Unveiling the molecular basis of *S. boulardii* probiotic activity: through the construction of a global promoter survey computational tool and transcriptomics analyses

Dias M., Teixeira M., Monteiro P.

Abstract

Although *S. cerevisiae* and *S. boulardii* share 95% homology, only *S. boulardii* seems to be probiotic. A better understanding of the underlying properties is crucial for the optimization of probiotic strain selection and usage. In this thesis, we have grown *S. boulardii* and *S. cerevisiae* in an intestinal like medium (ILM) and performed a transcriptional analysis in order to understand the different survival and probiotic characteristics that lead different phenotypic outcomes. It was concluded that *S. boulardii* is more sensitive to salt stress, but better adapted to human body temperature and oxidative stress. *S. boulardii* displayed up-regulation of genes associated with probiotic activity, including some involved in polyamine and acetate biosynthesis and cell wall and adhesion proteins. Consistently, *S. boulardii* displayed higher adherence to intestinal epithelium. Additionally, a new functionality was implemented in the ProBioYeastract platform, to enable a global evaluation of promoter regions in *S. boulardii* genes, when compared to *S. cerevisiae* homologs. This tool was used to predict the regulatory changes that underlie transcriptomic variations between the two strains, pointing out transcription factors Yap3 and Gcn4 as displaying a different set of regulated genes in the two species, with a potential impact in intestinal tract adaptation.

Keywords: *S. boulardii*, Probiotic, Intestinal like medium, Transcriptomics, ProBioYeastract

Introduction

Gut dysbiosis refers to the microbiome's composition alterations and probiotics are a promising treatment or adjuvant to diseases that lead to gut dysbiosis. Most common diseases of this kind are pediatric diarrhea, antibiotic-associated diarrhea, *H. pylori* infection, inflammatory bowel diseases, irritable colon syndrome, acute diarrhea in adults and *Clostridium difficile* infections and *S.boulardii* has shown very promising probiotic properties in most of these disesases without posing a safety risk. [1,2]

Although *S. cerevisiae* and *S. boulardii* share 95% homology (and in other studies 99%), only *S. boulardii* is considered a probiotic. Although they are, indeed, very close genetically (95% DNA homology) and have a very similar karyotype. *S. boulardii* is physiologically and metabolically different from *S. cerevisiae*. For example, it is asporogenous (but produces fertile hybrids with *S. cerevisiae*) and unable of switching to haploid form, assimilating galactose as a carbon source and of having α-glucosidase activity. It is more resistant to temperature and acidic stresses, similarly resistant to alkaline pH, but less resistant to bile salts. It lost its intact Ty1, Ty3 Ty4 elements and has an higher capability of pseudohyphal switching. However, phenotypic characteristics such as these cannot be used for identification. [1,3,]

Probiotics must be able to endure in adverse conditions. The main obstacles in the small intestine are the high concentrations of bile salts, pancreatic enzymes, hydrolytic enzymes, pancreatin, organic acids, the integrity of the epithelial and brush border, the immune defense and the native microbiota. Bile salts are toxic to some microorganism because they affect their cellular membranes' lipid bilayer structure.[4] *S. boulardii* grows faster than *S. cerevisiae* at both 30 and 37 °C, which is important in probiotic strains, since the human's body core temperature can go from 35.8 to 38.2 °C (medium of 37°C). [4] *S. cerevisiae* is more tolerant to bile salts than *S. boulardii*. However, since resistance to bile salts is considered for organisms that can support at least 0.3% (w/v), none of these strains can be considered resistant. [4]

Probiotics are believed to display a variety of mechanisms: antitoxin effects, physiological protection, modulation of the normal microbiome, metabolic regulation and signaling pathway modification, nutritional and trophic effects, immune system regulation, pathogen's inhibition, interactions with the brain-gut axis, cellular adhesion, cellular antagonism and mucin production. [1,5]

S. boulardii is capable of producing around 44 cell wall and/or adhesion proteins by which they mediate the adhesion of pathogenic bacteria to yeast cells. This interaction limits the ability of the pathogen to bind directly to the intestinal receptors and proceed with host invasion. Furthermore, since *S. boulardii* is unable to bind to epithelial cells of healthy individuals and is quickly flushed out, when pathogens bind to *S. boulardii* it is possible that they are flushed together with the yeast cells. [6]

S. boulardii produces a 54 kDa serine protease, which gradually degrades (by hydrolysis) directly toxin A and B originated from *C. difficile* and the cellular enterocytic surface's receptor to which the toxins bind. [7–10] *S. boulardii* also produces a 63 kDa alkaline phosphatase, which inhibits *E coli*'s endotoxins (by dephosphorylating their activation sites) and diminishes its toxicity. *Vibrio cholerae* increases cAMP levels and secretion and produces cholera toxin (CT). *S. boulardii* produces a 120 kDa protein that decreases water and sodium secretion in intestinal loops and counteract the increase in cAMP levels in rat intestinal cells done by *Vibrio cholerae*'s toxin, leading to this strains toxicity inhibition.[1, 11]

Materials and Methods

Strain and media

The strains used were *S. cerevisiae* strain BY4741 and *S. boulardii* strain CNCM I-745. Depending on the essay, each of the two strains was cultivated in SIEM liquid medium (pH 7)[12,13] or in modified YPD liquid medium. SIEM liquid medium was composed by BD bactotryptane (5.7 g/L), D-glucose (2.4 g/L), sodium chloride (6.14 g/L), monopotassium phosphate (0.68 g/L), monosodium phosphate (0.3 g/L), sodium bicarbonate (1.01 g/L), sodium cholate (2.8 g/L), sodium bicarbonate (2.8 g/L), lysozyme (0.2 g/L), α-amylase (1000 U/L), trypsin (110 U/L), chymotrypsin (380 U/L) and lipase (960 U/L). The same salts, bile salts and enzymes were added to normal YPD liquid media in the same concentrations as SIEM liquid medium.

Total RNA extraction and RNA sequencing

Three replicates of *S. cerevisiae* and of *S. boulardii* were grown in ILM medium until early-log phase (OD_{600nm} of 0.8±0.05). Cells were harvested by centrifugation. The resulting pellets were stored at -80ºC. Total RNA was isolated using an Ambion Ribopure-Yeast RNA kit, according to manufacturer's instructions.

Adhesion to human epithelial cell

FHs 74 Int small intestine cell line (ATCC® CCL-241™) were used. Cells were cultivated in 24-well polystyrene plates (Greiner), in Hybri-Care Medium (ATCC® 46-X™), until a density of $1.25x10^5$ cells/well was reached after 24h of incubation. *S. boulardii* and *S. cerevisiae* cells were added to each well with a density of $1.25x10^6$ cells/well (MOI=10). CFU (Colony Forming Units) were counted for each well, representing the proportion of cells adherent to the human epithelium. Statistical analysis of the results was performed using analysis of variance and differences were considered significant for $p < 0.05$.

Results and discussion

Optimization of in vitro intestinal tract-like growth media

In order to evaluate the global gene expression pattern in *S. boulardii* when compared with *S. cerevisiae*, cultivation conditions that mimic the gastrointestinal tract, but still enable the growth of both strains, were looked for. As starting point the SIEM liquid medium[12,13] was tested. Surprisingly, although *S. cerevisiae* did grow in this medium, *S. boulardii* was unable to do so (*Figure 1*).

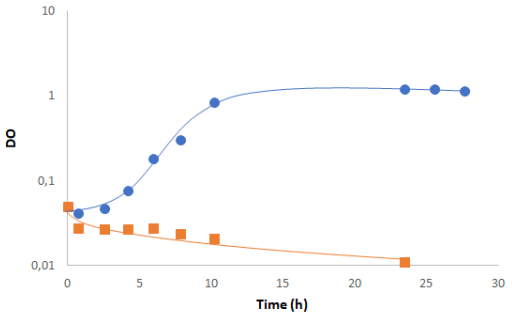

Figure 1: S. cerevisiae (●) and S. boulardii (■) growth curve in SIEM liquid medium.

To optimize the SIEM liquid medium to enable *S. boulardii* growth, several adaptations were tested. 35 media were tested. As a first hypothesis, it was considered that the gastrointestinal enzymes might inhibit the yeast's growth, but the results showed that either the enzymes were not responsible for the inhibition or they were not the only ones responsible. The yeast continued not to grow even though the medium was supplemented with different sources of nutrients (ammonium sulphate, YNB, peptone instead of tryptone; yeast extract, higher concentration of glucose). Supplementation with yeast extract and inhibition by enzymes was, but *S. boulardii* still didn't grow.

Modified YPD liquid medium was also analyzed and *S. boulardii* still did not grow, so it was concluded that the lack of growth was not due to lack of nutrients but to inhibition by one or several of the media's constituents. Inhibition of each group of substances (salts, bile salts and enzymes) was tested by themselves or in combination. It was found that bile salts were responsible for growth inhibition. All media without bile salts were able to sustain cell growth, however non-bile salts partially inhibited growth, as we can also see in *Figure 2*.

The final optimized growth medium contained bile salts and non-bile salts, but their concentration was reduced to half of the initial concentration (intestinal like medium (ILM) media). It was then possible to register growth of *S. boulardii*

cells in a medium that includes all components of the gastro intestinal tract (*Figure 3*).

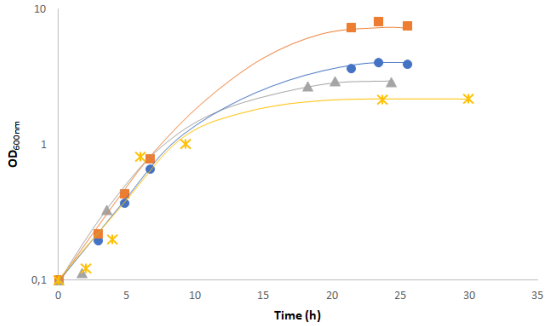

Figure 2: S. boulardii growth curve in SIEM liquid medium without salts, bile salts nor enzymes (●); SIEM liquid medium without salts nor bile salts (▲);modified liquid YPD medium without bile salts (■) and ILM medium (+).

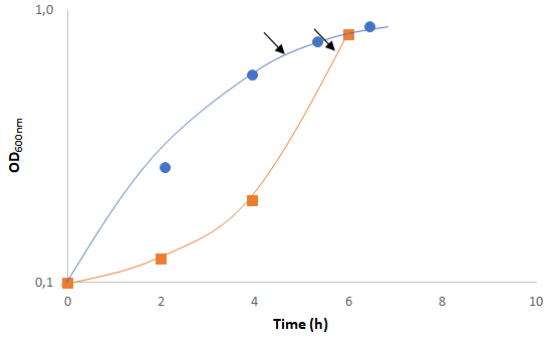

Figure3: S. cerevisiae (●) and *S. boulardii* (■) growth curve in ILM medium. The arrow indicates the time point and OD_{600nm} (0.8, approximately) at which RNA extraction was done.

It was concluded that *S. boulardii* is sensitive to salt stress, especially to bile salts, when compared to *S. cerevisiae*. However this sensitivity does not necessarily correlate to cell survival, since *S. boulardii* has been shown to be haloduric, meaning that it can survive at high salt concentration, but is unable to grow. [14]

Global gene expression analysis

Once the growth conditions were optimized, the transcriptome of *S. cerevisiae* cells was compared to that of *S. boulardii* cells cultivated for about 6 hours in ILM medium. ILM medium poses challenging conditions for both yeasts, such as high salt stress (especially bile salt stress), heat shock (especially for *S. cerevisiae*), enzyme activity (lysozyme, α-amylase, trypsin, chymotrypsin and lipase) and neutral pH.

Overall 577 and 523 genes were found to display higher and lower expression, respectively, in *S. boulardii*, when compared to *S. cerevisiae*. The list of genes was clustered according to functional categories, leading to the identification of biological functions that appear to be up- or down-regulated in *S. boulardii*, when compared to *S. cerevisiae* in the studied environment. Overall, the most overexpressed categories in *S. cerevisiae* were "Unknown function" and "Carbohydrate and energy metabolism", while the most up-regulated categories in *S. boulardii* were "RNA metabolism and translation" and "Nitrogen and amino acid metabolism". Smaller differences were also observed in the "Drug resistance", "Oxidative stress" and "Protein degradation" functions which were found to be over-represented within the *S. cerevisiae* over-expressed genes; and in "Probiotic activity", "Lipid metabolism" and "Osmotic stress", functions which were found to be over-represented among the *S. boulardii* over-expressed genes (*Figure 4*).

General stress response

There is a large set of genes whose expression changes in response to all stress conditions, irrespectively of the specific source of stress. An overall look at the genes whose expression is up-regulated in *S. cerevisiae* when compared to *S. boulardii* appears to suggest that *S. cerevisiae* is feeling a lot more stress than *S. boulardii*, especially in terms of heat shock and oxidative stress.

Indeed, many genes related to glycolysis (*GPM1*, *GPM2*, *MIG1*, *TYE7*, *ERR1*, *ERR2*, *ENO1*, *HXK1*, *GLK1*), pentose phosphate pathway and fructose and mannose metabolism (*PGI1*, *FBP1*, *FBA1*, *TDH2*, *TDH3*, *TDH1*, *ADH4*, *PGK1*, *GND1*, *TKL2*, *SOL4*, *XYL2*, *NQM1*, *DFS1*, *SOR2* and *TPI1*), TCA cycle (*PCK1*, *CIT3*, *SDH2*, *SDH3*, *SHH3*, *SHH4*, *LSC2* and *GRE2*), glyoxylate cycle (*CIT3*, *FDH1* and *YPL113C*), pyruvate (*HSP31*, *GLO4*, *CYB2*, *HSP32*, *SNO4*, *PDC6*, *PCK1* and *CDC19*) and glycogen (*GLG1*, *GSY2*, *GLC3*, *GPH1*, *GAC1*, *IGD1*, *SGA1* and *PIG2*) metabolism were found to be over-expressed in *S. cerevisiae*, when compared with *S. boulardii*, in cells grown in intestinal tract like medium. Also, genes required for mitochondrial respiration (*CYC1*, *COQ4*, *COQ9*, *COX5B*, *CYC7*, *QCR9*, *SDH2*, *SDH3* and *CYB2*), used as an energy source, are up-regulated in *S. cerevisiae* when compared with *S. boulardii*.

Significantly, genes involved in response to oxidative stress in mitochondria, in detoxifying reactive oxygen species and in general response to oxidative stress display up-regulated expression in *S. cerevisiae* when compared with *S. boulardii*. These include glutathione metabolism genes (*GTT1*, *GTT2*, *GPX1*, *IDP3*, *GND2*, *ECM4*, *GLO4* and *GRX2*), which represent a good part of the observed oxidative stress response, but also those encoding antioxidant enzymes (*GAD1*, *ALD2*, *GRE2*, *ALD3* and *SOD1*) and metabolic enzymes required for the

production of reducing power, mainly through the regeneration of NADPH. Under stress, yeast cells try to maintain their internal redox potential by synthesizing and recycling redox buffer molecules like glutathione, or more directly by detoxifying reactive oxygen species, through the activity of superoxide dismutases and catalases. As it has already been said, pentose phosphate pathway, which is up-regulated, regenerates NADPH, which may help restore NADPH reducing equivalents for combating oxidative stress. Interestingly, oxidative stress genes are significantly more up-regulated in *S. cerevisiae* when compared with *S. boulardii*, when grown in the intestinal tract like medium, than in general stress response [15], which suggests that the intestinal tract like medium used is perceived by *S. cerevisiae*, but not by *S. boulardii*, as a source of oxidative stress.

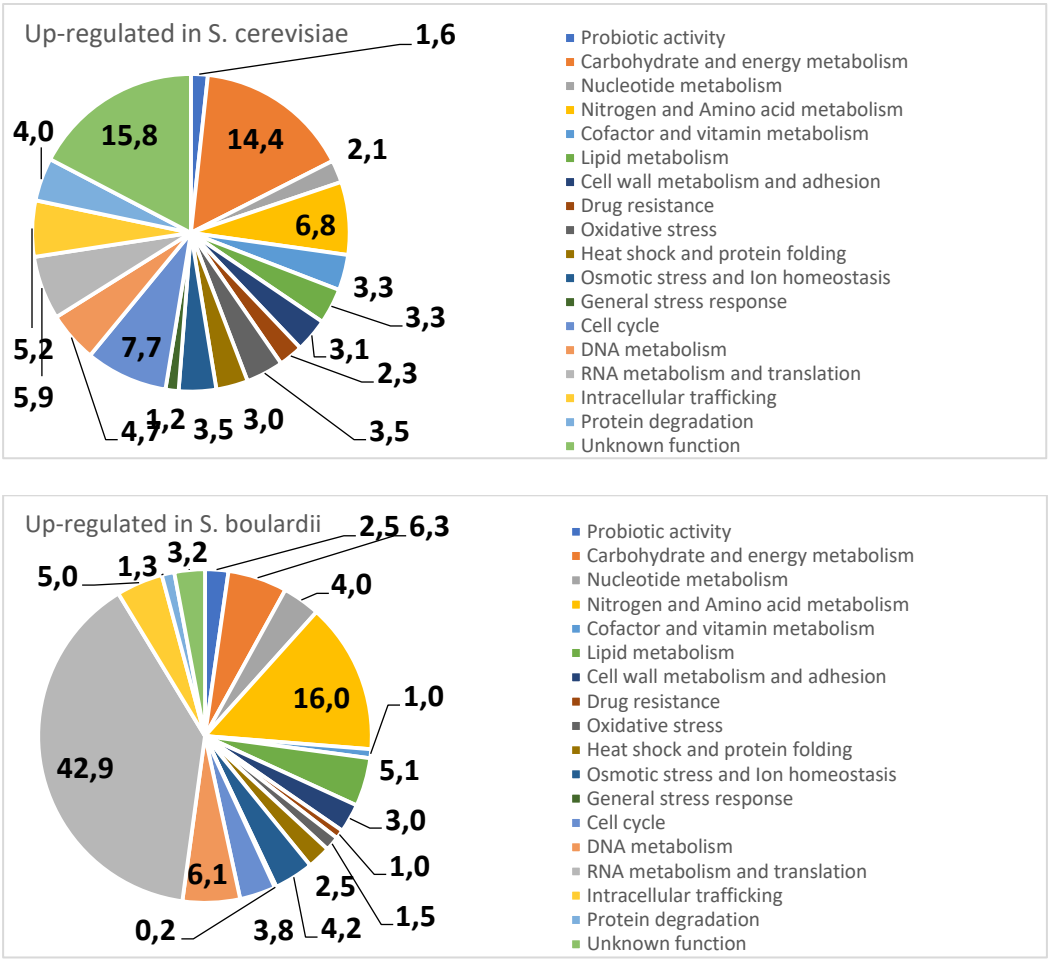

Figure 4: Main functional categories found to be over-expressed in S. cerevisiae or in S. boulardii.

Another feature of the general stress response, the expression of genes related to protein degradation, was also found to be up-regulated in *S. cerevisiae*, when compared with *S. boulardii*. Degradation is necessary for proteins that cannot be refolded or that are damaged or denatured in order to avoid their aggregation. Targeting of proteins for turnover can be done by ubiquitination (ubiquitin ligation and conjugation, polyubiquitin and deubiquitinating enzyme). Ubiquitination followed by autophagy, vacuolar protein degradation or proteasomal degradation may accelerate the cells ability to change its protein internal repertory in response to stress conditions.

Additionally, the down-regulation of ribosome proteins, RNA metabolism, translation and amino acid synthesis, which is a significant part of the general stress response, is clearly observed in *S. cerevisiae*, when compared to *S. boulardii* in the selected growth medium. Under stress, it is believed that these processes are down-regulated, probably because these are very energy consuming and their repression saves energy that can be relocated to synthesis of molecular chaperones, protein degradation and other mechanism involved in stress response. Also, amino acid synthesis, including that of phenylalanine, tyrosine, tryptophan, histidine, arginine, proline, glycine, serine, threonine, valine, leucine, isoleucine, lysine, cysteine and methionine,

was found to be down-regulated in *S. cerevisiae* when compared with *S. boulardii*. This is expected under stress, since protein synthesis is mostly repressed, while degradation of unfolded proteins is up-regulated [16–18, 19,20,21].

Response to Heat shock stress

Heat shock causes partial or total denaturation of proteins (and other molecules) and their consequent aggregation or separation of complexes. In response to heat stress, yeast responds with the so-called heat shock response (HSR) very rapidly (within minutes), which is controlled by the heat shock transcription factor Hsf1p.

Overall, heat shock gene expression was found to be up-regulated in *S. cerevisiae* when compared with *S. boulardii*. Indeed, the expression of *SSA4*, *HSP82*, *HSP104*, *SSA3*, *HSP30*, *HSP12* and *HSP26* genes, among other genes encoding chaperones that help in protein conformation, maintenance or refolding, was found to be over-expressed in *S. cerevisiae*, which is consistent with the notion that at 37ºC *S. cerevisiae* is suffering from heat shock, whereas *S. boulardii* is not. Synthesis of these chaperones and their activity consume lots of energy through ATP hydrolysis, which decreases ATP levels, affects central energy metabolism and leads to an increase in (or/and relocation of) energy production. Considering this effect, it is consistent to observe that together with the over-expression of HSP genes, carbohydrate and energy metabolism genes are also up-regulated in *S. cerevisiae*, when compared to *S. boulardii*, to an extent that appears far greater than the typical general response stress.

Response to hydrolytic enzymes

The selected medium contained the following hydrolytic enzymes, typically found in the intestinal tract: lysozyme, α-amylase, trypsin and lipase.

Lysozyme is an antimicrobial peptide that hydrolyzes β-linkage between the N-acetylglucosamine and N-acetylmuramic of peptidoglycan present in gram-positive bacteria cell wall. However, the yeast cell wall is not composed by peptidoglycan, but rather by mannoproteins, fibrous β(1→3) and β(1→6) glucans and chitin. Hence, lysozyme is unable to lyse yeast cells in the same way it does to gram-positive bacteria cells.

α-amylase is an enzyme that hydrolyses starch and glycogen into glucose and maltose. This includes hydrolysis of α(1-4) glycosidic bonds of exopolysaccharides present in extracellular polymeric substance secreted by bacteria to form biofilms. Hence, α-amylase is capable of degrading bacterial biofilms. Since yeast biofilms rely on a polymeric extracellular matrix composed by polysaccharides, it is possible that α-amylase might also be able to disrupt yeast biofilms. Since, in this thesis, biofilms were not analyzed in an intestinal simulated media, this disruption could not be evaluated. However, this enzyme may also be able to degrade chitin, when exposed. Usually, chitin is well protected by the rest of the cell wall above it, but during stress conditions it is possible that the protection offered by the cell wall is less effective and the enzyme might be able to affect chitin.

Trypsin is a serine protease that hydrolysis peptide chains in their lysine or arginine residues. Hence, this enzyme can possibly affect adhesion and flocculation proteins present in the surface of the yeast cell wall. Trypsin and α-chymotrypsin are tolerated by many cell types and, more specifically, do not seem to lyse or significantly affect *S. cerevisiae*'s cell wall even when used at the same time as glucanases. Although they do not seem to significantly affect cell viability, trypsin may still have an effect in external proteins, including those involved in functions such as flocculation, adhesion, antigen activation, recognition of mating partners, biofilm formation, pseudohyphal growth, iron retainment, uptake of sterols and cell wall hydrophobicity, synthesis and repair.

Lipase is an enzyme that hydrolyzes lipids, normally in their glycerol backbone. Yeast plasma membrane is mostly composed by lipids that are potential targets for lipases, it is protected by the cell wall. However, during stress conditions it is possible that the protection offered by the cell wall is less effective and this enzyme might affect the different lipids that compose the plasma membrane (mostly, phospholipids or glycerophospholipids). Additionally, bile salts as detergent like molecules that can also affect the membrane lipid bilayer of yeasts. Since the membrane is involved in Na^+, K^+ and pH homeostasis, nutrient and enzyme transport, mating, signal transduction, cytoskeleton anchoring, drug efflux, stress response, adhesion molecules and antigen activation, the presence of both lipase and bile salts in the GI tract medium can have strong repercussions at the level of plasma membrane related processes.

Altogether, the external cell wall proteins can be affected by both trypsin and α-chymotrypsin and, if the cell wall is disturbed during stress conditions and its protective effect is decreased, chitin might be affected by lysozyme and α-amylase and plasma membrane lipids might be affected by lipase and bile salts. Thus, depending on their sensitivity to these enzymes, *S. boulardii* and *S.*

cerevisiae may require specific changes in the transcription profile of cell wall and lipid metabolism related genes to adjust to them.

Indeed, lipid metabolism gene expression was found to be very different in *S. cerevisiae*, when compared to *S. boulardii*. Sphingolipids and fatty acid degradation were up-regulated, while ergosterol and glycerophospholipids metabolism were down-regulated in *S. cerevisiae*, when compared to *S. boulardii* (*Figure 5*). Glycerophospholipids and ergosterol biosynthesis up-regulation and fatty acid degradation down-regulation in *S. boulardii* when compared to *S. cerevisiae* might be deployed to fight membrane stress due to lipase and bile acid effects, suggesting that *S. boulardii* is more sensitive to the effect of these stress agents than *S. cerevisiae*. This observation is consistent with the higher sensitivity exhibited by *S. boulardii* to bile salt concentration.

On the other hand, up-regulation of sphingolipids in *S. cerevisiae* can be due to a response to heat shock. Indeed, inositol-phosphate, ceramide and phytosphingosine biosynthesis are all mostly up-regulated in *S. cerevisiae* when compared with *S. boulardii*.

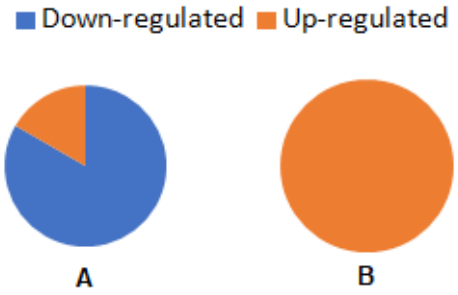

Figure 5: Comparison of up and down-regulated genes in S. boulardii when compared with S. cerevisiae for A: sphingolipids synthesis and fatty acid degradation and B: ergosterol and glycerophospholipids synthesis

Salt stress response

Osmotic stress may be balanced by adjusting the intracellular concentration of both osmolytes, such as glycerol or trehalose [16–18] , and of small ions such as Na^+, K^+ and H^+. Interestingly, most of the osmotic stress related genes showed higher expression in *S. cerevisiae*, when compared with *S. boulardii*. Indeed, the genes involved in the first steps of glycerol biosynthesis (*GPD1* and *GPD2*) were found to be up-regulated in *S. cerevisiae* when compared with *S. boulardii*. Glycerol synthesis and intracellular accumulation is increased during osmotic stress in order to adjust osmotic pressure along the cell membrane. By accumulating intracellularly, glycerol increases the uptake of water and consequently the cell swells to a certain size. This helps to fight water deficit and cell shrinkage and helps the yeast to restart growth. Additionally, glycerol is able to preserve the plasma membrane integrity and stabilize proteins, thus being a stress protectant, just like trehalose. Indeed, just like with trehalose, glycerol can be used as a compatible solute to counterbalance osmotic pressure. [14,16–18,20,22,19]

Altogether, the obtained results suggest that *S. cerevisiae* appears to be feeling more osmotic stress under ILM medium. Since apparently *S. cerevisiae* was found to grow better than *S. boulardii* in the presence of high salt concentrations, it is reasonable to ability to respond to osmotic stress that enables *S. cerevisiae* to survive in such conditions, when *S. boulardii* cannot.

Overexpressed probiotic functions

Only one potential anti-toxin gene was found to be overexpressed in *S. boulardii*, when compared to *S. cerevisiae*: *VAS1*. Vas1p is a 120 kDA protein that has been toxicity inhibitionof *Vibrio cholerae*'s toxin.

Key genes required for polyamine biosynthesis were also found to be up-regulated in *S. boulardii*, when compared to *S. cerevisiae*. Many digestive enzymes (sucrase-isomaltase, maltase-glucoamylase, lactase-phlorizin hydrolase, alanine aminopeptidase and alkaline phosphatase) and nutrient transporters (sodium-glucose transport proteins) activity may be induced by polyamines secreted by *S. boulardii*.[1,5,23,24] Thus, it appears very significant that, in the ILM medium used herein, the expression of *SPE2* and *SPE3*, encoding the enzymes that catalyse the steps for spermidine biosynthesis is higher in *S. boulardii*.

It has been shown that *S. boulardii* produces moderate amounts of acetic acid, whereas *S. cerevisiae* produces insignificant amounts. It has also been shown that acetic acid has significant antimicrobial properties. In this context, the expression of genes encoding acetic acid generating enzymes, namely *ALD5*, *MET17*, SFC1, *ACS2*, was found to be up-regulated in *S. boulardii* when compared with *S. cerevisiae*, while two other, *ALD2* and *ALD3*, were down-regulated. Down-regulation of *ALD2* and *ALD3* is not very significant, since these two genes encode aldehyde dehydrogenase that may participate unspecifically in acetate synthesis. *ALD5,* however, encodes a acetaldehyde dehydrogenase specifically involved in acetate synthesis, while Sfc1p is a mitochondrial succinate-fumarate transporter required for acetate utilization. Met17p is an O-acetyl homoserine-O-acetyl serine sulfhydrylase required for methionine

and cysteine biosynthesis that releases acetate in the process. Even though Acs2p, an acetyl-coA synthetase, degrades acetate, it is also involved in the production of propionate, another SCFA. It is thus reasonable to hypothesize that the overexpression of these SFCA biosynthesis related genes may contribute to the probiotic phenotype of *S. boulardii*.

Among the cell wall and adhesion related genes that have been predicted to play an eventual role in *S. boulardii* probiotic effect, only *HKR1* and *YPS1* genes display increased expression levels in *S. boulardii*, when compared to *S. cerevisiae*. Yps1p is a GPI-anchored plasma membrane attached aspartic protease with a role in cell wall integrity and adhesion and Hrk1p is a mucin family member that functions as an osmosensor in the HOG pathway, with a predicted role as an adhesin like protein. The over-expression of both these genes suggests *S. boulardii* may display increased adhesiveness then *S. cerevisiae*. On the contrary, however, the flocculin encoding genes *FLO5* and *FLO9* were found to have higher expression levels in *S. cerevisiae*, which suggests the opposite.

S. boulardii has been proposed to be capable of adhering to intestinal mucus membrane and avoiding adhesion of other pathogens flowing by to the intestine. Indeed, it has been shown that *S. boulardii* is able to colonize the intestine of gnotobiotic mice after single administration. However, other studies contradict this notion and say that *S. boulardii* does not strongly adhere to intestinal epithelial cells and is quickly removed from the gastrointestinal system in healthy individuals. This may mean that although *S. boulardii* can colonize the intestine, competition with intestinal microbiome limits it, unless the microbiome is depleted, such as under antibiotic therapy.[25]

In order to evaluate if *S. boulardii* is indeed able to adhere to epithelial intestinal cells, and if it is more able to do so than the non-probiotic yeast S. cerevisiae, adhesion assays were conducted for both species, against FHs 74 Int small intestine cell line (ATCC® CCL-241™). *S. boulardii* not only was able to successfully adhere to epithelial intestinal cells, but it did so exceedingly better than *S. cerevisiae*. From three replicate essays, *S. boulardii* had an average of 58% of adhesion, whereas *S. cerevisiae* had an average of 30% of adhesion, almost half of *S. boulardii*'s adhesion (*Figure 6*).

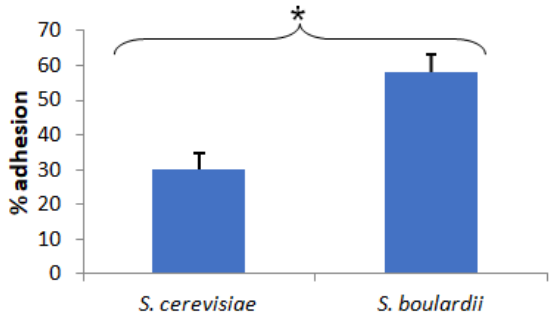

Figure 6: Adhesion percentage of S. boulardii and S. cerevisiae to FHs 74 Int small intestine cell line. Statistical analysis of the results was performed using analysis of variance and differences were considered significant for $p < 0.05$.

Altogether, these results support the notion that *S. boulardii* has a stronger capacity to adhere to intestinal epithelial cells than *S. cerevisiae*, a phenotype that may underlie its increased probiotic capacity.

ProBioYeastract platform's new computational functionality

Yeastract+ (Yeast Search for Transcriptional Regulators And Consensus Tracking +) is a comparative genomics platform that allows transcriptional regulatory networks analysis for certain yeast species. ProBioYeastract focus on probiotic species, currently including *S. boulardii* biocodex and unique28 strains. Each database is a curated repository of published transcriptional associations used to predict and visualize gene and genomic regulation through comparative genomics, considering orthologous regulatory associations from other yeast. [26–31]

In this thesis, a new functionality was implemented for the ProBioYeastract platform. This new tool finds transcription factors that specifically regulate the inputted genes. This is done by comparison of transcription factor's binding sites (TFBS) within the promoter's sequences and computation of which ones match. The species from which the TFBS are considered are selected by the user. At the moment, there is no such information available for *S. boulardii* in the literature, so only *S. cerevisiae* data can be used. Promoter's sequence is both from *S. cerevisiae* and *S. boulardii*. The matches are then separated in three groups: when they are found in both *S. cerevisiae* and *S. boulardii* or when they are found in only one of them. With this information a table is constructed with each group as a column and each transcription factor as a line. In each cell are the input genes or their homologs in the remaining species, regulated by that line's transcription factor.

It is now possible to predict which input genes a certain transcription factor regulates. However, we do not know if that transcription factor regulates only a small group of genes (as in, mostly the input genes and few others) or if it regulates a huge amount of genes besides the input genes. In the first case, the transcription factor would be highly interesting, since it would suggest a specific regulation, while in the latter case a general non-specific regulation is suggested. To analyze how specific the regulation is, a p-value is calculated for each pair of transcription factor and respective regulated genes through the hypergeometric distribution.[32] The hypergeometric distribution takes as input the population size (number of genes the species has), sample size (number of genes the user has inputted), number of successes in a given (species) population (number of genes whose promoter contains at least one binding site of a given transcription factor) and number of successes in the sample (number of inputted genes whose promoter of that species contains at least one binding site of a given transcription factor). This p-value is then displayed in the table for each transcription factor – target genes cluster; in other words, there is one p-value for each line and column that is independent of other p-values in other lines.

However, some p-values can be very similar to each other and it would be difficult to translate that value meaningfully. In order to create a bigger scale and separate the p-values from each other, the Bonferroni correction was used.[33] This correction relates p-values from different lines by taking into account the size of the table. This is done by dividing the p-value by the number of lines (or transcription factors).

Hence, we finally obtain a table with each species (or intersection of them) as a column, each transcription factor as a line and each cell with the input genes regulated by that line's transcription factor for that column's species or the respective corrected p-value. A pseudocode version of this new functionality's code can be found below (7).

Figure7: Pseudocode version of this new functionality's code.

User input:
UserGenes <- *S. boulardii* biocodex's or unique28's orfs/genes
TFBSSpecies <- species of transcription factors binding sites
UserSpecies <- GetSpecies(**UserGenes**)
UserHomologousSpecies <- GetHomologousSpecies(**userSpecies**)
TFandTFBS <- GetTFBS(**TFBSSpecies**)
For each SP in (**UserSpecies** U **UserHomologousSpecies**)
PromotersPerSpecies <- GetPromoters(SP)
Matches <- GetMatches(TFand**TFBS, PromotersPerSpecies)**
Intersect <- Get**RegulatedGenes**PresentIn**AllSpecies(Matches)**
UniquePerSpecies <- Get**RegulatedGenes**PresentOnlyIn**OneSpecies(Matches)**
For each TF in **TF**andTFBS
Foreach SP in **UserSpecies** U **UserHomologousSpecies**
P-value <- **CalculateP-Value** (**TF, SP, Matches**)
CorrectedP-Value <- Use**BonferroniCorrection**OnP-Value (**P-value**)

Howsoever, it should be noted that getMatchesBetweenTFBSandPromoters is a very computation heavy function. In total, the database has 124 transcription factors with documented binding sites belonging to *S. cerevisiae*, which corresponds to 401 binding sites. These 401 binding sites may be compared with the promoter's sequences of each species (6823 and 5482 promoter sequences for *S. cerevisiae*, *S. boulardii* Biocodex, respectively). If we consider that each comparison takes 0.01 seconds, all matches would take about twenty hours to run. This waiting time is not acceptable for a web page to load and, in order to avoid this, all matches between transcription factors and database's set of genes, for all species, were pre-computed and a table was added to the database with information about transcription factors and how many promoters' they regulate (how many matches they are associated to) for each species. Nonetheless, matches for the input genes still have to be calculated each time someone uses the new functionality, since each time different genes are inputted. Considering that differential gene expression analysis through total RNA-sequencing can lead to thousands of differently expressed genes, this could take about one hour and fifteen minutes to run.

As an example, down-regulated genes in *S. boulardii* when compared with *S.* cerevisiae from osmotic stress and ion metabolism were used as input to find out relevant differences in regulation by transcription factors.

The ideal would be to find a transcription factor with very low specificity for genes regulated in *S. cerevisiae* and very high specificity for genes regulated only in *S. boulardii*. Even better if there are few or no genes regulated in both strains. However, such transcription factor was not clearly found.

The most relevant results found were for transcription factors Yap3p and Gcn4p. Yap3p regulates only one gene (*AST2*) in both strains, 4 uniquely in *S. cerevisiae* (*GPD1*, *GRE2*, *VHS3* and *YML131W*) and 4 uniquely in *S. boulardii* (*CCC2*, *ATX2*, *PMA2* and *VMA22*). Gpd1p is an enzyme involved in glycerol synthesis that is essential for growth under osmotic stress. Its expression is regulated by high-osmolarity glycerol response pathway. Vhs3p is the negative regulatory subunit of protein phosphatase 1 (Ppz1p) involved in cellular metal ion homeostasis. Ppz1p is involved in regulation of potassium transport, which affects osmotic stability, cell cycle progression, and halotolerance. *GRE2* is a stress induced gene (osmotic, ionic, oxidative, heat shock and heavy metals) regulated by the HOG pathway. *YML131W* gene is induced by stresses including osmotic shock, DNA damaging agents, and other chemicals. Ccc2p is a Cu^{+2}-transporting P-type ATPase required for export of copper. Atx2p is involved in manganese homeostasis. Pma2p is a plasma membrane H^{+}-ATPase involved in pumping protons out of the cell and, hence, regulating cytoplasmic pH and plasma membrane potential. Vma22p is a peripheral membrane protein required for vacuolar H^{+}-ATPase (V-ATPase) function. All of these genes are down-regulated in *S. boulardii* when compared with *S. cerevisiae*. As we can see Yap3p not only regulates different genes in *S. cerevisiae* and *S. boulardii*, but up-regulates them in *S. cerevisiae*.

Yap3p is an AP-1 type transcription factor that belongs to the YAP protein family. This protein family is composed by stress response and metabolism control pathways transcription activators with similar yet identical DNA binding specificities. Normally they activate transcription from promoters containing a Yap recognition element. Although its regulatory targets aren't well known, Yap3p seems to be involved in multidrug resistance and unfolded protein response and possibly also in oxidative, hyperosmotic and alkaline pH stress tolerance. [34–37]

As it was seen, *GPD1*, *GRE2*, *VHS3* and *YML131W* genes are regulated by Yap3p uniquely in *S. cerevisiae*. This means that Yap3p lost these genes as targets in *S. boulardii*. This does not contradict the transcriptomics results, since Yap3p is a transcription factor that activates genes expression and these genes were shown to be down-regulated in *S. boulardii*. Altogether, this analysis suggests that the predicted loss of Yap3p targets in *S. boulardii*, when compared to *S. cerevisiae*, may explain its decreased osmotic stress resistance.

Gcn4p is predicted to regulate 7 genes (*BIT2*, *GPD1*, *MEP2*, *ZRT1*, *MPC3*, *SIP18* and *FRE4*) in both strains, 5 uniquely in *S. cerevisiae* (*FTR1*, *CTR3*, *GRE2*, *VHS3* and *YML131W*) and 5 uniquely in *S. boulardii* (*AST2*, *CCC2*, *ATX2*, *PMA2* and *VMA22*). Ftr1p and Ctr3p are high affinity iron and copper permease, respectively, involved in the transport of iron and copper, respectively, across the plasma membrane. All of the genes uniquely regulated in *S. cerevisiae* are up-regulated in *S. cerevisiae* when compared to *S. boulardii*. Ast2p is involved in targeting of plasma membrane H^{+}-ATPase (Pma1p) to the plasma membrane. All of the genes uniquely regulated in *S. boulardii* are down-regulated in *S. boulardii* when compared with *S. cerevisiae*. As we can see Gcn4p not only regulates different genes in *S. cerevisiae* and *S. boulardii*, but up-regulates them in *S. cerevisiae*, whereas it down-regulates them in *S. boulardii*. It is interesting to note that both Yap3p and Gcn4p regulate similar genes uniquely in *S. boulardii*.

Gcn4p is an amino acid synthesis' genes transcriptional activator, especially during amino acid starvation. It has been shown that this transcription factor is induced in other conditions besides amino acid starvation: purine starvation, glucose limitation, growth on ethanol, high salinity medium and treatment with methyl methanesulfonate or rapamycin.

Indeed, Gcn4p is crucial for high salt stress survival due to Hal1p transcriptional activation. Hal1p is crucial for maintaining Na^{+} and K^{+} ion homeostasis and has a CRE in its promoter. This protein is induced both during osmotic and salt stress. A *GCN4* mutant shows sensitivity to elevated potassium or sodium concentrations in the medium. Both Gcn4p and Sko1p compete antagonistically to bind to the CRE site. While Sko1p acts as a transcriptional repressor under normal growth conditions (causing down-regulation of *HAL1*), Gcn4p acts as a transcriptional activator during hyperosmotic stress (causing up-regulation of *HAL1*). However, only modest *GCN4* induction aids in salt stress survival, since exaggerated overexpression of these gene impedes growth. HOG pathway is responsible for stopping Sko1p repression. This repression is however not enough for increased *HAL1* expression, since an activator is necessary. During salt stress, Gcn4p represses the expression of translation factors and also leads to reduced protein synthesis.

It would be interesting to know if *AST2*, *CCC2*, *ATX2*, *PMA2* and *VMA22* gene expression is regulated at CRE motifs. In case they are, it may be possible that Sko1p was still repressing these genes expression in *S. boulardii* or that Gcn4p was not able to bind properly to the CRE site. The same could happen for genes *FTR1*, *CTR3*, *GRE2*, *VHS3*

and *YML131W.* In case they are regulated at CRE motifs, it may be possible that Sko1p was no longer repressing these genes expression in *S. cerevisiae* and that Gcn4p was able to bind properly to the CRE site. This would make sense since Sko1p transcription factor is up-regulated in *S. boulardii* when compared to *S. cerevisiae* and, hence, more available to repress gene expression in *S. boulardii*. Gcn4p is not differently expressed in *S. boulardii* and *S. cerevisiae*. This is interesting since it might explain why *S. cerevisiae* has an apparently higher ability to respond to osmotic stress than *S. boulardii* to survive in ILM medium.

As it was seen, *FTR1*, *CTR3*, *GRE2*, *VHS3* and *YML131W* genes are regulated by Gcn4p uniquely in *S. cerevisiae*. This means that Gcn4p lost these genes as targets in *S. boulardii*. Altogether, this analysis suggests that the predicted loss of Gcn4p targets in *S. boulardii*, when compared to *S. cerevisiae*, may explain its decreased osmotic stress resistance. It would, thus, be interesting to evaluate if by increasing the expression of these lost target genes in *S. boulardii*, it would be possible to increase its tolerance to osmotic stress and promote an even better probiotic performance of this species.

Conclusions and Perspectives

Clues for *S. boulardii* superior probiotic properties were found. This strain has a better adaptation to human temperature and perceives less heat and oxidative stress in an ILM medium. *S. boulardii* also showed up-regulation of certain genes associated with probiotic activity when compared with *S. cerevisiae*. This is the case of synthesis of Vas1p 120 kDa protein, polyamine (spermidine) and acetate and the higher expression of certain cell wall and adhesion related genes Cell wall's adhesion proteins up-regulation in *S. boulardii* was confirmed by higher adherence to intestinal epithelium in adhesion assays.

In this thesis, a new functionality was implemented for the ProBioYeastract platform. This tool allows a global evaluation of promoter regions in the ProBioYeastract database.

Analysis of protein Vas1p, spermidine and acetate concentration in the growth medium would also be very interesting in order to corroborate the transcriptional results that appear to show better probiotic properties for *S. boulardii.* In case of high quantities of acetate concentration of the medium, further tests of antimicrobial potential would be highly appealing.

It would also be interesting to assess if the different adhesion capacity of the two strains relies on the increased expression of *HKR1* and *YPS1* genes, observed in planktonic cultivation in intestinal tract like medium, or eventually of other adhesion related genes, whose expression can vary in biofilm formation assays or in the presence of intestinal epithelial cells.

Since the ProBioYeastract new functionality would still take about one hour and fifteen minutes to run using the whole set of differently expressed genes from the transcriptional analysis, a way to speed up the use of this function is still necessary. For that, the matches for each gene could be calculated and stored individually, in order to be easily accessible in the database. It would also be highly interesting to analyze Yap3p and Gcn4p transcription factor activity upon the expression of genes predicted to be differentially regulated in *S. boulardii* versus *S. cerevisiae*, under hyperosmotic and/or salt stress, and compare the results with the obtained with this bioinformatic tool. On the other hand, it would also be interesting to evaluate if an increase in the expression of lost target gene in *S. boulardii*, would also increase its tolerance to osmotic stress and promote an even better probiotic performance of this species.

Acknowledgments

This work was financially supported by "Fundação para a Ciência e a Tecnologia" (FCT) (Contract PTDC/BII-BIO/28216/2017) and is also integrated in the project LISBOA-01-0145-FEDER-022231 – the Biodata.pt Research Infrastructure, supported by Lisboa Portugal Regional Operational Programme (Lisboa 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). Funding received by Instituto de Bioengenharia e Biociências (iBB) from FCT (grant UID/BIO/04565/2013) and Programa Operacional Regional de Lisboa 2020 (project no. 007317) is also acknowledged.

References

1. McFarland L V. *Common Organisms and Probiotics: Saccharomyces Boulardii*. In: The Microbiota in Gastrointestinal Pathophysiology. Implications for Human Health, Prebiotics, Probiotics, and Dysbiosis. Academic Press, 2017:145-164.
2. Schrezenmeir J, de Vrese M. Probiotics, prebiotics, and synbiotics—approaching a definition. *Am J Clin Nutr*. 2001;73(2):361S-4S.
3. Arain MA, Li Y, Rajput IR, Baloch DM, Sun Y. Oral administration of Saccharomyces boulardii alters duodenal morphology, enzymatic activity and cytokine production response in broiler chickens . *Anim Sci J*. 2016;88(8):1204-1211.
4. Fietto JL., Araújo RS, Valadão FN, et al. Molecular and physiological comparisons between Saccharomyces cerevisiae and Saccharomyces boulardii . *Can J Microbiol*. 2004;50(8):615-621.
5. Plaza-Diaz J, Ruiz-Ojeda FJ, Gil-Campos M, Gil A. Mechanisms of Action of Probiotics. *Adv Nutr*. 2019;10(1):S49-S66.
6. Bajaj BK, Claes IJJ, Lebeer S. Functional mechanisms of probiotics. *J Microbiol Biotechnol Food Sci*. 2015;04(04):321-327.

7. Khatri I, Tomar R, Ganesan K, Prasad GS, Subramanian S. Complete genome sequence and comparative genomics of the probiotic yeast Saccharomyces boulardii. *Sci Rep*. 2017;7(1):1-13.
8. Qamar A, Aboudola S, Warny M, et al. Saccharomyces boulardii Stimulates Intestinal Immunoglobulin A Immune Response to Clostridium difficile Toxin A in Mice. 2001;69(4):2762-2765.
9. Castagliuolo I, Riegler MF, Valenick L, Mont JTLA. Saccharomyces boulardii Protease Inhibits the Effects of Clostridium difficile Toxins A and B in Human Colonic Mucosa. *Infect Immun*. 1999;67(1):302-307.
10. Khatri I, Akhtar A, Kaur K, Tomar R, Prasad GS. Gleaning evolutionary insights from the genome sequence of a probiotic yeast Saccharomyces boulardii. *Gut Pathog*. 2013:1-8.
11. Pontier-bres R, Rampal P, Peyron J, Munro P, Lemichez E, Czerucka D. The Saccharomyces boulardii CNCM I-745 Strain Shows Protective Effects against the B. anthracis LT Toxin. *Toxins (Basel).* 2015:4455-4467.
12. Polzin S, Huber C, Eylert E, Elsenhans I, Eisenreich W, Schmidt H. Growth Media Simulating Ileal and Colonic Environments Affect the Intracellular Proteome and Carbon Fluxes of Enterohemorrhagic Escherichia coli O157 : H7 Strain EDL933. *Applied and Environmental Microbiology*. 2013;79(12):3703-3715.
13. Cells E, Bielaszewska M, Greune L, et al. Anaerobic Conditions Promote Expression of Sfp Fimbriae and Adherence of Sorbitol-Fermenting Enterohemorrhagic Escherichia coli O157 : NM to Human Intestinal. 2008;74(4):1087-1093.
14. Avila-reyes SV. Effect of salt and sugar osmotic stress on the viability and morphology of Saccharomyces boulardii. I*nternational Journal of Environment, Agriculture and Biotechnology*. 2016;1(3)
15. Causton HC, Ren B, Koh SS, et al. Remodeling of Yeast Genome Expression in Response to Environmental Changes. *Mol Biol Cell.* 2001;12(2):323-37.
16. Melamed D, Pnueli L, Arava Y. Yeast translational response to high salinity : Global analysis reveals regulation at multiple levels. *RNA*. 2008;14(7):1337-51.
17. Yale J, Bohnert HJ. Transcript Expression in Saccharomyces cerevisiae at High Salinity. *J Biol Chem.* 2001;276(19):15996-6007.
18. Posas F, Chambers JR, Heyman JA, Hoeffler JP, Nadal E De. The Transcriptional Response of Yeast to Saline Stress. *J Biol Chem.* 2000 Jun 9;275(23):17249-55.
19. Albertyn J, Hohmann S, Thevelein JM, Prior BA. GPD1 , Which Encodes Glycerol-3-Phosphate Dehydrogenase , Is Essential for Growth under Osmotic Stress in Saccharomyces cerevisiae , and Its Expression Is Regulated by the High-Osmolarity Glycerol Response Pathway. *Mol Cell Biol.* 1994;14(6):4135-44.
20. Tekarslan-sahin SH, Alkim C, Sezgin T. Physiological and transcriptomic analysis of a salt- resistant Saccharomyces cerevisiae mutant obtained by evolutionary engineering. *Bosn J Basic Med Sci.* 2018 Feb 20;18(1):55-65.
21. Rep M, Krantz M, Thevelein JM, Hohmann S. The Transcriptional Response of Saccharomyces cerevisiae to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes. *J Biol Chem.* 2000;275(12):8290-300.
22. Hohmannn S, Prior BA. Role of trehalose in survival of Saccharomyces cerevisiae under osmotic stress. *Microbiology.* 1998;144:671-80.
23. Moré MI, Vandenplas Y. Saccharomyces boulardii CNCM I-745 Improves Intestinal Enzyme Function: A Trophic Effects Review. *Clin Med Insights Gastroenterol*. 2018;11: 1179552217752679.
24. Gastroenterology P, Nutrition JB, Hepatology P. Saccharomyces boulardii Produces in Rat Small Intestine a Novel Protein Phosphatase that Inhibits Escherichia coli Endotoxin by dephosphorylation. *Pediatr Res.* 2006;60(1):24-9.
25. Edwards-ingram L, Gitsham P, Burton N, et al. Genotypic and Physiological Characterization of Saccharomyces boulardii , the Probiotic Strain of Saccharomyces cerevisiae. *Appl Environ Microbiol*. 2007;73(8):2458-2467.
26. Monteiro PT, Pais P, Costa C, Manna S, Sa-Correia I, Teixeira MC. The PathoYeastract database: An information system for the analysis of gene and genomic transcription regulation in pathogenic yeasts. *Nucleic Acids Res*. 2017;45:D597-D603.
27. Teixeira MC, Monteiro PT, Palma M, et al. YEASTRACT: An upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2018;46:D348-D353.
28. Teixeira MC. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2006;34:D446-D451. doi:10.1093/nar/gkj013
29. Monteiro PT, Teixeira MC, Mira NP, et al. YEASTRACT : providing a programmatic access to curated transcriptional regulatory associations in Saccharomyces cerevisiae through a web services interface. *Nucleic Acids Res*. 2011;39:D136-D140.
30. Teixeira MC, Monteiro PT, Guerreiro JF, et al. The YEASTRACT database : an upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae. *Nucleic Acids Research.* 2014;42:D161-1D66.
31. Monteiro PT, Mendes ND, Teixeira MC, et al. YEASTRACT-DISCOVERER : new tools to improve the analysis of transcriptional regulatory associations in Saccharomyces cerevisiae Pattern matching. *Nucleic Acids Research.* 2008;36:D132-13D6.
32. Wolfram MathWorld. Hypergeometric Distribution. Wolfram Alpha Computational Intelligence. http://mathworld.wolfram.com/HypergeometricDistribution.html. Accessed November 6, 2019.
33. Wolfram MathWorld. Bonferroni Correction. Wolfram Alpha Computational Intelligence. http://mathworld.wolfram.com/BonferroniCorrection.html. Accessed November 6, 2019.
34. Fernandes L, Rodrigues-pousada C, Struhl K. Yap, a Novel Family of Eight bZIP Proteins in *Saccharomyces cerevisiae* with Distinct Biological Functions. *Molecular and Cellular Biology* 1997;17(12):6982-6993.
35. Rodrigues-Pousada C, Devaux F, Caetano SM, et al. Yeast AP-1 like transcription factors (Yap) and stress response: A current overview. *Microb Cell*. 2019;6(6):267-285.
36. Taymaz-Nikerel H, Cankorur-Cetinkaya A, Kirdar B. Genome-wide transcriptional response of Saccharomyces cerevisiae to stress-induced perturbations. *Front Bioeng Biotechnol*. 2016;4:17
37. Xu C, Bailly-maitre B, Reed JC, Xu C, Bailly-maitre B, Reed JC. Endoplasmic reticulum stress : cell life and death decisions. *J Clin Invest.* 2005;115(10):2656–2664