Identification and molecular characterization of the highly acetic acid tolerant *Zygosaccharomyces bailii* strain IST302

João Diogo André Peça

Thesis to obtain the Master of Science Degree in

Microbiology

Supervisor: Prof. Dr. Isabel Maria de Sá-Correia Leite de Almeida

Co-supervisor: Dr. Margarida Isabel Rosa Bento Palma

Examination Committee

Chairperson: Prof. Dr. Arsénio do Carmo Sales Mendes Fialho

Supervisor: Prof. Dr. Isabel Maria de Sá-Correia Leite de Almeida

Member of the committee: Dr. Paulo Jorge Moura Pinto da Costa Dias

July 2016

Agradecimentos

Apesar de perceber a globalização em que o mundo científico se insere onde o inglês domina como língua de comunicação entre todos, decido escrever apenas este capítulo em português pois entendo que os agradecimentos, sendo uma mensagem mais emocional e especificamente dirigida a pessoas próximas, só pode ser completamente entendida e percebida se for usada a minha língua materna.

Gostaria de começar por agradecer a duas pessoas muito importantes durante o processo de trabalho desta tese, quer na parte laboratorial quer na parte de escrita. À Professora Isabel Sá-Correia, primeiro, por me ter dado a oportunidade de trabalhar consigo e fazer parte do Grupo de Investigação para as Ciências Biológicas do IST (BSRG) tendo sempre palavras assertivas mas construtivas que me fizeram acreditar no meu trabalho. Em segundo, à Dr. Margarida Palma, minha coorientadora, que se mostrou incansável em me apoiar sempre que precisei e mesmo quando não pedia, sendo muito importante no processo experimental e de escrita deste documento. Ciente de que a minha desmotivação por vezes prejudicou este trabalho, sem estas duas pessoas e a paciência que mostraram para comigo, este trabalho não seria possível. A ambas o meu mais profundo agradecimento. Gostaria ainda de agradecer imenso a disponibilidade e vontade que a Dr. Cláudia Lobato Silva e o Ricardo Pereira dispensaram para me ajudar em toda a parte de citometria de fluxo que esta tese envolveu.

Para a realização desta tese é importante salientar e reconhecer o fundo recebido pelo IBB (Institute for Biotechnology and Bioengineering) a partir da Fundação para a Ciência e Tecnologia (FCT), (UID/BIO/04565/2013 e BBI14) e do Programa Operacional Regional de Lisboa 2020 (N. 007317).

Foram bastantes os amigos que ao longo destes anos fui fazendo e amizades que fui fortalecendo, e para estes quero também deixar uma grande palavra de apreço pela companhia, discussões, alegrias e tristezas que fomos partilhando. Aos meus colegas de mestrado e antigos colegas de curso em especial ao Tiago Pedreira, Bernardo Carriço, Laura Luzia, Miguel Santos e Pedro Moura aos meus amigos de longa data que embora conheçam áreas de estudo diferentes da minha nunca se abstiveram de me encorajar e dar o seu apoio, em especial ao André Gomes, Ana Raquel, João Agostinho, Vítor Vicente e Soraia Vieira e ao pessoal da anTUNiA (Tuna de Ciências e Tecnologia da Universidade Nova de Lisboa), grupo no qual se fazem amigos para vida dado o objetivo comum por que se luta, a música.

Por fim, à minha família, na pessoa dos meus pais e irmão, exemplos de vida para mim, não existem palavras que cheguem para agradecer o seu apoio incondicional mesmo quando tudo parecia difícil, sendo um suporte constante ao longo destes anos e que independentemente de qualquer decisão e consequência na minha vida estiveram sempre do meu lado dizendo o que precisava de ouvir e o que não queria mas devia de ouvir. Mais do que estas palavras, espero concretizar todo o sucesso e crença que depositam em mim de forma a deixá-los orgulhosos e que todos os seus sacrifícios não tenham sido em vão. A vocês, o mais profundo obrigado.

Abstract

Zygosaccharomyces bailii species is among the most threatening yeasts found in acidic foods and beverages. Its ability to proliferate under high concentrations of weak acid preservatives and fermentation products makes this spoilage yeast species an economical issue in foodstuff industry. In this work the yeast strain IST302 obtained from Douro wines was identified as *Z. bailii* through an *in silico* phylogenetic analysis using the partial gene sequence of D1/D2 region of LSU rRNA gene and nucleotide variations of highly conserved genes. Pulse-field gel electrophoresis and flow cytometry were used to characterize the karyotype and ploidy of strain IST302. Following comparison of *Z. bailii* IST302 with other *Z. bailii* and *S. cerevisiae* strains, results provided strong evidence that *Z bailii* IST302 is haploid with a genome size of approximately 11Mbp.

Although the mechanisms regarding *Z. bailii* tolerance under acetic acid stress remain unclear it is known that its metabolism, along with phenotypic heterogeneity within the population, contribute to its remarkable tolerance to acetic acid. The role of *Z. bailii* gene *ZbTPO2/3,* homologous to *S. cerevisiae TPO2* and *TPO3* that code for two plasma membrane transporters previously suggested to have an impact on the active export of acetic acid, thus reducing intracellular acid concentration, was studied in *Z. bailii* IST302 challenged with acetic acid. Heterologous expression of *ZbTPO2/*3 in S*. cerevisiae tpo2*Δ and *tpo3*Δ did not rescue the acetic acid susceptibility phenotype of these mutants, suggesting that this *Z. bailii* gene is not a functional homologue of *S. cerevisiae TPO2* and *TPO3*.

Keywords: *Zygosaccharomyces bailii*; food spoilage yeast; IST302 strain; acetic acid; *TPO2*; *TPO3*; *ZbTPO2/3*

Resumo

No quadro de espécies de levedura contaminantes de bebidas e alimentos ácidos, *Zygosaccharomyces bailii* é das mais problemáticas, muito devido à sua incrível habilidade para proliferar sob altas concentrações de conservantes e produtos da fermentação, constituindo uma problemática a nível económico na indústria alimentar. A estirpe de levedura IST302, obtida nas vinhas do Douro é identificada como *Z. bailii* através de uma análise filogenética, usando sequências parciais das regiões D1/D2 do gene LSU rRNA, e ainda analisando variações nucleotídicas de genes altamente conservados. A conjugação das técnicas moleculares, Electroforese de Campo Pulsado com Citometria de Fluxo, evidencia que a estirpe IST302 é haploide com um genoma de aproximadamente 11Mbp.

Apesar de não se compreender totalmente os mecanismos de tolerância de *Z. bailii* ao ácido acético, as suas características metabólicas, associada à heterogeneidade fenotípica dentro da população celular de *Z. bailii*, pensam-se contribuir para a tolerância desta espécie ao ácido acético. Em *Saccharomyces cerevisiae*, é sabido que os genes *TPO2* e *TPO3*, codificantes para transportadores de membrana da família de proteínas DHA1 da MFS-MDR, desempenham um importante papel no exporte activo de ácido acético, reduzindo a concentração intracelular do ácido. A função do gene de *Z. bailii* homólogo do *TPO2* e *TPO3* de *S. cerevisiae*, designado de *ZbTPO2/3*, é avaliada neste trabalho. A sua expressão heteróloga em *S. cerevisiae tpo2*Δ e *tpo3*Δ não restabeleceu o fenótipo de susceptibilidade ao ácido acético desses mutantes, sugerindo que o gene *ZbTPO2/3* não é um homólogo funcional dos genes *TPO2* e *TPO3* de *S. cerevisiae*.

Palavras-chave: *Zygosaccharomyces bailii*; leveduras; estirpe IST302; ácido acético; *TPO2*; *TPO3*; *ZbTPO2/3*

Contents

List of Figures

List of Tables

List of Abbreviations

LSU rRNA	Large Subunit of rRNA gene
ITS	Internal Transcribed Spacer region
TCA	Tri-Carboxylic Cycle
MDR	Multidrug Resistance
ABC	ATP-binding cassete
MFS	Major Facilitator Superfamily
TMS	Transmembrane spanner
DHA	$Drug^{+}$ H antiporter
WGD	Whole Genome Duplication
OD_{600nm}	Optical density at 600nm
YPD	Yeast extract Peptone Dextrose medium
EDTA	Ethylenediamine Tetraacetic Acid
SCE	Sorbitol Sodium Citrate EDTA buffer
ETB	EDTA-Tris β-Mercaptoethanol solution
TE	Tris-EDTA
PFGE	Pulse-field Gel Electrophoresis
CHEF	Clamped Homogeneous Electric Field Electrophoresis
TAE	Tris Acetic acid EDTA
EUROSCARF	European Saccharomyces cerevisiae archive for functional analysis
PCR	Polymerase Chain Reaction
MMB-U	Minimal medium broth supplement with aminoacids less uracil
*cox*II	Cytochrome Oxidase subunit 2
EF1-α	Translation Elongation Factor 1-*α*

mSSU rRNA	Mitochondrial Small Subunit 18S rRNA gene
RPB1	RNA polymerase II largest subunit
RPB2	RNA polymerase II second-largest subunit
BLAST	*Basic Local Alignment Sequence Tool*
FSC	Forward Scatter
SSC	Side Scatter
FL1-A	Fluorescence Light intensity area
FL1-H	Fluorescence Light intensity high
FL1-W	Fluorescence Light intensity width
RMSD	Root Mean Square Deviation or Error
CV	Coefficient of Variation

1. Introduction

1.1 Relevance of *Zygosaccharomyces bailii* in the spoilage of acidic foods and beverages.

Species within the *Zygosaccharomyces* genus are frequently referred in the literature as the most potentially threatening food spoilage yeasts [1–4]. Despite the legislation and high quality control enforced nowadays in food industry regarding the use of preservatives and other conservation methods, there are some species that exhibit an extraordinary resistance and capacity to proliferate in the manufactured foodstuff [5,6]. To inhibit the growth and proliferation of these microorganisms it is necessary to increase the concentration of additives which represents a legal barrier or it could mean, from the consumer point of view, a change in the natural aspect, texture and even flavor of the respective food [3,7]. This will turn the product unappealing to the consumer which translates in loss of industry reliability and consequent economic costs associated to product reproduction and replacement [4].

The concept of food spoilage yeast is in constant debate in the literature and it can have various definitions depending on the balance between the benefits and harms to the processed/treated foodstuff [8]. Through a practical point of view, it can be defined as an agent capable of spoil foodstuff that were processed and packaged according to standard protocols and measures of good manufacturing practices (GMPs) [4]. The food and beverage spoilage yeast *Z. bailii* is one of the main species responsible for acidic food and beverage spoilage, in particular salad dressings, sauces, mustards, pickles, vinegar, carbonated beverages, soft-drinks and some wines[1,9,10]. An explanation for *Z. bailii* spoilage in all these different products, is that they have some natural or implemented characteristics in common, such as low pH, low water activity (a_w), high concentrations of sugar or other metabolized carbon compounds (e.g., alcohol, glycerol or acetic acid), as well as preservatives such as weak acids (sorbic, benzoic and acetic acids) or sulphur dioxide [2,11], which represents an optimal environment for this species. To support this fact, it was recently shown that the natural ecological niches of *Z. bailii*, which are often found in dried, mummified fruits and tree exudates, orchards and vineyards, are very similar to those found in the spoiled foodstuff or manufactured products [1].

1.1.1 Taxonomic position of *Zygosaccharomyces bailii*

Several different criteria have been used for yeasts classification and identification over time. The first aspects that are often utilized to categorize and identify a species within yeasts are their morphological and histological characteristics such as, size and shape of their cells, the different kinds of structures comprehending them and their modes of asexual and, if observed, sexual reproduction [12]. However, these kind of diagnosing characters, frequently, do not allow the correct differentiation between microorganisms, especially, the ones from lowest taxa, such as species and strains. Other common criteria are based on the physiological and biochemical properties of yeasts, such as for example, their fermentation metabolism of different sugars, growth under different carbon and nitrogen sources, temperatures and antifungals [9]. Recently, a wide variety of new DNA and RNA-based techniques

conjugated with helpful bioinformatics tools, have become increasingly popular among the scientific community as methods for phylogenetic and taxonomic identification and differentiation at species and strain levels, providing more accurate results regarding taxonomic classification [13,14], given the comparison of highly conserved genetic domains. Although these recent approaches enable faster results regarding strains and species identification and classification, it is wise to combine the potential of all differentiating criteria.

Z. bailii was first discovered by Lindner in 1895 and described as a *Saccharomyces bailii*. The binomial name suffered few alterations until 1983 when Barnett reclassified it as a *Zygosaccharomyces bailii*, being this last nomenclature the one approved according the requirements of the *International Code of Botanical Nomenclature*. This eukaryotic species is classified, from lowest to highest taxa, in genus *Zygosaccharomyces*, family *Saccharomycetaceae*, order *Saccharomycetales*, class *Saccharomycetes*, subphylum *Saccharomycotina*, phylum *Ascomycota* of the *Fungi* kingdom. Presently, it is one of six species belonging to the genus, along *Z. bisporus, Z. kombunchaensis, Z. lentus, Z. mellis* and *Z. rouxii*. However, recent studies, based on conserved ITS (Internal Transcribed Spacer) regions and the D1/D2 domain regions of the LSU (Large Subunit) rRNA genes, and also based on phenotypic and physiological assays have proposed seven novel species to the genus, namely, *Z. machadoi* [15], *Z. gambellarensis* [16], *Z. siamensis* [17], *Z. sapae* [18], *Z. parabailii, Z. pseudobailii* [19] and *Z. favi* [20]. The wide use of rRNA gene domains in those studies can be considered for taxonomic identification of a species [21], given that it confers an accurate taxonomic tool to rapidly identify and evaluate conspecific strains, since these genetic elements have a sufficiently variable degree to allow the distinction between different species and are conserved enough to not show significant variations among strains of the same species [22]. In order to confirm the conspecificity the variations in those genetic domains usually correspond to less than 1% nucleotide substitutions [23].

Regarding the number of chromosomes of *Z. bailii* species the *Z. bailii* type strain CLIB213^T, whose genome was recently sequenced and annotated [24], has five chromosomes whilst strain ISA1307, previously considered a *Z. bailii* species, holds a total of 13 chromosomes [25]. Indeed, the genome sequencing and annotation of strain ISA1307 demonstrated that this strain is an interspecies hybrid between *Z. bailii* and a closely-related species due to the duplication of certain house-keeping genes, showing approximately 90% allelic divergence among those duplicated genes in ISA1307 strain. The second parental strain was suggested as being a *Z. parabailii* species, given that several of those orthologous alleles are practically identical to those found in *Z. parabailii* [25].

1.1.2 General morphological, physiological and metabolic characteristics

At a morphological level, *Z. bailii* vegetative cells, under the microscope observation, are usually large (5-8 x 3-5 µm), ovoidal, or ellipsoid in shape and sometimes forming pseudohyphae. Its cell cycle is often characterized by a long asexual phase where reproduction occurs by multilateral budding with the formation of mitotic spores [26], while sexual reproduction involves the formation of persistent asci, that generally are conjugated and contain one to four spherical or ellipsoidal, thin walled ascospores. On

several media, colonies are generally smooth, round, convex and white to cream colored [1,27]. Some strains tend to have a higher rate of flocculation causing a dense precipitate in liquid media.

During the past decades, the study of *Z. bailii* metabolism has been considered essential to understand its ability to spoil foodstuff. The understanding of the principal mechanisms, carbon sources used and behavior, either under aerobic or anaerobic conditions is an important topic of discussion among the literature in order to provide indications about how to overcome this yeast species as a food spoilage agent. *Z. bailii* can be characterized as fructophilic, osmotolerant, highly fermentative and extremely weak-acid preservative resistant yeast species [28]. As a fructophilic yeast, this species has a particular preference for consuming fructose over glucose, even when both are present in the medium [29,30]. This behavior could be explained by the fact that *Z. bailii* cells possess a high-capacity, low-affinity transport system specific for fructose; a second transporter is a low-capacity and high-affinity hexose uptake system. In this case, fructose and glucose compete for the hexose transport system, but high concentrations of fructose lead to its inactivation [29]. Other remarkable feature of this yeast species is its apparent capacity to simultaneously metabolize acetic acid and glucose, even when glucose is present in the medium. Although in several yeast species, such as in *Saccharomyces cerevisiae*, a glucose repression phenomenon, at transcription level, affects negatively the enzymes required to metabolize other sugars and non-fermentable carbon sources [31], in *Z. bailii* this trait is not observed. In mixed-substrate media, glucose and acetic acid are used as energy and carbon sources, with the dissociated form of the carboxylic acid (acetate) being transported through the membrane by an acetic acid carrier, controlled by the intracellular acid concentration [32]. Thus, acetic acid is used as an additional energy source to the TCA cycle (Tricarboxylic Acid) and/or as a precursor to the lipid biosynthetic pathway. Due to the high amount of acetyl-CoA produced via the acetic acid metabolism, the TCA cycle is replenished by the conversion of pyruvate into oxaloacetate [33,34]. It has already been shown that *Z. bailii* appears to have a remarkable ability to survive and grow under stressful conditions, specifically, tolerating low pHs (pH 2.2) an short-term variations of intracellular pH [27,35], concentrations up to 4M of glucose and nearly pasteurizing temperatures between 50 and 60 ºC [7].

1.1.3 *Z. bailii* mechanisms of tolerance

The low pH often present in beverages, salad dressings and sauces, due to the natural chemical composition of the initial foodstuff, is a characteristic that prevents or inhibits microbial growth, especially among bacteria, but yeasts can surpass this trait and grow at pH ranges far below those found in such foodstuff products (normally around pH 4.0) [10]. The large range of heat resistance temperatures in *Z. bailii* is explained by a synergic protective effect of high sugar concentrations in the medium, where higher sugar concentrations lead to tolerance to higher temperatures, because higher levels of heat-shock proteins being present in cells soon after the stress of inoculation into high glucose [7]. Its highly fermentative ability can also cause serious issues, either due to refermentations in wines with low alcohol content (generally white wines with less than 12% ethanol volume) [2] or, related to the gas production (mostly CO_2) from the fermentation of sugars present in the foodstuff products. This is severely problematic because it can lead to package damaging due to the gas pressure accumulated

during fermentation [4,7]. Although *Z. bailii* mechanisms of tolerance to weak acid preservatives and other conservation methods remain poorly characterized, recent studies suggest that there is a phenotypic phenomenon, non-inheritable (non-genotypic), in yeasts that explains the general high tolerance to acid agents as a result of the population heterogeneity [36]. This trait, is characterized as, within a population of *Z. bailii* cells, there are sub-populations, or even only few cells, that phenotypically differ from the average population [37], frequently having lower internal pH from the others, granting them the ability to resist to a varied or specific stressful conditions caused by different weak acids, sterilizing or decontamination methods. This means that *Z. bailii* tolerance to weak acids is not given by the population itself, but from single cells with higher levels of tolerance to those stressful conditions [37]. Also, for *Z. bailii*, it has been suggested that the high acetic acid tolerance of the species is due to an ability to largely rearrange its lipid profile granting the cell population a robust mechanism of adaptation to an acidic environment, by changing the permeability of the cells membrane, thus limiting the diffusion of the weak acids into the cytosol [38]. However, the high tolerance of *Z. bailii* to acetic acid and other weak acid preservatives may be due to a conjugation of the species mechanisms abovementioned and others yet not understood, that, at a very low inoculum, grants the spoilage of foodstuff by *Z. bailii* species.

In *Z. bailii*, the weak acid transport across the plasma membrane and its flux balance between extracellular media and the cytosol remain unknown. However, studies in the yeast experimental model *S. cerevisiae* have consistently indicated well-known metabolic pathways that could be related and/or compared to other yeast species such as *Z. bailii* [39]. At a pH below their pKa, the uptake of weak acid molecules in the undissociated form by yeast cells occurs by simple diffusion through the plasma membrane. Once in the near neutral cytosol environment, these uncharged molecules dissociate, causing an accumulation of protons and anions that leads to the acidification of the cytosol and therefore the inhibition of many essential enzymes [11]. To counteract this effect, *S. cerevisiae* plasma membrane and vacuolar H^+-ATPases are activated [40] and at least the plasma membrane multidrug resistance transporters Tpo2 and Tpo3 were hypothesized to play a role in the export of acetate from acetic acid-challenged *S. cerevisiae* cells [41].

1.2 *Saccharomyces cerevisiae* Multidrug resistance transporters

Throughout recent years, with the appearance of more sophisticated and accurate techniques that allow a more precise understanding of the molecular and genetic mechanisms underlying the role of the Multi-Drug Resistance (MDR) transporters, *Saccharomyces cerevisiae* has been widely used as a model in eukaryotic cells to understand their relevance. The MDR transporters are characterized as efflux pumps, found in the plasma membrane of living cells that presumably recognize a wide range of different unrelated drugs and actively export them from the cytosol to the surrounding environment of the microorganism [42]. This mechanism confers a significant advantage to yeast cells in acquiring resistance to several different xenobiotics. This trait is particularly important in agriculture and health, especially in pathogenic microorganisms that acquire the ability to export antifungals and other drug killing agents. In industry, the MDR transporters can be useful since they can provide a way to increase the yields of a certain metabolic product (e.g., ethanol) [43]. One of the main problems in achieving higher yields of the

desired product is that during cell growth, mainly in the stationary phase, the accumulation of secondary metabolites (e.g., acetic acid) can limit cell's capacity to continue the production of a desired metabolite. The use of engineered cells with the ability to specifically and actively export the high intracellular concentrations of these inhibitory compounds can help increase the economic revenue in biotechnological and food industries. The understanding, at a molecular basis, of the regulation and expression of these proteins in *S. cerevisiae* can help to unravel new strategies to overcome or take advantage of this mechanism in more complex and less accessible eukaryotes.

In yeasts, the MDR transporters are divided in two different superfamilies according to their structure and main function: the ATP-Binding Cassette (ABC) and the Major Facilitator Superfamily (MFS). The first superfamily comprises ATP-dependent transporters usually arranged in two transmembrane span (TMS) domains and two cytoplasmic nucleotide-binding domains in which ATP is hydrolyzed. These proteins require the spent of the energetic molecule (ATP) to carry out the efflux of the undesired xenobiotic to the environment. The second superfamily, the MFS, instead of using ATP as an energetic element, operates through an electrochemical proton-motive force, a sum of the electrical potential of the cell membrane and the proton gradient between intra and extracellular environments of the cell. According to their structure, the MFS transporters can be divided in the two families [44,45]. Specifically, the proteins in DHA1 family comprehend twelve TMS and those from the DHA2 family have fourteen TMS domains. These proteins are drug:H+ antiporters performing the influx of a proton with the simultaneous efflux of a substrate/xenobiotic [46].

1.2.1 The *S. cerevisiae TPO2* and *TPO3* genes

The recent years have been valuable for the uptake of information heretofore unknown about the importance of the DHA1 family proteins in response of different kinds of induced-stress, unraveling also the regulatory mechanisms and main transcription factors involved [42]. Regarding the stress induced by weak-acids, the major regulator, Haa1, is a transcription factor responsible for the main adaptive response of *S. cerevisiae* cells under acetic acid conditions [47,48]. In an acetic-acid induced stress, the Haa1 transcription factor possesses a complex regulatory network, activating a variety of genes in response to the adverse conditions caused, either by directly binding the promoter region of the target gene or indirectly through the activation of a set of other transcription factors implicated in the acetic-acid stress response. This is achieved by the presence of a Haa1 high affinity responsive element in the promoter region of all the Haa1-dependent genes [49]. The gene network expression of *HAA1* gene is also responsible for the encoding of proteins involved in lipid metabolism, regulation of gene expression by modulation of chromatin accessibility and in the reduction of the intracellular acetate concentration [47]. Among the Haa1-dependent weak acid response genes, *TPO2* and *TPO3* are the two MFS-MDR proteins encoding genes from the DHA1 family that play a central role in the early response to acetic acid [41,47,48,50]. *TPO2* and *TPO3* genes code for polyamine transporters, specific for spermine and spermidine [51,52]. The overexpression of these genes reduce polyamine toxicity and promote the accumulation of polyamine in vacuoles. These paralogous genes that arose from a WGD event and are required for tolerance to acetic acid stress [47]., The complementation of *tpo2*Δ and *tpo3*Δ deletion

mutants by the expression of the respective genes from recombinant plasmids leads to the reduction of the lag phase in the presence of acetic acid stress [39,41].

1.3 IST302 strain and the *ZbTPO2/3* gene

Originally isolated from screening of highly acetic acid tolerant strains from a spontaneous fermentation of wine must, the *Z. bailii* IST302 strain is on the focus of this thesis work. Based on a phylogenetic analysis of the DHA1 family of several hemiascomycetous species (provided by Dr. Paulo Dias), it was found that *Z. bailii* holds one gene homologous to *S. cerevisiae TPO2* and *TPO3* (Figure 1).

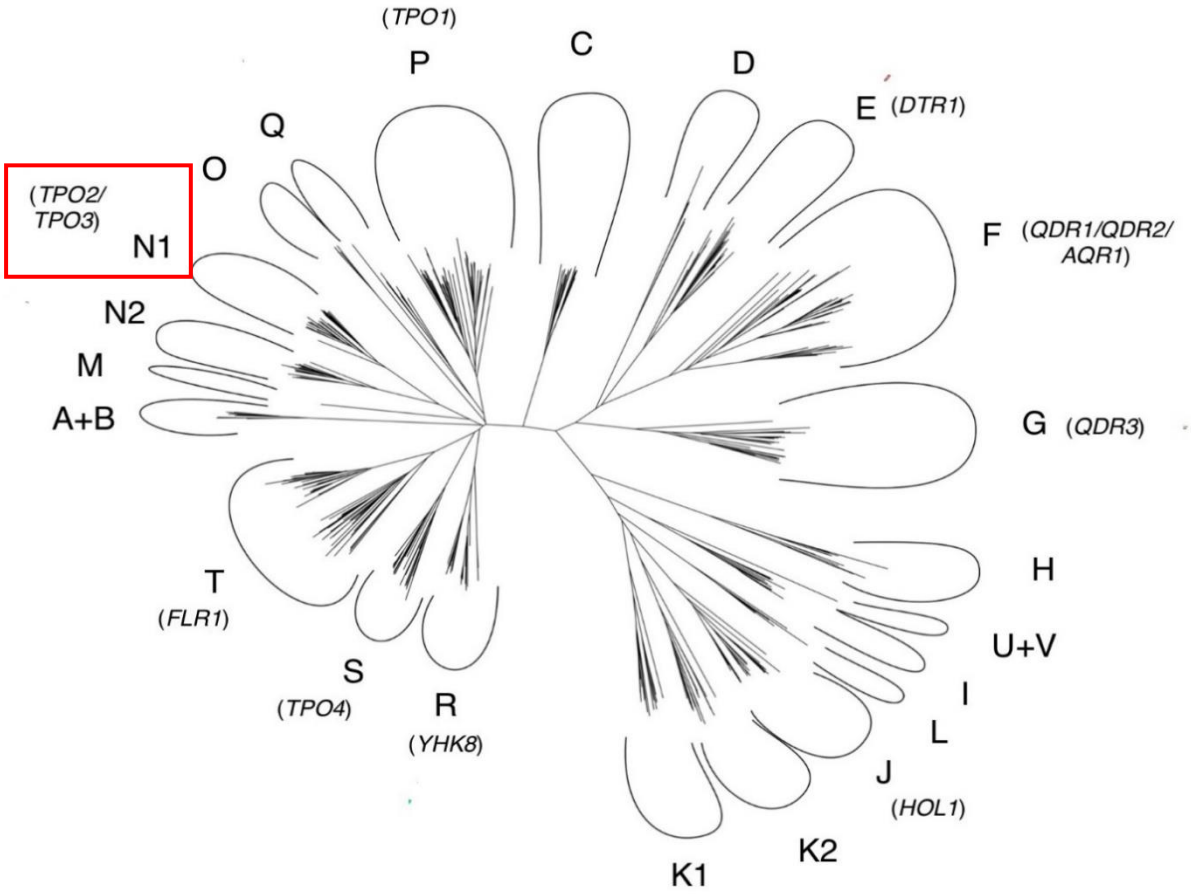

Figure 1 – Phylogenetic analysis of the DHA1 transporters family gathered from several hemiascomycetous species using the PROTDIST/NEIGHBOR packages of PHYLIP suite. Highlighted in the red square is the cluster N1 which comprises the phylogenetic relationship among the polyamine transporters encoding genes *TPO2* and *TPO3* in *S. cerevisiae* and their homologous sequences from the remainder hemiascomycetous species analyzed. (Provided by Dr. Paulo Dias).

Since the existence of *Z. bailii* MDR transporters involved in the export of acetic acid has not been characterized so far, one of the main objectives of this work is to study the function of the *Z. bailii* gene *ZbTPO2/3*, homologous to *S. cerevisiae TPO2* and *TPO3* (Fig. 1), regarding response and tolerance to acetic acid stress. The recently sequenced genomes of *Z. bailii* CLIB213^T and of a hybrid strain ISA1307

[24,25] and a third one, *Z. bailii* IST302 (unpublished results), provide the opportunity to characterize and find the genetic resemblances and differences between *Z. bailii* and *S. cerevisiae*. As a spoilage yeast, the interest in *Z. bailii* relies on the understanding of the mechanisms underlying its high resistance to the food preservative acetic acid.

1.4 Pulsed Field Gel Electrophoresis and Flow Cytometry: two powerful methodologies for karyotype analysis and quantification of total genomic DNA

Pulsed Field Gel Electrphoresis (PFGE) technique differs from the ordinary electrophoresis essentially for having a pulsed field capable of, throughout the run, change the migration direction of the DNA molecules sustained in the agarose gel due to the inversion of the poles set by the PFGE apparatus, thereby providing a way to successfully separate larger DNA molecules [53,54]. Several parameters have to be taken in consideration to obtain the best gel profile: gel run time, temperature, voltage, buffer and gel concentration and pulse duration and orientation. Slight changes in these parameters often cause poor gel resolution, due to the little/excessive migration and/or degradation of the DNA molecules [55]. Although some of these parameters seem straightforward in an ordinary electrophoresis, the pulse duration and orientation, also referred as switch time, and the buffer bath temperature are crucial parameters in a PFGE run. The switch time is the partial amount of time in the total gel run time in which the electric field is pulsed in a single direction [56]. The buffer bath temperature is another crucial parameter since its maintenance confers stability to the DNA molecules migration, preventing their degradation due to the rise of temperature by constant voltage [57].

In Flow Cytometry, the cytometer continuously emits a laser which interfere with the passage of single cells through a tube [58]. Several fluorophores can be used to label cells depending on the purpose of the analysis. In this work, the cells were labelled with SYBR® Green I, a fluorophore that is more suitable for cell DNA content determination, once its particular amount incorporation is proportional to the amount of DNA due to the stoichiometric bind that forms with it [59]. At each cell passage through a tube, the cytometer will gather the surrounding deflected light by a sensor which will translate into an electric pulse called the Forward Scatter Channel (FSC) which is a rough measure of the cell size. At the same time, the cells deflect the light through the inner metabolites and organelles and that deflected light will be translated into an electronic pulse denominated as the Side Scatter Channel (SSC) which gives information about the cell complexity [60]. The stained material is then measured in the flow cytometer and the emitted fluorescent signal yields an electronic pulse with a height (amplitude) proportional to the total fluorescence emission from the cell. The emitted fluorescent light of the SYBR® Green I (FL1) generates an electronic signal that can be recorded as high (FL1-H) for the intensity of the staining, as well as measured as pulsed-area (FL1-A) and pulse-width (FL1-W) of the samples [59]. This indirect DNA measurement needs reference cells with different and known amounts of DNA in order to correctly identify the ploidy and estimate the DNA content. Regarding the cell cycle analysis, there are five phases

that can be recognized in a proliferating cell population: G0 (Quiescence) and G1 (Cell Growth), S (DNA replication), G2 and M (Mitosis) phases. In a normal haploid/diploid cell, for the first two phases (G0 and G1) it is expected to encounter half the amount of DNA that is predictable in the last two phases (G2 and M) after the DNA replication.

1.5 Thesis outline

The main objective of this Master thesis is the taxonomic and molecular characterization of the highly acetic acid tolerant strain IST302 isolated from a spontaneous fermentation of grape wine must. In the past years the taxonomic identification of *Zygosaccharomyces* species has suffered several alterations either by inclusion of new species or by redefinition of previously established species. With this work it is expected to confirm that strain IST302 belongs to the *Z. bailii* species. Moreover, due to the relevance of *Z. bailii* species regarding to its high tolerance to weak acids, in particular to acetic acid, it is also intended to demonstrate, from a wide perspective, the putative role of the DHA1 protein encoded by the *ZbTPO2/3* gene, based on the demonstrated function of its homologs in *S. cerevisiae* (*TPO2* and *TPO3*) and, therefore, to elucidate its eventual involvement in *Z. bailii* mechanisms of tolerance to acetic acid stress.

The taxonomic identification will be performed based on a bioinformatics approach that will have in consideration the comparison of highly conserved regions of rRNA from different species of the Zygosaccharomyces genus in order to perform a phylogenetic tree, in which strain IST302 is included. Alongside, comparative analysis with highly conserved genes, will be performed in order to differentiate the *Z. bailii* IST302 strain from the closely related *Z. parabailii* and *Z. pseudobailii* species. To complement the results regarding the characterization of the *Z. bailii* IST302 karyotype and total DNA quantity, molecular approaches will be performed to support the *in silico* data gathered. Comparisons with the *Z. bailii* strain type (CLIB213^T) [24] will be important to confirm the correct taxonomic position of the *Z. bailii* IST302. The PFGE methodology is expected to allow the determination of the total number of chromosomes and the estimation of DNA content. Flow Cytometry will be used to, more accurately, through cell cycle analysis, determine the total DNA and the ploidy of the strain. The combination of both techniques is expected to confirm taxonomic and genomic data obtained for *Z. bailii* IST302 strain. Considering the relevance of *Z. bailii* metabolic activity in weak acid rich environments, the role of *Z. bailii ZbTPO2/3,* homologous gene to *S. cerevisiae TPO2* and *TPO3,* in acetic acid tolerance will be examined. A recombinant vector containing the *ZbTPO2/3* gene will be constructed and transformed in *S. cerevisiae* BY4741 parental strain and in the respective derived deletion mutants *tpo2*Δ and *tpo3*Δ. The susceptibility to acetic acid will be compared in the recombinant strains cultivated with and without acetic acid.

2. Materials and methods

2.1 Molecular phylogenetic analysis of strain IST302

The taxonomic identity of strain IST302, isolated from a spontaneous fermentation of grape must and provided by Prof. Alexandra Mendes-Ferreira (Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal), as from *Zygosaccharomyces bailii* species was confirmed by an *in silico* phylogenetic analysis that was performed based on the sequence of the D1/D2 region of the Large Subunit rRNA gene (LSU) of several strains from *Z. bailii*, *Z. parabailii*, *Z. pseudobailii* and other closely related species [19] (gene accession codes and strains in table A1). Since the genome sequencing and annotation of strain IST302 was carried out at the Biological Sciences Research Group (BSRG) laboratory, the partial sequence of the D1/D2 region of the LSU rRNA gene from strain IST302 was obtained directly from its genome sequence. Using the package of programs provided in http://mobyle.pasteur.fr, a multiple sequence alignment was made using ClustalW2.0.12 program followed by a phylogenetic tree construction using the program fastdnaml 1.2.2 of the PHYLIP package to estimate the phylogenetic relationships among the sequences using the maximum likelihood method. The phylogenetic tree was visualized in Dendroscope. Default parameters were used in all programs above mentioned.

2.2 Karyotyping and estimation of total DNA content of *Z. bailii* IST302

2.2.1 Strains and Growth conditions

The karyotype of *Z. bailii* IST302 was performed by comparing the chromosomal profile of this strain with the chromosomal profile of other *Z. bailii* or *Z. bailii* related strains and also with *S. cerevisiae*. The following strains were used: the hybrid strain ISA1307 [25], isolated from a sparkling wine continuous production plant [61] and provided by Prof. Cecilia Leão (Universidade do Minho, Braga, Portugal); the type strain of *Z. bailii*, strain CLIB213^T, isolated from a brewery [24] and provided by Prof. Sylvie Dequin; the laboratorial auxotrophic *Z. parabailii* ATCC 36947 *Leu2*Δ obtained by targeted gene deletion [62] (genotype *leu2*Δ) and provided by Prof. Paola Branduardi (Università degli Studi di Milano-Bicocca, Milano, Lombardy, Italy).

All strains were cultivated in rich YPD (Yeast extract Peptone Dextrose) growth medium which contains, per liter, 2% glucose (Merck), 1% yeast extract (Difco) and 2% peptone (Difco), at 30 ºC and 250 rpm (orbital agitation).

2.2.2 Karyotype analysis of strain *Z. bailii* IST302 by pulsed-field gel electrophoresis (PFGE)

Construction of the plugs – To obtain the total genomic DNA, all strains were cultivated overnight in rich YPD growth medium, until an optical density (OD_{600nm}) of 1.5±0.01 was achieved and harvested by centrifugation to obtain a total pellet volume of 50µl for each strain and washed with 1ml of EDTA (Ethylenediamine Tetraacetic Acid) 50mM, pH8.0. Samples were top spin to remove the supernatant

and resuspended in 100µl of EDTA 50mM, pH8.0. A total of 50µl of a Zymolyase solution (SCE buffer, 9µl Zymolyase™ (Zymo Research Group) and 15µl β-mercaptoethanol) was added to each sample followed by the addition of 300µl of low-melting point agarose solution (1% Low melting point agarose in 10mL EDTA 0,125mM, pH7.0). The mixture was split into appropriate plastic molds and placed at 4 ºC during 30 minutes until total solidification of the agarose. The plugs were then carefully removed, placed in 2,5mL of an EDTA-Tris-β-Mercaptoethanol (ETB) solution (EDTA 0,5M, pH8.0, Tris-HCL 1M, pH8.0, β-Mercaptoethanol) and then incubated overnight at 37 ºC with orbital agitation (50 rpm).

The ETB solution was removed and the samples were washed 3 times with TE buffer 1X (Tris-HCL 1M, pH8.0, EDTA 0.5M, pH8.0) during 1h each time, at 37 ºC and 100 rpm. To digest the cellular membrane, proteins and RNA it was added 2,5 mL of a proteinase solution (10% proteinase K (Thermo Fisher), 1% (v/v) RNAse (Thermo Fisher), EDTA 0,5M, pH8.0 and Laurosylsarcosine 10%) to each sample and incubated overnight at 37 ºC and 50 rpm. The proteinase solution was removed and plugs were first washed with EDTA 50mM and then with TE 1X buffer at 37 ºC and 100 rpm during 15 minutes and 1h, respectively. At this point the plugs can be stored in a solution containing EDTA 0,5M, pH8.0 and Tris-HCL 1M, pH8.0, at 4ºC for several weeks. Prior to use this solution should be removed by washing the plugs with TE 1X during 1h at 37 ºC and 100 rpm.

PFGE parameters and run conditions – The following parameters and conditions were carefully chosen given the expected large band size of DNA chromosomes in *Z. bailii or Z. bailii*-related strains. A 3L solution of TAE 1X (Tris Base, Glacial Acetic Acid and EDTA) buffer was prepared from a 50X stock solution of which 140mL were reserved to make an 0,8% agarose gel. The remaining buffer was used to fill the PFGE apparatus (PFGE-CHEF system: Pharmacia LKB gene Navigator). In gel assembly and plug insert, *Hansenulla wingei* (BioRad) was used as marker.

The DNA fragments of all strains were separated in 0,8% agarose gel at 81 V constant voltage in 1X TAE buffer at 14 ºC during 48h. The 48h run was separated into 5 phases, with different durations (in hours) and switch times (in seconds) in order to achieve a better resolution of the migrant DNA. The first phase took 25h and was followed by 4 phases of 15h, 5h, 2h and 1h. For each of the 5 phases, the corresponding switch time was, 250s, 400s, 550s, 700s and 850s. All these parameters were set by the Pharmacia LKB GN Controller.

2.3 Quantification of *Z. bailii* IST302 total genomic DNA through flow cytometry and ploidy confirmation

In order to confirm the ploidy and total genomic DNA of *Z. bailii* IST302 another *Z. bailii* strain with expected different DNA content was used, the hybrid ISA1307 strain. As haploid model it was used *S. cerevisiae* BY4741 and as diploid model the *S. cerevisiae* BY4743. All these four strains were pre-cultured overnight in rich YPD growth medium and then reinoculated at an initial OD600nm of 0.1±0.01, in 50 ml of the fresh YPD medium until mid-exponential phase (OD_{600nm} of 0.6) followed by cell fixation, staining and preparation for flow cytometry analysis [63].

Cell fixation: The harvested cells were centrifuged at 4ºC and 8000rpm following a washed step with cold ddH_2O, twice and the pellet ressuspended in ddH_2O. A final concentration of 10^7 cells/mL for each sample strain was required for a better acquisition of the staining and signal reading. The cell aliquots were centrifuged again with the same conditions already mentioned, fixated with 1ml of 70% ethanol and stored overnight at 4ºC.

Cell staining with SYBR® Green I (Thermo Fisher): The cells fixated in the previous step were centrifuged twice at 4 ºC and 8000rpm with a wash step in between with Tri-sodium citrate 50mM, pH7,5 and resuspended in 750µl with the same buffer. A total of 250µl of an RNAse solution (1mg/ml) were added to each sample and incubated at 50 ºC during 1 hour in a water bath. This step was followed by the addition of 50µl of proteinase K (20mg/ml) and incubation during 1h at 50 ºC in a water bath. A total of 20µl of SYBR® Green I (500 times diluted from a SYBR® Green I stock solution) were added to the reaction solution samples and carefully kept away from light exposure, overnight, at 4ºC.

Cell preparation and flow cytometry setup: 0,25% of Triton X-100 (v/v) from total samples volume were added and a quick vortex was made before a light sonication during 2 to 5 seconds, in order to ensure a maximum number of cells fully disaggregated, for a more rigorous reading by the Flow cytometer. The samples were diluted 2 times with tri-sodium citrate 50mM, pH7.5 followed by exposition to a 488nm excitation light and the fluorescence emission was collected at 525nm by the Flow Cytometer.

Data analysis: The analysis of the histograms generated was performed using FLOWJO® flow cytometry software v.10.0.8 (Tree Star Inc., Ashland, OR, USA).

2.4 Cloning and expression of *ZbTPO2/3* in a *Saccharomyces cerevisiae*

2.4.1 Construction of pGREG_ZBTPO2/3 vector

The pGREG506 plasmid from the DRAG&DROP collection was used to individually clone by homologous recombination and express the gene *ZbTPO2/3* identified as the ZBIST_0758 DNA fragment scaffold002 from *Z. bailii* IST302. This plasmid was acquired from EUROSCARF and contains a *HIS3* gene under the control of a galactose inducible promoter (*GAL1*) and the yeast selectable marker *URA3*. During homologous recombination the *HIS3* gene is replaced by the gene of interest. *ZbTPO2/3* DNA was generated by PCR using genomic DNA extracted from the strain IST302 and the following specific primers:

*ZbTPO2/3*_REC-FWD:

5’ – *GAATTCGATATCAAGCTTATCGATACCGTCGACA*ATGTCCACTCGCAACAGC – 3’

*ZbTPO2/3*_REC-REV:

5’ – *GCGTGACATAACTAATTACATGACTCGAGGTCGAC*TTACACGTCATTGGCTACTTCTTCC – 3’

The designed primers contain the regions with homology to the first 18 nucleotides for primer forward and the last 25 for the primer reverse of the *ZbTPO2/3* coding region (underlined). The remaining nucleotides from both primers have homology with the corresponding cloning site flanking regions of the pGREG506 vector (*italic*).

PCR reagents: before the transformation of the vector with the gene of interest in the desired host strain, it was necessary to amplify the *ZbTPO2/3* gene through Polymerase Chain Reaction (PCR). The following conditions and reagents listed in table 2 were used in all assays to achieve the amplification product.

Table 1 – List of reagents and respectively volumes applied to each reaction of a total six.

Reagents	Volume/reaction (µl)	Master Mix (6x)
H_2O	10,8	64,8
Buffer HF	4	24
$MgCl_2$	1	6
10mM dNTP’s	0,4	2,4
*ZbTPO2/3*_REC-FWD	1	6
*ZbTPO2/3*_REC-REV	1	6
DMSO	0,6	3,6
Enzyme (Phusion)	0,2	1,2
Z. bailii IST302 total gDNA	1	---

PCR settings: it was necessary to evaluate the best annealing temperature for the designed primers to achieve better product amplification quality and quantity. The PCR amplification of *ZBTPO2/3* cycle was composed by a first denaturation step at 98ºC during 30 seconds followed by 30 cycles of a three step phase which included a denaturation step at 95ºC during 10 seconds, an annealing step at 56ºC during 20 seconds and an extension step at 72ºC along 1 minute and 30 seconds. To finalize, a final extension step was performed at 72ºC during 7 minutes followed by a cooling step at 8ºC. The amplified product was then mixed with loading buffer and an electrophoresis gel was made to verify the correct amplification of *ZBTPO2/3* using a 1Kb Ladder DNA marker.

Vector cloning and yeast transformation: it was necessary to prepare the vector pGREG506 to correctly recombine with gene *ZBTPO2/3* digesting the plasmid in a reaction mixture containing 1µl of SalI enzyme, 1µl of CIAP, 3µl of Buffer H 10x, 3µl of pGREG506(null) DNA and 22µl of H_2O to a total volume of 30µl during 3 hours at 37ºC. The reaction was stopped by restriction enzyme inactivation at 60ºC during 20 minutes. The gene *ZBTPO2/3* as well as pGREG506 previously digested were cotransformed into the parental strain BY4741 and derived deletion mutants *tpo2*Δ and *tpo3*Δ, using the Alkali-Cation™ Yeast Transformation kit. The recombinant plasmid pGREG506_*ZBTPO2/3* was obtained through homologous recombination in yeast. Correct cloning of the gene was verified by DNA sequencing. To have the adequate control cells, *S. cerevisiae* BY4741 parental and derived deletion mutants Δ*tpo2* and Δ*tpo3* strains were transformed with the empty vector pGREG506 with the *HIS3*

gene deleted (pGREG506_no*HIS3*), since in the recombinant vector the *HIS3* gene is substituted by the insert of interest and it was found that the expression of the *HIS3* gene in the cloning vector led to a remarkable increase of acetic acid tolerance of the auxotrophic Δ*haa1* host strain (his^-) [64].

2.4.2 Expression of *ZbTPO2/3* in *S. cerevisiae* parental and respective derived deletion mutants *tpo2*Δ and *tpo3*Δ

In order to compare the susceptibility to acetic acid of *S. cerevisiae* parental BY4741 and derived deletion mutants Δ*tpo2* and Δ*tpo3* transformed either with pGREG506(null) or with pGREG_*ZBTPO2/3*, yeast cells were cultivated in MMB-U medium at pH4.0 (minimal medium supplemented with amino acids minus uracil). To test the advantage function of promoter *GAL1* present in vector pGREG506, the strains were also cultivated in the same MMB-U medium supplemented with 1% galactose. For the two types of media a control group without acetic acid and a test group with 60mM of acetic acid, from a stock solution of 5M, in a total volume of 50mL were made. All cultures were grown until mid-exponential phase (OD_{600nm} = 0,5) and inoculated at an initial OD of 0,05 in MMB-U or MMB-U+1%Galactose, either or not supplemented with acetic acid. Yeast cells were cultivated at 30ºC and 250rpm (orbital agitation).

Culture growth was followed by periodically measuring culture optical density until cells reach their stationary phase of growth, in order to understand the behavior of each transformed strain under the presence or absence of the acetic acid.

3. Results

3.1 *Z. bailii* IST302 molecular phylogenetic analysis

The taxonomic identification of strain IST302 as *Z. bailii* species was based on the phylogenetic comparison of the highly conserved genetic D1/D2 region from the large subunit (LSU) 26S ribosomal DNA from closely related species. This genetic region provides a useful tool to successfully distinguish the taxonomic relationships among strains and species due to the few differences in those DNA sequences [23]. Based on the recently reported reallocation of *Z. bailii* strains to the new *Z. parabailii* and *Z. pseudobailii* species[19] a phylogenetic tree was constructed using the D1/D2 partial sequence from strain IST302 and the corresponding sequences from other *Z. bailii*, *Z. parabailii* and *Z. pseudobailii* strains (Table A1). Other species from *Zygosaccharomyces* genus were also included in this phylogenetic analysis and *Lachancea thermotolerans* NRRL Y-8284^T was used as an outgroup (Figure 3). The strain IST302 forms a well-established monophyletic clade with all the other *Z. bailii* strains, in which the type strain *Z. bailii* ATCC 58445^T (=CLIB 213^T) is also included. Also, it is possible to identify a clades from *Z. pseudobailii* strains, which form a monophyletic group within the paraphyletic clade of *Z. parabailii* species.

In order to completely distinguish the IST302 strain from the recently identified novel sp. *Z. parabailii* and *Z. pseudobailii,* an additional *in silico* analysis was carried out based on pairwise sequence alignment of several housekeeping genes from each of the three different closely related species *Z. bailii*, *Z. parabailii* and *Z. pseudobailii* against strain IST302 (Table 2).

Table 2 – IST302 homologous sequence genes accessed through a *BLAST* of *Z. bailii* type strain ATCC 58445^T (=CLIB213^T) genes considered against the IST302 genome available at pedant.helmholtz-muenchen.de.

Gene	Scaffold	Coordinates	Length (bp)
coxII	ZBU_scaffold089	5172 – 5787	616
mSSU rRNA	ZBU_scaffold089	10603 – 11162	559
EF-1α	ZBU_scaffold018	143879 – 144810	932
β-tubulin	ZBU_scaffold043	92439 – 93221	782
RPB1	ZBU_scaffold044	201433 – 202092	649
RPB2	ZBU_scaffold020	56135 – 57167	1032

The nucleotide sequences of the genes encoding a cytochrome oxidase subunit 2 (*coxII*), a mitochondrial small-subunit 18 rRNA gene (mSSU rRNA), a translation elongation factor 1-*α* (*EF1-α*), a *β*-tubulin, a RNA polymerase II largest subunit (*RPB1*) and a RNA polymerase second-largest subunit (*RPB2*), from several strains (Table A2) were compared using the EMBOSS Needle global alignment tool and the identity percentage and nucleotide variations were calculated (Figure 2). Regarding the genes *coxII*, mSSU rRNA, *EF1-α* and *β-tubulin* only the type strains from each species were compared with the homologous sequences from IST302 strain (Figure 2.A). The results show that the identity

percentage is higher than 95% for all the genes compared with IST302 strain. Remarkably, with the exception of the gene encoding a *β-tubulin*, all the genes from *Z. bailii* share a perfect homology (100% identity) with IST302 homologous sequences. In what concerns the comparison of the genes from IST302 and *Z. parabailii* the identity of the sequences is lower than the one obtained in *Z. bailii* pairwise alignments, with values ranging from 99,52% (mSSU rRNA gene) and 95,69% (*β-tubulin* gene). Finally, the *Z. pseudobailii EF1-α* and *β-tubulin* genes showed the lowest identity with the IST302 homologous sequences, although a perfect homology with the *coxII* gene and an identity of 99,52% with the mSSU rRNA gene were observed.

The *RPB1* and *RPB2* genes were also used to calculate the number of nucleotide variations between several strains from the *Z. bailii, Z. parabailii* and *Z. pseudobailii* species and IST302 strain (Figure 2.B). The results show a small number of variations amongst the *Z. bailii* strains and the IST302 strain for both genes (maximum of 3 nucleotide variations in *RPB1* and 8 in *RPB2* in a total of 6 different strains). The differences between *Z. parabailii* strains and IST302 is larger with an average of 22 nucleotide variations among the six strains for *RPB1* and an average of 45 nucleotide variations for the seven strains for *RPB2* gene. However, one strain of *Z. parabailii* (ATCC 60484) shows the same nucleotide variation than the maximum observed for *RPB1* in *Z. bailii*. The *RPB1* and *RPB2* genes of the two strains of *Z. pseudobailii* show the highest nucleotide variations (38 and 71 respectively) with the homologous sequences of IST302 (Table A3). The results obtained using both the *in silico* approaches phylogeny and molecular comparison of several housekeeping genes from *Z. bailii*, *Z. parabailii* and *Z. pseudobailii* provide good indications that strain IST302 is from *Z. bailii* species.

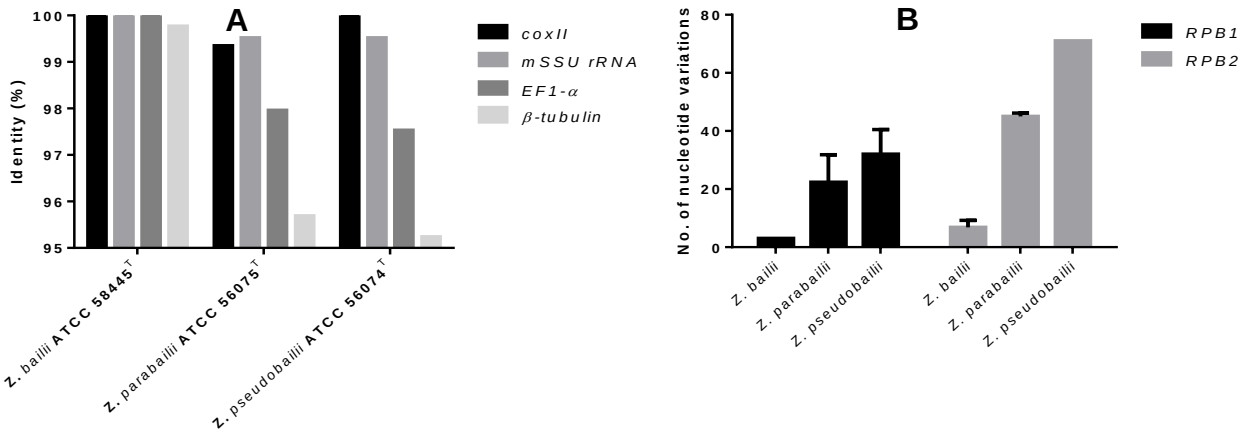

Figure 2 – (A) Percentage of identity calculated based on a pairwise sequence alignment of the genes *coxII*, mSSU rRNA, *EF1-α* and *β-tubulin* from the type strain of *Z. bailii*, *Z. parabailii* and *Z. pseudobailii* against the respective homologous sequences in IST302. **(B)** Mean and standard deviation (error bars) of the number of nucleotide variations identified between 6 different strains of *Z. bailii*, 6 strains of *Z. parabailii* and 2 strains of *Z. pseudobailii* for *RPB1* gene and 6 different strains of *Z. bailii*, 7 strains of *Z. parabailii* and 2 strains of *Z. pseudobailii* for *RPB2* gene against the respective found in IST302. The pairwise alignments used to calculate identity and number of nucleotide variations were performed using the EMBOSS Needle global alignment tool from EMBL-EBI packages.

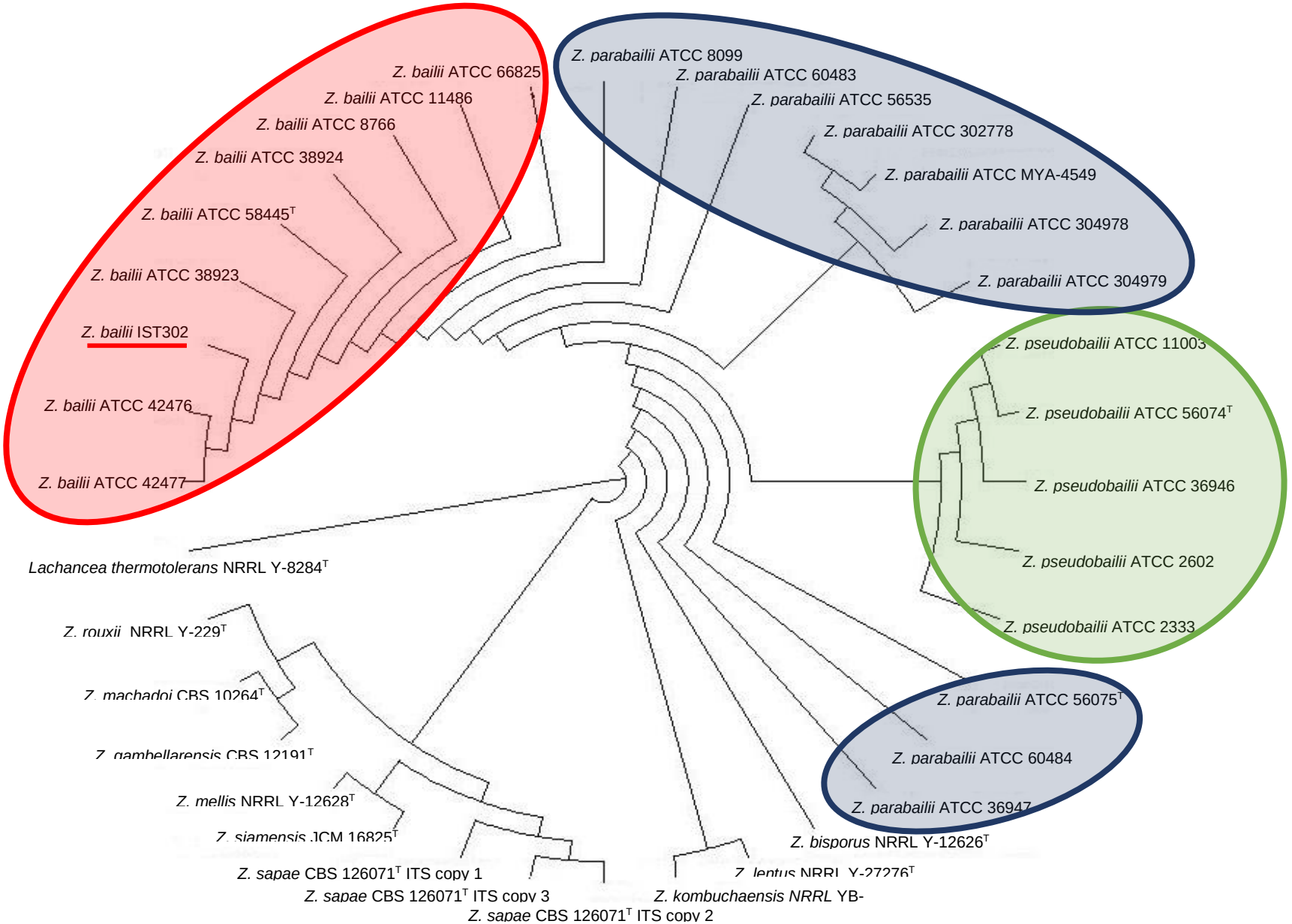

Figure 3 – Phylogenetic analysis based on D1/D2 region of the LSU rDNA gene using CLUSTALW for multiple alignment and FASDNAML for tree construction, of PHYLIP packages through a maximum likelihood method available on mobile.pasteur.fr site. Cladogram provided by the Dendroscope program. Highlighted in: RED - *Z. bailii* clade; BLUE – *Z. parabailii* clades; GREEN – *Z. pseudobailii* clade. Underlined in red is the strain *Z. bailii* IST302. *Lachancea thermotolerans* NRRL Y-8284^T was used as an outgroup.

3.2 Characterization of *Z. bailii* IST302 karyotype by Pulsed-Field Gel Electrophoresis

A Pulsed-Field Gel Electrophoresis (PFGE) technique was used to compare the karyotype of *Z. bailii* IST302 with *Z. bailii*-related strains. Given that total intact DNA is necessary to obtain a clear separation of the chromosomes no restriction enzyme digestion step was used in the preparation of the DNA. Several attempts were made until a compromise between resolution and distinct separation of the chromosomes was reached. The best PFGE result (Figure 4) was obtained by using the parameters listed in Table 2.

Table 3 – List of parameters used in the PFGE technique that enable the best gel resolution and visualization of the strains' karyotype in this assay. The duration (in hours) of each phase is a subdivision of the total run time (48h). For each phase it is displayed the respective switch time used.

Run temperature		14ºC
Voltage		81V
Buffer		TAE1X
Agarose gel		0,8%
Total Run Time		48h
Nº phases	**Time (h)**	**Switch time (s)**
1	25	250
2	15	400
3	5	550
4	2	700
5	1	850

The gel obtained (Figure 4) has the sufficient resolution to identify the number of chromosomes present in the strain IST302 (lane 2) and in the type strain CLIB213^T (lane 3) with, respectively six and five chromosomes each. Considering that the chromosomes' molecular size of *Z. bailii* CLIB213^T range between 1 and 3 Mbp [25], the chromosomal DNA marker from *Hansenula wingei*, with DNA bands ranging from 1.05Mbp to 3.13Mbp was used in this assay to estimate the relative size of each chromosome. In *Z. bailii* IST302 PFGE profile it is possible to identify six distinct chromosomes, each with a molecular size of approximately 1.3Mbp, 1.4Mbp, 1.6Mbp, 1.8Mbp, 2.0Mbp and 2.2Mbp. Based on this approximation, it was also possible to roughly predict the IST302 genome size as approximately 10.3Mbp. IST302 chromosomal profile shows some differences in comparison with the type strain CLIB213^T, which possesses five confirmed chromosomes with molecular sizes of approximately, 1.3Mbp, 1.4Mbp, 1.6Mbp, 1.8Mbp and 2.3Mbp for a total of 8.4Mbp genome size prediction. However, the DNA band of lowest molecular size has a higher intensity than the others, thereby suggesting it can be an overlap of two different chromosomes with similar sizes. It was also included in this analysis the hybrid strain ISA1307 (lane 4) and the *Z. parabailii* ATCC 36947 derived mutant *Leu2*Δ strain (lane 5) which exhibit similar chromosome profiles although completely different from those showed by IST302 and CLIB213^T strains. The resemblance of the ISA1307 hybrid strain DNA profile with the *Leu2*Δ strain (lane 4 and 5) could indicate the possibility of the last one, also be an interspecies hybrid [19,25].

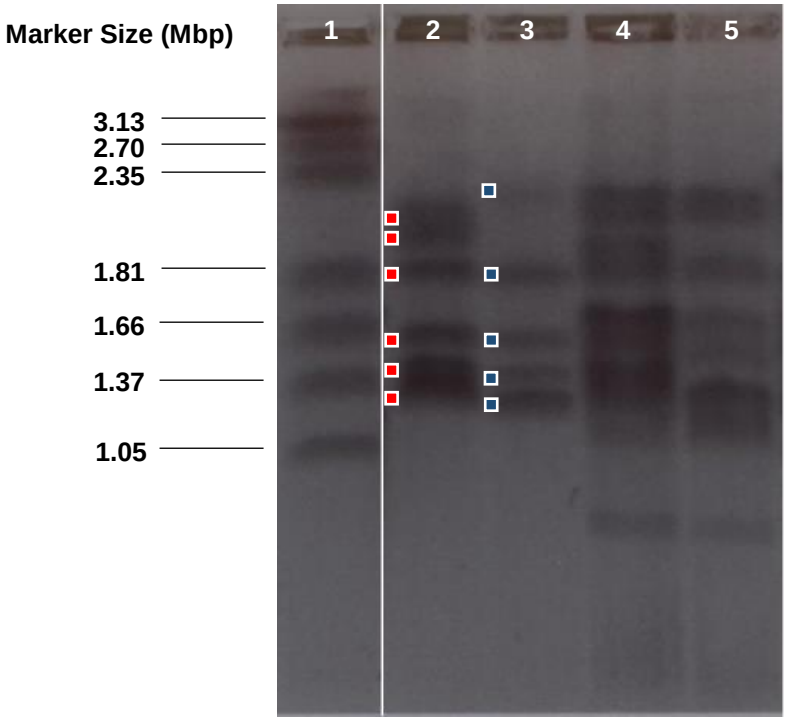

Figure 4 – Partial 0,8% agarose gel result from a 48h run split in five different phases, at 81V, 14 ºC in TAE 1X buffer. Lanes: **(1)** marker *Hansenula wingei*; **(2)** IST302 strain; **(3)** *Z. bailii* CLIB213^{T}; **(4)** Hybrid ISA1307 strain; **(5)** *Z. parabailii* ATCC 36947 derived mutant *Leu2*Δ; The marker sizes correspond to the *Hansenula wingei* marker in lane 1. The red and blue dots indicate the visible chromosomes in IST302 and CLIB213^{T} strains, respectively.

3.3 Determination of *Z. bailii* IST302 ploidy and estimation of total genomic DNA content by Flow Cytometry

In order to estimate *Z. bailii* IST302 genome size using a different approach and also to determine the ploidy of this strain, exponential cells of strains IST302 and ISA1307, as well as *S. cerevisiae* haploid BY4741 and diploid BY4743 strains, were fixed and stained with the fluorescent probe SYBR® Green I. DNA was quantified by flow-cytometry and the acquired data was analyzed using FLOWJO® v10.0.8 software. For this assay, *S. cerevisiae* BY4741 was used as a known haploid model strain and *S. cerevisiae* BY4743 as a known diploid model [65]. Since the genome content and ploidy of the hybrid strain ISA1307 was recently estimated [25], this strain was also used in this approach for comparison purposes. In addition to the estimation of total DNA content, data acquired by flow cytometry was also used to determine the ploidy of *Z. bailii* IST302 by cell cycle analysis.

Data acquisition was performed by the measurement of the mean fluorescence values (FL1-A) of 50000 total events (single cells) in each sample. In order to avoid the unwanted subpopulations of doublets or aggregates in the sample, it was necessary to gate the correct population (Figure 5.A). Since the electronic signal generated by the emitted fluorescent light can be recorded in different forms, when it conjugates in a cell density plot [pulsed area (FL1-A) vs. pulsed width (FL1-W)] it is possible to observe two different hot spots in the plot due to the fluorescence exhibited by the G0/G1 phase cells and the G2/M phase cells (double fluorescence, double DNA) (Figure 5.B). All the cells with a higher FL1-W are likely to be doublets or aggregates since the pulse-width is directly related with the cell diameter [66].

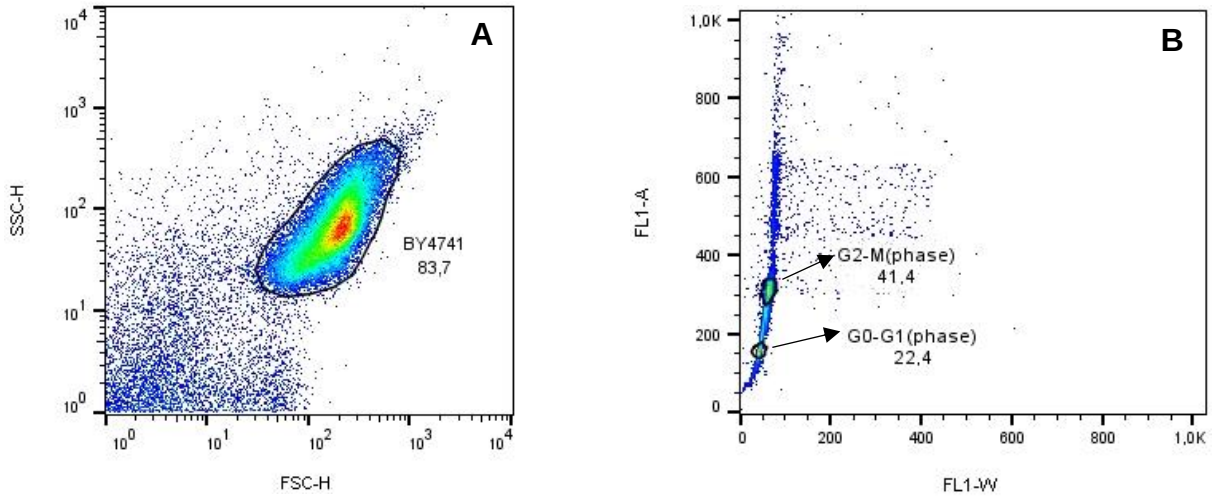

Figure 5 – Sample of *S. cerevisiae* BY4741 cell density plot represented by **(A)** Cell size (FSC-H, Forward Scatter High) and complexity (SSC-H, Side Scatter High). 83,7 is the percentage of the gated cells over the total. **(B)** Fluorescence light area (FL1-A) and fluorescence light width (FL1-W). The two gated subpopulations represent each one of the states of the cell cycle by means of the fluorescence measured. The values below the gated descriptions are the percentage cell subpopulation of the total sample.

As anticipated, the fluorescent light intensity of the reference haploid strain *S. cerevisiae* BY4741 is half the fluorescent light intensity of the diploid BY4743 strain, since strain BY4741 possesses half the DNA content of strain BY4743 (Figure 7.A). The cell cycle analysis of the strain IST302 shows two fluorescence peaks highly similar to those observed in *S. cerevisiae* BY4741 haploid strain (Figure 7.C) and very different from those evidenced by the diploid strain BY4743 and the hybrid strain ISA1307 (Figure 7.B).

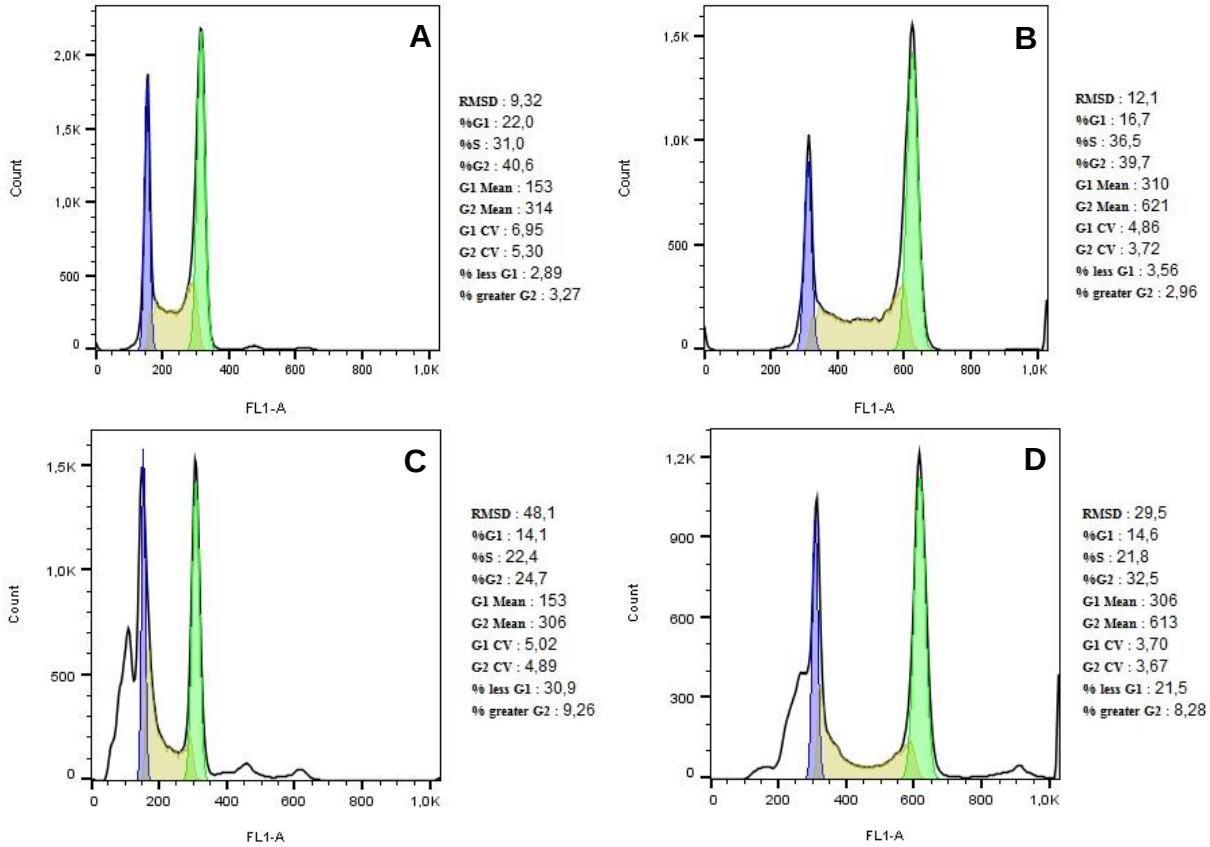

Figure 6 – Cell cycle analysis histograms statistically analyzed by the Watson (pragmatic) model. At the right of each graphic there are the corresponding values for the root mean square deviation (RMSD), percentage of cells in G0/G1 (%G1), percentage of cells in S-phase (%S), percentage of cells in G2/M (%G2), mean pulsed-area fluorescence light intensity in G0/G1 (G1 mean), mean pulsed-area fluorescence light in G2/M (G2 mean), percentage of cells below the gated G0/G1 cells (% less G1) and percentage of cells above the gated G2/M cells (% greater G2) of each sample. Highlighted in blue and green are the corresponding G0/G1 and G2/M areas measured by the Watson model. **(A)** *S. cerevisiae* BY4741 **(B)** *S. cerevisiae* BY4743 **(C)** *Z. bailii* IST302 **(D)** Hybrid ISA1307 strain. All data collected are a representation of at least 3 independent assays.

Indeed, both these strains show a similar fluorescent light intensity pattern. At the same time a cell cycle analysis was performed in order to collect the mean fluorescence light area values for each strain at each G0/G1 and G2/M phases (Figure 6.A-D). The statistical analysis by the Watson model provided in the FLOWJO® software retrieve a series of values in which CV% values represent the quality resolution

of the histograms presented. The lower the CV% values the higher the histogram resolution. All histograms in Figure 6 have a CV% value ≤7% which represents a good indicator, and therefore reliable mean fluorescence light area values represented by the G1 and G2 mean values. These values of *Z. bailii* ISA1307 (G1 mean – 306; G2 mean – 613), *S. cerevisiae* BY4741 (G1 mean – 153; G2 mean – 314) and *S. cerevisiae* BY4743 (G1 mean – 310; G2 mean – 621) were incorporated in a plot along their respective DNA content in G0/G1 and G2/M phases already known. The equation of the regression line resultant with a good R^2 (=0,99) was then used with the IST302 mean fluorescence light area values (Figure 6.C; G1 mean – 153; G2 mean – 306) and the DNA content for both G1 and G2 phases was estimated according with the displayed equation (Figure 8). An approximate DNA content of 11.35Mbp was calculated for the G0/G1 phase of the IST302 strain. These results suggest that *Z. bailii* IST302 is an haploid strain with a DNA content similar to the one from *S. cerevisiae* BY4741 haploid strain (11.68Mbp).

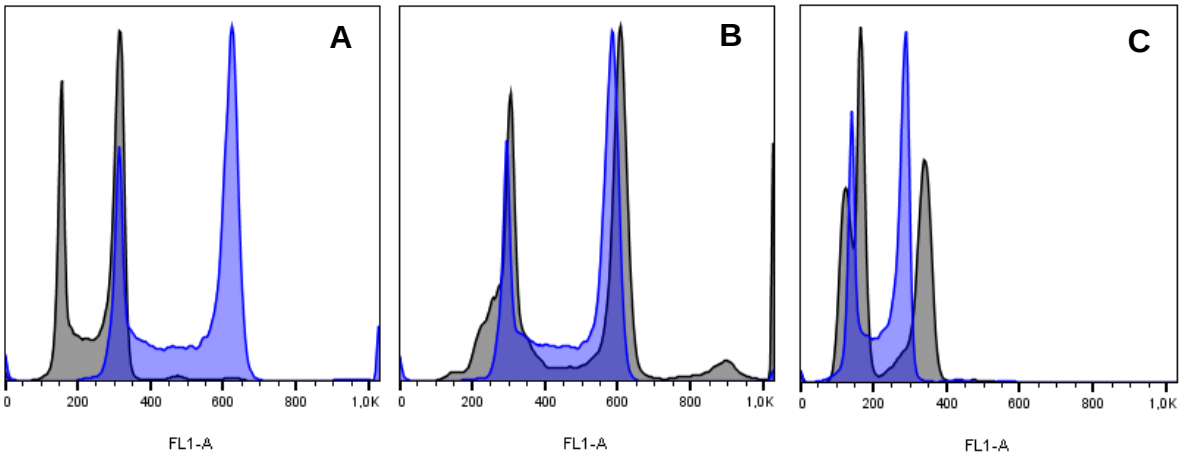

Figure 7 – Histograms of different yeast cell populations regarding their fluorescent light intensity pulsed-area (FL1-A). The analysis is automatically computed by the FLOWJO® v10.0.8 software with the compare tool available. **(A)** Comparison between *S. cerevisiae* BY4741 (grey) and *S. cerevisiae* BY4743 (blue) **(B)** Comparison between *S. cerevisiae* BY4743 (blue) and the hybrid ISA1307 strain (grey) **(C)** Comparison between *S. cerevisiae* BY4741 (blue) and *Z. bailii* IST302 (grey).

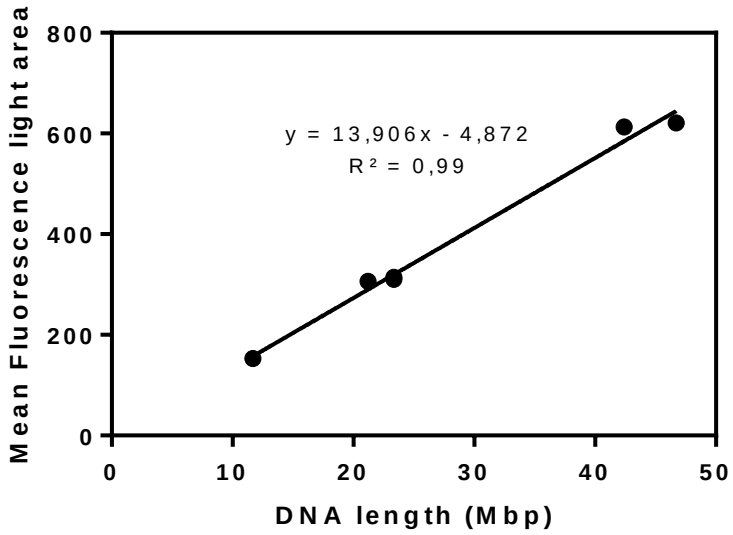

Figure 8 – Calibration curve designed through the mean fluorescence light area values (G1 and G2 mean values) displayed in Figure 7.A, 7.B and 7.D description. DNA length in G0/G1 and DNA replication phase of the *S. cerevisiae* BY4741, *S. cerevisiae* BY4743 and *Z. bailii* ISA1307. The equation on the chart represents the linear regression of the data plot with the respective R-square.

3.4 Comparison of acetic acid susceptibility of *S. cerevisiae* BY4741 parental and derived deletion mutants *tpo2*Δ and *tpo3*Δ expressing *TPO2* and *TPO3* homologue from *Z. bailii* IST302, the gene *ZbTPO2/3*

The main objective in this assay was to investigate the presumable role of the *S. cerevisiae TPO2* and *TPO3* homologous gene *ZbTPO2/3* from *Z. bailii* IST302, under acetic acid conditions.

Z. bailii IST302 gene *ZbTPO2/3* (1917bp) was amplified by PCR with specific primers (Section 2.4.1 in Materials and Methods) and confirmation of the insertion was carried out by sequencing the recombinant vector with only one specific primer for the *ZbTPO2/3* gene (*ZbTPO2/3*_REC-FWD). To evaluate the effect of the *ZbTPO2/3* gene under acetic acid conditions, regarding the predictable role of the Tpo2 and Tpo3 proteins, the *ZbTPO2/3* gene was cloned into pGREG506 by homologous recombination in parental *S. cerevisiae* BY4741 and its derived deletion mutants *tpo2*Δ and *tpo3*Δ. Alongside, pGREG506 vector lacking only the *HIS3* gene (empty vector) and without *ZbTPO2/3* was also transformed into *S. cerevisiae* BY4741 and its derived deletion mutants *tpo2*Δ and *tpo3*Δ in order to obtain plasmid control transformants. Acetic acid susceptibility was compared in the three background strains expressing *ZbTPO2/3* by following their growth when cultivated in a medium supplemented or not with acetic acid.

The results, displayed in Figure 9, globally, do not suggest a positive effect of the expression of *ZbTPO2/3* gene in the derived deletion mutants *tpo2*Δ and *tpo3*Δ transformed with the recombinant vector with *ZbTPO2/3*. Also, the use of galactose in the medium did not enhance the growth of those strains (Figure 9B) when compared with the same strains inoculated in medium without galactose (Figure 9A) suggesting that the enhanced transcription of the *ZbTPO2/3* by the *GAL1* gene, in medium with galactose, does not have any effect on the tolerance under acetic acid in each of the derived deletion mutants used. The derived deletion mutants *tpo3*Δ transformed with the empty vector or with the recombinant vector with *ZbTPO2/3* in both media, present a similar growth with a long lag phase (≈25h) suggesting that these mutants have a higher susceptibility to acetic acid than the *tpo2*Δ mutants and that *ZbTPO2/3*, possible, do not comprehend the same function as the *S. cerevisiae TPO3* homologue. However, in both media, it is possible to observe that, both the parental strain and the derived deletion mutant strains, transformed with the recombinant vector with the *ZbTPO2/3* gene, present a slight increase in tolerance to acetic acid than its respective strains transformed with the empty vector. Nevertheless, once the levels of tolerance to acetic acid presented by the derived deletion mutant strains in both media, do not achieved, at least, the ones in the parental strain, the *ZbTPO2/3* do not complement the susceptibility phenotype of both deletion mutants, suggesting that, this gene appear to not contribute to a positive effect in the acetic acid induce stress.

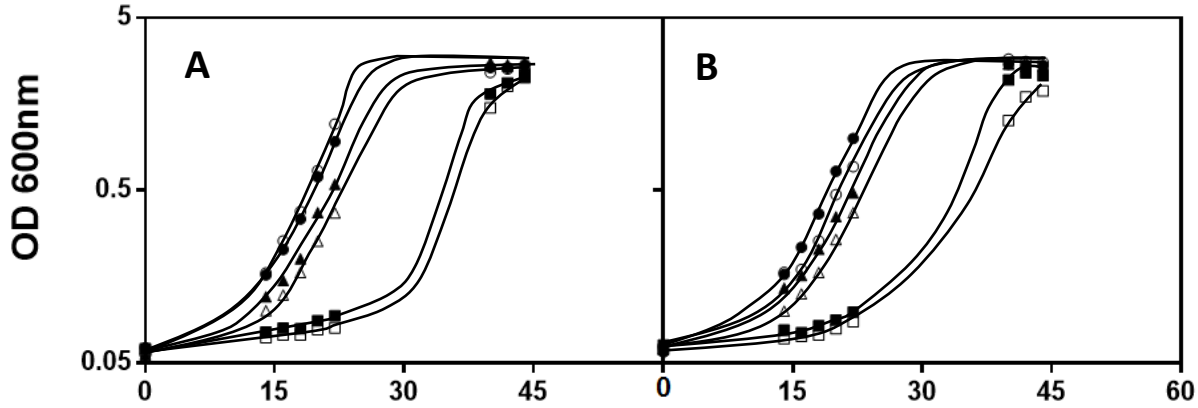

Figure 9 – Growth of *S. cerevisiae* BY4741 wild type, *tpo2*Δ and *tpo3*Δ mutant transformants under 60mM of acetic acid, at 30ºC in **(A)** MMB-U minimum medium supplemented with amino acids less uracil at pH4.0 **(B)** MMB-U supplemented with an additional 1% Galactose at pH4.0. The filled symbols represent the *S. cerevisiae* strains transformed with the recombinant vector that possesses the *ZbTPO2/3* gene from IST302 strain. The blank symbols are the *S. cerevisiae* strains transformed with the empty vector (pGREG506 without *HIS3* gene). Circles (*S. cerevisiae* BY4741 parental strain), triangles (*tpo2*Δ deletion mutant), squares (*tpo3*Δ deletion mutant).

4. Discussion

4.1 *In silico* analysis of IST302 strain

In the past several years the *Zygosaccharomyces* genus has been taxonomically rearranged due to the identification of new species and to the use of new molecular methods to differentiate them [16–20,67,68]. A recent study based on the molecular analysis of D1/D2 regions of the LSU rRNA gene allowed the taxonomic reallocation of former *Z. bailii* strains into two novel species, the *Zygosaccharomyces parabailii* and *Zygosaccharomyces pseudobailii* [19]. Also, the already sequenced and annotated hybrid ISA1307 strain, widely used in the past few years in investigation for the understanding of the tolerance mechanisms of *Z. bailii* species under acetic acid stress, was recently characterized as an interspecies hybrid between a *Z. bailii* with a close related species [25]. These evidences have raised some concerns about the taxonomic identification of the presumable *Z. bailii* IST302 strain. Based on this, *in silico* phylogenetic and molecular analyses were carried out in this work to confirm the taxonomic position of the IST302 strain.

The *in silico* analysis results were based on a previous work [19], with the addition of the respective homologous genetic elements of IST302 strain. For the construction of the phylogenetic tree, a maximum likelihood method was applied to the multiple alignment of the D1/D2 regions of the LSU rRNA gene sequences amongst the *Zygosaccharomyces* species considered plus the IST302 strain. A similar analysis with the ITS regions of the ribosomal cistron was also intended, however the homologous region in IST302 was not available in this strain's genome sequence. These genetic elements (nucleotide sequences from D1/D2 and ITS regions) constitute a well-established and accepted diagnostic characters (barcode) for fungal species, which allows the identification of a given strain at species level due to the few alterations in the DNA sequence found between species [69]. In the resulting circular phylogenetic tree strain IST302 is located within the *Z. bailii* monophyletic clade with high proximity with the type strain (CLIB213^T), thereby providing the first evidence that strain IST302 is a *Z. bailii* species. The housekeeping genes, usually highly conserved genetic elements with low rate of variability, used in nucleotide variation analysis are also a useful tool to successfully distinguish different species where differences ≤1% in the DNA sequence refer to conspecific strains [22,23]. Although this percentage criterion was observed for some of the genes analyzed when comparing strain IST302 with *Z. parabailii* and *Z. pseudobailii* strains, only the analysis between the *Z. bailii* strains exhibit ≥99% of total identity with *coxII*, *EF1-α*, mSSU RNA gene, *RPB1* and *RPB2* altogether. The use of different reliable molecular characters in different analysis, such as the D1/D2 regions of the LSU rRNA and few highly conservative protein-coding genes provide a more strong and successful result instead of using one simple approach [69,70].

4.2 Characterization of *Z. bailii* IST302 karyotype and ploidy

In order to confirm the results gathered in the *in silico* analysis, the chromosomal profile of strain IST302 was compared with the one from *Z. bailii* CLIB213^T using a PFGE technique, commonly used as a

molecular approach for species identification and typing because it provides a visible profile of the species' chromosomes [71].

One difficulty in PFGE method is the optimization of the gel resolution regarding the DNA samples of the strains used. In this case, it was required an apparatus capable of separating a long range of sizable DNA molecules (approximately between 1 and 3Mbp). The apparatus used to perform the PFGE is based on the Clamped Homogeneous Electric Field Electrophoresis (CHEF) system, which provides a constant voltage gradient produced across the gel, allowing large molecules up to 7Mbp to be separated in straight lines [53,72]. Also, the parameters implied in the gel run have an important impact on the gel resolution. The buffer activity and capacity of TAE used grants a faster and farther migration of the DNA molecules when compared to the commonly used TBE buffer due to its lower ionic strength [57]. The maintenance of the temperature by the PFGE apparatus of the buffer is also a crucial factor [73]. The higher the temperature, the faster the gel run, although with increasingly loss of resolution. On the other hand, lowering the temperature will enhance the gel resolution, yet in longer run times. Given this, the chosen temperature of 14 ºC is a compromise between run time and resolution [74]. The combination of voltage gradient, total run time and respective switch times used in the gel run are important factors that contribute to the improvement of DNA molecules separation. Since the strains used have presumable long size DNA chromosomes (≥900Kbp) [25], a long run time (48h) at low voltage (81V) were the suitable settings. The long run allows the larger molecules to, at low voltage, slowly separate with better resolution. Linked to these features is also the switch times, in a ramping format, which means that for strains with a wide range of DNA molecule sizes the better separation is achieved by gradually increasing the switch times [56].

Since the objective of this study was the separation of *Z. bailii* IST302 chromosomes and the comparison of its chromosomal profile with other *Z. bailii* or *Z. bailii* related strains, no digestion of genomic DNA with restriction enzymes was performed. The chromosomal profile of *Z. bailii* IST302 is similar to the chromosomal profile of *Z. bailii* type strain (CLIB213^T) although with slight differences. The PFGE gel obtained in this study identified six chromosomes in strain IST302 while only 5 chromosomes were identified in strain CLIB213^T with similar molecular DNA sizes between the two strains. Although the resolution of the gel did not allow to identify the same number of chromosomes in both strains, it is possible to observe that the lowest DNA molecule in CLIB213^T is thicker and has higher intensity in the gel than the others, indicating that it can be the result of an overlap of two different chromosomes with similar size. A previous study [25] showed the chromosomal profile of the CLIB213^T and, although it had been performed at different conditions, it has allowed the separation of two DNA molecules of approximately 1.2Mbp but not the highest presented in this work (≈2.3Mbp), suggesting that the *Z. bailii* CLIB213^T strain, might have the same number of chromosomes as IST302. Therefore, the rough total DNA estimation of both CLIB213^T and IST302 strains, with a total sum of 9.5Mbp and 10.3Mbp, respectively, became closer to the one already described in the literature for CLIB213^T genome sequencing (≈10.3Mbp) [24]. However, the difference found for CLIB213^T (≈0.8Mbp) and the poor separation of the smallest DNA molecules for both strains, is significant enough to eventually all the chromosomes had not been correctly separated. The differences found in literature regarding the

number of chromosomes and total DNA amount of these strains (Table 5) is also an insight to support the hypothesis that CLIB213^T and IST302 strains possibly have more chromosomes than the ones observed in the PFGE results. Also the *Leu2*Δ strain [62,75,76], recently identified as a hybrid strain from *Z. parabailii* species shows a similar DNA pattern with the hybrid ISA1307 strain chromosome profile, which testifies the hypothesis of being a hybrid strain of *Z. bailii* and *Z. parabailii* species [19,25].

Table 4 – Summary of the data obtained in this work and the data available in literature regarding the number of chromosomes and DNA estimation of *Z. bailii* IST302 and CLIB213^T and of the hybrid strain ISA1307, through Pulse-Field Gel Electrophoresis and DNA sequencing. **(a)** data from Mira et. al 2014; **(b)** data from Galeote et. al 2013. ND (Not Defined).

Strain	**PFGE**				**Sequencing**
	N.º chromosomes		DNA estimation (Mbp)		DNA estimation (Mbp)
Z. bailii IST302	6		10.3		10.8
Z. bailii CLIB213^T	5	5 [a]	9.5	7.25 [a]	10.36 [b]
ISA1307	ND	13 [a]	ND	18.982 [a]	21.2 [a]

The need of readjusting some PFGE parameters, specifically, the total run time and its division in a certain number of phases and corresponding switch times, can be required to achieve a better chromosome separation and resolution of the gel. Other molecular approach, such as flow cytometry, is also a good method for, with more accuracy, estimate the DNA content of a certain strain and in this work an adaptation of a flow cytometry protocol for yeast cell cycle analysis was used [63].

Together with the total DNA estimation, another important genomic feature that was considered in this work was the prediction of the ploidy of the *Z. bailii* IST302 strain. *S. cerevisiae* BY4741 and BY4743 strains were used as haploid and diploid models [65], respectively. Also for comparison, the *Z. bailii* type strain and the hybrid ISA1307 strain were included in order to fully understand the major differences with IST302.

The SYBR® Green I dye has been successfully used to stain the nuclear DNA of the strains, thus providing an accurate method for the determination of ploidy and total DNA content of the cells [63,77]. A flow cytometry analysis is largely dependent on the correct dye usage and the good sample quality with the maximum number of single cells that successfully incorporate the dye [78]. A major problem often present is the formation of doublets or cellular aggregates [79]. A doublet is formed when two cells with a G0/G1-phase DNA content are recorded by the flow cytometer as one event with a cellular DNA content similar to the one of a cell in G2/M-phase [59]. Considering that *Z. bailii* CLIB213^T is a highly flocculating strain [9] and that even after sonication the cells remain aggregated (data not shown), this strain was not considered in flow cytometry analysis. To correctly assume the total DNA content and ploidy of a certain sample, a known genome size of comparable strains (*S. cerevisiae* BY4741 and BY4743 strains) should be employed [80] and therefore conclusions based on the similar values obtain in the mean fluorescence intensity values of the G1 and G2 peak could be done [77]. Regarding the reliability of the values obtained,

the low coefficient of variation (CV) exhibited in the cell cycle analysis (≤7%) is an indicator of the good quality of the histograms [81], which is a consequence of the high selectivity of the stain bind to the dsDNA [63]. Although the root mean square deviation (RMSD) values are considerably high for the CV [82] in both IST302 and ISA1307 strains, and the presence of lower peaks before the G0/G1 peak (Figure 6.A) can possibly be interpreted as aneuploid cells with lower DNA content in the cell population, it could be explained by the lack of quality of the samples, where possible debris and dead cells incorporate the dye [77,80,83]. This could be improved by changing the sonication conditions to a proper magnitude that allows the separation of the aggregates yet does not cause the cells disruption and therefore stain unspecific binding [63]. Hence, the values and the histogram profile comparisons obtained for the haploid *S. cerevisiae* BY4741 with the IST302 strain revealed similar profiles and mean fluorescent light intensity values for both G1 and G2 phases of each sample. In contrast, the diploid *S. cerevisiae* BY4743 and the hybrid ISA1307 strain show an approximate double value from those last two strains. Also, the histogram comparison between the haploid and diploid *S. cerevisiae* strains shows that the G2 peak of the haploid strain is coincident with the G1 peak of the diploid strain, which proves the double DNA content for the different populations in the same cell cycle phase.

The 11.35Mbp IST302 strain total genome size estimated through flow cytometry is close to the one known for the *S. cerevisiae* BY4741 haploid strain (11.68Mbp). Since the genome of the *Z. bailii* type strain comprises a total of approximately 10.3Mbp [24], IST302 could possibly have more chromosomes than the six obtained in the PFGE (estimation in 10.3Mbp) and therefore the optimization of PFGE running conditions should also be considered in future works in order to improve these results. Altogether, the *in silico* and molecular approaches carried out to characterize the genome of IST302 strain strongly suggest that it is a haploid strain of the *Z. bailii* species.

4.3 Heterologous expression of *ZbTPO2/3* gene in *S. cerevisiae* BY4741 parental and derived deletion mutants *tpo2*Δ and *tpo3*Δ

S. cerevisiae TPO2 and *TPO3* genes are activated in response to acetic acid induced-stress [46] and the corresponding deletion mutants are susceptible to acetic acid when compared to *S. cerevisiae* parental strain [50]. Given that *Z. bailii* species is remarkably tolerant to high concentrations of acetic acid, the role of the single *Z. bailii* gene homologous to *S. cerevisiae TPO2* and *TPO3*, the gene *ZbTPO2/3* in acetic acid tolerance was investigated in this thesis. Hence, the *ZbTPO2/3* gene was expressed in *S. cerevisiae tpo2*Δ and *tpo3*Δ deletion mutants.

One of the objectives of this work was the construction of a recombinant vector containing *ZbTPO2/3* under the control of *TPO2* promoter region by replacement of the *GAL1* promoter in pGREG506 [84]. However, despite repeated efforts, it was not possible to successfully engineer the intended recombinant vector (data not shown) and a recombinant vector containing *ZbTPO2/3* under the control of *GAL1* promoter was used in further studies (see Section 2.4.1 in Materials and Methods). For this reason, a second selective medium with the same components plus 1% galactose was added to the experiment

in order to understand whether different levels of *ZbTPO2/3* activation would induce differences in yeast transformants' susceptibility to acetic acid.

The assay only comprehended the susceptibility phenotype complementation to acetic acid of the transformed *S. cerevisiae* deletion mutants *tpo2*Δ and *tpo3*Δ in order to evaluate if the *ZbTPO2/3* gene could be a functional homologue of *S. cerevisiae TPO2* or *TPO3* genes under acetic acid induced stress. The results were not as expected since no evident phenotype complementation was observed, although a slight increase to acetic acid tolerance for both of the *S. cerevisiae* derived deletion mutants was observed with the expression of *ZbTPO2/3* gene [41]. Also, strains' growth profile was very similar either in a medium containing glucose and galactose or only glucose as a carbon source. Considering that the presence of galactose in the growth medium would lead to higher *ZbTPO2/3* transcript levels controlled by *GAL1* promoter, it would be expected to find some differences in strains growth under acetic acid. It would be interesting to determine the mRNA levels of *ZBTPO2/3* in the recombinant strains to check whether the different growth media lead to different *ZBTPO2/3 transcript level*s. The results also showed that *tpo3*Δ deletion mutants presented a longer lag phase than the *tpo2*Δ mutants which is in agreement with the observed in previous results that showed that the *S. cerevisiae tpo3*Δ mutant is more susceptible to acetic acid than *tpo2*Δ mutant [41,47]. In general the results presented in this work show that the expression of *ZbTPO2/3* does not complement the acetic acid susceptibility phenotype of both deletion mutants up to the level of the parental strain, strongly suggesting that the *ZbTpo2/3* is not a crucial player in *Z. bailii* tolerance to acetic acid.

However, further studies in this matter should be performed to better understand these results. For example, a recombinant plasmid expressing *ZbTPO2/3*, *TPO2* or *TPO3* under the control of the same promoter should be constructed in order to have the appropriate positive controls. Additionally, the deletion of *ZbTPO2/3* gene should be considered in *Z. bailii*. *Z. bailii* IST302 would be a promising candidate strain for that purpose given that is very easy to manipulate in the laboratory (does not flocculate) and was found to be haploid in the present study, which would imply the elimination of only one gene allele from the chromosome. A similar essay to the one described in this work with those derived deletion mutants, should be carried out in order to understand the true role and the importance of the *ZbTPO2/3* under acetic acid stress. The obtained mutant would be compared with *Z. bailii* parental strain regarding their susceptibility to acetic acid and to other weak acids. The intracellular accumulation of acetic acid in both strain under acetic acid stress would also provide good indications of the role of ZbTpo2/3 in acetic acid tolerance. It is quite possible that ZbTpo2/3 is not an acetate exporter in *Z. bailii* as Tpo2 and Tpo3 were suggested to be in *S. cerevisiae*. Considering that *Z. bailii* cells have higher basal levels of sphingolipids than the ones exhibited in *S. cerevisiae* cells that have been proposed to contribute to a thicker and less acetic acid permeable membrane, leading to less accumulation of the weak acid into the cytosol [38,85], it would be expected that *Z. bailii* cells do not have to deal with high intracellular acetic acid concentrations. Also, *Z. bailii* cells are able to use acetic acid as a carbon source even in the presence of glucose, with acetate being channeled to both Krebs cycle and Lipid synthesis, which might contribute as well to the reduction of intracellular acetic acid

concentration [32,33]. Both physiological strategies might make unnecessary the export of acetate by a plasma membrane transporter such as ZbTpo2/3.

4.4 Conclusions and prospects

Although the results in this work confirm the correct identification of IST302 strain as a *Z. bailii* species through the *in silico* analysis, its karyotype and total genomic DNA are not yet clearly defined. The difference of the total DNA estimation in both PFGE and flow cytometry methods used in this work, suggests that the total number of chromosomes in IST302 strain is possibly higher than initially predicted by these results, since the flow cytometry data regarding the DNA quantification is more accurate. Obtaining the full DNA sequence of strain IST302 could be an important step to support these results and clarify its karyotype and also achieve the total amount of DNA from IST302 strain. In addition, the improvement of PFGE conditions here defined by changing the run parameters could be a suitable alternative to access the total number of chromosomes and elucidate the IST302 karyotype. The confirmation of the haploid state in IST302 cells is an important result, once it provides new future insights regarding the use of this strain as a model for spoilage yeasts, studying its molecular mechanisms under weak-acid induced stress by genetic engineering or other possible approaches.

5. Bibliography

1. Sa-correia, I., Guerreiro, J. F., Loureiro-Dias, M. C., Leao, C. & Côrte-Real, M. *Encyclopedia of Food Microbiology*. *Elsevier Ltd, Academic Press* **3,** (2014).

2. Thomas, D. S. & Davenport, R. R. Zygosaccharomyces bailii - a profile of characteristics and spoilage activities. *Food Microbiol.* **2,** 157–169 (1985).

3. Fleet, G. Spoilage yeasts. *Crit. Rev. Biotechnol.* **12,** 1–44 (1992).

4. Loureiro, V. & Malfeito-Ferreira, M. Spoilage yeasts in the wine industry. *Int. J. Food Microbiol.* **86,** 23–50 (2003).

5. Kregiel, D. Health safety of soft drinks: Contents, containers, and microorganisms. *Biomed Res. Int.* **2015,** (2015).

6. Lambert, R. J. & Stratford, M. Weak-acid preservatives: Modelling microbial inhibition and response. *J. Appl. Microbiol.* **86,** 157–164 (1999).

7. Martorell, P., Stratford, M., Steels, H., Fernández-espinar, M. T. & Querol, A. Physiological characterization of spoilage strains of Zygosaccharomyces bailii and Zygosaccharomyces rouxii isolated from high sugar environments. *Int. J. Food Microbiol.* **114,** 234–242 (2007).

8. Pitt, J. I. & Hocking, A. D. *Fungi and Food Spoilage*. *Springer Science* **53,** (Springer, 2009).

9. Kurtzman, C. P., Fell, J. W. & Boekhout, T. *The Yeasts, A Taxonomic Study*. (Elsevier B.V., 2011).

10. Lund, B. M., Baird-Parker, T. C. & Gould, G. W. *The Microbiological Safety and Quality of Food*. (Aspen Publishers, Inc., 2000). doi:10.1097/00000433-198206000-00020

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

12. Pincus, D. H., Orenga, S. & Chatellier, S. Yeast identification – past, present, and future methods. *Med. Mycol.* **45,** 97–121 (2007).

13. Barnett, J. A. in *Yeast* **21,** 1141–1193 (John Wiley & Sons, Inc., 2004).

14. Kurtzman, C. P. Yeast species recognition from gene sequence analyses and other molecular methods. *Mycoscience* **47,** 65–71 (2006).

15. Rosa, C. a. & Lachance, M. A. Zygosaccharomyces machadoi sp. n., a yeast species isolated

from a nest of the stingless bee tetragonisca angustula. *Lundiana* **6,** 27–29 (2005).

16. Torriani, S., Lorenzini, M., Salvetti, E. & Felis, G. E. Zygosaccharomyces gambellarensis sp. nov., an ascosporogenous yeast isolated from an Italian 'passito' style wine. *Int. J. Syst. Evol. Microbiol.* **61,** 3084–3088 (2011).

17. Saksinchai, S., Suzuki, M., Chantawannakul, P., Ohkuma, M. & Lumyong, S. A novel ascosporogenous yeast species, Zygosaccharomyces siamensis, and the sugar tolerant yeasts associated with raw honey collected in Thailand. *Fungal Divers.* **52,** 123–139 (2012).

18. Solieri, L., Dakal, T. C. & Giudici, P. Zygosaccharomyces sapae sp. nov., isolated from Italian traditional balsamic vinegar. *Int. J. Syst. Evol. Microbiol.* **63,** 364–371 (2013).

19. Suh, S. O., Gujjari, P., Beres, C., Beck, B. & Zhou, J. 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 (2013).

20. Čadež, N., Fülöp, L., Dlauchy, D. & Péter, G. Zygosaccharomyces favi sp. nov., an obligate osmophilic yeast species from bee bread and honey. *Antonie Van Leeuwenhoek* **107,** 645–654 (2014).

21. Fleet, G. H. Yeasts in foods and beverages: impact on product quality and safety. *Curr. Opin. Biotechnol.* **18,** 170–175 (2007).

22. Peterson, S. W. & Kurtzman, C. P. Ribosomal RNA Sequence Divergence Among Sibling Species of Yeasts. *Syst. Appl. Microbiol.* **14,** 124–129 (1991).

23. Kurtzman, C. P. & Robnett, C. J. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **73,** 331–371 (1998).

24. Galeote, V., Bigey, F., Devillers, H., Neuvéglise, C. & Dequin, S. Genome Sequence of the Food Spoilage Yeast Zygosaccharomyces bailii CLIB 213T. *Genome Announc.* **1,** 2010–2011 (2013).

25. Mira, N. P. *et al.* 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 (2014).

26. Rodrigues, F. *et al.* The Spoilage Yeast Zygosaccharomyces bailii Forms Mitotic Spores : a Screening Method for Haploidization. *Appl. Environ. Microbiol.* **69,** 649–653 (2003).

27. Pitt, J. I. & Hocking, A. D. *Fungi and Food Spoilage*. *Blackie Academic & Professional* (Blackie Academic & Professional, 1997).

28. Steels, H., James, S. A., Roberts, I. N. & Stratford, M. Sorbic acid resistance: the inoculum effect. *Yeast* **16,** 1173–1183 (2000).

29. Sousa-Dias, S., Gonçalves, T., Leyva, J. S., Peinado, J. M. & Loureiro-Dias, M. C. Kinetics and regulation of fructose and glucose transport systems are responsible for fructophily in Zygosaccharomyces bailii. *Microbiology* **142,** 1733–1738 (1996).

30. Leandro, M. J., Cabral, S., Prista, C., Loureiro-Dias, M. C. & Sychrova, H. The High-Capacity Specific Fructose Facilitator ZrFfz1 Is Essential for the Fructophilic Behavior of Zygosaccharomyces rouxii CBS 732T. *Eukaryot. Cell* **13,** 1371–1379 (2014).

31. Trumbly, R. J. Glucose repression in the yeast Saccharomyces cerevisiae. *Mol. Microbiol.* **6,** 15–21 (1992).

32. Sousa, M. J., Rodrigues, F., Côrte-Real, M. & Leăo, C. Mechanisms underlying the transport and intracellular metabolism of acetic acid in the presence of glucose in the yeast Zygosaccharomyces bailii. *Microbiology* **144,** 665–670 (1998).

33. Rodrigues, F. *et al.* The Fate of Acetic Acid during Glucose Co-Metabolism by the Spoilage Yeast Zygosaccharomyces bailii. *PLoS One* **7,** 1–7 (2012).

34. Guerreiro, J. F., Mira, N. P. & Sá-Correia, I. Adaptive response to acetic acid in the highly resistant yeast species Zygosaccharomyces bailii revealed by quantitative proteomics. *Proteomics* **12,** 2303–2318 (2012).

35. Arneborg, N., Jespersen, L. & Jakobsen, M. Individual cells of Saccharomyces cerevisiae and Zygosaccharomyces bailii exhibit different short-term intracellular pH responses to acetic acid. *Arch. Microbiol.* **174,** 125–128 (2000).

36. Avery, S. V. Microbial cell individuality and the underlying sources of heterogeneity. *Nat. Rev. Microbiol.* **4,** 577–587 (2006).

37. Stratford, M. *et al.* International Journal of Food Microbiology Extreme resistance to weak-acid preservatives in the spoilage yeast Zygosaccharomyces bailii. *Int. J. Food Microbiol.* **166,** 126–134 (2013).

38. Lindberg, L., Santos, A. X. S., Riezman, H., Olsson, L. & Bettiga, M. Lipidomic Profiling of Saccharomyces cerevisiae and Zygosaccharomyces bailii Reveals Critical Changes in Lipid Composition in Response to Acetic Acid Stress. *PLoS One* **8,** 1–12 (2013).

39. Casal, M., Cardoso, H. & Leão, C. Mechanisms regulating the transport of acetic acid in Saccharomyces cerevisiae. *Microbiology* **142,** 1385–1390 (1996).

40. Carmelo, V., Santos, H. & Sá-Correia, I. Effect of extracellular acidification on the activity of

plasma membrane ATPase and on the cytosolic and vacuolar pH of Saccharomyces cerevisiae. *Biochim. Biophys. Acta - Biomembr.* **1325,** 63–70 (1997).

41. Fernandes, a. R., Mira, N. P., Vargas, R. C., Canelhas, I. & Sá-Correia, I. Saccharomyces cerevisiae adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. *Biochem. Biophys. Res. Commun.* **337,** 95–103 (2005).

42. dos Santos, S. C., Teixeira, M. C., Dias, P. J. & Sá-Correia, I. MFS transporters required for multidrug/multixenobiotic (MD/MX) resistance in the model yeast: Understanding their physiological function through post-genomic approaches. *Front. Physiol.* **5 MAY,** 1–15 (2014).

43. Teixeira, M. C., Godinho, C. P., Cabrito, T. R., Mira, N. P. & Sá-Correia, I. Increased expression of the yeast multidrug resistance ABC transporter Pdr18 leads to increased ethanol tolerance and ethanol production in high gravity alcoholic fermentation. *Microb. Cell Fact.* **11,** 98 (2012).

44. Sá-Correia, I., dos Santos, S. C., Teixeira, M. C., Cabrito, T. R. & Mira, N. P. Drug:H+ antiporters in chemical stress response in yeast. *Trends Microbiol.* **17,** 22–31 (2009).

45. Reddy, V. S., Shlykov, M. a, Castillo, R., Sun, E. I. & Saier, M. H. The Major Facilitator Superfamily (MFS) Revisited. *FEBS J.* **279,** 2022–2035 (2012).

46. Dias, P. J., Seret, M.-L., Goffeau, A., Correia, I. S. & Baret, P. V. Evolution of the 12-spanner drug:H+ antiporter DHA1 family in hemiascomycetous yeasts. *OMICS* **14,** 701–710 (2010).

47. Mira, N. P., Becker, J. D. & Sá-Correia, I. Genomic expression program involving the Haa1p-regulon in Saccharomyces cerevisiae response to acetic acid. *OMICS* **14,** 587–601 (2010).

48. Mira, N. P., Teixeira, M. C. & Sá-Correia, I. Adaptive response and tolerance to weak acids in Saccharomyces cerevisiae: a genome-wide view. *OMICS* **14,** 525–540 (2010).

49. Mira, N. P. *et al.* Identification of a DNA-binding site for the transcription factor Haa1, required for Saccharomyces cerevisiae response to acetic acid stress. *Nucleic Acids Res.* **39,** 6896–6907 (2011).

50. Tanaka, K., Ishii, Y., Ogawa, J. & Shima, J. Enhancement of acetic acid tolerance in Saccharomyces cerevisiae by overexpression of the Haa1 gene, encoding a transcriptional activator. *Appl. Environ. Microbiol.* **78,** 8161–8163 (2012).

51. Tomitori, H. *et al.* Multiple polyamine transport systems on the vacuolar membrane in yeast. *Biochem. J.* **353,** 681–688 (2001).

52. Uemura, T., Tachihara, K., Tomitori, H., Kashiwagi, K. & Igarashi, K. Characteristics of the polyamine transporter TPO1 and regulation of its activity and cellular localization by

phosphorylation. *J. Biol. Chem.* **280,** 9646–9652 (2005).

53. Birren, B. & Lai, E. *Pulsed Field Gel Electrophoresis: A Practical Guide*. **9,** (Academic Press, Inc., 1993).

54. Mills, D. & McCluskey, K. Electrophoretic Karyotypes of Fungi: The New Cytology. *Mol. Plant-Microbe Interact.* **3,** 351–357 (1990).

55. Birren, W. B., Lai, E., Clark, M. C., Hood, L. & Simon, I. M. Optimized conditions for pulsed field gel electrophoretic separations of DNA. *Nucleic Acids Res.* **16,** 7563–7582 (1988).

56. Herschleb, J., Ananiev, G. & Schwartz, D. C. in *Nature Protocols* **2,** 677–684 (2007).

57. Sanderson, B. A., Naoko, A., Jennifer, L. L., Gilberto, G. & Kevin, L. L. Modification of gel architecture and TBE/TAE buffer composition to minimize heating during agarose gel electrophoresis. *Anal Biochem.* **8,** 425–429 (2014).

58. Brown, M. & Wittwer, C. Flow Cytometry: Principles and Clinical Applications in Hematology. *Clin. Chem.* **46,** 1221–1229 (2000).

59. Nunez, R. DNA Measurement and Cell Cycle Analysis by Flow Cytometry. *Curr. Issues Mol. Biol.* **3,** 67–70 (2001).

60. Picot, J., Guerin, C. L., Le Van Kim, C. & Boulanger, C. M. Flow cytometry: Retrospective, fundamentals and recent instrumentation. *Cytotechnology* **64,** 109–130 (2012).

61. Ferreira, M. M., Loureiro-Dias, M. C. & Loureiro, V. Weak acid inhibition of fermentation by Zygosaccharomyces bailii and Saccharomyces cerevisiae. *Int. J. Food Microbiol.* **36,** 145–153 (1997).

62. Dato, L. *et al.* 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 (2010).

63. Fortuna, M. *et al.* Cell cycle analysis of yeasts. *Curr. Protoc. Cytom.* **Chapter 11,** Unit 11.13 (2001).

64. Palma, M. *et al.* Search for genes responsible for the remarkably high acetic acid tolerance of a Zygosaccharomyces bailii-derived interspecies hybrid strain. *BMC Genomics* **16,** 1070 (2015).

65. Brachmann, C. B. *et al.* Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14,** 115–132 (1998).

66. Hammond, T. G., Majewski, R. R., Morré, D. J., Schell, K. & Morrissey, L. W. Forward Scatter

Pulse Width Signals Resolve Multiple Populations of Endosomes. *Cytometry* **14,** 411–420 (1993).

67. Steels, H., James, S. a., Bond, C. J., Roberts, I. N. & Stratford, M. Zygosaccharomyces kombuchaensis: The physiology of a new species related to the spoilage yeasts Zygosaccharomyces lentus and Zygosaccharomyces bailii. *FEMS Yeast Res.* **2,** 113–121 (2002).

68. Steels, H. *et al.* Zygosaccharomyces lentus sp. nov., a new member of the yeast genus Zygosaccharomyces Barker. *Int. J. Syst. Bacteriol.* **49,** 319–327 (1999).

69. Schoch, C. L. *et al.* Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. **109,** 6241–6246 (2012).

70. Liu, K. L., Porras-Alfaro, A., Kuske, C. R., Eichorst, S. A. & Xie, G. Accurate, rapid taxonomic classification of fungal large-subunit rRNA Genes. *Appl. Environ. Microbiol.* **78,** 1523–1533 (2012).

71. Alcoba-flórez, J., Arévalo-morales, M. P., Pérez-roth, E., Rivero-pérez, B. & Méndez-álvarez, S. Yeast molecular identification and typing. 535–546 (2007).

72. Basim, HŸ. Pulsed-Field Gel Electrophoresis (PFGE) Technique and its use in Molecular Biology. *Turk J Biol* **25,** 405–418 (2001).

73. Basim, E. & Basim, HŸ. H. Pulse-Field Gel Electrophoresis (PFGE) Technique and its use in Molecular Biology. *Turkish J. Biol.* **25,** 405–418 (2001).

74. Krogerus, K., Magalhães, F., Vidgren, V. & Gibson, B. New lager yeast strains generated by interspecific hybridization. *J Ind Microbiol Biotechnol* 769–778 (2015). doi:10.1007/s10295-015-1597-6

75. Branduardi, P. *et al.* The yeast Zygosaccharomyces bailii: A new host for heterologous protein production, secretion and for metabolic engineering applications. *FEMS Yeast Res.* **4,** 493–504 (2004).

76. Vigentini, I. *et al.* Heterologous protein production in Zygosaccharomyces bailii: Physiological effects and fermentative strategies. *FEMS Yeast Res.* **5,** 647–652 (2005).

77. Delobel, P. & Tesnière, C. A Simple FCM Method to Avoid Misinterpretation in Saccharomyces cerevisiae Cell Cycle Assessment between G0 and Sub-G1. *PLoS One* **9,** e84645 (2014).

78. Brown, M. *et al.* Improved Flow Cytometric Analysis of the Budding Yeast Cell Cycle. *Cell Cycle* **13,** 117–121 (2013).

79. Wersto, R. P. *et al.* Doublet discrimination in DNA cell-cycle analysis. *Commun. Clin. Cytom.* **46,** 296–306 (2001).

80. Dolezel, J. Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size. *Ann. Bot.* **95,** 99–110 (2005).

81. Marie, D., Simon, N., Guillou, L., Partensky, F. & Vaulot, D. in *Current Protocols in Cytometry* 1–21 (John Wiley & Sons, Inc., 2000).

82. Ormerod, M. G., Payne, A. W. R. & Watson, J. V. Improved Program for the Analysis of DNA Histograms. *Cytometry* **8,** 637–641 (1987).

83. Zamai, L., Falcieri, E., Zauli, G., Cataldi, A. & Vitale, M. Optimal Detection of Apoptosis by Flow Cytometry Depends on Cell Morphology. *Cytometry* **14,** 891–897 (1993).

84. Jansen, G., Wu, C., Schade, B., Thomas, D. Y. & Whiteway, M. Drag&Drop cloning in yeast. *Gene* **344,** 43–51 (2005).

85. Lindahl, L., Genheden, S., Eriksson, L. A., Olsson, L. & Bettiga, M. Sphingolipids contribute to acetic acid resistance in Zygosaccharomyces bailii. *Biotechnol. Bioeng.* **113,** 744–753 (2016).

6. Attachments

Table A1 – List of all strains and respective D1/D2 region of the LSU rRNA gene accession codes used in the consctruction of a phylogenetic tree (except the respective for strain Z. bailii IST302). This data was acquired from the Song-Oui et al. 2013 supplementary article.

Species	Strain	Acession Code
Lachancea thermotolerans	NRRL Y-8284^T	U69581
Zygosaccharomyces mellis	NRRL Y-12628^T	U72164
Zygosaccharomyces siamensis	JCM 16825^T	AB565756
Zygosaccharomyces gambellarensis	CBS 12191^T	AF432228
Zygosacchraomyces machadoi	CBS 10264^T	FR725931
Zygosaccharomyces rouxii	NRRL Y-229^T	U72163
Zygosaccharomyces sapae	CBS 12607 ITS copy 3	AJ966342
	CBS 12607 ITS copy 2	AJ966342
	CBS 12607 ITS copy 1	AJ966342
Zygosaccharomyces bisporus	NRRL Y-12626^T	U72162
Zygosaccharomyces parabailii	ATCC MYA-4549	FJ914902
	ATCC 302778	JX458132
	ATCC 304978	JX458133
	ATCC 60483	JX458110
	ATCC 8099	JX458114
	ATCC 60484	JX458111
	ATCC 304979	JX458134
	ATCC 56535	JX458112
	ATCC 56075^T	JX458113
	ATCC 36947	JX458115
Zygosaccharomyces bailii	ATCC 38923	JX458116
	ATCC 58445^T	JX458117
	ATCC 38924	JX458118
	ATCC 8766	JX458119
	ATCC 11486	JX458120
	ATCC 42476	JX458121
	ATCC 42477	JX458122
	ATCC 66825	JX458123
Zygosaccharomyces pseudobailii	ATCC 2333	JX458124
	ATCC 2602	JX458125
	ATCC 36946	JX458126
	ATCC 11003	JX458127
	ATCC 56074^T	JX458128
Zygosaccharomyces kombuchaensis	NRRL YB-4811^T	AF339904
Zygosaccharomyces lentus	NRRL Y-27276^T	AF339888

Table A2 – List of gene accession numbers for all of the strains used in in silico comparison analysis with the homologous sequences found in IST302 strain for identity percentage and nucleotide variation. The *coxII*, mSSU rRNA, *β-tubulin* and *EF-1α* genes were considered only the type strain of each species. For *Z. parabailii* ATCC 60484 the *RPB1* gene is not available (not sequenced).

Species	Strain	Gene (Accession number)					
		coxII	**mSSU rRNA**	***β-tubulin***	***EF-1α***	***RPB1***	***RPB2***
Zygosaccharomyces bailii	ATCC 58445^{T}	KC678086	KC678092	KC678083	KC678089	KC678099	KC678068
	ATCC 8766					KC678095	KC678069
	ATCC 42476					KC678098	KC678070
	ATCC 66825					KC678100	KC678071
	ATCC 38924					KC678097	KC678072
	ATCC 38923					KC678096	KC678073
Zygosaccharomyces parabailii	ATCC 56075^{T}	KC678088	KC678094	KC678085	KC678090	KC678102	KC678076
	ATCC 8099					KC678103	KC678080
	ATCC 56535					KC678105	KC678078
	ATCC 60483					ND	KC678079
	ATCC 60484					KC678101	KC678080
	ATCC MYA-4549					KC678106	KC678081
	ATCC 36947					KC678104	KC678082
Zygosaccharomyces pseudobailii	ATCC 56074^{T}	KC678087	KC678093	KC678084	KC678091	KC678107	KC678074
	ATCC 2333					KC678108	KC678075

Table A3 – List of all strains and respective protein-coding genes used in nucleotide variation analysis. In the "Nucleotide Variation" column is represented in bold the number of nucleotide variations exhibited between *Z. bailii*, *Z. parabailii* and *Z. pseudobailii* against IST302 homologous genes and in parenthesis are the nucleotide identity over the total of the respective gene sequence considered.

Genes	Strains	Nucleotide Variation
coxII	*Z. bailii* ATCC 58445^T	**0** (616/616)
	Z. parabailii ATCC 56075^T	**4** (612/616)
	Z. pseudobailii ATCC 56074^T	**0** (616/616)
mSSU rRNA	*Z. bailii* ATCC 58445^T	**0** (560/560)
	Z. parabailii ATCC 56075^T	**2** (421/423)
	Z. pseudobailii ATCC 56074^T	**2** (421/423)
β-tubulin	*Z. bailii* ATCC 58445^T	**2** (881/883)
	Z. parabailii ATCC 56075^T	**38** (845/883)
	Z. pseudobailii ATCC 56074^T	**42** (841/883)
EF1-α	*Z. bailii* ATCC 58445^T	**0** (932/932)
	Z. parabailii ATCC 56075^T	**19** (913/932)
	Z. pseudobailii ATCC 56074^T	**23** (909/932)
RPB1	*Z. bailii* ATCC 58445^T	**3** (657/660)
	Z. bailii ATCC 8766	**3** (657/660)
	Z. bailii ATCC 42476	**3** (657/660)
	Z. bailii ATCC 66825	**3** (657/660)
	Z. bailii ATCC 38924	**3** (657/660)
	Z. bailii ATCC 38923	**3** (657/660)
	Z. parabailii ATCC 56075^T	**25** (635/660)
	Z. parabailii ATCC 8099	**27** (633/660)
	Z. parabailii ATCC 56535	**27** (633/660)
	Z. parabailii ATCC 60484	**3** (657/660)
	Z. parabailii ATCC MYA-4549	**26** (634/660)
	Z. parabailii ATCC 36947	**26** (634/660)
	Z. pseudobailii ATCC 56074^T	**38** (622/660)
	Z. pseudobailii ATCC 2333	**26** (622/660)
RPB2	*Z. bailii* ATCC 58445^T	**8** (1025/1033)
	Z. bailii ATCC 8766	**8** (1025/1033)
	Z. bailii ATCC 42476	**8** (1025/1033)
	Z. bailii ATCC 66825	**8** (1025/1033)
	Z. bailii ATCC 38924	**7** (1026/1033)
	Z. bailii ATCC 38923	**2** (1031/1033)
	Z. parabailii ATCC 56075^T	**45** (988/1033)
	Z. parabailii ATCC 8099	**44** (989/1033)
	Z. parabailii ATCC 56535	**46** (987/1033)
	Z. parabailii ATCC 60483	**43** (990/1033)
	Z. parabailii ATCC 60484	**45** (988/1033)
	Z. parabailii ATCC MYA-4549	**46** (987/1033)
	Z. parabailii ATCC 36947	**46** (987/1033)
	Z. pseudobailii ATCC 56074^T	**71** (962/1033)
	Z. pseudobailii ATCC 2333	**71** (962/1033)