrane transport BP GO:0006827
mitochondrial RNA metabolic process BP GO:0000959
age-dependent response to reactive oxigen species BP GO:0001320
response to copper ion BP GO:0046688
protein complex oligomerization BP GO:0035786
mitochondrial respiratory chain complex III CC GO:0005750
mitochondrial proton-transporting ATP synthase, stator stalk CC GO:0000274
mitochondrial ATP synthase complex, F(o) CC GO:0000276
mitochondrial proton-transporting ATP synthase, central stalk CC GO:0005756
mitochondrial intermembrane space CC GO:0005758
high-affinity iron permease complex CC GO:0033573
glyoxysome CC GO:0009514
proton-transporting ATP synthase activity, rotational mechanism MF GO:0046933
ubiquinol-cytochrome-c reductase activity MF GO:0008121
proton-transporting ATPase activity, rotational mechanism MF GO:0046961
formate dehydrogenase (NAD+) activity MF GO:0008863
NAD binding MF GO:0051287
phosphoglycerate dehydrogenase activity MF GO:0004617
siderophore transmembrane transporter activity MF GO:0015343
biotin binding MF GO:0009374
pyruvate carboxylase activity MF GO:0004736
inorganic phosphate transmembrane transporter activity MF GO:0005315
biotin carboxylase activity MF GO:0004075
siderophore uptake transmembrane transporter activity MF GO:0015344
acetate CoA-transferase activity MF GO:0008775
acetyl-CoA hydrolase activity MF GO:0003986
glutamate N-acetyltransferase activity MF GO:0004358
B
cytoplasmic translation BP GO:0002181
rRNA export from nucleus BP GO:0006407
translation BP GO:0006412
pentose-phosphate shunt BP GO:0006098
succinyl-CoA metabolic process BP GO:0006104
negative regulation of transcription from RNAPolIII promoter BP GO:0007070
positive regulation of transcription involved in G1/S transition BP GO:0071931
cytosolic small ribosomal subunit CC GO:0022627
cytosolic large ribosomal subunit CC GO:0022625
90S preribosome CC GO:0030686
cytoplasmic stress granule CC GO:0010494
extracellular region CC GO:0005576
preribosome, large subunit precursor CC GO:0030687
P-body CC GO:0000932
cell cortex CC GO:0005938
cell wall CC GO:0005618
fungal-type cell wall CC GO:0009277
actin cortical patch CC GO:0030479
structural constituent of ribosome MF GO:0003735
rRNA binding MF GO:0019843
structural constituent of cell wall MF GO:0005199
pyruvate decarboxylase activity MF GO:0004737
succinate-CoA ligase (ADP-forming) activity MF GO:0004775
chitin binding MF GO:0008061
thiamine pyrophosphate binding MF GO:0030976
glucan endo-1,3-beta-D-glucosidase activity MF GO:0042973

Figure 3 Enriched GO terms among differentially expressed genes. Bar plots show the numbers of differentially expressed genes associated with a GO term (dots) or with a group of GO terms (dots connected by vertical lines). Upregulated genes are shown in panel **A** and downregulated genes in panel **B**. For example, among the 33 upregulated genes with the term GO:0055114 for oxidation reduction process in panel A, 19 show only this term, and 2 also show the term GO:0001320 for age−dependent response to reactive oxygen species. The GO terms are ordered by ontology type (BP biological process, CC cellular component and MF molecular function) and by decreasing adjusted P value, always < 0.05 (values are in Table S3).

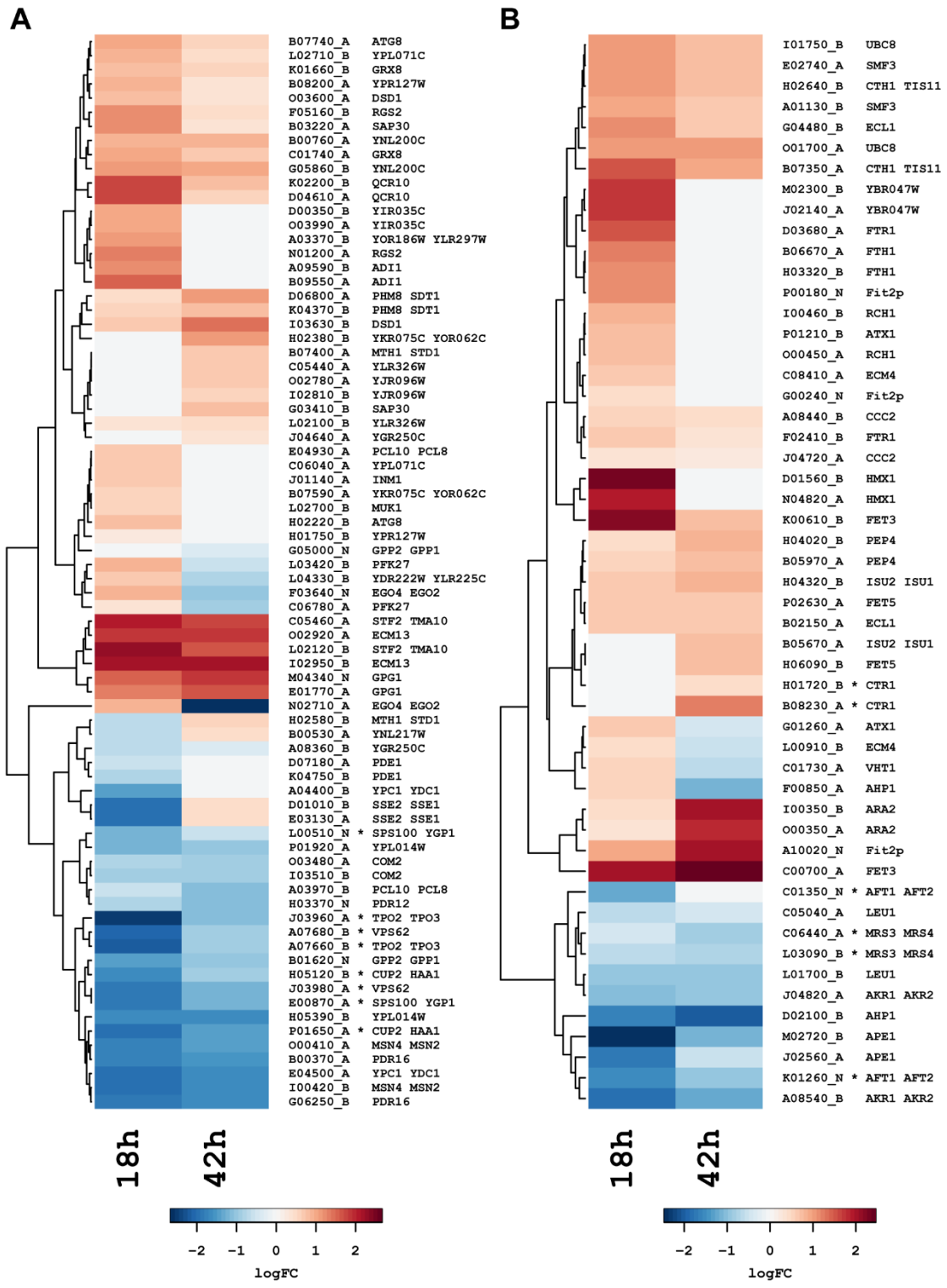


Figure 4 Log-fold changes for *Z. parabailii* genes putatively controlled by the Haa1 or the Aft1/2 transcription factors. Genes under Haa1p control are shown in panel A and genes controlled by Aft1/2 are shown in panel B. Asterisks (*) are used to mark *S. cerevisiae* genes reported as lactic acid-responsive by Abbot *et al.*, whose *Z. parabailii* homologs display an opposite response profile (*i.e.*

upregulated in *S. cerevisiae* and downregulated in *Z. parabailii*). Positive log-fold change values in lactic acid vs control are coloured in red as a sign of upregulation whereas negative values are blue. The colour scale is shown at the bottom of the corresponding panel and all the changes shown have an adjusted P value < 0.05.

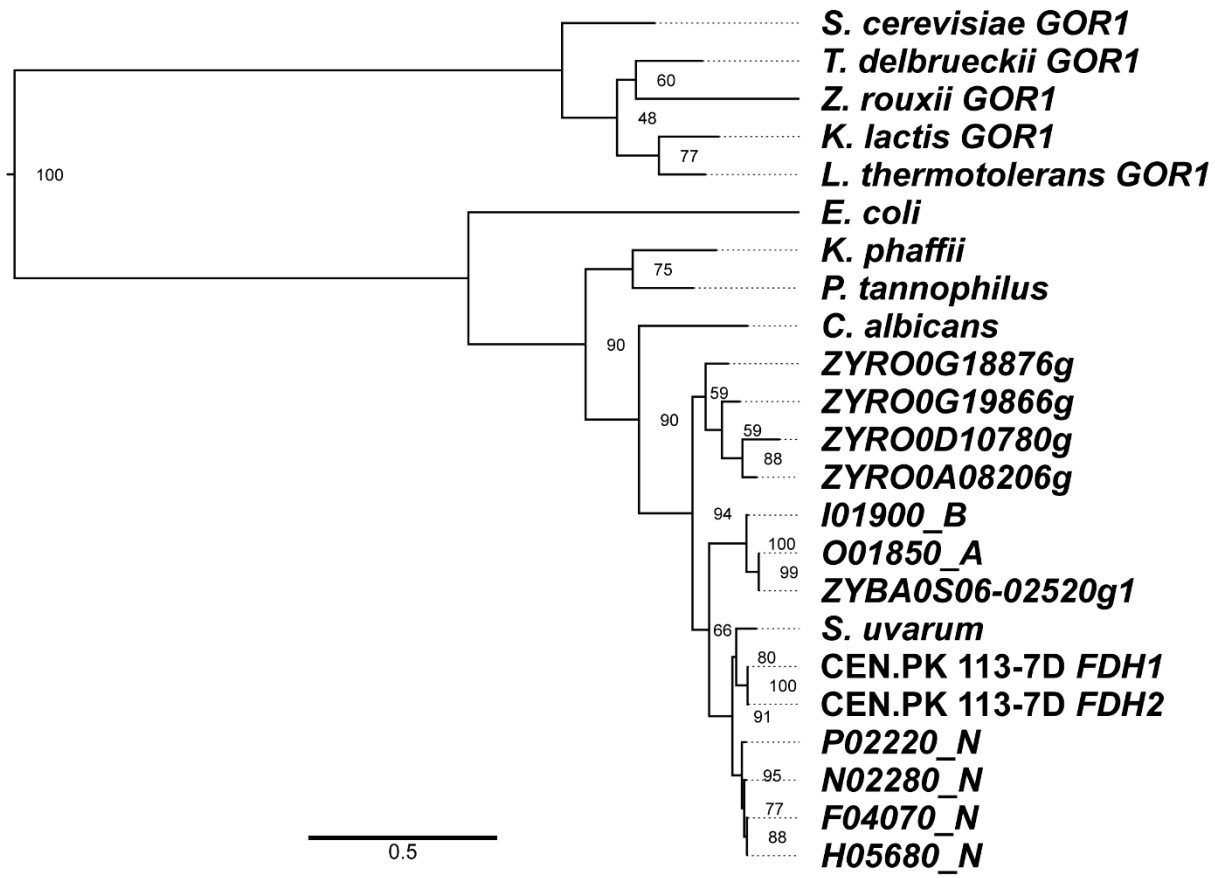


Figure 5. PhyML phylogenetic tree of formate dehydrogenase aminoacid sequences across yeast species. Homologs to *S. cerevisiae* glyoxylate reductase were added as outgroup and are indicated with the GOR1 suffix. The *Z. parabailii* genes are shown with their corresponding gene codes whereas the *Z. rouxii* homologs are labelled with a code starting with ZYRO and the only one for *Z. bailii* has ZYBA as a prefix. Nodes show bootstrap coefficients with 100 replicates.

References

ABBOTT, DA, SUIR, E, DUONG, GH*, et al.* 2009. Catalase overexpression reduces lactic acid-induced oxidative stress in *Saccharomyces cerevisiae*. *Appl Environ Microbiol,* 75**,** 2320-2325.

ABBOTT, DA, SUIR, E, VAN MARIS, AJ & PRONK, JT 2008. Physiological and transcriptional responses to high concentrations of lactic acid in anaerobic chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol,* 74**,** 5759-5768.

AGUILAR-USCANGA, B & FRANCOIS, JM 2003. A study of the yeast cell wall composition and structure in response to growth conditions and mode of cultivation. *Lett Appl Microbiol,* 37**,** 268-274.

ALHASAWI, A, CASTONGUAY, Z, APPANNA, ND*, et al.* 2015. Glycine metabolism and anti-oxidative defence mechanisms in *Pseudomonas fluorescens*. *Microbiol Res,* 171**,** 26-31.

ALI, MA, YASUI, F, MATSUGO, S & KONISHI, T 2000. The lactate-dependent enhancement of hydroxyl radical generation by the Fenton reaction. *Free Radic Res,* 32**,** 429-438.

ALTSCHUL, SF, GISH, W, MILLER, W*, et al.* 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-410.

ANDERS, S, PYL, PT & HUBER, W 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics,* 31**,** 166-169.

BECKER, J, LANGE, A, FABARIUS, J & WITTMANN, C 2015. Top value platform chemicals: bio-based production of organic acids. *Curr Opin Biotechnol,* 36**,** 168-175.

BERTERAME, NM, 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,* 15**,** 39.

BRANDUARDI, P, FOSSATI, T, SAUER, M*, et al.* 2007. Biosynthesis of vitamin C by yeast leads to increased stress resistance. *PLoS One,* 2**,** e1092.

CABRAL, S, PRISTA, C, LOUREIRO-DIAS, MC & LEANDRO, MJ 2015. Occurrence of FFZ genes in yeasts and correlation with fructophilic behaviour. *Microbiology,* 161**,** 2008-2018.

CASTILLO MARTINEZ, FA, BALCIUNAS, EM, SALGADO, JM*, et al.* 2013. Lactic acid properties, applications and production: A review. *Trends in Food Science & Technology,* 30**,** 70-83.

CHEN, Y & NIELSEN, J 2016. Biobased organic acids production by metabolically engineered microorganisms. *Curr Opin Biotechnol,* 37**,** 165-172.

COMBES, MC, HUEBER, Y, DEREEPER, A*, et al.* 2015. Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biol Evol,* 7**,** 1110-1121.

CONWAY, JR, LEX, A & GEHLENBORG, N 2017. UpSetR: An R Package for the Visualization of Intersecting Sets and their Properties. *Bioinformatics*.

COX, MP, DONG, T, SHEN, G*, et al.* 2014. An interspecific fungal hybrid reveals cross-kingdom rules for allopolyploid gene expression patterns. *PLoS Genet,* 10**,** e1004180.

DATTA, R & HENRY, M 2006. Lactic acid: recent advances in products, processes and technologies — a review. *Journal of Chemical Technology & Biotechnology,* 81**,** 1119-1129.

DECKER, CJ & PARKER, R 2012. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb Perspect Biol,* 4**,** a012286.

EDGAR, R, DOMRACHEV, M & LASH, AE 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res,* 30**,** 207-210.

ELBOURNE, LD, TETU, SG, HASSAN, KA & PAULSEN, IT 2017. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res,* 45**,** D320-D324.

FINN, RD, BATEMAN, A, CLEMENTS, J*, et al.* 2014. Pfam: the protein families database. *Nucleic Acids Res,* 42**,** D222-230.

FINN, RD, CLEMENTS, J & EDDY, SR 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res,* 39**,** W29-37.

FITZPATRICK, JJ, MURPHY, C, MOTA, FM & PAULI, T 2003. Impurity and cost considerations for nutrient supplementation of whey permeate fermentations to produce lactic acid for biodegradable plastics. *International Dairy Journal,* 13**,** 575-580.

GALEOTE, V, BIGEY, F, DEVILLERS, H*, et al.* 2013. Genome Sequence of the Food Spoilage Yeast *Zygosaccharomyces bailii* CLIB 213T. *Genome Announc,* 1.

GIANNATTASIO, S, GUARAGNELLA, N, ZDRALEVIC, M & MARRA, E 2013. Molecular mechanisms of *Saccharomyces cerevisiae* stress adaptation and programmed cell death in response to acetic acid. *Front Microbiol,* 4**,** 33.

GLOVER, NM, REDESTIG, H & DESSIMOZ, C 2016. Homoeologs: What Are They and How Do We Infer Them? *Trends Plant Sci,* 21**,** 609-621.

GOTZ, S, GARCIA-GOMEZ, JM, TEROL, J*, et al.* 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res,* 36**,** 3420-3435.

IWAKI, A & IZAWA, S 2012. Acidic stress induces the formation of P-bodies, but not stress granules, with mild attenuation of bulk translation in *Saccharomyces cerevisiae*. *Biochem J,* 446**,** 225-233.

KAPTEYN, JC, TER RIET, B, VINK, E*, et al.* 2001. Low external pH induces HOG1-dependent changes in the organization of the *Saccharomyces cerevisiae* cell wall. *Mol Microbiol,* 39**,** 469-479.

KAWAHATA, M, MASAKI, K, FUJII, T & IEFUJI, H 2006. Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in *Saccharomyces cerevisiae* cultures induce expression of intracellular metal metabolism genes regulated by Aft1p. *FEMS Yeast Res,* 6**,** 924-936.

KELLER, G, RAY, E, BROWN, PO & WINGE, DR 2001. Haa1, a protein homologous to the copper-regulated transcription factor Ace1, is a novel transcriptional activator. *J Biol Chem,* 276**,** 38697-38702.

KROGH, A, LARSSON, B, VON HEIJNE, G & SONNHAMMER, EL 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol,* 305**,** 567-580.

KUANYSHEV, N, ADAMO, GM, PORRO, D & BRANDUARDI, P 2017. The spoilage yeast *Zygosaccharomyces bailii*: foe or friend? *Yeast*.

KUANYSHEV, N, AMI, D, SIGNORI, L*, et al.* 2016. Assessing physio-macromolecular effects of lactic acid on *Zygosaccharomyces bailii* cells during microaerobic fermentation. *FEMS Yeast Res,* 16.

LANGMEAD, B 2010. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics,* Chapter 11**,** Unit 11 17.

LAW, CW, CHEN, Y, SHI, W & SMYTH, GK 2014. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol,* 15**,** R29.

LEANDRO, MJ, SYCHROVA, H, PRISTA, C & LOUREIRO-DIAS, MC 2011. The osmotolerant fructophilic yeast *Zygosaccharomyces rouxii* employs two plasma-membrane fructose uptake systems belonging to a new family of yeast sugar transporters. *Microbiology,* 157**,** 601-608.

LESAGE, G & BUSSEY, H 2006. Cell wall assembly in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev,* 70**,** 317-343.

LIMAYEM, A & RICKE, SC 2012. Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects. *Progress in Energy and Combustion Science,* 38**,** 449-467.

LINDBERG, L, SANTOS, AX, RIEZMAN, H*, et al.* 2013. Lipidomic profiling of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* reveals critical changes in lipid composition in response to acetic acid stress. *PLoS One,* 8**,** e73936.

MARTANI, F, MARANO, F, BERTACCHI, S*, et al.* 2015. The *Saccharomyces cerevisiae* poly(A) binding protein Pab1 as a target for eliciting stress tolerant phenotypes. *Sci Rep,* 5**,** 18318.

MARTORELL, P, STRATFORD, M, STEELS, H*, et al.* 2007. Physiological characterization of spoilage strains of *Zygosaccharomyces bailii* and *Zygosaccharomyces rouxii* isolated from high sugar environments. *Int J Food Microbiol,* 114**,** 234-242.

MCCARTHY, DJ, CHEN, Y & SMYTH, GK 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res,* 40**,** 4288-4297.

MIRA, NP, MUNSTERKOTTER, M, DIAS-VALADA, F*, et al.* 2014. The genome sequence of the highly acetic acid-tolerant *Zygosaccharomyces bailii*-derived interspecies hybrid strain ISA1307, isolated from a sparkling wine plant. *DNA Res,* 21**,** 299-313.

MIRA, NP, TEIXEIRA, MC & SA-CORREIA, I 2010. Adaptive response and tolerance to weak acids in *Saccharomyces cerevisiae*: a genome-wide view. *OMICS,* 14**,** 525-540.

MOLLAPOUR, M & PIPER, PW 2001. The ZbYME2 gene from the food spoilage yeast *Zygosaccharomyces bailii* confers not only YME2 functions in *Saccharomyces cerevisiae*, but also the capacity for catabolism of sorbate and benzoate, two major weak organic acid preservatives. *Mol Microbiol,* 42**,** 919-930.

ORTIZ-MERINO, RA, KUANYSHEV, N, BRAUN-GALLEANI, S*, et al.* 2017. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biol,* 15**,** e2002128.

OVERKAMP, KM, KOTTER, P, VAN DER HOEK, R*, et al.* 2002. Functional analysis of structural genes for NAD(+)-dependent formate dehydrogenase in *Saccharomyces cerevisiae*. *Yeast,* 19**,** 509-520.

PALMA, M, DIAS, PJ, ROQUE, FC*, et al.* 2017a. The *Zygosaccharomyces bailii* transcription factor Haa1 is required for acetic acid and copper stress responses suggesting subfunctionalization of the ancestral bifunctional protein Haa1/Cup2. *BMC Genomics,* 18**,** 75.

PALMA, M, MUNSTERKOTTER, M, PECA, J*, et al.* 2017b. Genome sequence of the highly weak-acid-tolerant *Zygosaccharomyces bailii* IST302, amenable to genetic manipulations and physiological studies. *FEMS Yeast Res*.

PIPER, P, CALDERON, CO, HATZIXANTHIS, K & MOLLAPOUR, M 2001. Weak acid adaptation: the stress response that confers yeasts with resistance to organic acid food preservatives. *Microbiology,* 147**,** 2635-2642.

PIPER, PW 1999. Yeast superoxide dismutase mutants reveal a pro-oxidant action of weak organic acid food preservatives. *Free Radic Biol Med,* 27**,** 1219-1227.

PROUX-WERA, E, ARMISEN, D, BYRNE, KP & WOLFE, KH 2012. A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics,* 13**,** 237.

RAPP, RA, UDALL, JA & WENDEL, JF 2009. Genomic expression dominance in allopolyploids. *BMC Biol,* 7**,** 18.

REGO, A, DUARTE, AM, AZEVEDO, F*, et al.* 2014. Cell wall dynamics modulate acetic acid-induced apoptotic cell death of *Saccharomyces cerevisiae*. *Microb Cell,* 1**,** 303-314.

REVELLE, W 2017. psych: Procedures for Personality and Psychological Research. Northwestern University.

RITCHIE, ME, PHIPSON, B, WU, D*, et al.* 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research,* 43**,** e47-e47.

ROBINSON, MD & OSHLACK, A 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol,* 11**,** R25.

RODRIGUES, F, SOUSA, MJ, LUDOVICO, P*, et al.* 2012. The fate of acetic acid during glucose co-metabolism by the spoilage yeast *Zygosaccharomyces bailii*. *PLoS One,* 7**,** e52402.

RUTHERFORD, JC, JARON, S & WINGE, DR 2003. Aft1p and Aft2p mediate iron-responsive gene expression in yeast through related promoter elements. *J Biol Chem,* 278**,** 27636-27643.

SAUER, M, BRANDUARDI, P, GASSER, B*, et al.* 2004. Differential gene expression in recombinant Pichia pastoris analysed by heterologous DNA microarray hybridisation. *Microb Cell Fact,* 3**,** 17.

SAUER, M, PORRO, D, MATTANOVICH, D & BRANDUARDI, P 2010. 16 years research on lactic acid production with yeast – ready for the market? *Biotechnology and Genetic Engineering Reviews,* 27**,** 229-256.
SIMOES, T, MIRA, NP, FERNANDES, AR & SA-CORREIA, I 2006. The SPI1 gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. *Appl Environ Microbiol,* 72**,** 7168-7175.
SIMPSON, CLARE E & ASHE, MARK P 2012. Adaptation to stress in yeast: to translate or not? *Biochemical Society Transactions,* 40**,** 794.
STRATFORD, M, STEELS, H, NEBE-VON-CARON, G*, et al.* 2013. Extreme resistance to weak-acid preservatives in the spoilage yeast *Zygosaccharomyces bailii*. *Int J Food Microbiol,* 166**,** 126-134.
SUH, SO, GUJJARI, P, BERES, C*, et al.* 2013. Proposal of *Zygosaccharomyces parabailii* sp. nov. and *Zygosaccharomyces pseudobailii* sp. nov., novel species closely related to *Zygosaccharomyces bailii*. *Int J Syst Evol Microbiol,* 63**,** 1922-1929.
SWISHER, KD & PARKER, R 2010. Localization to, and effects of Pbp1, Pbp4, Lsm12, Dhh1, and Pab1 on stress granules in *Saccharomyces cerevisiae*. *PLoS One,* 5**,** e10006.
TEIXEIRA, MC, MONTEIRO, P, JAIN, P*, et al.* 2006. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res,* 34**,** D446-451.
THOMAS, SC, ALHASAWI, A, AUGER, C*, et al.* 2016. The role of formate in combatting oxidative stress. *Antonie Van Leeuwenhoek,* 109**,** 263-271.
TIROSH, I, REIKHAV, S, LEVY, AA & BARKAI, N 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science,* 324**,** 659-662.
TISHKOV, VI & POPOV, VO 2004. Catalytic mechanism and application of formate dehydrogenase. *Biochemistry (Mosc),* 69**,** 1252-1267.
VERDUYN, C, POSTMA, E, SCHEFFERS, WA & VAN DIJKEN, JP 1992. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast,* 8**,** 501-517.
WANG, Z, SUN, X, ZHAO, Y*, et al.* 2015. Evolution of gene regulation during transcription and translation. *Genome Biol Evol,* 7**,** 1155-1167.
YOKOTA, A, KAWABATA, A & KITAOKA, S 1983. Mechanism of Glyoxylate Decarboxylation in the Glycolate Pathway in *Euglena gracilis* Z : Participation of Mn-Dependent NADPH Oxidase in Chloroplasts. *Plant Physiol,* 71**,** 772-776.
YOO, MJ, SZADKOWSKI, E & WENDEL, JF 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity (Edinb),* 110**,** 171-180.
YOUNG, MD, WAKEFIELD, MJ, SMYTH, GK & OSHLACK, A 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol,* 11**,** R14.

Chapter 5

Proposed role of the plasma membrane transporter Pdr12p in the *Kluyveromyces marxianus* weak acid response

Finalizing the experiments

Proposed role of the plasma membrane transporter Pdr12p in the *Kluyveromyces marxianus* weak acid response

Nurzhan Kuanyshev[1,2], Javier Alberto Varela [2], Danilo Porro[1], Paola Branduardi[1], John P. Morrissey[2*]

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

[2]School of Microbiology, University College Cork, Cork, Ireland

Abstract

In present study, we investigated the effect of acetic and lactic acid on two different *K. marxianus* strains, CBS6556 and CBS397. The results revealed a strain-specific response toward these weak acids. The expression of the MDR genes *AQR1*, *TPO2* and *PDR12* in response to both acids was assessed. The most notable finding was that *PDR12* expression in CBS6556 was highly-induced in response to treatment with either acid although the strain is more sensitive than CBS397 to both acids. Further investigation revealed that there are two copies of *PDR12* in most *K. marxianus* strains and these arose from a duplication event after the divergence of *K. lactis* and *K. marxianus*. There is only one copy in CBS6556, however, and this arose from a recombination between *PDR12-1* and *PDR12-2* in this strain. The hypothesis is proposed that this recombination has created a non-functional protein and the induced expression is a futile effort to combat the toxic effects of acetic and lactic acid. Molecular approaches to test the hypothesis are proposed.

Introduction

The emerging tendency to switch from oil-based to bio-production of fuels and chemicals requires development of robust microbial hosts. Biotechnology has been exploiting bacterial hosts for production of various chemicals including pharmaceuticals, biofuels and amino acids (Chen, et al., 2013). However, the bacterial platform suffers from low tolerance to organic acids, acidic pH, furan derivatives and phenolic compounds. These limitations are not a serious problem when growing on synthetic medium to produce high value products for biopharma, but they greatly restrict the applications of bacterial hosts in industrial biotechnology. Therefore, the yeast *Saccharomyces cerevisiae,* which has been used for centuries in different fields, represents a major microbial platform for industrial biotechnology. Although *S. cerevisiae* remains the first-choice organism for bio-based production of bulk chemicals, modern biotechnology is exploring alternative non-Saccharomyces yeasts, with native unique traits. Among those yeasts, *Kluyveromyces marxianus* has great potential to substitute *S. cerevisiae* in production of value added compounds and chemicals (Lane and Morrissey, 2010).

K. marxianus is an emerging non-Saccharomyces yeast relevant for industrial applications. It possesses important traits such as thermotolerance (up to 52 C^{o}), fast growth and ability to utilize various substrate including C5 sugars (xylose in particular) and lactose (Lane and Morrissey, 2010). In addition, a high secretory ability along with high biomass yield, compared to *S. cerevisiae,* makes it an excellent candidate for protein production (Gombert *et al.*, 2016). The yeast already has been proposed for production of various value-added chemicals, such as flavour and fragrance molecules and the solvent ethyl acetate (Morrissey *et al.*, 2015, Radecka *et al.*, 2015, Loser *et al.*, 2015). In addition, the respiro-fermentative metabolism led to many studies to exploit the yeast in bioethanol production from lignocellulosic (Goshima *et al.*, 2013) and whey biomass (Guimaraes *et al.*, 2010). There is now considerable interest in extending the applications of *K. marxianus* to other bio-based chemicals, but this will require better understanding of how the yeast performs in industrial settings.

Bio-based chemical production from renewable biomass has drawn the attention of scientists to address issues with excessive carbon emission and fossil resource usage. Several different renewable resources have been used as a substrate for microbial fermentation. Lignocellulosic biomass represents a cheap, ubiquitous and renewable source, consisting of cellulose, hemicellulose and lignin. To release the free sugars (mainly glucose and xylose) for subsequent fermentation, lignocellulose needs to be pretreated and hydrolyzed (Jonsson *et al.*, 2013). Whey biomass, primarily consisting of lactose, is a major byproduct of the cheese and dairy industry. However, these feedstocks are not easily fermentable and contain several inhibitors, which can hamper the fermentation. For example, during lignocellulose pretreatment and hydrolysis, the process generates considerable amounts of fermentation inhibitors, including acetic acid (Jonsson *et al.*, 2013). Whey may contain an excess of lactic acid, due to lactic acid bacteria contamination, which can be potentially harmful for the yeast (Christensen *et al.*, 2011). Thus, in different industrial fermentations, the presence of weak organic acids (WOA) like acetic or lactic acid may become a problem for *K. marxianus* performance.

WOAs have been used as food preservatives for a long time, and their physiological effect on cell growth have been well studied. The main mechanism of WOA toxicity derives from the ability of the protonated acid to diffuse into the cytosol, where, at near neutral pH, WOAs effectively dissociate, releasing a proton and respective anion (Piper *et al.*, 2001). Since charged molecules cannot escape the cytosol, accumulation of protons decreases internal pH, while the (counter)anion exerts toxicity depending on its nature. For example, lactate anions increase reactive oxygen generation and affect cell membrane fatty acid composition (Ali *et al.*, 2000, Narendranath *et al.*, 2001), while acetate anions induce programmed cell death (Giannattasio *et al.*, 2013).

The weak organic acid response mechanism has been well studied in *S. cerevisiae*. To decrease the concentration of intracellular weak acids, the yeast activates different membrane transporters to extrude protons and anions. Exposure of yeast to inhibitory concentrations of acetic or lactic acid causes rapid acidification of the cytosol and induces the plasma membrane proton pumps H^+-ATPase (Pma1p) and V-ATPase (Schuller *et al.*, 2004, Mollapour *et al.*, 2004, Carmelo *et al.*, 1997). The pumps exports protons at the cost of ATP to maintain the

optimum internal pH. Respective anions are removed via multi drug resistance (MDR) transporters, which are regulated by the Haa1p and War1p transcriptional factors. Haa1p is required for weak acid adaptation to relatively less-lipophilic weak acids like acetic, lactic and propionic acids (Fernandes *et al.*, 2005, Abbott *et al.*, 2008). The regulon is responsible for activation of a wide range of genes associated with acetic and lactic acid adaptation (Mira *et al.*, 2010), among them MDR transporters *TPO2/3* and *AQR1*, previously reported as direct Haa1p targets (Keller *et al.*, 2001). The expression of *TPO2/3* is highly-induced upon acetic and lactic acid exposure, however deletion of the genes has no effect on lactic acid sensitivity, while rendering sensitivity to acetic acid. Yet, the *haa1* null mutant exhibits a strong growth defect in the presence of inhibitory concentrations of either acid (Fernandes *et al.*, 2005, Abbott *et al.*, 2008). War1p solely controls expression of *PDR12*, which encodes Pdr12p, an ATP binding cassette transporter family implicated in weak organic acid tolerance. Transcription of *PDR12* is strongly induced by sorbic and benzoic acids, but not by acetic and formic acids (Kren *et al.*, 2003, Hatzixanthis *et al.*, 2003). Pdr12p lowers the intracellular concentration of weak acids by active extrusion of anions out of the cell (Mollapour *et al.*, 2008). Although some of the earlier studies suggested an involvement of Pdr12p in acetic acid tolerance (Bauer *et al.*, 2003), more recent studies demonstrated that *PDR12* deletion did not affect acetic, lactic and formic acid sensitivity (Nygard *et al.*, 2014).

Exposure to weak organic acids are predicted to have similar toxic effects on different yeasts, however the response and adaptation to the stress can be yeast species-specific. Moreover, the reported wide genetic diversity of *K. marxianus*, together with limited research on its physiology, biochemistry and genetics, create difficulties to understand mechanisms of weak acid resistance (Rocha *et al.*, 2011, Fonseca *et al.*, 2008). Therefore, further research on the genetics and physiology of the weak acid response in *K. marxianus* is required to facilitate development of the yeast as an alternative bio-based production platform. This study is an initial exploration of the involvement of major MDR transporters in acetic and lactic acid resistance in *K. marxianus*.

Materials and Methods

Yeast strains

The *K. marxianus* strains used in this study (CBS 397, CBS 6556) were obtained as lyophilized stocks from The Centraalbureau voor Schimmelcultures, Delft, The Netherlands. Strains were revived in yeast, peptone, glucose (YPD) media (1% yeast extract, 2% bactopeptone, 2% glucose) at 30ºC with agitation at 180 rpm. Yeast strains were routinely cultured at 30ºC in YP medium (1% yeast extract; 2% bactopeptone; 2% lactose).

Drop plate assays and flask fermentation

Serial dilutions of *K. marxianus* strains on YNB and 2 % glucose (YNB) agar plates were used to assess its tolerance to weak organic acids. For these assays, strains were grown in YNB medium shaking at 30ºC until mid-exponential phase and diluted to OD_{600} 1 in fresh YNB medium. Serial dilutions down to 10^{-4} were prepared in 96 well microtitre plates and spotted aseptically onto YNB agar plates at pH3 using a 48 pin replicator. YNB agar plates were pH adjusted using 1M HCL and supplemented with increasing concentrations of acetic acid/lactic acid. Plates were incubated at 30ºC and growth inhibition was recorded after 48 h incubation. The flask fermentation was performed in 125mL shake flasks with YNB (2% glucose) at pH 3 with or without weak acid. The cells were inoculated at mid exponential phase and OD_{600} 0.1 and grown at 30ºC and 180 RPM.

Table 1 Primer sequences used in this study

Gene/Name	Primer	Primer sequence
PDR12-1	F	TCGTTCATGGATGCATTTCCC
	R	CACCAACTCACTGGTTTCACC
PDR12-2	F	TGCCAGGTTTCTGGAAGAGG
	R	GTTTGGCCTTCCGGAGGATT
TPO2	F	TGAGTTCCCTAGCCACCTGGA
	R	TCCACGGTGAACAAACCACCT
AQR1	F	TTTGGTCATCGTTGGGTTCG
	R	CCGTAGGCGCAATACCTTGG
ACT1	F	GGCTGAACGTGGTTACTCCT
	R	AGAAGCGGTTTGCATTTCTT
PDR12_Copynumber	F	TGGTCCGCTTTGCTAATATGCCA
	R	GATATGTCACCTGTGACTTCCACCAAC

Gene expression analysis

K. marxianus strains were grown in YNB to A600 of 1. Cells were then recovered by centrifugation and re-innucluated at A600 of 0.5 to YNB pH 3 and incubated for 1 hr. After initial incubation either acetic or lactic acid was added to the medium. The cells were collected at 0 min, 5 min, 15 min and 30 min and immediately frozen in liquid nitrogen. The total RNA was extracted using Zymo Research Fungal/Bacterial RNA MiniPrep™ kit (Irvine, USA). RNA quantification was done by nanodrop spectrophotometry. cDNA was synthesised from 800 ng of RNA, using the ProtoScript First Strand cDNA Synthesis Kit (New England Biolabs Inc., MA, USA). cDNA samples were diluted by adding 30 µL of nuclease-free water, resulting in a final volume of 50 µL per cDNA sample. Aliquots were then diluted 1/10 and stored at –

20◦C. Real-time quantitative PCR (qPCR) was used to measure gene expression. cDNA samples were amplified by using LightCycler 480 SYBR Green I Master qPCR and ran in a LightCycler

480 real-time PCR machine. Primers were designed to amplify the *ACT1*, *AQR1*, *TPO2* and *PDR12* (Table 1). Standard curves were performed to check for amplification efficiency with all primers showing >90% efficiency. Additionally, melting curves were carried out to check reaction specificity and to ensure that each primer pair produced a single amplicon. *ACT1* (coding for actin) were used as reference genes. To compare expression of the MDR genes during time course experiment, the data were normalised against *ACT1* expression. Data were further calculated as $2^{-\Delta Ct}$ as described by (Pfaffl, 2001).

Results

The effect of acetic and lactic acid stress at low pH in *K. marxianus*

Knowing the phenotypic diversity of *K. marxianus*, we chose to compare acetic and lactic acid tolerance in two different strains. The comparison was made by a serial dilution on YNB plates plus the appropriate weak organic acid at pH 3 (Fig 1A). Because the pKa of acetic and lactic acids are 4.76 and 3.86 respectively, at pH 3 the acids are predominantly present in the toxic undissociated form. As shown in Fig. 1A, the cell growth at YNB pH 3 plate without acid is comparable and both strains show sensitivity to both acids. Addition of 20 gL^{-1} or 30 gL^{-1} of lactic acid inhibits the growth of both strains, with CBS6556 being more sensitive. The presence of 1 gL^{-1} of acetic acid inhibits growth of CBS6556, while CBS397 shows similar growth as in the control condition. However, at 1.5 gL^{-1} of acetic acid both strains are strongly inhibited, with CBS6556 showing no growth at all. In summary, the drop plate assay showed that although both strains are sensitive to these WOA, CBS6556 is more sensitive than CBS397 to acetic and lactic acid. These patterns were largely replicated when the yeasts were grown in liquid medium at 30 gL^{-1} of lactic acid or 1 gL^{-1} of acetic acid (Fig 1B).

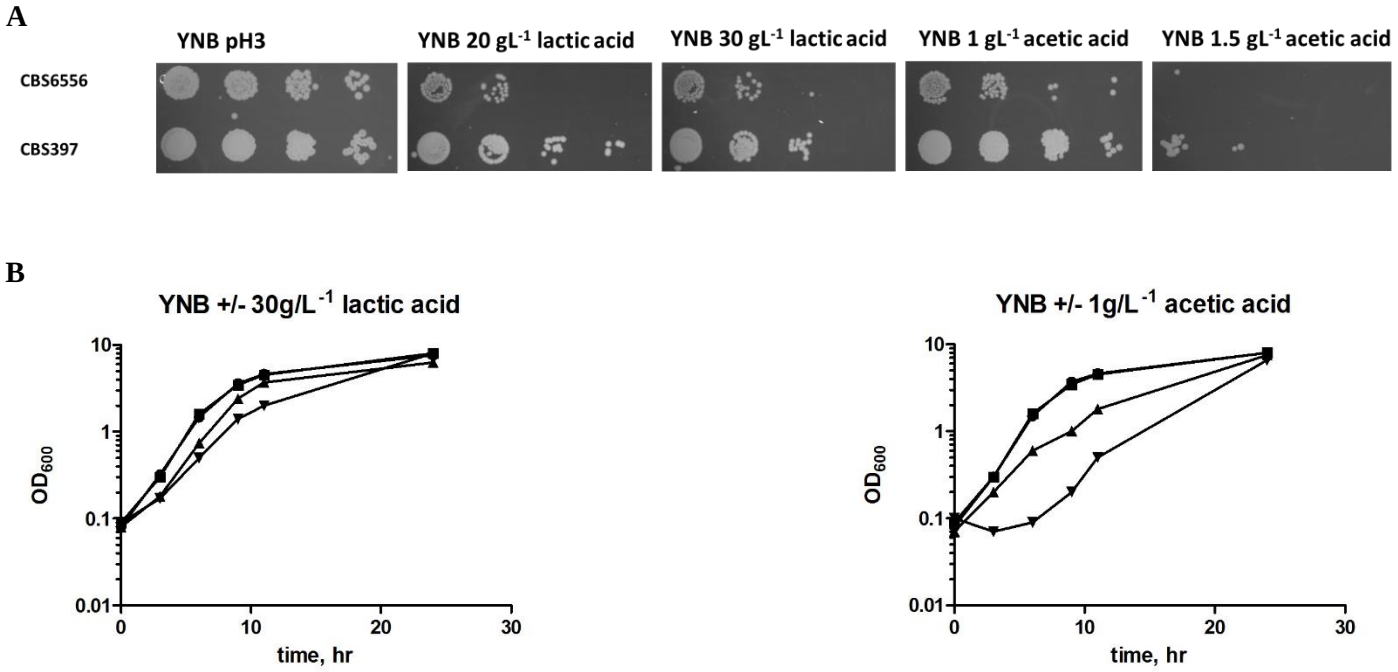


Figure 1 A. Growth plate assay of CBS397 and CBS6556. The cells were cultivated until log-phase and spotted to YNB plates (2% glucose) at pH3 with different concentration of acetic acid and lactic acid. Cells were 10-fold serially diluted and incubated at 30 °C for 2 days. **B.** Growth curve of CBS397

and CBS6556. The cells were pre-cultured in YNB (2% glucose) until log phase and were inoculated at OD_{600} 0.1 in YNB +/- weak acid at pH3 (control without weak acid). Closed circle, CBS397 control (no acid); closed square, CBS6556 control (no acid); closed triangle, CBS397 (+weak acid); closed inverted triangle, CBS6556 (+weak acid).

Expression profile of MDR transporters during pulse exposure to acetic or lactic acid

Multidrug resistance transporters are known to be weak acid tolerance determinants in the model yeast *S. cerevisiae*, therefore, we studied the expression level of major MDR transporters by pulse exposure to either 30 gL^{-1} lactic acid or 1 gL^{-1} acetic acid. We specifically wanted to investigate if the level of MDR transporter expression determines the sensitivity difference between the strains. The concentrations were empirically chosen for their ability to moderately inhibit the growth during flask fermentation (Fig. 1B). To ensure that MDR genes response is due to a presence of the weak organic acid (as opposed to low pH), prior to the pulse shock, cells were inoculated at mid-exponential phase to YNB (2% glucose) at pH 3 and incubated for 1 hr at 30 °C. After the incubation, acetic or lactic acid were added, and triplicate samples for RTqPCR were collected at 0 min, 5 min, 15 min and 30 min. The level of expression of *PDR12*, *TPO2* and *AQR1* homologs of *K. marxianus* was assessed by RTqPCR using the *ACT1* gene as a reference. CBS6556 and CBS397 have one copy of each gene, except for *PDR12*, which is duplicated in CBS397 (designated here as *PDR12-1* and *PDR12-2*). The *PDR12* duplication was confirmed with PCR (not shown). The RTqPCR analysis showed that while CBS6556 strongly induced *PDR12* during acetic and lactic acid shock, the response of CBS397 was much less apparent (Fig 2A). In CBS 397, the induction by lactic acid of both copies of *PDR12* was moderate in comparison to CBS6556, while acetic acid resulted in a gradual induction of *PDR12-1* and little change in expression of *PDR12-2* (Fig. 2B). This result demonstrated a strain-specific response to both acids. The expression of *AQR1* and *TPO2* was also assessed. In CBS6556, both genes were induced to similar levels by both acids, though the response-time for *AQR1* induction by lactic acid was delayed (Fig. S1A). Once again, the overall induced response of these genes was lower in CBS397 and there were also some specific differences between the effects of lactic acid and acetic acid (Fig. S1B). Most notably, neither *AQR1* nor *TPO2* were induced by lactic acid, whereas both genes were induced by acetic acid. In CBS397, the level of induction by acetic acid of *AQR1* and *TPO2* were in the order 2-3 fold, compared to >10 fold in CBS6556. Summarising, the

expression profile showed that in CBS6556, all tested MDR transporters were upregulated in both acid conditions, with *PDR12* having the highest level of expression. In contrast, the expression profile of MDR in CBS397 showed an intermediate response to acetic acid, but with *PDR12* still having the highest level of expression.

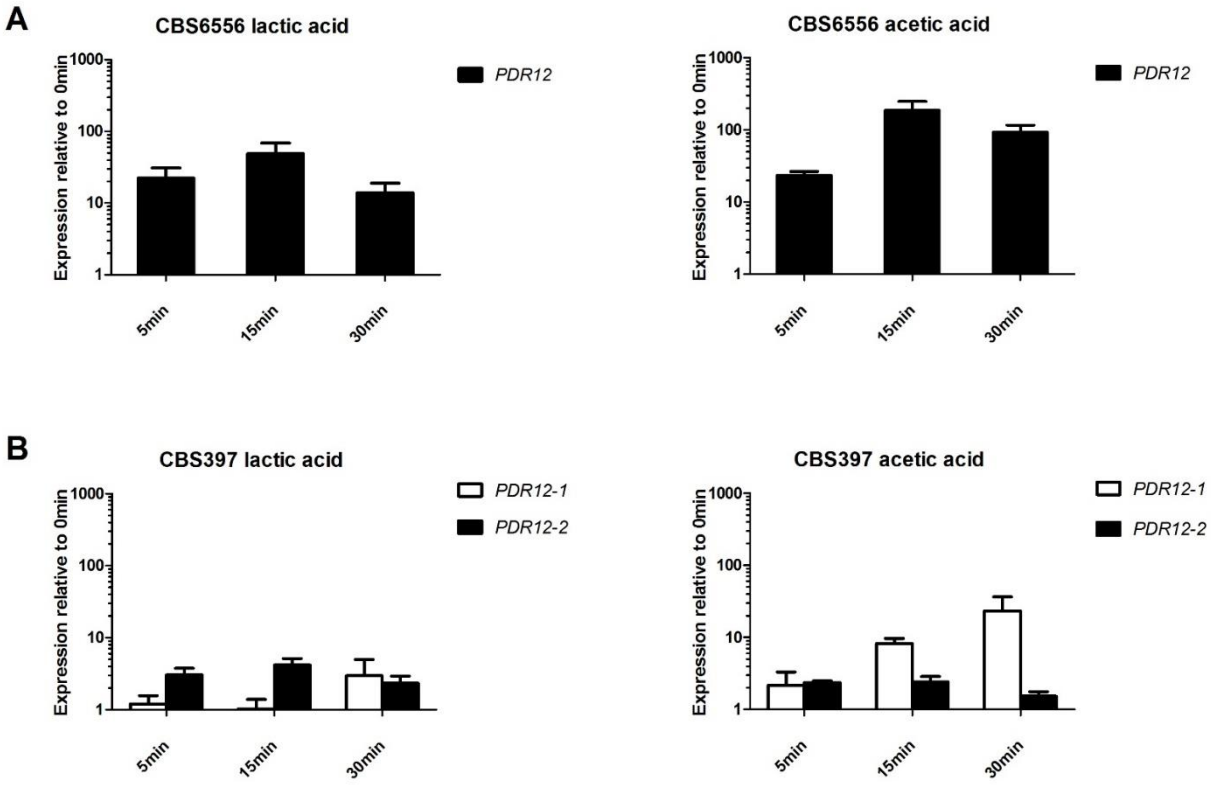


Figure 2 Relative expression profile of PDR12 in CBS397 and CBS6556 shock exposed to acetic acid and lactic acid. The cells were cultivated until mid-exponential phase and resuspended at OD660 = 0.5 in fresh YNB medium (2% glucose) at pH3 (three biological replicates). The cells were incubated for 30mins prior addition of 1gL-1 acetic acid or 30gL-1 lactic acid (all adjusted to pH3). The samples for RNA extraction were taken at 0min, 5min,15min and 30min after addition of indicated acids. The synthesized cDNA was used for subsequent RTqPCR analysis. **A.** Expression profile of CBS6556 PDR12 **B.** Expression profile of CBS397 PDR12-1 and PDR12-2

Comparison of *PDR12* sequences in *K. marxianus* strains

The expression results showed high expression of *PDR12* in comparison to other MDR genes in both strains and condition. We therefore compared the nucleotide sequence identity of CBS6556 *PDR12* to both the CBS397 *PDR12-1* and *PDR12-2* genes. We established that several other strains also carry a duplication of *PDR12*, therefore the *PDR12* sequences from

K. marxianus NBRC 1777 and DMB were also included in the analysis. The phylogenetic tree of NBRC 1777, DMB, CBS6556 and CBS397 nucleotide sequences showed that both *PDR12-1* and *PDR12-2* arose from an ancestral duplication event after the divergence of *K. lactis* and *K. marxianus* (Fig. 3). The CBS6556 *PDR12* sequence clustered with *PDR12-1* but was still more divergent than would be expected. More detailed comparison of CBS6556 *PDR12* to CBS397 *PDR12-1/PDR12-2*, however, revealed that CBS6556 *PDR12* is a recombined version of the gene, having N-terminal (5') sequence identity to *PDR12-1* and C-terminal (3') to *PDR12-2* (Fig. 4A). The recombination point took place in the middle of the ORF, between nucleotides 1887 and 1893 (Fig. 4B).

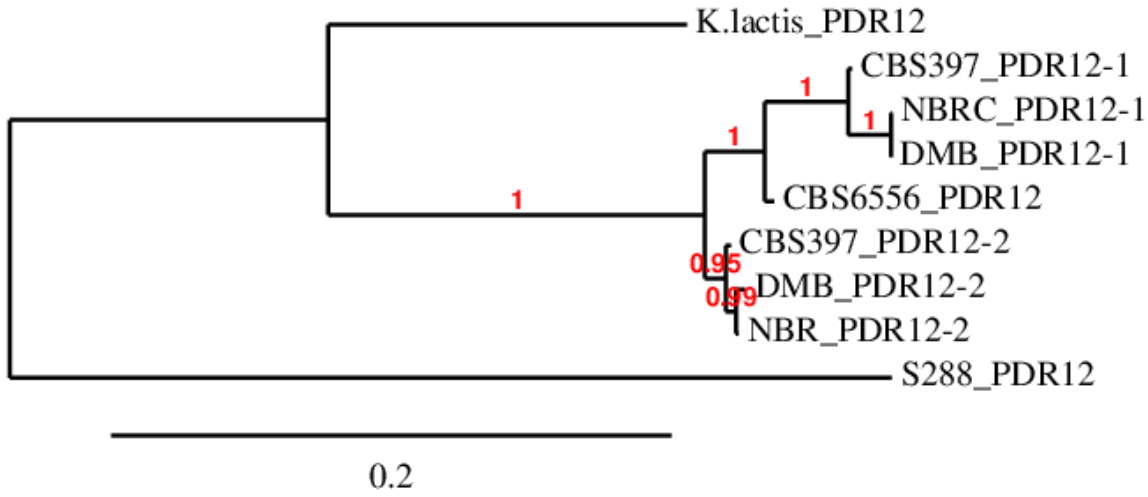
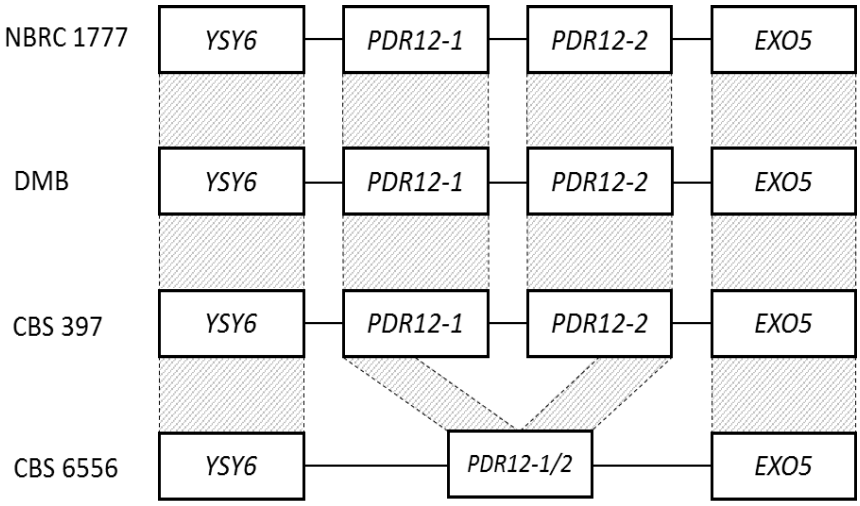


Figure 3 Phylogenic tree of *PDR12* sequences from *S. cerevisiae, K. lactis* and *K. marxianus.* Maximum-likelihood phylogenetic tree generated by comparing nucleotide sequences of *PDR12* from *K. marxianus, K. lactis* and *S. cerevisiae.*

A



B

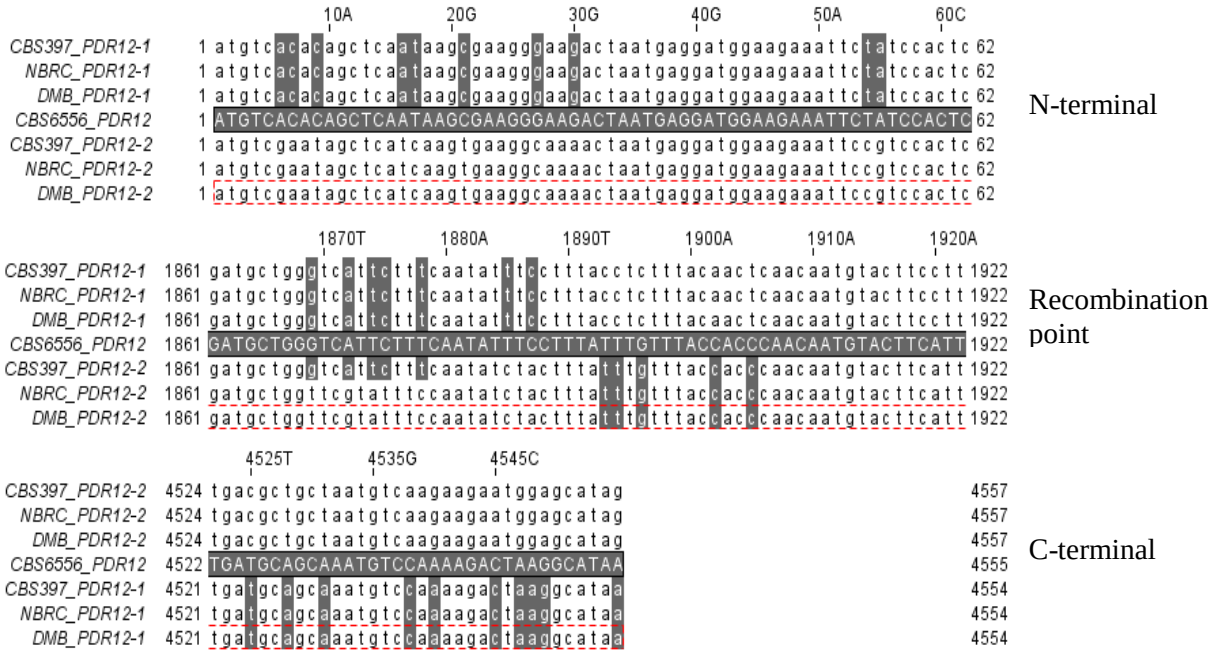


Figure 4 Genomic organisation at the *PDR12* locus in *K. marxianus*. A. Schematic overview showing the *PDR12* locus in *K. marxianus* strains. The recombination between *PDR12*-1 and *PDR12-2* that gave rise to CBS6556 *PDR12* is shown. **B.** Multiple sequence alignment (MSA) of the *PDR12* nucleotide sequences. The CBS6556 *PDR12* sequence showed in upper case and the dark shading in vertical columns highlights the nucleotides that distinguish *PDR12-1* and *PDR12-2.*

Discussion

The aim of this study was to assess acetic and lactic acid tolerance in CBS6556 and CBS397 and to evaluate MDR transporters expression level during the stress. Acetic and lactic acid are mainly encountered during industrial application and can affect the performance of the yeast during fermentation. Previous studies on *K. marxianus* demonstrated wide phenotypic diversity toward stressful condition between strains (Lane *et al.*, 2011, Rocha *et al.*, 2011, Fonseca *et al.*, 2008). In our study, we used CBS 6556 and CBS397, which are among the most studied in the field (Fonseca *et al.*, 2008, Gombert *et al.*, 2016, Jeong *et al.*, 2012, Etschmann *et al.*, 2003). The data collected allowed to confirm strain specific difference to the stress. Moreover, our study demonstrated strain-specific induction of the transporters. The expression profile of *AQR1* and *TPO2* is consistent with *S. cerevisiae* studies, where both strain showed induction during acetic acid stress (Fernandes *et al.*, 2005). However, the presence of lactic acid induced expression of *AQR1* and *TPO2* only in CBS6556. We focused our attention on *PDR12* since CBS397 has two copies of the gene and CBS6556 just one. In addition, the RTqPCR showed that *PDR12* was responsive to both acetic and lactic acid, while in *S. cerevisiae PDR12* is only responsive to moderately lipophilic weak organic acids (sorbic and benzoic acids) (Nygard *et al.*, 2014, Holyoak *et al.*, 1999). However, *PDR12* induction was not uniform in *K. marxianus* strains. CBS397 gradually induced *PDR12-1* during acetic acid shock, whereas *PDR12-2* showed weak induction following exposure to both acetic and lactic acid. In contrast to CBS397, CBS6556 highly induced *PDR12* at both conditions. Despite, the high induction, CBS6556 growth was strongly inhibited by presence of both acids.

Detailed examination of the *PDR12* sequence from both strains, revealed a divergence in the nucleotide sequence. The sequence showed that CBS6556 *PDR12* is a recombined version of CBS397 *PDR12-1* and *PDR12-2*, which probably affects the functionality of the transporter and stress resistance.

Proposed role of Pdr12p in weak acid tolerance

The rapid high level of induction of *PDR12* (up to 100-fold increase) in CBS6556 in response to either lactic or acetic acid indicates that these acids directly induce expression via a

transcription factor binding to *PDR12* promoter elements. In contrast, CBS397 *PDR12-1*, which shares the same promoter, shows a slower and more modest induction to acetic acid and a very small response to lactic acid. CBS397 *PDR12-2* does not respond to acetic acid and has a rapid, but modest, induction in response to acetic acid. CBS397 is also more tolerant to both lactic acid and acetic acid than CBS6556. These observations lead us to present a hypothesis that Pdr12-1p and Pdr12-2p may have complementary roles in transporting acetate and lactate ions, respectively from the cell in CBS397, but the recombinant Pdr12p in CBS6556 can do neither due to structural issues. The very high induction in CBS6556 would be a consequence of a futile effort to induce expression of a detoxification system. This hypothesis can be tested by molecular means such as disruption of *PDR12* genes in CBS397 or heterologous expression of *PDR12*, *PDR12-1* and *PDR12-2* genes in CBS6556. The fact that CBS397 is a diploid yeast and the limited molecular tools available for *K. marxianus* create challenges for these experiments. Nonetheless, construction of the required reagents is ongoing and the proposed role of Pdr12p will be tested.

Acknowledgments

This work was supported by the YEASTCELL Marie-Curie ITN project (REA Grant No. 606795) under the EU's Seventh Framework Programme for Research (FP7).

References

ABBOTT, DA, SUIR, E, VAN MARIS, AJ & PRONK, JT 2008. Physiological and transcriptional responses to high concentrations of lactic acid in anaerobic chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol,* 74**,** 5759-5768.

ALI, MA, YASUI, F, MATSUGO, S & KONISHI, T 2000. The lactate-dependent enhancement of hydroxyl radical generation by the Fenton reaction. *Free Radic Res,* 32**,** 429-438.

BAUER, BE, ROSSINGTON, D, MOLLAPOUR, M*, et al.* 2003. Weak organic acid stress inhibits aromatic amino acid uptake by yeast, causing a strong influence of amino acid auxotrophies on the phenotypes of membrane transporter mutants. *Eur J Biochem,* 270**,** 3189-3195.

CARMELO, V, SANTOS, H & SA-CORREIA, I 1997. Effect of extracellular acidification on the activity of plasma membrane ATPase and on the cytosolic and vacuolar pH of *Saccharomyces cerevisiae*. *Biochim Biophys Acta,* 1325**,** 63-70.

CHRISTENSEN, AD, KADAR, Z, OLESKOWICZ-POPIEL, P & THOMSEN, MH 2011. Production of bioethanol from organic whey using *Kluyveromyces marxianus*. *J Ind Microbiol Biotechnol,* 38**,** 283-289.

ETSCHMANN, MM, SELL, D & SCHRADER, J 2003. Screening of yeasts for the production of the aroma compound 2-phenylethanol in a molasses-based medium. *Biotechnol Lett,* 25**,** 531-536.

FERNANDES, AR, MIRA, NP, VARGAS, RC*, et al.* 2005. *Saccharomyces cerevisiae* adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. *Biochem Biophys Res Commun,* 337**,** 95-103.

FONSECA, GG, HEINZLE, E, WITTMANN, C & GOMBERT, AK 2008. The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl Microbiol Biotechnol,* 79**,** 339-354.

GIANNATTASIO, S, GUARAGNELLA, N, ZDRALEVIC, M & MARRA, E 2013. Molecular mechanisms of *Saccharomyces cerevisiae* stress adaptation and programmed cell death in response to acetic acid. *Front Microbiol,* 4**,** 33.

GOMBERT, AK, MADEIRA, JV, JR., CERDAN, ME & GONZALEZ-SISO, MI 2016. *Kluyveromyces marxianus* as a host for heterologous protein synthesis. *Appl Microbiol Biotechnol,* 100**,** 6193-6208.

GOSHIMA, T, TSUJI, M, INOUE, H*, et al.* 2013. Bioethanol production from Lignocellulosic biomass by a novel *Kluyveromyces marxianus* strain. *Biosci Biotechnol Biochem,* 77**,** 1505-1510.

GUIMARAES, PM, TEIXEIRA, JA & DOMINGUES, L 2010. Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. *Biotechnol Adv,* 28**,** 375-384.

HATZIXANTHIS, K, MOLLAPOUR, M, SEYMOUR, I*, et al.* 2003. Moderately lipophilic carboxylate compounds are the selective inducers of the *Saccharomyces cerevisiae* Pdr12p ATP-binding cassette transporter. *Yeast,* 20**,** 575-585.

HOLYOAK, CD, BRACEY, D, PIPER, PW*, et al.* 1999. The *Saccharomyces cerevisiae* weak-acid-inducible ABC transporter Pdr12 transports fluorescein and preservative anions from the cytosol by an energy-dependent mechanism. *J Bacteriol,* 181**,** 4644-4652.

JEONG, H, LEE, DH, KIM, SH*, et al.* 2012. Genome sequence of the thermotolerant yeast *Kluyveromyces marxianus var. marxianus* KCTC 17555. *Eukaryot Cell,* 11**,** 1584-1585.

JONSSON, LJ, ALRIKSSON, B & NILVEBRANT, NO 2013. Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol Biofuels,* 6**,** 16.

KELLER, G, RAY, E, BROWN, PO & WINGE, DR 2001. Haa1, a protein homologous to the copper-regulated transcription factor Ace1, is a novel transcriptional activator. *J Biol Chem,* 276**,** 38697-38702.

KREN, A, MAMNUN, YM, BAUER, BE*, et al.* 2003. War1p, a novel transcription factor controlling weak acid stress response in yeast. *Mol Cell Biol,* 23**,** 1775-1785.

LANE, MM, BURKE, N, KARREMAN, R*, et al.* 2011. Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. *Antonie Van Leeuwenhoek,* 100**,** 507-519.
LANE, MM & MORRISSEY, JP 2010. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. *Fungal Biology Reviews,* 24**,** 17-26.
LOSER, C, URIT, T, KEIL, P & BLEY, T 2015. Studies on the mechanism of synthesis of ethyl acetate in *Kluyveromyces marxianus* DSM 5422. *Appl Microbiol Biotechnol,* 99**,** 1131-1144.
MIRA, NP, BECKER, JD & SA-CORREIA, I 2010. Genomic expression program involving the Haa1p-regulon in *Saccharomyces cerevisiae* response to acetic acid. *OMICS,* 14**,** 587-601.
MOLLAPOUR, M, FONG, D, BALAKRISHNAN, K*, et al.* 2004. Screening the yeast deletant mutant collection for hypersensitivity and hyper-resistance to sorbate, a weak organic acid food preservative. *Yeast,* 21**,** 927-946.
MOLLAPOUR, M, SHEPHERD, A & PIPER, PW 2008. Novel stress responses facilitate *Saccharomyces cerevisiae* growth in the presence of the monocarboxylate preservatives. *Yeast,* 25**,** 169-177.
MORRISSEY, JP, ETSCHMANN, MM, SCHRADER, J & DE BILLERBECK, GM 2015. Cell factory applications of the yeast *Kluyveromyces marxianus* for the biotechnological production of natural flavour and fragrance molecules. *Yeast,* 32**,** 3-16.
NARENDRANATH, NV, THOMAS, KC & INGLEDEW, WM 2001. Effects of acetic acid and lactic acid on the growth of *Saccharomyces cerevisiae* in a minimal medium. *J Ind Microbiol Biotechnol,* 26**,** 171-177.
NYGARD, Y, MOJZITA, D, TOIVARI, M*, et al.* 2014. The diverse role of Pdr12 in resistance to weak organic acids. *Yeast,* 31**,** 219-232.
PFAFFL, MW 2001. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research,* 29**,** e45-e45.
PIPER, P, CALDERON, CO, HATZIXANTHIS, K & MOLLAPOUR, M 2001. Weak acid adaptation: the stress response that confers yeasts with resistance to organic acid food preservatives. *Microbiology,* 147**,** 2635-2642.
RADECKA, D, MUKHERJEE, V, MATEO, RQ*, et al.* 2015. Looking beyond Saccharomyces: the potential of non-conventional yeast species for desirable traits in bioethanol fermentation. *FEMS Yeast Res,* 15.
ROCHA, SN, ABRAHAO-NETO, J & GOMBERT, AK 2011. Physiological diversity within the *Kluyveromyces marxianus* species [corrected]. *Antonie Van Leeuwenhoek,* 100**,** 619-630.
SCHULLER, C, MAMNUN, YM, MOLLAPOUR, M*, et al.* 2004. Global phenotypic analysis and transcriptional profiling defines the weak acid stress response regulon in *Saccharomyces cerevisiae*. *Mol Biol Cell,* 15**,** 706-720.

Conclusion

The ongoing need to substitute the majority of fossil derived chemicals and fuels with bio-based products requires a biorefinery to be efficient and economically viable. Therefore, it is important to develop effective and robust microbial cell factories, which can be applied for lignocellulosic or other residual biomass fermentations and produce economically viable bio-based products. Weak organic acids are among the interesting products obtained by microbial fermentation that have been already introduced to the market. The production of weak organic acids requires a robust microbial host able to cope with the high titer of the product, and in most cases at low pH, which favors the desired undissociated form of the product. Next generation hosts need to be robust also against inhibitors present in the lignocellulosic biomass. The availability of robust strains for organic acid production can therefore reduce process cost and contribute to the success of second generation industrial-scale production. For some of the organic acids on the market, natural producers are used (e.g. citric acid), but for others, in addition to natural producers, engineered hosts have been developed, matching specific industrial requirements to peculiar physiological traits. This is, for example, the case of lactic acid, where starting from the mid , scientists proposed substituting lactobacilli, the natural producers, with yeast, among which was *Saccharomyces cerevisiae* (Porro *et al.*, 1995). Yeast was preferred over bacteria because it can produce lactic acid at low pH and in microaerobic conditions. After intensive cycles of strain engineering and selection, superior lactic acid-producing strains of *S. cerevisiae* were generated (Porro *et al.*, 1995, Pacheco *et al.*, 2012, Valli *et al.*, 2006, Ishida *et al.*, 2005, Bianchi *et al.*, 2001).

There is, however, always room for further improvement of producing strain by exploiting yeasts that naturally more robust and versatile than *S. cerevisiae.* Given the wide diversity of yeasts, it is clear that alternative ways to develop or to improve existing microbial hosts is to further explore non-Saccharomyces yeasts, as already proven with *Kluyveromyces lactis* (Bianchi *et al.*, 2001), *Candida utilis* and *Candida boidinii* (Osawa *et al.*, 2009, Ikushima *et al.*, 2009). Furthermore, the company Cargill has developed yeast-based production of commercial lactic acid using an undisclosed yeast species. This positive result is paving the way for the development of additional alternative hosts.

The present study investigates the potential of *Z. parabailii* and *K. marxianus* as microbial cell factories from a weak acid stress resistance perspective. *Z. parabailii* is a well-known food

spoilage yeast with a native ability to resist high concentrations of weak acids at low pH. *K. marxianus* is a thermotolerant, fast growing yeast with an innate ability to ferment pentose and hexose sugars. Both yeasts are relevant for biotechnology exploitation, therefore understanding the weak acid resistance will provide an opportunity for their efficient exploitation.

Z. bailii sensu lato has been extensively studied in last decades by food scientists to understand the spoilage mechanism and develop preventative measures. In Chapter 1, we discussed the important findings on *Z. bailii* and its relatedness to biotechnology exploitation. We emphasized the studies done from food science moving then to perspectives for biotechnology. Chapter 1 clearly indicates the potential of *Z. bailii* as a promising host for various bio-based products, where resilience to low pH and weak organic acids is required. Knowing the potential of *Z. bailii*, in Chapter 2 we studied the lactic acid stress response of the yeast. Our study revealed that lactic acid does not cause an extended lag phase in growth and does not affect viability of the cells at a concentration that impairs growth in *S. cerevisiae*. These intriguing results were further expanded to understand the macromolecular changes induced by lactic acid, to depict possible unique response network. We observed that *Z. bailii* cells indeed respond to the lactic acid by reorganizing the composition of cell wall and cell membrane. These findings are important to understand the lactic acid tolerance in this species, although it would be desirable to narrow down the analysis to the level of the specific molecular changes. In addition, the lack of a completely annotated and assembled genome sequence of *Z. bailii* limited our study. Therefore, Chapter 3 was dedicated to the genomic study of the *Z. bailii* strain selected for our research using PacBio sequencing technology. The study revealed that this strain is in fact a hybrid *Z. parabailii*. This finding confirms, in agreement with previous studies, that *Z. bailii sensu lato* is highly divergent. Intriguingly, the genome study of *Z. parabailii* paved the way to understand and possibly to answer important fundamental questions related to the whole genome duplication event that occurred in other yeasts (Wolfe, 2015). The study provided a basis to hypothesize that *Z. parabailii* is undergoing a similar fate as it was with the lineage leading to *S. cerevisiae*. This case can be a perfect example of the tight connection existing between fundamental and applied science. Knowing the physio-morphological response to lactic acid and genomic data, in Chapter 4 we investigated the transcriptomic response of *Z. parabailii* to lactic acid. The study revealed that *Z. parabailii* to some extent has similarities and differences in transcriptomic response if

compared with previous studies performed in *S. cerevisiae*. The similarities include activation of iron transport related genes and of ROS detoxifying genes. Differences are related to the downregulation of cell wall genes and MDR transporter genes associated with weak organic acid tolerance. Moreover, we identified multifamily *Z. parabailii* specific genes that are responsive to lactic acid presence. The observation clearly indicates that *Z. parabailii* still has some secrets to be unveiled, and RNA-seq results can be a starting point to discover them.

The last part of the study was dedicated to the other industrially relevant yeast selected – *K. marxianus*. The yeast is getting a lot of attention from biotechnology companies and scientists due to its unique combination of abilities. Our study was focused on understanding the weak acid stress response of MDR transporters. As for *Z. bailii*, strong strain variability is known (Lane *et al.*, 2011), and our study confirmed that *K. marxianus* has strain-specific stress tolerance also in respect to organic acids. We focused our attention on lactic and acetic acid, in agreement with the general aim of the work.

Our study revealed that the strain variability could be related to the copy number and sequence identity of the *PDR12* gene, which encodes for one of the major membrane transporters responsible for counteracting acetic acid toxicity. Furthermore, we found that sequence difference in *PDR12* of the two strains studied might affect the tolerance level to acetic and lactic acid. With the development of modern molecular tools, we can continue studying this observation by deleting the genes and inserting the gene from the tolerant strain to the weak one.

Indeed, both *Z. parabailii* and *K. marxianus* represent interesting species due to industrial relevance and exploitation potential. This work was focused in getting insight into the weak acid resistance mechanisms in different yeast species to understand how biodiversity and habitat tailor intrinsic adaptation to this stress. This knowledge can be further improved and exploited to transfer the traits into chosen microbial cell factory strains.

This work highlighted the importance of yeast biodiversity exploration. There are still yeasts with unique traits that should be studied and characterized. Therefore, we should not limit our interest only to the organism we know the best for industrial application, but expand and exploit nature’s diversity.

References

BIANCHI, MM, BRAMBILLA, L, PROTANI, F*, et al.* 2001. Efficient homolactic fermentation by *Kluyveromyces lactis* strains defective in pyruvate utilization and transformed with the heterologous LDH gene. *Appl Environ Microbiol,* 67**,** 5621-5625.

IKUSHIMA, S, FUJII, T, KOBAYASHI, O*, et al.* 2009. Genetic engineering of *Candida utilis* yeast for efficient production of L-lactic acid. *Biosci Biotechnol Biochem,* 73**,** 1818-1824.

ISHIDA, N, SAITOH, S, TOKUHIRO, K*, et al.* 2005. Efficient production of L-Lactic acid by metabolically engineered *Saccharomyces cerevisiae* with a genome-integrated L-lactate dehydrogenase gene. *Appl Environ Microbiol,* 71**,** 1964-1970.

LANE, MM, BURKE, N, KARREMAN, R*, et al.* 2011. Physiological and metabolic diversity in the yeast *Kluyveromyces marxianus*. *Antonie Van Leeuwenhoek,* 100**,** 507-519.

OSAWA, F, FUJII, T, NISHIDA, T*, et al.* 2009. Efficient production of L-lactic acid by Crabtree-negative yeast *Candida boidinii*. *Yeast,* 26**,** 485-496.

PACHECO, A, TALAIA, G, SA-PESSOA, J*, et al.* 2012. Lactic acid production in *Saccharomyces cerevisiae* is modulated by expression of the monocarboxylate transporters Jen1 and Ady2. *FEMS Yeast Res,* 12**,** 375-381.

PORRO, D, BRAMBILLA, L, RANZI, BM*, et al.* 1995. Development of metabolically engineered *Saccharomyces cerevisiae* cells for the production of lactic acid. *Biotechnol Prog,* 11**,** 294-298.

VALLI, M, SAUER, M, BRANDUARDI, P*, et al.* 2006. Improvement of lactic acid production in *Saccharomyces cerevisiae* by cell sorting for high intracellular pH. *Appl Environ Microbiol,* 72**,** 5492-5499.

WOLFE, KH 2015. Origin of the Yeast Whole-Genome Duplication. *PLoS Biol,* 13**,** e1002221.

Acknowledgments

My PhD course within YEASCELL program was amazing adventure both from personal and scientific perspective. I am deeply grateful to people who shared with me this wonderful experience.

First, I would like to express my deepest gratitude and respect to my supervisors **Prof. Paola Branduardi** and **Dr. John P. Morrissey** for giving me opportunity to work and study at your labs. You helped me to build the foundations for my scientific career, through encouragement and support, mentoring, patience and constructive criticism. I truly admire your passion for the science and care for each student working at your labs. Also, I greatly appreciate support and assistance from **Dr. Francesca Doonan,** who did a great job in organizing meetings, answering inquiries and just being very nice to awesome YEASTCELL group.

My research wouldn't have been the same if I had not met my YEASCELL fellow, friend and great collaborator **Raul Ortiz-Merino**. We did a great work together on the genome and transcriptome study. You helped me to develop an interest in bioinformatics and in drinking good craft beer.

Also, I would like to thank my friend **Javier Varela** who helped me a lot during my stay at UCC. I appreciate your ability to quickly grasp the research topic and provide valuable discussion for the topic. I really enjoyed your guitar play.

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. The advices and guidance during experimental procedures like bioreactor techniques (**Dr. Lorenzo Signori**, **Ricardo Posteri** and **Dr. Stefano Bertagnoli**), molecular biology (**Dr. Francesca Martani**, **Francesca Marano** and **Dr. Nadia Berterame**) and scientific discussion (**Marco Brambilla**, **Stefano Bertacchi and Dr. Jan Knudsen**) was crucial for my PhD course. Thanks to the bright Master's students who are always around to help or to clean glassware ☺.

At last, I give my deepest gratitude to my wife **Dinara** and **Mum** and **Dad**. I greatly appreciate your love and never-ending support that helped me to go through this difficult but amazing journey. I love you with all my hearth.

UNIVERSITA' DEGLI STUDI DI MILANO-BICOCCA

DIPARTIMENTO DI BIOTECNOLOGIE E BIOSCIENZE

The revision of the thesis:
" Elucidation of weak organic acid resistance mechanisms in non-Saccharomyces yeast
A case study of *Zygosaccharomyces parabailii* and *Kluyveromyces marxianus*"

Dear PhD defence committee members,

I am very grateful to the Referees for positive comments and constructive suggestions about the thesis. I would like to resubmit to your attention this revised version. The changes in the text are highlighted in yellow.

In the following, you can find a point-by-point response to the Referees' comments.
I addressed all the points of the Referees.

Therefore, I hope that the thesis in the present form will be acceptable to fulfil the PhD degree requirements of UNIMIB/UCC.

Looking forward for a positive reply,

Yours sincerely,

Nurzhan Kuanyshev

Referee 1. Prof. Sauer

Point 1:

"The introduction goes very quickly to the single points of the individual works, without giving the whole picture. Biorefineries for example, comprise many more technologies than industrial microbiology. Functional concepts must inevitably combine physical, chemical and biological processes, in order to be efficient and competitive. This has not even been mentioned"

AUTHOR:

The following paragraphs were added to introduction section to address this point:

"The technologies that are jointly applied to efficiently convert biomass into valuable products can be divided into four groups:

1. Thermochemical processes: The aim of thermochemical process is to convert biomass into energy and chemical products by gasification and pyrolysis. Gasification is biomass treatment at high temperature (above 700°C) with low oxygen levels to produce syngas. Pyrolysis is biomass treatment at intermediate temperature (300-600°C) without oxygen to produce pyrolytic oil (or bio-oil), solid charcoal and light gases like syngas. Both products can be directly used as a stationary biofuel or as precursors to produce other fuels (Spath and Dayton, 2003, Bridgwater and Peacocke, 2000).
2. Mechanical/physical processes: The aim of a mechanical process is to reduce the size and separate the components of the biomass, without changing the state and composition of it. This process is usually performed before biomass utilization to reduce the size of biomass within specific ranges for an efficient subsequent processing (Sun and Cheng, 2002).
3. Chemical processes: The aim of chemical processes is to change the chemical composition and structure of biomass by hydrolysis. Hydrolysis uses acids, bases and enzymes to depolymerize polysaccharides into monomers (cellulose into glucose) for subsequent processing of the monomers into valuable products (Sun and Cheng, 2002).
4. Biochemical processes: Biochemical processes involve application of fermentation via microbial cell factories to convert fermentable substrate into desired product. The process occurs at low temperature and slower rate in comparison to thermochemical processing (Hamelinck et al., 2005)."

Point 2:

"Its relationship with Z. parabailii has been mentioned only in one sentence. Clearly, the starting idea was to work on Z. bailii, but analyzing the genome sequence it became clear that the strain was sth. else. This fact is interesting and allows many conclusions, but it requires explanation and contemplation of the PhD candidate, which is unfortunately absent"

AUTHOR:

The following paragraphs were added to introduction section to address this point:

"Despite the importance of the yeast for industrial and fundamental microbiology, accurate identification of *Z. bailii* and related strains is problematic. The phylogenetic relationships of many industrial isolates formerly known as *Z. bailii* have been re-evaluated, and significant differences in rRNA gene sequences were found. These led to the proposal that there are two novel species closely related to *Z. bailii*, namely *Zygosaccharomyces parabailii* and *Zygosaccharomyces pseudobailii* (Suh et al., 2013). Indeed, further genome sequencing projects of commonly-used strains confirmed that in fact ISA1307 and ATCC60483 are interspecies hybrids of *Z. bailii* and closely related species rather than pure strains (Mira et al., 2014, Ortiz-Merino et al., 2017). Regardless of molecular differences between *Z. bailii sensu lato* species, physiological traits are undistinguishable."

Point 3:

"Of course this part of the work fits very well into the context of this thesis. However, the chapter stands isolated and Mr. Kuanyshev didn't even try to connect it in any way – which is a pity"

AUTHOR:

The following paragraphs were added to conclusion section to address this point:

"Indeed, both *Z. parabailii* and *K. marxianus* represent interesting species due to industrial relevance and exploitation potential. This work was focused in getting insight into weak acid resistance mechanism in different yeast species to understand how biodiversity and habitat tailor intrinsic adaptation to this stress. This knowledge can be further improved and exploited to transfer the traits into chosen microbial cell factory strains."

Referee 2 Prof. Villaverde

Point 1:

"As a minor issue for consideration, the author indicates in the publication list (page 7 of the MS), the references of three published papers on which the thesis is based. Two of these references are incomplete (paper 1 and paper 2), what should be amended for the final presentation as both have been already published."

AUTHOR:

This point was duly addressed and references were modified accordingly

Additional changes made by author:

Chapter 4 has been revised and changed to final submission form to the journal. General concept and goal of the chapter are not changed.