TÉCNICO LISBOA

Functional analysis and evolution of *Zygosaccharomyces bailii* transcription factor Haa1, involved in acetic acid tolerance

Laura Raquel Guilherme Luzia

Thesis to obtain the Master of Science Degree in

Microbiology

Supervisors: Prof. Dr. Isabel Maria de Sá Correia Leite de Almeida

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

Members of the Committee: Dr. Paulo Jorge Moura Pinto da Costa Dias

December 2015

Acknowledgments

First of all, I would like to acknowledge my supervisor, Professor Isabel Sá-Correia for giving me the opportunity to join the Group of the Biological Sciences Research Group (BSRG), for her guidance, patience and brainstorming moments. May I sincerely thank my co-supervisor Dr. Margarida Palma for guiding me throughout this work during the last year, helping me to develop the thesis's experimental part and my analytical and writing scientific skills. I would also like to acknowledge PhD student Joana Guerreiro for introducing me in my first steps in the laboratory and Dr. Paulo Dias, who guided me in the bioinformatics part of this work.

The experimental and bioinformatics work was carried out within the Sá-Correia's BSRG, a research unit of the Institute for Bioengineering and Biosciences (iBB), Instituto Superior Técnico (IST), Universidade de Lisboa. I am greatly indebted to this research group for providing technical support, access to equipment, materials and other resources and also to the good working environment. A special thanks to my laboratory colleagues Ana Vila Santa and João Peça for making the laboratory days less stressful and for all the breaks.

Funding received by iBB-Institute for Bioengineering and Biosciences from FCT-Portuguese Foundation for Science and Technology (UID/BIO/04565/2013) and from Programa Operacional Regional de Lisboa 2020 (Project N. 007317) is acknowledged.

I would like to acknowledge Serviços de Acção Social, Universidade de Lisboa, for guaranteeing me accommodation during this year and also to my housemates and friends Ânia Ferreira, Cátia Dias, Daniela Cruz and Andreia Gomes for all support and relaxation moments.

Finally, I would like to express my gratitude to my parents and grandmother for the unconditional support and encouragement to do my best and for always believing in me. There are no words to describe all their efforts, care and love. I hope I will make them proud.

A special thanks to Tiago Pedreira for making me fall in love for Microbiology and for his patience and support during my entire academic journey. An acknowledgement is not enough to show how lucky I am for having him in my life.

Abstract

The transcription factor (TF) Haa1 was described as a key regulator in *Saccharomyces cerevisiae* adaptive response to acetic acid. The aim of the present work was the functional analysis of the *S. cerevisiae* Haa1 homologue in *Zygosaccharomyces bailii* strains CLIB213^{T} and IST302 (*ZbHAA1*). In this context, the *ZbHAA1* ORFs were expressed in *S. cerevisiae* BY4741 and in the derived deletion mutant *haa1*Δ under the control of *HAA1 S. cerevisiae* BY4741 promoter, as also the *HAA1* from *S. cerevisiae* BY4741 (*ScHAA1*) as an internal control. The expression of *ZbHAA1* ORFs analysed was found to lead to a faster adaptation to acetic acid by the complementation of the *haa1*Δ mutant phenotype or the improvement of tolerance of the parental strain. These results strongly suggest that these *Z. bailii* genes are functional homologues of *ScHAA1*. Also, it was found that the expression of the *ZbHAA1* sequences under study lead to a higher increase of yeast tolerance to acetic acid than the expression of *ScHAA1* sequence. Combined phylogenetic and synteny analyses allowed to confirm that the copper-responsive Ace1 TF is the *S. cerevisiae* Haa1 paralogue with origin in the Whole Genome Duplication event. A number of amino acid variations in Haa1 homologues were consistently observed in efficient fermentative *S. cerevisiae* strains. These are promising candidates for subsequent studies envisaging acetic acid tolerance improvement. However, the *Zb*Haa1 sequences from *Z. bailii* strains examined in this work appear to be the most promising candidates for further studies in the increase yeast robustness against acetic acid tolerance.

Keywords: Haa1, *Saccharomyces cerevisiae*, *Zygosaccharomyces bailii*, acetic acid tolerance, Whole Genome Duplication.

Resumo

O fator de transcrição (FT) Haa1 foi descrito como o principal agente regulador na resposta ao ácido acético em *Saccharomyces cerevisiae*. O objetivo deste trabalho consistiu na análise funcional do homólogo do Haa1 de *S. cerevisiae* nas estirpes de *Zygossacharomyces bailii* CLIB213^{T} e IST302 (*ZbHAA1*). Neste contexto, as ORFs *ZbHAA1* foram expressas em *S. cerevisiae* BY4741 e no mutante de deleção *haa1*Δ sob o controlo do promotor do *HAA1* de *S. cerevisiae* BY4741, bem como o *HAA1* de *S. cerevisiae* BY4741 (*ScHAA1*) enquanto controlo interno. A expressão das *ZbHAA1* ORFs levou a uma resposta mais eficiente ao ácido acético pela complementação do fenótipo de susceptibilidade do mutante *haa1*Δ e pelo aumento da tolerância ao ácido na estirpe parental, sugerindo os *ZbHAA1* testados como homólogos funcionais do *ScHAA1*. Simultaneamente, a expressão dos *ZbHAA1* levou a um aumento da tolerância ao ácido acético em leveduras em relação ao *ScHAA1*. Uma análise combinada envolvendo o estudo da filogenia e da sintenia dos homólogos Haa1 de *S. cerevisiae* confirmou que o FT Ace1 envolvido na resposta ao cobre é o seu parálogo com origem num evento de duplicação do genoma. Adicionalmente foram identificadas várias variações aminoacídicas no Haa1 de estirpes de *S. cerevisiae* anteriormente descritas como altamente fermentativas. Estas proteínas são atractivos candidatos para subsquentes estudos no aumento da tolerância ao ácido acético. No entanto, as sequências *Zb*Haa1 das estirpes de *Z. bailii* incluídas neste trabalho parecem ser os candidatos mais promissores para estudos posteriores no aumento da robustez de leveduras ao ácido acético.

Palavras-chave: Haa1, *Saccharomyces cerevisiae*, *Zygosaccharomyces bailii*, tolerância ao ácido acético, duplicação completa do genoma completo

Contents

List of figures

List of tables

List of abbreviations

ATP	Adenosine triphosphate
ABC	ATP-binding cassette
BLASTP	Basic local alignment search tool
BSRG	Biological Sciences Research Group
BLASTP	Protein-protein BLAST
C. glabrata	*Candida glabrata*
CIAP	Calf intestinal alkaline phosphatase
CuRD	Copper regulatory domain
DBD	DNA binding domain
D. bruxellensis	*Dekkera bruxellensis*
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
DOE	Department of Energy
EBI	European Bioinformatics Institute
EMBL	European Molecular Biology Laboratory
Euroscarf	European *Saccharomyces cerevisiae* archive for functional analysis
JGI	Joint Genome Institute
GOLD	Genomes Online Database
H. valbyensis	*Hanseniaspora valbyensis*
iBB	Institute for Bioengineering and Biosciences
IST	Instituto Superior Técnico
LBA	Luria Bertani amplicilin
MDR	Multidrug resistance
MFS	Major facilitator superfamily
N. dairenensis	*Naumovozyma dairenensis*
MMB	Mineral basal medium
mRNA	Messenger RNA
MUSCLE	Multiple sequence comparison by log-expectation
NAD(P)H	Nicotinamide adenine dinucleotide phosphate
N. castellii	*Naumovozyma castellii*
NCBI	National Center for Biotechnology Information

OD600nm	Optical density at 600nm
ORF	Open reading frame
PCR	Polymerase chain reaction
PDH	Pyruvate dehydrogenase
PDR	Pleiotropic drug resistance
PEG	Polyethylene glycol
pHi	Intracellular pH
PHYLIP	Phylogeny inference package
pKa	-log(Ka), where Ka is the acid dissociation constant
rpm	Rotations per minute
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulfate
SGD	*Saccharomyces* genome database
TAD	Transactivation Domain
TF	Transcription factor
TFBS	Transcription factor binding site
TCA	Tricarboxylic acid
T. blattae	*Tetrapisispora blattae*
V. polyspora	*Vanderwaltozyma polyspora*
WGD	Whole genome duplication
Z. bailii	*Zygosaccharomyces bailii*
Zn	Zinc
Z. rouxii	*Zygosaccharomyces rouxii*
YPD	Yeast peptone dextrose

1. Introduction to the thesis work and thesis outline

Weak acid tolerance exhibited by some yeast species is a very important trait in the biotechnological industry, which depending on the field of application can be considered as an advantage or a disadvantage[1]. Weak acids are broadly used in food and beverage industries as antimicrobial agents, especially in those products that have high sugar concentration. However, some yeast species, like those belonging to the genus *Zygosaccharomyces* and some isolates of *Saccharomyces cerevisiae*, are able to survive and proliferate when weak acid concentrations are near the legally permitted limits, causing substantial economic losses in this industry sector[2–4]. The yeast species *Z. bailii* was early recognized as the most problematic spoilage agent in acidic foods and beverages due to its exceptional tolerance to weak acids used as food preservatives, including acetic acid[4–6]. This species frequently appears at the end point of wine fermentations, when acetic acid and ethanol concentrations are very high to admit the survival of *S. cerevisiae*[2,4,7]. At the same time, weak acid tolerant species can provide an interesting resource in genetic terms since they can be used as cell factories in the production of valuable compounds, like bioethanol or other 'building blocks' molecules, from lignocellulosic non-feedstock substrates. Yeast cells better adapted to weak acids will be able to tolerate higher concentrations of the inhibitors released during the pretreatment of lignocellulosic substrates, such as acetic and formic acids, which are responsible for the inhibition of yeast growth and, thus, fermentation arrest[8–10]. Understanding the molecular determinants behind yeast tolerance to acetic acid will give us further insights to design more efficient food preservation strategies, as well as to develop industrial yeast strains more resistant to lignocellulosic-derived inhibitors and therefore with higher fermentation performances crucial to industrial scale-up and to the economic sustainability of processes[11,12]. Despite all the efforts and the current knowledge regarding acetic acid tolerance in *S. cerevisiae* and *Z. bailii*, which have been addressed to several mechanisms, the entire system involving the yeast response to the acid is not completely elucidated, especially in *Z. bailii* species. Indeed, seventeen years separate the genome sequencing of *Z. bailii* from *S. cerevisiae* laboratory strain S288c, the first genome to be sequenced[13,14], thus hampering the molecular and functional dissection of the first one.

The transcription factor Haa1 was described as a key regulator in *S. cerevisiae* adaptation response to weak acid stress, including to acetic acid, since it modulates the transcription of multiple genes involved in weak acid tolerance[15,16]. A recent study performed in *S. cerevisiae*

demonstrated that variations in the Haa1 amino acid sequence may improve acetic acid tolerance in yeast strains carrying the mutated Haa1[17]. In evolutionary terms, transcription factors (TFs) are subject to severe functional constrains which contribute to the conservation of their basic features. However, a remarkable rewire was observed among regulatory networks, frequently involving successive mutations in TFs and their target genes and being responsible for the evolutionary divergence demonstrated in TFs[18,19]. In this context, and since a putative Haa1 sequence was identified in *Z. bailii* genome based on amino acid similarity with the Haa1 sequence previously described in *S. cerevisiae*, it was hypothesized that *Z. bailii* genome could include a functional homologue of *S. cerevisiae* Haa1 (*Sc*Haa1). This assumption was based on protein functional prediction, *i.e.*, the functional annotation addressed to a specific protein-coding gene, from a given species, can be transferred to another protein-coding gene, from another species, if the corresponding amino acid sequences share a significant similarity[20]. The different amino acid conservation between these sequences raised the hypothesis that the Haa1 from *Z. bailii* (*Zb*Haa1) shares a common signature behaviour with its *S. cerevisiae* homologue and grants a similar or even more efficient acetic acid protection needed to high fermentation performances. Taking into account all of the previously mentioned, genetic engineering approaches using the *ZbHAA1* gene seems a promising strategy to explore the design of new-engineered species of *S. cerevisiae* by protein heterologous expression.

The present work aims to study the role of the putative *ZbHAA1* gene in acetic acid tolerance using the experimental model *S. cerevisiae* BY4741 and the derived deletion mutant *haa1*Δ as host strains for the heterologous expression of *ZbHAA1* ORFs. Plasmid construction procedures included the amplification of the *ZbHAA1* ORFs sequences from the *Z. bailii* strains CLIB213^{T}, IST302 and the hybrid ISA1307, followed by their cloning into the pGREG506 vector, under the control of *S. cerevisiae HAA1* promoter. Additionally, the *HAA1* from *S. cerevisiae* was also cloned as an internal control for expression profiles comparisons. Acetic acid tolerance profiles of yeast cells carrying the recombinant vectors were analysed by spot assays and comparing the growth curves in basal medium supplemented or not with this acid. It is expected that the heterologous expression of different *Zb*Haa1 proteins can complement the acetic acid susceptibility phenotype of the mutant *haa1*Δ and increase the acetic acid tolerance of the parental strain. In an ideal scenario, the *Zb*Haa1 would provide a better tolerance to acetic acid in comparison to the *Sc*Haa1. In order to obtain a more integrated view about the Haa1 and its homologue Ace1, involved in yeast copper sensing, we studied the evolutionary history of the protein family Haa1/Ace1 in species of the Saccharomycetaceae family by phylogenetic and

comparative genomic (synteny) analyses. Moreover, we examined the Haa1 amino acid sequence from several *S. cerevisiae* isolates relatively well known, aiming to correlate the amino acid variation with the ecological niches and traits of biotechnological relevance already described to strains under study.

2. Introduction

2.1. Mechanisms of toxicity and tolerance to acetic acid in *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*: role of the transcription factor Haa1

2.1.1. Acetic acid as a stress agent

In a modern world where consumers have a much more critical attitude in relation to the freshness of food products, and where a large percentage of food products have to be preserved for long periods of time, microbial food conservation and safety is certainly a challenge[21]. Acetic acid and other weak acids such as sorbic, propionic and benzoic acids are broadly used as food preservatives, especially in products that have a low pH and a high sugar concentration[2,4,21]. However, despite all the efforts to prevent microorganism contamination, there are few species of fungus able to adapt and proliferate in such conditions where the concentration of weak acids is usually very close or even exceed those legally accepted, causing significant economic losses[4]. Among these species, the budding yeast *Z. bailii* was early recognized as the most problematic spoilage agent due to its exceptional tolerance to acetic acid and ethanol at low pH[4,21,22]. In fact, the acetic acid minimal inhibitory concentration for *Z. bailii* of 26 g/L (433 mM) contrasts to the 9 g/L (150 mM) observed for *S. cerevisiae*[23]. Knowledge on the main mechanisms involved in yeast tolerance to acetic acid stress could be applied in bioprocess conditions optimization as a way to avoid microbial contaminations in food industry[1].

Acetic acid is also one of the most relevant carboxylic acids released during the pretreatment of lignocellulosic substrates applied in the production of bioethanol and other chemical compounds of economic interest. The use of renewable and non-food feedstocks appears as an important alternative to fossil fuels to deal with the world's energy demand while preventing global warming[24]. Since acetic acid is an inhibitor of microbial growth and fermentative metabolism, decreasing ethanol yield and volumetric productivity, and also because the detoxification processes of pretreated substrates has very high associated costs, the use of fermentative microorganisms more resistant to lignocellulosic derived inhibitors as cell factories seems a promising strategy to improve the bioethanol production cost-efficiency[8,25–27]. More robust organisms include those obtained under fermentative conditions, like the natural isolates well adapted to industrial environments, the strains developed by evolutionary engineering under identical selective pressure, and also the recombinant strains

developed by metabolic engineering that are able to tolerate high inhibitory concentrations of toxic compounds. *S. cerevisiae* has been one of the preferred candidate species to be applied in bioethanol production from lignocellulosic substrates, which can be associated to its high growth rate and ethanol productive yield, considerable resistance to fermentation stresses and the great amount of genetic and bioinformatics tools available to this species[1]. Some studies in the biorefinery context have already demonstrated the potential of genetic engineering approaches performed using this suitable eukaryotic model[27–29].

2.1.2. Effect of acetic acid in *S. cerevisiae* and *Z. bailii* metabolic pathways

In general, yeasts exhibit a fermentative facultative metabolism, alternating between full fermentative and respiratory metabolism which can be simultaneously found in a mixed respiratory-fermentative metabolism. The type and concentration of different carbon sources, as well as the biochemical oxygen demand and growth conditions trigger the metabolic pathway carried out by yeasts, suggesting the importance of the environment in yeast metabolic strategies[30]. The glucose consumption by *S. cerevisiae* is favoured in a complex mixture of carbon sources that can include other sugars, polyols, alcohols, amino acids and organic acids, such as acetic acid, being described as a catabolic repression phenomenon. In batch cultivation or continuous culture under aerobic conditions most yeasts consume glucose mainly by fermentation, a strategy known as Crabtree effect that is caused by high glucose concentrations. When glucose concentration decreases above critical dilution rates, yeast cells switch their metabolism to a respiratory-fermentative metabolism, resulting in ethanol and acetic acid production as fermentation byproducts, and also glycerol[31]. When simultaneously exposed to a medium with glucose and acetic acid, in aerobic conditions, *S. cerevisiae* exhibits a diauxic growth totally consuming glucose by fermentation, as above mentioned, followed by the respiration of acetic acid and ethanol produced in fermentative phase[30,32]. In contrast, *Z. bailii* exhibits the ability of co-consume sugars in combination to acetic acid, displaying a biphasic growth which is associated to its constant presence in wine fermentations and some acidic foods[31,33]. In aerobic conditions when glucose or other metabolizable sugars are available in a suitable concentration, yeast cells withdraw all the sugars energy and molecular components through the glycolysis, tricarboxylic acid cycle (TCA) and oxidative phosphorylation pathways, repressing the glyoxylate cycle and gluconeogenesis. Ethanol and acetate could also be converted to acetyl-CoA and follow the TCA and oxidative phosphorylation pathways, representing an additional source of energy supply, or could even

be guided to pentose phosphate pathway that is responsible for the maintenance of anabolic metabolites biosynthesis such as NADPH, required for fatty acid biosynthesis and nucleotides (**figure 1**). Fatty acids biosynthesis through acetyl-coA is essential to plasma membrane remodeling under acetic acid stress. Acetyl-CoA synthesis from ethanol is ensured by its conversion to acetate, via acetaldehyde, followed by activation by coenzyme A[34,35].

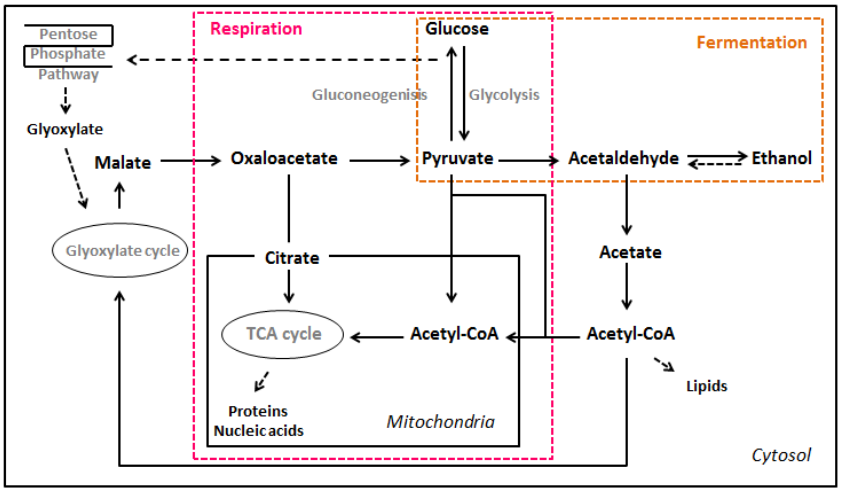

Figure 1. Major metabolic pathways involved in the simultaneous utilization of glucose and acetic acid in yeasts. (Figure adapted[30,34]).

2.1.3. Molecular mechanisms underlying acetic acid toxicity in yeasts

Acetic acid was shown to induce phenotypic stress effects in unadapted populations of *S. cerevisiae* and *Z. bailii* in exponential growth when suddenly exposed to an inhibitory concentration of this acid. Negative effects include growth inhibition with latency phase arrest and loss of cell viability by programmed death[15,35–38]. Previous studies highlight the importance of acid lipophilicity in the magnitude of negative stress effects induced in *S. cerevisiae* cells, particularly at the level of membrane, suggesting that more lipophilic acids have a less toxicity potential[15,39,40]. Additionally, weak acids' toxicity is also affected by other chemical properties of these molecules such as its pK_a and volatility. Environmental conditions such as pH, temperature, oxygen and nutrient availability, acid concentration, eventual metals that could be present in the medium, as well as the metabolic yeast cell status, are still important in the severity of the acid phenotype[7,12,30,39,41]. At a pH equal or below acetic acid pKa (4.7), the undissociated form of acetic acid (RCOOH) can easily cross the plasma membrane lipid bilayer by simple diffusion and enter into the yeast cells[12,41]. Besides this mechanism, specific transporters were described in *S. cerevisiae* and *Z. bailii* as being involved in the import of weak acids through the plasma membrane[42,43]. Once in the cytosol, where the pH is near-

neutral, the acid dissociates in its deprotonated form ($RCOO^-$) and its counter-ion (H^+), which due to its electrical charge cannot cross the hydrophobic plasma membrane and leave the cell, accumulating inside it and causing a decrease of intracellular pH. At low internal pH, electrochemical gradient of cell membranes is disrupted leading to the loss of membrane systems integrity, including embedded proteins[12,39,41]. Once secondary transport capacity and selectivity permeabilization has been impaired there is a tendency to increase free proton influx rate from surrounding medium with consequent acetate accumulation inside the yeast cells in a co-work that amplifies cellular damage[12].

Considering that most of the cellular processes require energy, the decrease of energy pool arises as a threat to yeast cells. Furthermore, energy requirements enhance with the activation of mechanisms that mediate acid detoxification, being imperative to ensure energy-consuming defense responses[2,12,41].

2.1.3.1. Molecular mechanisms underlying acetic acid tolerance in *S. cerevisiae*

In *S. cerevisiae*, three main processes can be associated to overcome acetic acid effects: the release of dissociated acid forms (H^+ and $RCOOH^-$) out of the cell, the surpassing of negative effects induced by acid and the prevention of undissociated acid form to reenter the cell. Proton (H^+) extrusion out of the cell aims to reestablish the cytosolic pH, in a pumping mechanism carried out by the H^+-ATPase (Pma1) and V-ATPase transporters present in the plasma and vacuolar membranes, respectively[2]. Elimination of the negatively charged weak acid counterions relies on the activation of specific Multidrug Resistance (MDR) transporters[12] (**figure 2**).

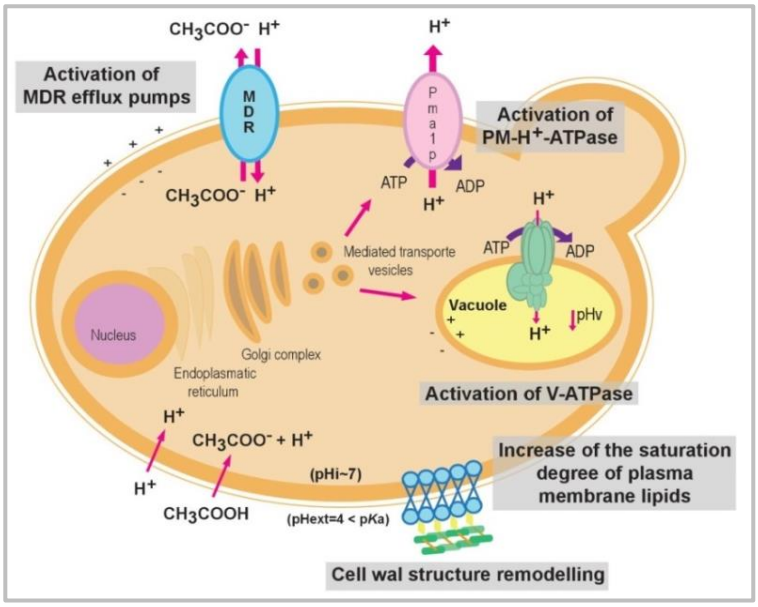

Figure 2. Schematic representation of *S. cerevisiae* cellular mechanisms used to counteract acetic acid stress effects. Acetic acid and protons present in the acidic medium enter into the *S. cerevisiae* cells via plasma membrane, decreasing the intracellular pH. The reestablishment of cell homeostasis includes the pumping of protons and acetate by the Pma1 and V-ATPase membrane proteins and MDR transporters, respectively, as well as the reshaping of cell wall and plasma membrane (figure adapted[12]).

Previous studies support the idea that MDR transporters have an intrinsic physiological role that can be combined with its ability to transport a wide range of structural and functional unrelated cytotoxic compounds. MDR transporters comprise several classes of transporters of which the ATP-binding cassette (ABC) Superfamily and Major Facilitator Superfamily (MFS) are the most successful transporters implicated in weak acids' tolerance[44,45]. Concerning acetic acid protective effect it was recognized a higher number of MFS transporters in comparison to ABC transporters. Among the MFS transporters found to confer acetic acid tolerance are the Aqr1, Azr1, Tpo2 and Tpo3 transporters[12,41,46,47]. The ABC transporter Pdr18, responsible for plasma membrane ergosterol incorporation, was also shown to be determinant to overcame acetic acid negative effects (unpublished results).

To fight the negative effects induced by weak acids, yeast cells activate a general response that includes the upregulation of a set of genes involved in primary functional classes like carbohydrate metabolism, ion transport, intracellular trafficking and cell wall remodeling. However, there are specific genes triggering these pathways that frequently vary according to weak acid, or more specifically to weak acid lipophilicity[12].

Weak acid response and tolerance in *S. cerevisiae* can be assigned to the transcription factors (TFs) Haa1, Rim101, Msn2/Msn4 and War1, being interestingly demonstrated that a coordinated cross-talk occurs between the regulatory pathways controlled by these regulator proteins. Haa1 was recognized as the strongest regulator of acetic acid resistance, activating, directly or indirectly, 80% of genes expressed under acetic acid conditions[16]. More recently, the transcription factors Skn7 and Stb5 were identified as promising regulators in acetic acid tolerance[41].

2.1.3.2. Molecular mechanisms underlying acetic acid tolerance in *Z. bailii*

Although the molecular mechanisms behind the acetic acid tolerance by *Z. bailii* are not completely elucidated previous studies have reported numerous mechanisms that might be responsible for the exceptional ability of this yeast to withstand acetic acid negative effects. This coordinate response encompasses the plasma membrane remodeling, the co-consumption of acetic acid in the presence of sugars and the maintenance of a slight acidic physiological intracellular pH (pH_i)[4,23].

Plasma membrane is the first protective barrier to be strengthened against acetic acid, suffering a number of lipid rearrangements, largely determined by the high basal levels of complex sphingolipids that promotes a thicker and less permeable plasma membrane to the undissociated acid form, while increases membrane anchored proteins support and keeps their active fold. Additionally, the enrichment of saturated acyl chains, especially monosaturated, in glycerophospholipids and complex sphingolipids, together with their long-chain fatty acids, also contribute to prevent acetic acid uptake. When compared to *Z. bailii*, in the presence of acetic acid, *S. cerevisiae* cells revealed much less pronounced lipidomic changes, showing a slight decrease in sphingolipids and ergosterol amounts that are maintained in *Z. bailii*, although the total amount of sterol and glycerophospholipids is lower in *Z. bailii* than in *S. cerevisiae*[23].

When in the presence of glucose, acetic acid is co-consumed by *Z. bailii* cells, as previously noticed, a strategy that allows the reduction of intracellular acetate concentration while yeast cells can take advantage of the acid as a carbon and energy source[5,23,34]. A proteomic study performed in our laboratory showed that acetic acid stressed *Z. bailii* cells express a set of proteins in the presence of this acid that are involved in cellular carbohydrates metabolism

and energy supply, supporting the previously reported results regarding glucose and acetic acid co-metabolism by *Z. bailii*[35].

Regarding pH_i homeostasis, *Z. bailii* was shown to be more resistant to pH_i changes than *S. cerevisiae* in high acetic acid concentrations, being able to maintain a better pH homeostasis whether in short or long term and thus, a greater tolerance to acetic acid (since acid dissociation is dependent from pH_i)[36,48]. Recent studies have attributed this phenomenon to specific cells of a *Z. bailii* population that naturaly present a lower pH_i, whithout compromising cellular processes. This resistance mechanism is supported by *Z. bailii* population heterogeneity[3,4,49].

Ethanol produced by yeast cells during fermentation metabolism has a synergistic effect in acetic acid toxicity in *S. cerevisiae* cells, leading to a sluggish fermentation and, consequently, a decrease in ethanol volumetric production[30,50]. Interestingly, no significant amplification of acetic acid severity was observed in *Z. bailii* cells simultaneously exposed to acetic acid and ethanol[22]. In fact ethanol was found to inhibit acetic acid uptake in *Z. bailii* cells, since it promotes the plasma membrane disruption and then the destabilization of protein carriers that favour acid transport[42]. This strategy was demonstrated to contribute to prevent high acetic acid levels inside the yeast cells that in certain cases could even be maintained below to toxic limits[42], corroborating recent studies which proposed that acetic acid overall uptake by *Z. bailii* cells is essentially determined by protein channels carriers, instead of passive diffusion through plasma membrane[23].

2.1.4. The role of the transcription factor Haa1 in yeast response to acetic acid stress

Transcription factors are proteins involved in the control of the expression of multiple genes under specific environmental conditions. Structurally, these proteins are composed by a DNA Binding Domain (DBD) and a Transactivation Domain (TAD). The DBD recognizes specific motifs and amino acid residues (well conserved) in the promoter region of TFs target genes (Transcription Factor Binding Sites, TFBSs), being responsible for the TFs assembly to these DNA regions. Frequently, multiple TFBSs are recognized by the same TF[51,52]. The TAD domain is involved in TF activation and frequently includes binding sites to other proteins or ligands like cofactors[52].

Haa1 was first identified based on its homology with the DBD of the copper-metalloregulated TF Ace1 (also known as Cup2)[53]. The DBD of Haa1 and Ace1 comprises the first 124 amino acids of these protein sequences and includes a conserved zinc (Zn) module, a (R/K)GRP motif and a set of four cysteine-cysteine clusters organized in a consensus sequence that composes the Copper Regulatory Domain (CuRD)[53,54] (**figure 3**). So far, all of the Haa1 and Ace1 protein domains identified are reported to their DBD.

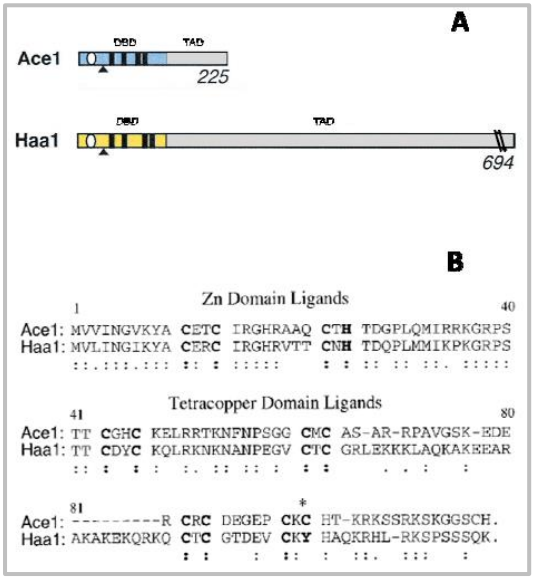

Figure 3. Comparison of Haa1 and Ace1 protein sequences. A) Ace1 and Haa1 protein sequences representing the DNA Binding Domain (DBD) that includes the first 124 amino acid residues (represented in blue and yellow colours, respectively); and the Transactivation Domain (TAD) (represented in grey). **B)** Alignment of the DBD of Ace1 and Haa1 showing the zinc (Zn) and CuRD domains[53,59].

The similarities found in the DBD of Haa1 and Ace1 led to initially hypothesize that the Haa1 could have an action in copper sensing, although posterior works demonstrated that only Ace1 is sensible to copper cell status. Moreover, it was not observed any cross-talk between Haa1 and Ace1[53]. Indeed, the Haa1 and Ace1 recognize different promoter regions, activating a complete different set of genes[55,56]. Whereas the Haa1 is involved in *S. cerevisiae* tolerance to short-chain weak acids, Ace1 mediates the response to copper nutritional status in this species[15,53]. The recognition of different TFBSs by the Haa1 and Ace1 was demonstrated to be a consequence of the structural differences found in the DBD of these proteins as the result of specific amino acid variations[55]. Previous works also demonstrated the importance of specific amino acid residues in Haa1 activity, for instance, the amino acid residues Arg38, Arg50 and Ser113 were shown to be essential in *S. cerevisiae* for the recognition and binding of the Haa1 to the promoter region of its target genes[55].

The action of Haa1 regulon in *S. cerevisiae* was initially demonstrated by the comparison of yeast cells phenotype exhibited by the parental strain BY4741 and the derived deletion mutant *haa1*Δ exposed to acetic acid inhibitory concentrations. Additional, the Haa1 was demonstrated to be necessary for a faster adaptation of a *S. cerevisiae* cell population suddenly exposed to weak acid stress by the upregulation of a set of target genes encoding effector proteins involved in yeast cell detoxification. This protective effect was shown for the less hydrophilic short-chain weak acids, especially against acetic acid, but also to propionic and benzoic acids and more recently to lactic acid[15,57]. The first Haa1 target genes to be identified included the *TPO2* and *TPO3* that encode two MFS transporters responsible for mediating acetate export and the *YGP1* that encodes a cell wall remodeling related glycoprotein activated under nutrient constraints and other stress conditions[15]. Actually, this list of genes almost comprises 120 genes of which 85 genes are related to acetic acid tolerance. Most of these genes encode proteins related to lipid metabolism, regulation of gene expression and reduction of acetate concentration; the remaining genes encode proteins of unknown functions[16] (**table 1**). Among these genes, *SAP30* and *HRK1* are considered those that ensure the most powerful protective effect in acetic acid challenged yeast cells. *SAP30* encodes a component of the RPd3L histone deacetylase complex and the *HRK1* encodes a protein kinase responsible for posttranscriptional regulation of plasma membrane H^+-ATPase Pma1 activity[16,58].

Table 1. Biological functional classes and respective target genes activated by Haa1 in response to acetic acid stress[16].

Functional class	Target genes
Carbohydrate metabolism	*TOS3, NGR1, HRK1, PCL10*
Transcription factors	*MSN4, NRG1, FKH2, STP3, STP4*
Lipid metabolism	*SUR2, YPC1, INM1*
Cell wall remodeling	*YGP1, SPI1, ECM13*
Nucleic acid processing	*SAP30, SYC1*
Protein folding	*HSP30, HSP26, SSE2*
MDR transporters	*TPO2, TPO3, PDR12, PDR16, AQR1*
Amino acid metabolism	*ADI1*
Stress response	*HOR2, SSE2*

The majority of functional classes observed in **table 1** reinforce the idea of a broad spectrum response against acetic acid induced stress, based on genomic reprogramming of main cell processes, in order to maintain cell homeostasis. This detoxification strategy aims to minimize the negative effects induced by acetic acid in yeast cells, either by acetate and protons export

or by the replenishment of cell wall integrity that prevents acid reentry, or even by recovering cellular damage[16].

Although it had been hypothesized that the Haa1 presence in the cell nucleus is independent from the acetic acid levels in *S. cerevisiae* cells[53,55,59], recent results obtained in our laboratory indicate that in the absence of acetic acid stress the TF Haa1 is widely spread in the cytoplasm, but in the presence of acetic acid is phosphorylated and imported to the nucleus where activates its target genes (unpublished results). This activation mechanism of the Haa1 was also demonstrated in *S. cerevisiae* cells cultivated in lactic acid conditions where Haa1 suffers a dephosphorylation reaction required for its export from the nucleus to the cytoplasm through the Msn5 transporter[60]. Interestingly, the protein-protein interaction between Haa1 and Msn5 was also proven upon *S. cerevisiae* cells exposure to acetic acid, even though it is not clear the complete mechanism responsible for the Haa1 activation and the major players involved[60]. The Haa1 activation by phosphorylation is in agreement with a recent patent that showed that some mutations in Haa1 phosphorylation sites could be in the origin of acetic acid more tolerant strains[17].

Recent studies evaluated the effect of an *HAA1* extra copy in acetic acid tolerance (0.5% W/V, pH 4.5) which resulted in lower intracellular levels of acetic acid in overexpressed mutant, when compared to parental strain, and thus in a higher acetic acid tolerance and fermentative performance[61,62].

2.2. Comparative genomics and phylogeny applied to yeast

2.2.1. Yeast genome evolution

At first sight yeasts seem morphological and physiologically very similar, and even in genomic terms, but a closer look shows that the divergence is greater than expected. In fact, the evolutionary variations among protein sequences from Hemiascomycetes can be equivalent to the corresponding in chordates[63]. The progress of molecular biology and bioinformatics sciences in the last decade has allowed the rapid sequencing of short nucleotide sequences, and more recently complete genomes that facilitated the reconstruction of yeasts' evolutionary history. Definitely, next-generation sequencing completely changed the way to face genomic challenges through the establishment of genome large-sequencing approaches

which promoted the extension of genome sequence programs beyond *Saccharomyces* to other Hemiascomycetes genera[63]. This achievement generated a massive amount of genomic information revealing the deeper genome intricacies and favouring comparative genomics studies that made possible the development of numerous phylogenetic and comparative genomic analyses[63,64]. Nevertheless, the amount of entire genomes completed and well annotated is still low, limiting comparative studies, whereas the multigene sequence is the most commonly used methodology to phylogenetic reconstructions[65].

Phylogenetic trees often emerge as the ultimate goal of a phylogenetic analysis allowing the description of evolutionary relationships among species or genes under study. In 2003, Kurtzman and Robnett proposed the first yeast phylogenetic analysis based on some unlinked multigenes to the Hemiascomycetes subdivision '*Saccharomyces* complex', leading to resolution in a very efficient way of 75 species in 14 clades (**figure 4**)[66,67].

The particular interest on the '*Saccharomyces* complex' is associated with the fact that this complex includes the eukaryotic model *S. cerevisiae*, the small and non-repetitive genomes of the yeast included in the complex, the biotechnological potential displayed by a number of species belonging to this group or its medical relevance[68].

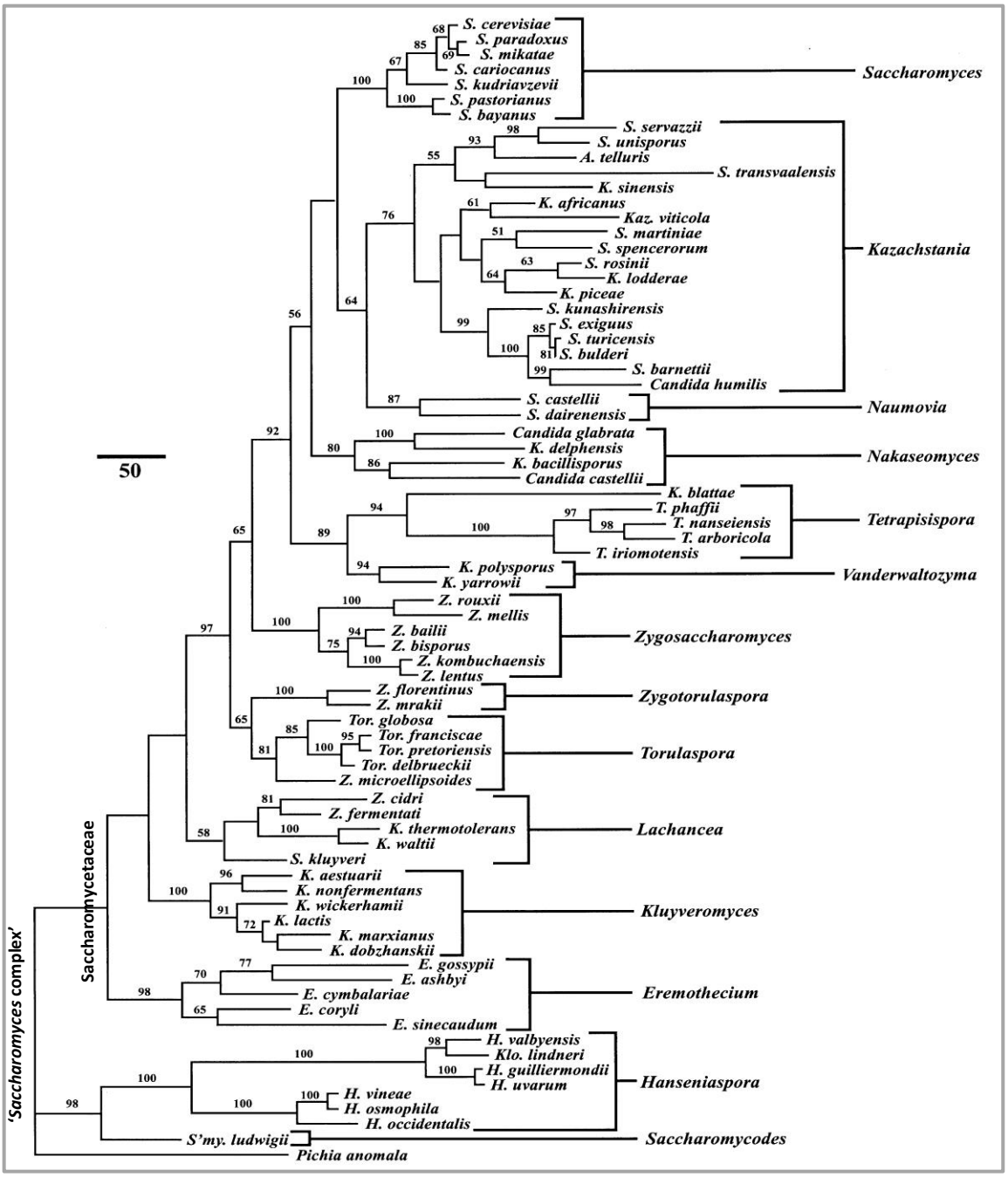

Figure 4. Phylogenetic tree among '*Saccharomyces* complex' yeast species. This tree was proposed by Kurtzman and allowed to resolve 75 species into 14 clades representing phylogenetically circumscribed genera. The maximum-parsimony method was used to generate the tree according to the analysis of nucleotide sequences from 18S, 5.8S/alignable ITS, and 26S (three regions) rDNAs, EF-1K, mitochondrial small-subunit rDNA and COX II[67].

Visually, a phylogenetic tree comprises the nodes, connected by branches (edges) corresponding to speciation events[69,70]. The size of the branches represents the evolutionary divergence between each node and could vary according to the method applied in tree construction[71]. The tree branch which contains the root of the tree encloses the most dissimilar member and the common ancestor of all tree members. One of the aims of a

phylogenetic analysis is to group its members in different clusters that might share specific properties[69].

Numerous studies have narrowed down phylogenetic analysis to a single gene, trying to reconstruct the evolutionary history of orthologous genes (*i.e.*, genes originated from a common ancestor *locus* which diverged by a speciation event) in a gene family[70,72–76]. Interestingly, in some cases the evolutionary history of genomes and specific genes do not match, denoting the differences behind the time scale assigned to speciation and mutation events that trigger genomic and gene evolution, respectively[77]. Furthermore, different genes enclose a series of unique properties which in combination with selective pressure shape their particular evolutionary course and fate. Indeed, we must be careful in establishing some assumptions, because orthologous genes when exposed to extremely divergence forces could no longer encode the same expected function[68]. The phylogenetic history of a given gene is frequently assessed by the corresponding protein evolution, as a function of synonymous and nonsynonymous amino acid substitutions along tree branches[70].

Genome sequencing and annotation made possible the construction of whole-genome maps contributing to elucidate synteny relationships in yeasts whose genomes were extremely shaped by the Whole Genome Duplication (WGD) and subsequent chromosomal rearrangements. Synteny analysis is based on the assumption that two genes of distinct yeast species whose protein sequences belong to the same sequence cluster (homologous by similarity) will be members of the same gene linage if they share at least one pair of neighbour genes (also homologous by similarity)[74,78] (note that homologous genes are defined as genes sharing an ancestral common origin)[72]. This last assumption is founded on the idea that after WGD gene blocks orientation would be conserved if they are originated by reciprocal translocations among duplicated chromosomes[79]. Gene blocks can be defined as structural blocks that enclose a set of genes and behave as individual units in the genomic remodeling process[64,80]. One of the most powerful tools given by the synteny analysis is the definition of gene lineages through the distinction of orthologous from paralogous genes (*i.e.*, genes derived from a duplication event) in related species[81–83]. Regarding Hemiascomycetes, the number of conserved gene blocks is underestimated relatively to chromosomal rearrangements, an inconsistency explained by the superposition of evolutionary events and the accumulation of microrearrangements involving few genes[84]. Despite the wide genome reshuffling observed in hemiascomycetous yeasts, a similar synteny often conserved among

sister chromosomes of the same species and even in pre- and post-WGD lineages was demonstrated[82].

The combination of both phylogenetic and synteny analyses seems to lead to more realistic results, since the results from each analysis often assist each other, even though it has been demonstrated a nonlinear relationship between sequence divergence rates and chromosomal rearrangements in Hemiascomycetes[82]. This relationship, also observed in Saccharomycetaceae family, evidences the extremely high amount of genome reshuffling after WGD. Moreover, as the evolutionary distance increases, the phylogenetic information that we can gather from synteny analysis tends to decrease due to the multiple rounds of genome shuffling, becoming more difficult to achieved reliable genomic information and congruent results[68,82,84].

Whole Genome Duplication was first described in Saccharomycetaceae as an unique event responsible for the duplication of the entire genome that is in the origin of the current paralogous genes (*i.e.*, duplicated genes from the same ancestor) and the actual diversity found in budding yeasts[85] (**figure 5**). However, due to the genomic and epigenetic instability as a result of genomic redundancy and size increase, WGD was followed by a rapid disruptive genome reduction in such a way that only approximately 13% of duplicated genes were retained[85,86]. Genomic reduction combined with rearrangements in gene content and changes in its expression levels allowed the re-establishment of genomic functionality and the divergence of yeast lineages in an efficient specialization process[86].

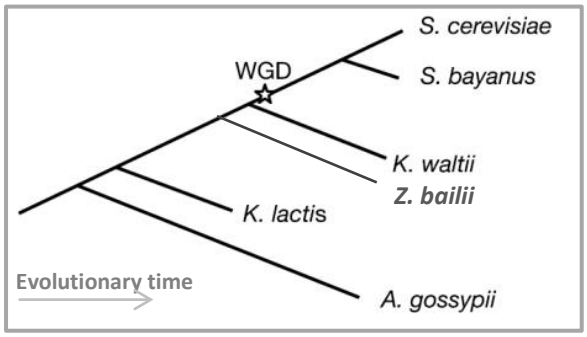

Figure 5. Divergence of paralogous genes over evolutionary time in Saccharomycetaceae. Whole Genome duplication event is highlighted (figure adapted[86]).

Until now there is not a consensus beyond the molecular mechanisms involved in the duplication of the ancestral yeast diploid genome. Some studies proposed an autotetraploidization (duplication of the entire chromosome set) while other indicated a more

probable allotetraploidization (hybridization between two distinct but related species). Transient tetraploidization was initially characterized by high sequence similarity among paralogous gene pairs that was quickly dissipated across the evolutionary divergence[63,85]. Indeed, WGD appears as a powerful mechanism of evolutionary innovation revealing a great opportunity of genome remodeling in response to the adaptation process driven by genetic drift and natural selection. Natural selection is defined as the differential survival and maintenance of specific phenotypic traits in a population in response to environmental conditions[87]. The genetic drift can be defined as the random fluctuations in the allele frequencies of a population over time due to the sampling effect in small populations[88].

Functional divergence in paralogous genes can occur by different ways, including genetic dosage changes in response to gene loss biases and neo/subfunctionalization[88]. Gene loss can be considered the major responsible for the divergence among different yeast lineages and despite its independent nature it seems to act by fostering a common pattern favouring elimination of specific 'obsolete' ancestral functions[73]. Neo/subfunctionalization processes comprise three principal life stages of paralogous genes: creation, fixation-preservation and subsequent optimization. Functional innovation is responsible for the development of a new selectively beneficial function and can be addressed to post-WGD mutations that affect paralogous genes in a distinct way. Alternatively, remaining new genes can be fixed by subfunctionalization, as a result of purifying selection, in which paralogous genes assist each other by sharing differential functional components of the ancestral gene function, being this one of the most frequently phenomena observed in yeast genome divergence. The fixation and optimization of novel duplicated genes require that all the members of a species retain the new extra-copy which may be maintained in a population through the genetic drift and positive selection if the novel feature favour the population[88]. Moreover, the fixation of paralogous genes seems to rely on a duplication event behind their generation, which can be assigned to WGD or small-scale duplications, as a result of the complexity underlying each evolutionary mechanism[73,89]. Most of the remaining duplicated genes pairs reveal a strong divergence in nucleotide sequences supported by asymmetrical substitution mutations, resulting in the optimization of duplicated genes to specific selective constraints[63]. In fact, genome remodeling after WGD and subsequent extensive deletion, in 95% of cases only involve one of the two paralogous genes, which in theory is sufficient to counteract genetic dosage and fix a new function, without losing the ancestral information (maintained unchanged or as a pseudogene, *i.e.*, a remaining gene that lost its biological function)[86,90,91]. Notably, it has been recognized that remaining paralogous genes, in average, tend to evolve

slower than genes that have lost its paralogue copy probably due to the stronger genetic neighborhood associated to the first ones[72].

2.2.2. Transcription factors networks evolution

There are 180 transcription factors described in the YEASTRACT database for *S. cerevisiae*[56]. The evolution of transcription factors as the main agents of transcriptional regulation in response to environmental changes is crucial to functional and morphological innovation. Because the transcription factors make part of intricate regulatory circuits, the evolutionary mechanisms assigned to these proteins are much more complex to those observed for other proteins[18,19]. Indeed, it was demonstrated that TFs and their target and neighbour genes tend to evolve together, at the same rate, evidencing a tendency of regulatory networks to evolve like modules[92].

As already mentioned, the WGD and subsequent evolutionary events provided an increase in genomic diversity giving yeasts the flexibility necessary to face new challenges, in particular the development of newly regulatory circuits or the rewiring of the previously ones. After WGD, ancestral genes coding for specific TFs gave rise to two new identical genes which frequently diverged during evolution. Depending on the remodeling mechanism involved in the divergence process, TFs faced diverse destinies driven by the acquisition of a novel function by one or both paralogous copies, or in the case of an ancestral promiscuous protein able to recognize multiple promoter target sequences, the differential split off of ancestral functions over the two new copies[19] (**figure 6**). Several studies indicate a high tendency to one of the paralogous gene to fix the ancestral traits while the other evolves towards a novel function[18,19].

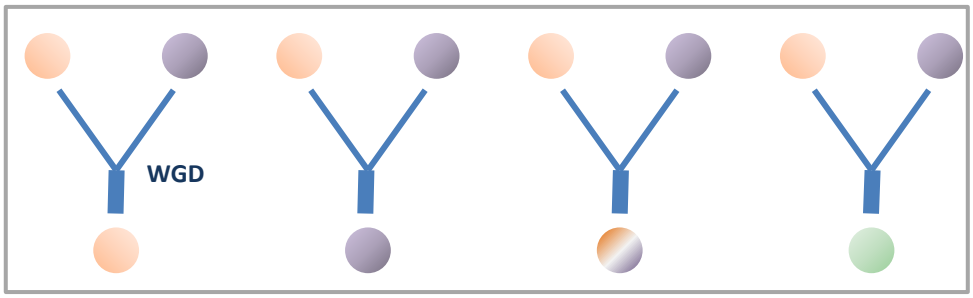

Figure 6. Evolutionary hypotheses for paralogues divergence from a pre-WGD ancestor. Genes are represented as circles and a different colour which represents different functions. The bifurcation illustrates the WGD event.

Since paralogous genes often code for proteins with different functions, it is clear the importance of establishing independent regulatory systems for each TF paralogue that allows the simultaneous activation of the different transcription factor-regulons, while preventing the undesirable cross-regulation. This functional divergence can be achieved through the different molecular mechanisms that include mutations in both TFs (*trans* mutations) and corresponding target genes promoter sequences (*cis* mutations), chromatin reorganization and interactions between specific proteins/ligands and the TFs or their target genes[19]. Several studies demonstrated that only few mutations in the transcription factor DBD or in the TFBSs are sufficient to ensure new patterns of gene expression[19,93]. *Cis* mutations, generally linked to TFBSs are responsible for the disruption of the ancestor binding sites or the creation of new ones, sometimes only being recognized by a novel TF. *Trans* mutations, frequently found in transcription factor DBD, could change the transcription factor's binding specificity and/or eventual interactions with cofactors, which ultimately could imply the replacement of the original TF, in a given regulatory network, by other with a distinct function[18,19]. In order to support the evolutionary genomic adjustments imposed by the WGD it was recognized an extensive loss of *cis*-regulatory elements in the promoter regions of many TF target genes after the entire genome duplication event, however paralogous genes often retain the total number of these elements which means that novel *cis* elements were incorporated during evolution. *Trans* mutations also have significantly contributed to WGD remodeling despite the major efforts that have been highlighting *cis* mutations[19]. An interesting fact is that paralogous TFs resulting from a WGD event display a preferential retention in comparison to those originated from a single gene duplication[73].

All of these reported changes illustrate the extraordinary plasticity of transcription networks over evolution. Nevertheless, transcription factors' DBD and TFBSs are remarkably conserved *loci* within yeast genome evolution[18,93]. Accordingly, it seems clear that the regulatory reshaping mechanisms of TFs and their target genes is an interdependent complex process. On one hand, TFs act as principal players in regulatory pathways by triggering a general response through the on/off switch of the entire set of target genes; on the other hand, target genes trigger specific responses[92].

3. Materials and Methods

3.1. Cloning and expression of *HAA1* homologous genes from strains CLI213^T, IST302 and ISA1307 (ZBAI_B) and the *HAA1* from *S. cerevisiae* BY4741 in *S. cerevisiae* BY4741 and derived deletion mutant *haa1*Δ

3.1.1. Cloning of *HAA1* homologous genes from *Z. bailii* and *Z. bailii*-derived strain and the *HAA1* gene from *S. cerevisiae* BY4741

3.1.1.1. Strains and growth media

S. cerevisiae strains BY4741 (*MATa*, *his3Δ1*, *leu2Δ0*, *met15Δ0*, and *ura3Δ0*) and the derived deletion mutant *haa1*Δ from EUROSCARF Collection were used as hosts for heterologous expression of *HAA1* homologues from CLI213^T, IST302 and ISA1307 (ZBAI_B). *Z. bailii* strains used in this work include *Z. bailii* CLIB213^T isolated from a brewery[14] and provided by Prof. Sylvie Dequin and *Z. bailii* IST302 isolated from a spontaneous fermentation of grape must that was provided by Prof. Alexandra Mendes-Ferreira (Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal). The hybrid strain ISA1307[94] was isolated from a sparkling wine continuous production plant[95] and received from Prof. Cecilia Leão (Universidade do Minho, Braga, Portugal). This last strain is an interspecies hybrid between *Z. bailii* and a closely related species[94]. Yeast cells were maintained in Yeast Peptone Dextrose (YPD) medium (containing, per liter, 20 g glucose (Merck), 20 g bactopeptone (Difco) and 10 g yeast extract (Difco)) or in Mineral Basal (MMB) medium (containing, per liter, 20 g glucose (Merck), 2.65 g $(NH_4)_2SO_4$ (Merck), 1.7 g yeast nitrogen base without amino acids (Difco), 20 mg L-methionine, 20 mg L-histidine, 20 mg L-uracil and 60 mg L-leucine (all amino acids from Sigma)). MMB medium pH adjustment was performed by addition of HCL (1 and 5M) until pH 4.0 or 4.5 was reached. Solid medium include 20 g/L agar (Iberagar) in addition to the above components indicated. MMB medium without uracil (MMB-U) was used to grow yeast cells holding the pGREG506 or the derived recombinant cloning vectors, which required a constant selective pressure to uracil prototrophy given by *URA3* genetic marker present in these vectors.

E. coli XL-1 blue strain (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacIqZ_M15 Tn10* (Tetr)]) was the bacterial strain chosen for maintenance of plasmid constructions as a common model used in molecular biology procedures. *E. coli* cells were maintained in selective Luria Bertani amplicilin (LBA) enriched medium containing, per liter, 10 g tryptone, 5 g yeast extract, 5 g sodium chloride (Pronadisa) and supplemented with

amplicilin (150 mg/mL) since pGREG506 has the ampicillin resistance gene cassette (*AmpR*). Solid medium include 20 g/L agar (Iberagar) in addition to the above mentioned components.

Yeast and *E. coli* strains used in cloning procedures were preserved at -80°C in adequate media enriched with 15% of glycerol (v/v). Prior to use, cells from frozen cultures were transferred to appropriate fresh media agar plates (20 g/L agar); bacterial cells were incubated at 37°C overnight and yeast cells at 30°C until a generous amount of culture was reached.

All culture media were sterilized by autoclaving during 15 minutes at 121°C and 1 atm, except the amino acids and ampicillin solutions that were filter-sterilized and added to the basal medium.

3.1.1.2. Plasmid constructions

The pGREG506 (P_{GAL1} :: *HIS*3; *URA*3) from the DRAG & DROP Collection was the shuttle vector chosen for cloning and expression of the *HAA1* homologues from *Z. bailii* CLIB213^T and IST302 and from the hybrid ISA1307 (ZBAI_B) and the *HAA1* gene from *S. cerevisiae* BY4741 (**figure 7**).

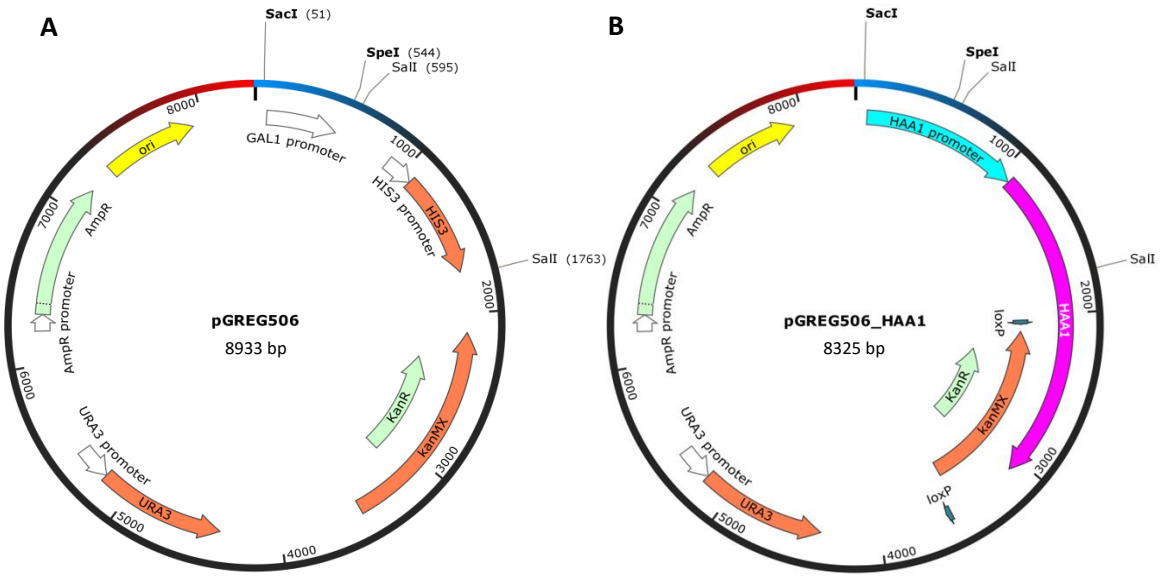

Figure 7. Cloning maps of the pGREG506 vector system before (A) and after (B) the recombination event that led to the insertion of *S. cerevisiae* BY4741 *HAA1* promotor and the *HAA1* homologous genes under study to be expressed from this common promotor. The native pGREG506 *GAL1* promoter was replaced by *HAA1* promoter from *S. cerevisiae* BY4741 and the *HIS3* reporter gene by the *HAA1* ORFs from *Z. bailii* strains CLIB213^T and IST302 and from the hybrid strain ISA1307 (ZBAI_B) and the *HAA1* gene from *S. cerevisiae* BY4741. The used restriction enzymes and the corresponding restriction sites (*SacI*, *SpeI* and *SalI*), as well as the uracil selectable yeast marker (*URA3*) and an additional ampicillin resistance cassette (*AmpR*) for *E. coli* selection are also represented (figure adapted[96]; representation performed using SnapGene® Viewer 2.5).

The incorporation of novel DNA fragments into the cloning vector was based on the efficient homologous recombination system of *S. cerevisiae* that allows the direct exchange of recombinant sequences (*rec* sequences) and these molecules[96]. In this context, to make possible the *in vivo* recombination it is essential the addition of specific homologous sequences to cloning sites flanking regions (*rec* sequences) in amplification primers. All *rec* sequences contain a restriction site to verify at the end of cloning procedure the insertion of the desired sequence into the vector. To allow identical expression levels of the recombinant genes, all the ORFs were expressed under the regulation of the *S. cerevisiae* BY4741 *HAA1* promoter and in the same cloning vector (pGREG506).

The nucleotide sequences of *HAA1* ORFs from CLI213^T, IST302, ISA1307 and the *HAA1* sequence from BY4741 were obtained as follows. The *HAA1* gene sequence from *S. cerevisiae* BY4741 and the respective promoter sequence (comprising 1kb region upstream of the codon START) were retrieved from the *Saccharomyces* Genome Database SGD (http://www.yeastgenome.org/). The coding sequence of the *HAA1* ORF from ISA1307 hybrid strain (ZBAI_B also known as ZBAI_5761) was obtained from Pedant3 database (http://pedant.gsf.de/). The nucleotide sequence of the *HAA1* ORF from *Z. bailii* IST302 was obtained based on the sequencing and annotation of this strain's genome[94], a work recently performed in our laboratory. The *Z. bailii* CLIB213^T *HAA1* ORF was assessed from a Basic Local Alignment Search Tool (BLAST) in the National Center for Biotechnology Information NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and was performed using as query the nucleotide sequence of the *HAA1* ORF from *Z. bailii* IST302. The nucleotide sequence from the *HAA1* ORF from *Z. bailii* strain IST302 (not published) is shown in **figure 22** from **annex I**.

All of the designed primers used to amplify these sequences were analysed on OligoAnalyzer 3.1 (https://eu.idtdna.com/calc/analyzer) and Thermo Scientific tm calculator (http://tinyurl.com/nfgyfsf) tools and purchased from STAB Vida.

3.1.1.3. *GAL1* promoter replacement by *S. cerevisiae* BY4741 *HAA1* promoter

In order to insert the *S. cerevisiae* BY4741 *HAA1* promoter sequence into the pGREG506 cloning vector, the *GAL1* promoter was removed from the vector by a double restriction reaction carried out by *SacI* (5' - GAGCT/C - 3' recognition site) and *SpeI* (5' - A/CTAGT - 3' recognition site) enzymes. The procedure was performed according manufacturer's

specifications (Takara) with addition of Calf Intestinal Alkaline Phosphatase (CIAP) reagent (1 µl/20 µl of total reaction) to prevent vector recircularization. Correct digestion of the cloning vector was confirmed by separation profile of 5 µl of the enzymatic reaction product on agarose gel electrophoresis. The remaining digested and linearized plasmid was reserved for *S. cerevisiae* BY4741 *HAA1* promoter insertion by homologous recombination.

S. cerevisiae BY4741 *HAA1* promoter DNA fragment was obtained by Polymerase Chain Reaction (PCR) amplification using the genomic DNA from *S. cerevisiae* BY4741 as a template and the specific primers listed in **table 2**. Genomic DNA from *S. cerevisiae* BY4741 was previously extracted by *Yeast Genomic DNA Extraction with LiOAc-SDS* protocol[97]. For primers design it was considered the region 1kb upstream to *HAA1* gene, followed by the addition of *rec5* and *rec1* reverse complement nucleotide sequences to forward and reverse primers, respectively.

Table 2. Primers used to amplify *S. cerevisiae* BY4741 *HAA1* promoter. The designed primers contain a region with homology to the ends of *S. cerevisiae* BY4741 *HAA1* promoter sequence (underlined) and a region with homology to the promoter cloning site flanking sequences of pGREG506 vector (not underlined). *SacI* and *SalI* restriction sites are highlighted in red and blue, respectively.

Primer	Recombination and amplification sequences	Length (nt)	Length of fragment (bp)
Forward	5' - AACAAAAGCTGGAGCT/CGTTTAAACGGCGCGCC CCTACTAAATTATCAATCCTTGCAATTAGCTTC - 3'	66	995
Reverse	5' - TG/TCGACGGTATCGATAAGCTTGATATCGAATTC TATTTTTTAGGTTTTTTTTTTCTATACTTTTGTTTCC - 3'	71	

The PCR reaction setup was prepared by using the components and concentrations shown in **table 3**. During the preparation of the PCR reaction all the components were maintained in ice.

Table 3. PCR mixture composition used to *S. cerevisiae* BY4741 *HAA1* promoter amplification.

Reaction component	25 µ reaction	Final concentration
5X Phusion HF Buffer (with 7.5 mM $MgCl_2$)	5 µL	1X
$MgCl_2$ (50 mM)	1 µL	2 mM
DMSO (100%)	0.75 µL	3%
dNTPs (10 mM)	0.5 µL	0.2 mM
Forward primer (10 mM)	0.5 µL	0.2 mM
Reverse primer (10 mM)	0.5 µL	0.2 mM
Template DNA	0.5 µL	variable (1 ng - 1 µg)
Phusion DNA polymerase (2 U/µl)	0.25 µL	0.02 U/µl
Water	16 µL	-

The PCR reaction was run on a C1000 Thermal Cycler (Bio-Rad) and its program was adjusted to 30 cycles (step 2 to 4) and 25 µl of reaction volume (**table 4**). Annealing temperature was optimized for the set of primers used taking into account primers' GC content and the specifications of the high proofreading Phusion DNA polymerase. The extension period was estimated according to the size of the sequence yet to be amplified (995 bp) and the Phusion DNA polymerase deoxyribonucleotide triphosphate (dNTP) incorporation rate. In order to confirm the correct amplification of the *HAA1* promoter the separation pattern of PCR product was analysed on an agarose gel electrophoresis.

Table 4. Thermocycling conditions to PCR amplification of *S. cerevisiae* BY4741 *HAA1* promoter sequence.

Step reaction	Temperature (°C)	Time
1. Initial denaturation	98	30 ''
2. Denaturation (x 30 cycles)	98	10''
3. Annealing (x 30 cycles)	64	20''
4. Extension (x 30 cycles)	72	80 ''
5. Final extension	72	7'
6. Hold	8	-

The recombination between the *S. cerevisiae* BY4741 *HAA1* promoter and the previously obtained pGREG506 without the *GAL1* promoter was performed by transformation of both the linearized plasmid and *HAA1* promoter region in *S. cerevisiae* BY4741 using the Alkali-Cation™ Yeast Transformation Kit (MP Biomedicals) and MMB-U as a selection medium. The obtained colonies were streaked in fresh MMB-U solid medium plates in order to obtain enough biomass to extract DNA from yeast transformants through the *Yeast total DNA Rapid Extraction* protocol that was performed as follows. Total DNA from yeast cells were extracted by resuspending a generous loop of biomass from positive candidates in 200 mL of lysis buffer (50 mM Tris, 50 mM EDTA, 250 mM NaCl and 0.3% SDS), followed by the addiction of 100 µL of glass beads and a short spin of 2 minutes in vortex at maximum speed in order to destroy yeast cell membranes and release cellular components. The sample was then incubated at 65°C during 1 hour and immediately put on ice for 2 minutes. Yeast material was centrifuged for 15 minutes at 13000 rpm and 4°C in the Centrifuge 5810 R (Eppendorf) and the obtained supernatant recovered. A total of 20 µL sodium acetate 3 M at 4.8 pH and 440 µL of ethanol absolute were added to the sample and the new mixture was incubated during 20 minutes at 20°C to induce plasmid DNA precipitation. Afterwards, the sample was centrifuged for 20 minutes at 13000 rpm and 4°C and the formed supernatant was removed. The remaining pellet was washed with 70% ethanol, dried in speed vacuum Concentrate Plus (Eppendorf) at 45°C during 15 minutes and finally resuspended in 100 µL of water. This DNA was used as DNA

template to *HAA1* promoter amplification by PCR reaction in order to validate the success of recombination. To ensure the exclusive amplification of *HAA1* promoter cloned into the pGREG506 vector, instead of genomic yeast *HAA1* promoter copy, it was used a specific set of primers, one from *HAA1* promoter (reverse primer) (**table 2**) and another from pGREG506 vector (pGREG506_PROM_up (5' - GCGCGCAATTAACCCTCA - 3')). Amplification reaction was performed according to conditions expressed in **tables 3** and **4**, except the annealing temperature that was reduced to 58°C due to forward primer specificities.

Total DNA from the positive candidates was used to transform *E. coli* by electroporation. An aliquot with 100 µl of electrocompetent *E. coli* XL-1 Blue cells, maintained at -80°C, were thawed on ice and 20 µl of DNA was added followed by gentle pipette mixing. The mixture was transferred to an electroporation icy cuvette and inserted in Bio-Rad Gene Pulser™ Electroporation Porator system, being submitted to 2.5 kV, 400 Ω and 25 µFD during a short time pulse. 1 mL of LB liquid medium was added to the electroporated cells and the mixture was incubated with orbital agitation (250 rpm) at 37°C during 1 hour. Bacterial cells were pelleted by centrifugation at 1300 rpm during 2 minutes and spread in LBA solid medium plate followed by overnight incubation at 37°C. The new colonies were transferred to fresh LBA solid medium plates in order to obtain enough bacteria biomass for plasmid DNA extraction according to the *QIAprep Spin Miniprep* Kit (250) (QIAGEN) manufacturer's instructions. The positive candidates were confirmed again by plasmid digestion with *SacI* and *SalI* restriction enzymes (**figure 8**) and PCR amplification with *S. cerevisiae* BY4741 *HAA1* promoter primers (**table 2**) followed by agarose gel electrophoresis. Since *SpeI* restriction site is upstream of *SalI* restriction site, when homologous recombination occurs the first sequence is eliminated remaining the *SalI* enzyme site to confirm the success of *HAA1* promoter sequence insertion (**figure 8**).

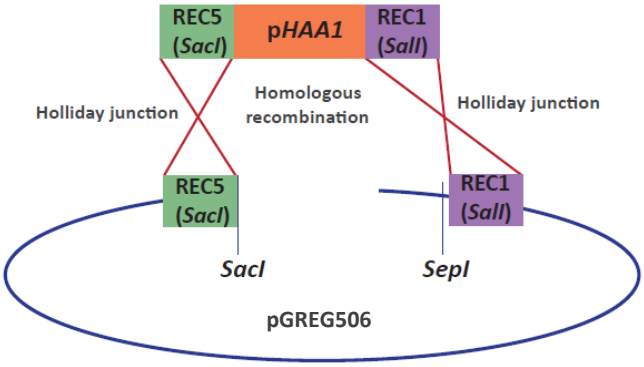

Figure 8. Schematic representation of *HAA1* promoter (p*HAA1*) cloning strategy in the pGREG506 vector. Homologous recombination between corresponding *rec* sequences (regions of the same colour) are herein represented by Holliday junctions. *SacI* and *SpeI* enzymes used to linearize vector are also showed.

In fact, the use of *SalI* enzyme to cut the empty vector, instead of *SpeI*, seems more obvious. Nevertheless both the restriction enzymes can be used to cut the vector without affecting cloning efficiency. The correct integration was finally confirmed by sequencing the obtained plasmid[98].

3.1.1.4. *HIS3* replacement by *HAA1* coding sequence

The deletion of *HIS*3 gene from the PGREG506 cloning vector harbouring the *S. cerevisiae* BY4741 *HAA1* promoter was performed by a restriction reaction carried out by *SalI* enzyme (5' - G/TCGAC - 3' recognition site) according manufacturer's specifications (Takara). Part of the digested product obtained was focused by agarose gel electrophoresis to confirm *HIS*3 elimination and the remaining was reserved to the next cloning procedures.

To amplify the *HAA1* nucleotide sequences by PCR amplification, genomic DNA from *S. cerevisiae* BY4741, *Z. bailii* CLIB213^{T} and IST302 and hybrid ISA1307 was extracted according the *Classical Phenol DNA Extraction* protocol[99] to be used as DNA template. Specific primers for each *HAA1* ORF under study were designed and *rec1* and *rec2* recombinant sequences incorporated to the forward and reverse primers, respectively (**table 5**). Since ISA1307 is a *Z. bailii*-derived interspecies hybrid strain it holds two *HAA1* copies, encoding two Haa1 proteins (with different size). In this work it was only investigated the ORF ZBAI_B, instead of the ORF ZBAI_A (also known as ZBAI_01494), because the first one, although encoding a smaller protein, has a higher identity percentage when compared to the corresponding protein from type strain *Z. bailii* CLIB213^{T}[14].

Table 5. Primers used to amplify the nucleotide sequences of *HAA1* ORFs from *Z. bailii* CLIB213^T and IST302, the ZBAI_B ORF from hybrid ISA1307 and the *HAA1* from *S. cerevisiae* BY4741 species. The designed primers contain a region with homology to the ends of *HAA1* homologous genes (underlined), and a region with homology to the cloning site flanking sequences of pGREG506 cloning vector (not underlined). *SalI* restriction site are highlighted in green.

Species	Annealing region	Recombination and amplification sequences	Length (nt)	Length of fragment (bp)
***S. cerevisiae* BY4741**	Forward	5' – GAATTCGATATCAAGCTTATCGATACCG/TCGACA ATGGTCTTGATAAATGGCATAAAGTATGCC - 3'	64	2085
	Reverse	5' – GCGTGACATAACTAATTACATGACTCGAGG/TCGAC TCATAACGAAGACATGAAATTATCCAAATCC - 3'	66	
***Z. bailii* CLIB213^T**	Forward	5' – GAATTCGATATCAAGCTTATCGATACCG/TCGACA ATGGTGTTGATAAACGGTGTCAAG - 3'	58	2085
	Reverse	5' – GCGTGACATAACTAATTACATGACTCGAGG/TCGAC TCACAATGTGGACATGAAATCG - 3'	57	
***Z. bailii* IST302**	Forward	5' – GAATTCGATATCAAGCTTATCGATACCG/TCGACA ATGGTGTTGATAAACGGTGTCAAG - 3'	58	2085
	Reverse	5' – GCGTGACATAACTAATTACATGACTCGAGG/TCGAC TCACAATGTGGACATGAAATCG - 3'	57	
Hybrid ISA1307	Forward	5' – GAATTCGATATCAAGCTTATCGATACCG/TCGACA ATGATGATTAAGCCGAAGGGGAG - 3'	57	1995
	Reverse	5'- GCGTGACATAACTAATTACATGACTCGAGG/TCGAC TCACAATGTGGACATGAAATCGTCC - 3'	60	

The PCR reaction components, the volume amounts per 25 µl of reaction and the final concentrations used were the same as defined for the *S. cerevisiae* BY4741 *HAA1* promoter amplification (**table 3**), as well as the PCR program (**table 4**), except the annealing temperature (62°C) and extension time (90''). These two parameters were optimized for each set of primers as described for the *HAA1* promoter amplification. Due to unspecific PCR amplification of the ZBAI_B ORF from hybrid strain ISA1307 the corresponding PCR product was submitted to an agarose gel electrophoresis with the purpose of obtaining only the PCR product with the desired size by extracting the corresponding band and purifying the DNA from the gel using the *Gel Extraction Spin* (250) Kit (GENOMED).

pGREG506 harbouring the *S. cerevisiae* BY4741 *HAA1* promoter, previously obtained, was used for a second round of recombination events in which the *HAA1* ORF sequences amplified were inserted into this cloning vector. The transformation protocol as well as the next procedures to select yeast positive candidates and confirm the insertion of each *HAA1* ORF under study, with specific differences reported to *HAA1* ORFs intrinsic proprieties, are the same as described to

the *HAA1* promoter. The differences include the primers used to amplify each *HAA1* ORF from yeast positive transformants (forward primer (**table 5**) and reverse primer pGREG506_REV (5' – GTCGACCTCGAGTCATG TAA – 3')); and *E. coli* positive candidates (**table 5**). Moreover, the restriction enzyme used to confirm the *HAA1* ORFs correct insertion, was also different (*SalI*).

Quality and concentration of purified DNA from several approaches were analysed by agarose gel electrophoresis profile and Nanodrop Spectrophotometer ND-1000 (Alfagene) values. Plasmid DNA extracted from *E.coli* positive candidates carrying the recombinant vector was used to sequencing the *S. cerevisiae HAA1* promoter and *HAA1* ORFs under study at STAB Vida.

3.1.1.5. Transformation of *S. cerevisiae* BY4741 and derived mutant *haa1*Δ with the recombinant vectors

The *S. cerevisiae* BY4741 parental and the derived mutant *haa1*Δ strains were transformed with the vectors previously obtained (**table 6**) thought the *High-efficiency Yeast Transformation Method* using the LiAc/SS carrier DNA/PEG method[100]. The *S. cerevisiae* BY4741 and the derived mutant *haa1*Δ strains transformed with the pGREG506 vector without the *HIS*3 gene (gently granted by BSRG colleagues) were used as a control since the recombinant plasmids also loose the *HIS*3 gene active mark.

Table 6. Summary of plasmid constructions obtained in this work. All the vectors express the *HAA1* promotor from *S. cerevisiae* BY4741, however in order to simplify the nomenclature this information was suppressed.

Vector construction	Gene inserted	Selection mark in *S. cerevisiae*	Selection mark in *E. coli*
pGREG506_*Zb*CLIB213^T_*HAA1*	*HAA1* ORF from *Z. bailii* CLIB213^T	*URA*3	*AmpR*
pGREG506_*Zb*IST302_*HAA1*	*HAA1* ORF from *Z. bailii* IST302		
pGREG506_ISA1307_ZBAI_B	ZBAI_B ORF from hybrid ISA1307		
pGREG506_*ScHAA1*	*HAA1* from *S. cerevisiae* BY4741		

3.1.2. Acetic acid susceptibility assays to compare the tolerance of different recombinant strains

The *HAA1* protective effect towards acetic acid was tested by susceptibility assays in solid and liquid media. The parental strain *S. cerevisiae* BY4741 and the derived mutant *haa1*Δ

previously transformed with the recombinant vectors were cultivated in MMB-U containing, or not, inhibitory concentrations of acetic acid. This analysis of cultures growth was performed either by spot assays, in which the yeast growth pattern was assessed at the surface of agarized medium, or following the yeast growth in liquid medium by measuring culture OD_{600nm}. Both assays were performed at a pH slightly below to the acetic acid pKa value (pH 4.5 or 4 for spot or growth assays, respectively) in order to avoid acid dissociation.

Yeast cells suspensions used as inoculum in spot assays were prepared from batch cultivated cells in the mid-exponential phase of growth in MMB-U (pH 4.0) medium, at 30°C, with orbital agitation (250 rpm) until culture's OD_{600nm} = 0.5 ± 0.05 was reached, followed by dilution in sterile water to an OD_{600nm} = 0.05 ± 0.0005. These suspensions and the subsequent dilutions (1:5 and 1:10) were spotted as 4 µL drops into the surface of MMB-U (pH 4.5) agarized medium either or not supplemented with acetic acid (60 and 70 mM). The plates were incubated at 30 °C during 3-6 days, depending on the severity of inhibition.

Yeasts cells used to compare growth were also cultured in MMB-U (pH 4.0) liquid medium, at 30°C with orbital agitation (250 rpm), until mid-exponential phase (OD_{600nm} = 0.5 ± 0.05) was reached and then reinoculated after removal of cells by filtration in 50 mL of fresh medium, supplemented or not with 60 and 80 mM of acetic acid, to achieve an initial OD_{600nm} of 0.05 ± 0.05. Cells were batch cultivated at 30°C with orbital agitation (250 rpm) and growth was followed by measuring culture OD_{600nm}.

3.2. Evolutionary and comparative analysis of the Haa1 and its homologue Ace1 in Hemiascomycetes

3.2.1. Identification of *S. cerevisiae* Haa1 and Ace1 homologues in Hemiascomycetes

Amino acid sequences from Haa1 and Ace1 homologues were retrieved from the local genome database MySQL developed by the BSRG, which contains 172 hemiascomycetous yeast genomes from 68 yeast species, using the BLASTP (Protein-Protein BLAST) algorithm that compare the translated ORFs all-against-all. The Haa1 sequence from *S. cerevisiae* S288c was used as 'starting node' for traversing the BLASTP network constrained at different e-value thresholds, ranging from E-60 to E-0. This procedure contributed to gain further insights about

the homologue sequence clusters which comprise the Haa1/Ace1 protein family, as well as eventual sub-networks belonging to other protein families that show distant homology with the Haa1. The e-value of E-13 was chosen as threshold. Sequences corresponding to frameshift mutations were removed from the analysis. The Haa1 and Ace1 homologues were analysed according to one of the three phylogenetic analyses performed to the phylogenetic groups of yeasts: i) Hemiascomycetes; ii) Saccharomycetaceae family; iii) and *S. cerevisiae* strains. Concerning the analysis limited to Saccharomycetaceae the amino acid sequences represented (in number) by more than two strains, from the same species, were also removed as a way to avoid repeated sequences that does not add novel information. For the analysis performed to Hemiascomycetes, it was considered the initial portion of N-terminal domain of Haa1 and Ace1 homologues defined using the first 65 amino acid residues of the Haa1 protein from *S. cerevisiae* S288c as the cut-off mark in the multiple alignment.

3.2.2. Alignment of *S. cerevisiae* Haa1 and Ace1 homologous proteins and phylogenetic trees construction

The amino acid sequences identified as homologues of the *S. cerevisiae* Haa1 and Ace1 proteins were used to construct a multiple alignment using MUSCLE (Multiple Sequence Comparison by Log-Expectation) algorithm[101]. The obtained fasta file was imported to the Jalview software environment in order to compare the amino acid differences and similarities among the aligned protein sequences. The alignment step was followed by processing using the PHYLIP (PHYLogeny Inference Package) phylogenetic software suite to the construction of the phylogenetic trees according to the amino acid sequences conservation among sequences under study[102]. The analysis performed to hemiascomycetous species which involved the Haa1 and Ace1 homologues N-terminal domain, as well as the analysis accomplished to *S. cerevisiae* strains that involved the Haa1 homologues complete sequences were processed using the PROTDIST/NEIGHBOUR algorithm; the analysis performed to Saccharomycetaceae family focused on the Haa1 and Ace1 homologues complete sequences was processed using the PROML algorithm. The previous approaches are based on distance and maximum likelihood methods, respectively. Both methods use heuristic approaches to calculate the best phylogenetic tree in which the initial tree returned, calculated by a fast algorithm that simultaneously compare all the sequences from multiple alignment, is continuously improved by local arrangements until best scored is attained[70]. All phylogenetic trees obtained in this study were visualized in Dendroscope software[103].

Once this analysis consists in a preliminary study of the Haa1 and its homologue Ace1 evolutionary history, the entire dataset was not analysed at the detail level necessary to resolve some dubious cases regarding fragment and frameshift sequences that can be responsible for misrepresented results. In this context, we opted to classify the different groups identified in the phylogenetic trees (which do not reflect the tree topology) as 'groups' instead of 'clusters'.

3.2.3. Gene neighbourhood analysis of the *S. cerevisiae HAA1* and *ACE1* homologous genes in Saccharomycetaceae

The analysis of the chromosome environment where the *S. cerevisiae HAA1* and *ACE1* homologous genes reside was based on the fifteen neighbour genes immediately down and upstream to these genes in Saccharomycetaceae. Neighbour genes were retrieved from the local genome database (describe above) using the package 'sqldf' and complementary scripting in R language. An e-value threshold of E-50 was used to define clusters of protein similarity. The synteny output consisted on the homologues of the *S. cerevisiae HAA1* and *ACE1* genes (nodes) connected by pairs of neighbour genes (edges). The Cytoscape software was used to visualize and analyse the network. Synteny relationships are represented by the amount of shared neighbour pairs between two query genes (given by the number of connections between two specific nodes). The use of Cytoscape to assess the synteny lineages is supported by a dynamic framework that allows to manually explore the network through the selection of subnetworks and the application of filters. This approach is a powerful tool able to discriminate the orthologue/paralogue status of the majority of the *S. cerevisiae HAA1* and *ACE1* homologous genes encoded in the genome of the post-WGD yeast species. Biological information imported into the Cytoscape (as edges attributes) contributed to enhance network understanding since it enabled to determine the strength of each homologous neighbour pair. This additional information included the sequence clusters of neighbour genes, the frequency of the neighbour genes per sequence cluster, the chromosome position of a given neighbour gene in relation to a specific *HAA1* and *ACE1* homologue and the percentage of identity and similarity between homologous neighbour genes (obtaining using the BLASTP and Needleman-Wunsch alignment algorithms). The results obtained allowed to reconstruction the *HAA1* and *ACE1* lineage in yeast species of the Saccharomycetaceae, as well as the neighbour genes order in the chromosome.

4. Results

4.1. Effect in tolerance to acetic acid stress of the expression of *HAA1* from strains CLIB213^T, IST302 and ISA1307 in *S. cerevisiae* BY4741 and derived deletion mutant *haa1Δ*

4.1.1. Cloning of the *S. cerevisiae HAA1* and *HAA1* homologues from *Z. bailii* strains (*ZbHAA1*) or *Z. bailii*-derived hybrid strain

In order to compare the increase of acetic acid tolerance conferred by the expression of *HAA1* ORFs from *Z. bailii* CLI213^T, IST302 and the hybrid ISA1307 (ZBAI_B) with the expression of *S. cerevisiae HAA1*, their corresponding nucleotide sequences were amplified using the genome DNA of these strains (obtained as mentioned in Materials and Methods section). These sequences were cloned into the pGREG506 vector under the control of the *HAA1* promoter from *S. cerevisiae* (previously inserted into the pGREG506 cloning vector). Recombinant plasmids were introduced into the parental strain BY4741 or in the derived deletion mutant *haa1*Δ and the new recombinant strains submitted to acetic acid susceptibility assays. The *HAA1* from *S. cerevisiae* BY4741 was cloned in the same expression vector, under the control of its native promoter, as an internal control for expression comparison experiments. Figure 9 shows a schematic representation of the strategy used.

The replacement of the *GAL1* promoter by the *HAA1* promoter from *S. cerevisiae*, as well as the replacement of the *HIS3* gene by the *HAA1* ORFs from *Z. bailii* CLI213^T, IST302, the ZBAI_B ORF from hybrid ISA1307 and the *HAA1* from *S. cerevisiae* BY4741 in the pGREG506 cloning vector were verified based on the DNA electrophoresis profiles obtained for the PCR amplification of DNA inserts and the restriction digestion of the recombinant vectors constructed (according to the procedures in Materials and Methods section) (results not shown). The confirmation of the correct insertions was performed by sequencing the DNA inserts cloned into the pGREG506 cloning vector[98].

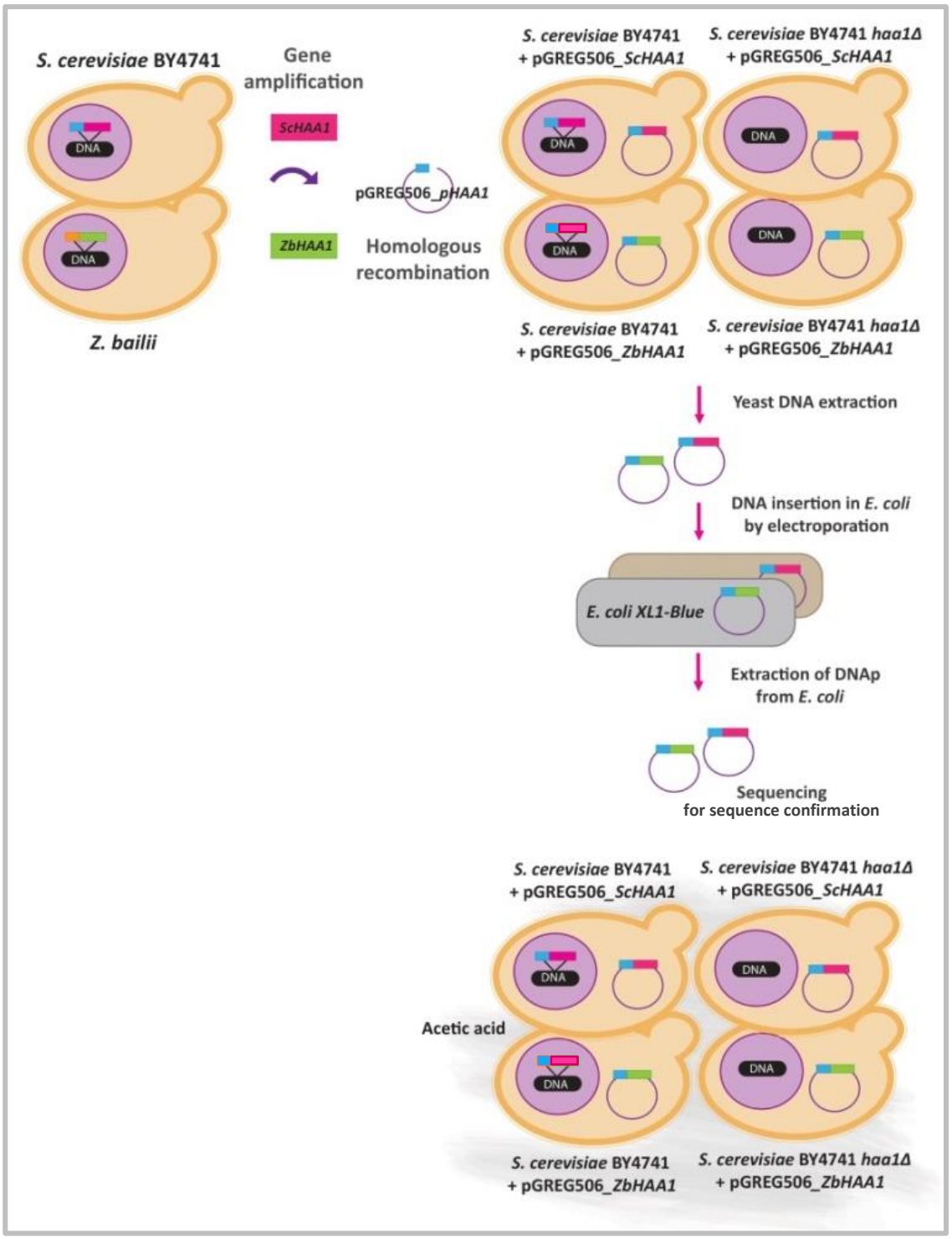

Figure 9. Schematic overview of the cloning strategy used to obtain the *S. cerevisiae* BY4741 and the derived deletion mutant *haa1*Δ strains expressing the *HAA1* ORFs from *Z. bailii* CLIB213^T, IST302 and from hybrid strain ISA1307 (ZBAI_B) and the *HAA1* gene from *S. cerevisiae* BY4741, subsequently submitted to acetic acid susceptibility assays. The *HAA1* cloning approach, as well as the subsequent procedures to verify its successful insertion, are here shown. To simplify the scheme the *S. cerevisiae* BY4741 *HAA1* promoter cloning procedures were omitted.

The protein sequence of Haa1 from *S. cerevisiae* S288c and the Haa1 homologous sequences from the hybrid strain ISA1307 were retrieved from the databases SGD and Pedant3, respectively. The sequence of the Haa1 homologue from *Z. bailii* IST302 was obtained from the translation of the corresponding nucleotide sequence that was obtained by our laboratory following the full sequencing and annotation of IST302 genome. The sequence of Haa1 homologue from *Z. bailii* CLIB213^T was obtained using the IST302 Haa1 homologous sequence as the query sequence in a BLASTP search. In order to compare the amino acid similarity among these sequences, a multiple alignment was performed with Geneious software (**annex II**, **figure 23**).

The comparison of Haa1 homologous sequences from *Z. bailii* CLIB213^T and IST1302 (*Zb*Haa1) and the *S. cerevisiae* Haa1 evidenced high amino acid conservation in the N-terminal domain that is essentially maintained for the first 100 amino acid residues. Such conservation was not observed at the level of the C-terminal domain, which considerably differs among *Zb*Haa1 homologues and the Haa1 from *S. cerevisiae* (the gap percentage is around 20.9%, **annex II**, **table 13**). Indeed, *Zb*Haa1 homologues only share around 37% of identity with the Haa1 sequence from *S. cerevisiae*, which contrasts with the high identity percentage between the two *Zb*Haa1 homologues under study (99.6% of identity) (**annex II**, **table 14**). The amino acid residues variations among the *Zb*Haa1 homologues are highlighted in **table 15** of **annex II**. Due to the high differences found between the *Zb*Haa1 homologues and the Haa1 from *S. cerevisiae* at the level of C-terminal domain the amino acid variation at specific positions among these sequences is not compared here.

A prediction for the presence of phosphorylation sites in *S. cerevisiae* Haa1 and *Zb*Haa1 homologues was also performed. A considerable difference was found regarding the number and position of amino acid residues predicted to be phosphorylatable among Haa1 homologues from *Z. bailii* strains and the Haa1 from *S. cerevisiae* (**annex II**, **figure 24** and **table 16**). However, the number and positions of predicted phosphorylatable residues are conserved between the *Zb*Haa1 homologues, which is expected since these sequences share a high amino acid similarity.

As mentioned above, the hybrid strain ISA1307 possesses two *HAA1* ORFs, one derived from the *Z. bailii* species and the other derived from a *Z. bailii*-related species. These ORFs will be referred here as ZBAI_A and ZBAI_B. Based on the identity percentage given by pairwise alignments of these two ORFs with *ZbHAA1* homologues (**annex II, table 14**), it was

hypothesized that the ZBAI_B is the *Z.bailii*-derived ORF, although 90 nucleotides are missing in its sequence at the N-terminal domain when the ZBAI_A sequence is compared to the *ZbHAA1* and *ScHAA1* sequences. In order to investigate the reason behind the annotation of a shorter ZBAI_B ORF, this ORF was aligned with scaffold 15 (that includes ZBAI_A ORF) and scaffold 17 (that includes ZBAI_B ORF) using Geneious software (**figure 10**). Results indicate that the two scaffolds are relatively well conserved (95.5% of identity), but it was found that scaffold 15 has a single nucleotide difference (ACG) at the position corresponding to the START codon (ATG) of ZBAI_A in scaffold 17. This nucleotide variation led to the annotation of a shorter ZBAI_B ORF.

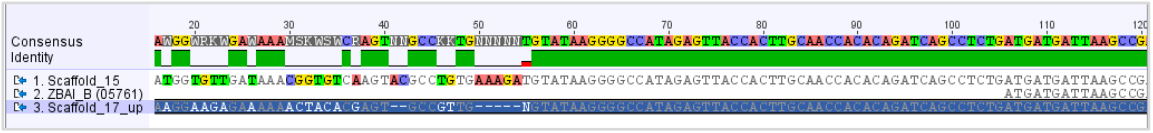

Figure 10. Alignment of the first 120 nucleotide residues of scaffolds 15 and 17 and the initial region of HAA1 ORF sequence from hybrid ISA1307 (ZBAI_B). Results were obtained in Geneious software.

In this work, it was decided to analyse solely the role of ZBAI_B ORF hypothesized as being the *Z. bailii*-derived *HAA1* homologue and also because the analysis of this shorter ORF would provide some indications about the importance of the missing N-terminal domain (encoding the first 30 amino acids residues of the protein) in the role of this Haa1 homologue. Amino acid conservation of the Haa1 homologues encoded by ZBAI_A and ZBAI_B ORFs was also examined. The amino acid similarity among ISA1307 Haa1 and *Zb*Haa1 homologues is higher in comparison to the Haa1 from *S. cerevisiae* (**annex II**, **table 14**). This relationship was also observed at the level of the number and position of predicted phosphorylatable residues among Haa1 homologues from ISA1307 and from *Z. bailii* strains in study and the Haa1 from *S. cerevisiae* BY4741 (**annex II**, **figure 24** and **table 16**).

4.1.2. Comparison of the acetic acid susceptibility of *S. cerevisiae* cells expressing the Haa1 proteins from *S. cerevisiae* BY4741 or *Z. bailii* and *Z. bailii*-derived strains

In order to examine the hypothesized role of *ZbHAA1* from CLIB213^T and IST302 strains and the effect of ZBAI_B ORF from hybrid strain ISA1307 in acetic acid tolerance, the *S. cerevisiae* BY4741 parental strain and derived deletion mutant *haa1*Δ were used as cell hosts where the effect of the expression of each *ZbHAA1* gene under study and the ZBAI_B ORF from ISA1307 were experimentally compared with the *HAA1* from *S. cerevisiae* BY4741 (used as an internal control). For this, all the *HAA1* homologues under study and the *HAA1* from *S. cerevisiae* were cloned into the same expression plasmid (pGREG506) under the control of a common promoter (the *HAA1* promoter from *S. cerevisiae* BY4741). A spot analysis was performed in MMB-U solid medium (pH 4.5) supplemented or not with acetic acid. In parallel, the growth of the same strains in MMB-U liquid medium (pH 4.0) supplemented or not with acetic acid was followed by measuring culture optical density (OD_{600nm}).

Results from spot assays showed that the expression of *ZbHAA1* from CLIB213^T and from IST302, as well as the *HAA1* from *S. cerevisiae* BY4714, complemented the *S. cerevisiae haa1*Δ mutant acetic acid susceptibility phenotype (**figure 11 B**), as well as increased the acetic acid tolerance of the *S. cerevisiae* BY4741 parental strain (**figure 11 A**). Moreover, *S. cerevisiae* cells expressing the *ZbHAA1* genes from the two strains under study led to a higher increase of the levels of tolerance to acetic acid of the host yeast cells when compared to the *S. cerevisiae* cells expressing the *HAA1* gene from *S. cerevisiae* BY4741. However, the complementation of the *haa1*Δ mutant phenotype when yeast cells expressed the ZBAI_B ORF from ISA1307 was not observed (**figure 11 B**) and the expression of this ORF in the parental strain did not improve its tolerance to acetic acid (**figure 11 A**). The results clearly indicate that the cloned ZBAI_B ORF is not functional considering the hypothesized function of this homologue to *S. cerevisiae* Haa1. It is likely that the ZBAI_B of ISA1307 (which encodes a protein with less 30 amino acids compared to *Zb*Haa1 and *Sc*Haa1) is the reason for such lack of function since there are no indications, at this time, that the expression of the predicted complete ORF is not able to confer acetic acid tolerance.

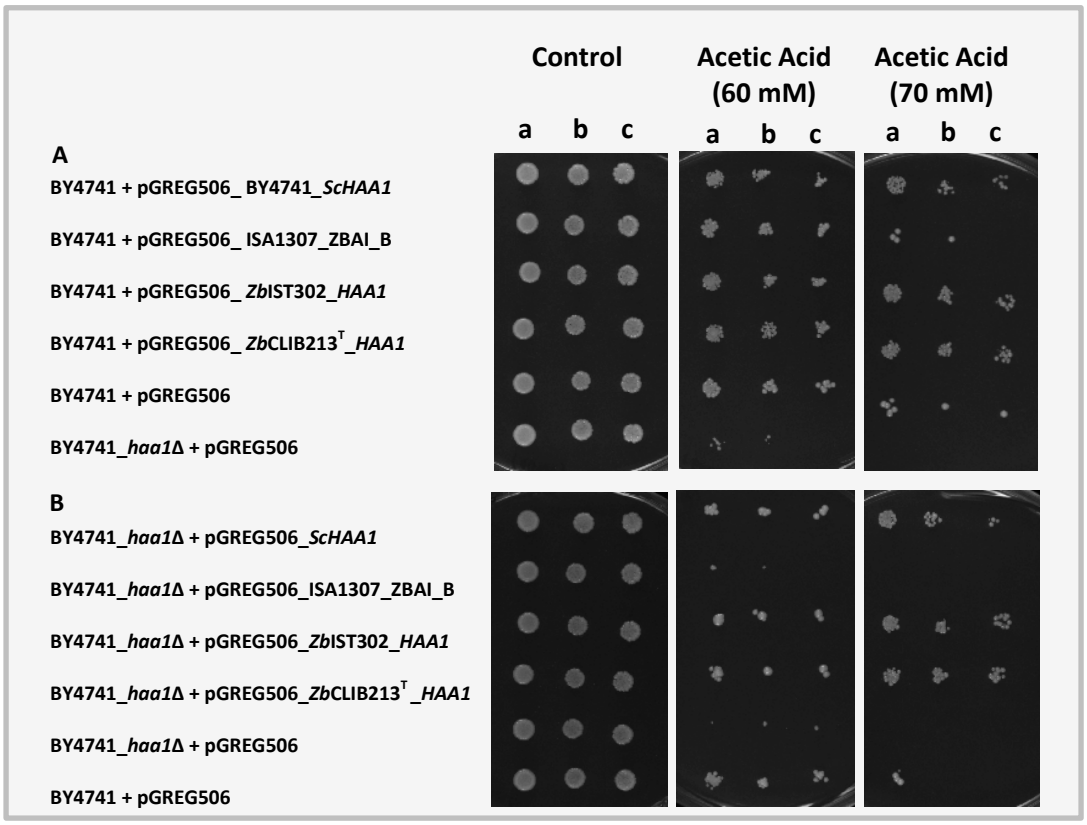

Figure 11. Results from spot assays to assess the susceptibility to acetic acid of parental strain *S. cerevisiae* BY4741 (A) and derived deletion mutant *haa1*Δ (B) expressing the *HAA1* homologues from *Z. bailii* CLIB213^T or IST302, or the ISA1307 ZBAI_B ORF or the *S. cerevisiae* BY4741 *HAA1* gene. Yeast cells were grown in MMB-U liquid medium (pH 4.0) until mid-exponential phase (OD_{600nm} = 0.5 ± 0.05) was reached and used to prepare the initial cells suspension (a) from which diluted suspensions (1:5 (b), 1:10 (c) dilutions) were spotted onto the MMB-U solid medium (pH 4.5) supplemented with acetic acid (60 or 70 mM). Parental and derived *haa1*Δ mutant strains carrying the expressing vector without the histidine marker (as referred in Materials and Methods section) were used as the negative control. The results shown were obtained after 5 days of incubation at 30°C of spotted cultures and are representative of one of three independent experiments.

Regarding acetic acid susceptibility assays performed in liquid medium it was observed the complementation of the *S. cerevisiae haa1*Δ mutant phenotype in mutant cells expressing the *ZbHAA1* from CLIB213^T or IST302 or the *HAA1* from *S. cerevisiae* (**figure 12 B**), as registered in solid medium for the same strains (**figure 11 B**). Mutant cells expressing the *HAA1* from these *Z. bailii* strains showed a latency period of 39.5 hours and those carrying the *ScHAA1* of 43 hours at 80 mM of acetic acid. These latency periods contrast with the increased duration of the latency phase (of around 100 hours) observed for the mutant strain carrying the ZBAI_B ORF from ISA1307 (**figure 12 B**). At the same acetic acid concentration the severity of this inhibitor was much higher for the *haa1*Δ mutant carrying the empty vector that even after 336 hours (14 days) of incubation did not grow (**figure 12 B**). When the parental strain was transformed with expression vector cloned with the *ZbHAA1* from CLIB213^T or IST302 or the

HAA1 from *S. cerevisiae* BY4741 the lower latency periods were 40.5, 38.5 and 43 hours in 80 mM of acetic acid, respectively, confirming the more marked positive effect in acetic acid tolerance conferred by the *ZbHAA1* from CLIB213^T or IST302 when compared to the *ScHAA1*. The parental strain expressing the ZBAI_B from ISA1307 or the empty vector at 80 mM showed more extended latency periods (69 and 65 hours, respectively) (**figure 12 A**). Although, the role of the expression of the *HAA1* homologue from ISA1307 towards acetic acid was negligible in the acetic acid tolerance of the parental strain, compared with the parental strain carrying the empty vector (**figure 12 A**), the effect in growth improvement under acetic acid stress was observed in the *haa1*Δ mutant strain harbouring this gene in comparison to *haa1*Δ mutant strain harbouring the cloning vector (**figure 12 B**).

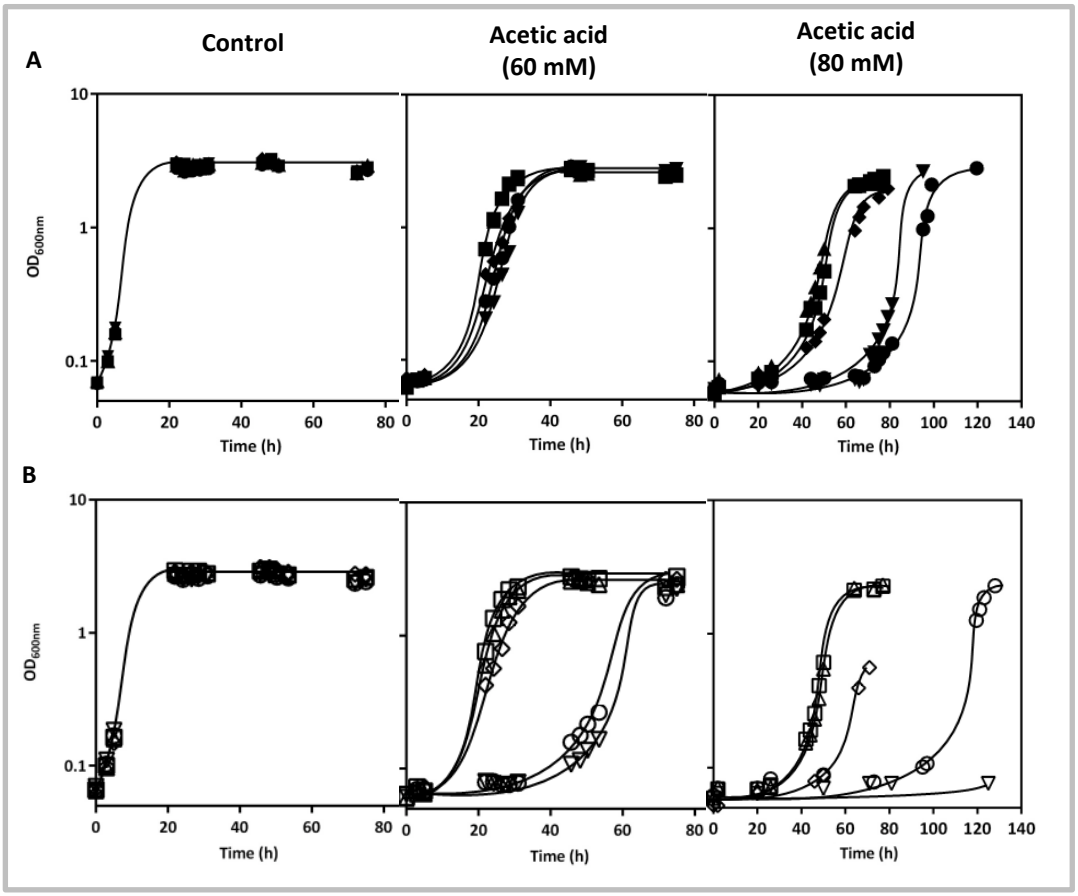

Figure 12. Growth curves obtained by measuring the OD_{600nm} of parental strain *S. cerevisiae* BY4741 (A) and derived deletion mutant *haa1*Δ (B) expressing an *HAA1* ORF from *Z. bailii* CLIB213^T or IST302, hybrid ISA1307 (ZBAI_B) or *S. cerevisiae* BY4741, in culture medium either or not supplemented with acetic acid (60 and 80 mM). The strains BY4741 (closed symbols) and derived deletion mutant *haa1*Δ (open symbols) expressing the cloning vector pGREG506 (▼;▽); or the recombinant plasmids pGREG506_*Zb*CLIB213^T_*HAA1* (■;□); pGREG506_*Zb*IST302_*HAA1* (▲;△); pGREG506_ISA1307_ZBAI_B (●;○); pGREG506_*ScHAA1* (◆;◇) are here tested. Cells suspensions used to reinoculate an acetic acid enriched medium were grown in MMB-U medium (pH 4.0) in the absence of this inhibitor until exponential phase. Parental and derived *haa1*Δ mutant strains carrying the vector without the histidine marker were used as the negative control. Presented results are representative from two independent assays that give rise to identical growth curves.

The negative effect detected in the parental strain expressing the ZBAI_B from ISA1307 is possibly related with the physiological burden associated to the expression of a non-functional protein in *S. cerevisiae* expressing *HAA1*. At 80 mM of acetic acid, the cellular aggregation of BY4741 and the derived mutant *haa1*Δ cells expressing the recombinant plasmids was observed, a phenomenon not registered at 60 mM of acetic acid or in the absence of acid (control conditions). Cellular aggregation has been described as a strategy to counteract environmental stress[104] and in this particular case can be attributed to the high acetic acid concentration which yeast cells were exposed (80 mM). Notably, growth curves seem to fit in a very representative way the yeast growth profile in the presence of acetic acid inhibitor.

Concerning the maximal biomasses achieved at the stationary phase of growth (corresponding to the maximal OD_{600nm} measured) for the BY4741 parental strain and the corresponding *haa1*Δ mutant strain carrying the *ZbHAA1* from CLIB213^T or IST302, no significant differences were observed. Parental and derived mutant *haa1*Δ cells growing in the presence of acetic acid and carrying an extra *HAA1* copy (from the cloning vector) showed a lower biomass at stationary phase in comparison to the same strains growing in control conditions and also to the parental strain carrying the empty vector growing in acid acetic conditions (**table 7**). Although the positive effect of the different *HAA1* functional sequences in the specific growth rate under acetic acid stress was observed, the experimental data were considered insufficient to calculate the corresponding values.

Table 7. Duration of the latency period and maximum OD_{600nm} attained for the parental and derived *haa1*Δ mutant strains expressing the *HAA1* homologous genes under study and the *HAA1* from *S. cerevisiae* in control conditions and in the presence of acetic acid. Values are reported to yeast cell suspensions cultivated in liquid medium MMB-U (pH 4) in the absence (control conditions) or presence of acetic acid (80 mM). The maximum OD_{600nm} was measured at stationary phase. Results are representative from at least two independent assays (with the exception of results gathered for the BY4741 + *haa1*Δ_*ScHAA1* strain that result from a single assay).

Condition	Strain	Latency period (h)	Final OD_{600nm}
Control (without acid)	BY4741 + pGREG506	(2.5±0.87)	2.70 ± 0.06
	BY4741_*haa1*Δ + pGREG506	(2.5±0.87)	2.67 ± 0.05
Acetic acid (80 mM)	BY4741 + pGREG506	65	2.64 ± 0.03
	BY4741 + pGREG506_*Zb*CLIB213^T_*HAA1*	40.5±4.5	2.35 ± 0.02
	BY4741 + pGREG506_*Zb*IST302_*HAA1*	38.5±3.5	2.29 ± 0.13
	BY4741 + pGREG506_ISA1307_ZBAI_B	69±5	2.60 ± 0.23
	BY4741 + pGREG506_Sc*HAA1*	43±3	2.28 ± 0.24
	BY4741+ pGREG506_*haa1*Δ_pGREG506	(did not grow)	(did not grow)
	BY4741 + pGREG506_*haa1*Δ_*Zb*CLIB213^T_*HAA1*	39.5±0.5	2.22 ± 0.08
	BY4741 + pGREG506_*haa1*Δ_*Zb*IST302_*HAA1*	39.5±0.5	2.27 ± 0.10
	BY4741 + pGREG506_*haa1*Δ_ISA1307_ZBAI_B	101.5 ±5	2.18 ± 0.11
	BY4741 + pGREG506_*haa1*Δ_*ScHAA1*	49.5±0.5	2.59

4.2. Analysis of the *S. cerevisiae* Haa1 and Ace1 homologues in Hemiascomycetes

4.2.1. Identification of *S. cerevisiae* Haa1 and Ace1 homologues in Hemiascomycetes

The evolutionary history of the *S. cerevisiae* genes encoding the transcription factor Haa1 and its homologue Ace1 was reconstructed in yeast species belonging to the Saccharomycetaceae family and in several other families of the Hemiascomycetes, leading to the phylogenetic resolution of the *HAA1*/*ACE1* gene family. This analysis was performed using the homologues of the *S. cerevisiae* Haa1 and Ace1 amino acid complete sequences of yeast species from Saccharomycetaceae family and also the N-terminal domain, corresponding to the first 65 amino acids of the Haa1 sequence from *S. cerevisiae* that was used as the cut-off mark in the multiple alignment, in hemiascomycetous species. The phylogenetic analysis of Haa1 and its homologue Ace1 of species of the Saccharomycetaceae family was complemented by a synteny approach based on the analysis of the gene neighbourhood of the *S. cerevisiae HAA1* and *ACE1* homologous genes. The Haa1 amino acid sequence similarity was also evaluated in several *S. cerevisiae* strains isolated from different habitats with the goal of trying to correlate specific amino acid variations found in this protein with the ecological niches associate to specific strains. In this context, it was essential the information from a previous phylogenetic work that correlated phenotypic variation and ecological niches from several *S. cerevisiae* isolates with the genome-wide phylogenetic relationships among these strains[105]. Moreover, the Haa1 sequence codified in the genome of the *S. cerevisiae* (that includes several strains) were compared with the Haa1 mutated sequences described in a patent[17] that were reported to improve acetic acid tolerance.

In order to reconstructed the evolutionary history of the *S. cerevisiae* transcription factors Haa1 and Ace1 in hemiascomycetous species the Haa1 and Ace1 homologues of *S. cerevisiae* were retrieved from the local genome database MySQL using the BLASTP algorithm. The inclusion of the Ace1 homologues in this analysis is a consequence of the high homology shared by the N-terminal domain of this transcription factor and the corresponding from Haa1 (as described in Introduction section)[53]. The amino acid sequences identified were used to construct a multiple alignment and a phylogenetic tree focusing the first 65 amino acid residues of *S. cerevisiae* Haa1 and Ace1 homologues in hemiascomycetous species. Additionally, the retrieved sequences were also used to perform a phylogenetic and synteny analysis in the Saccharomycetaceae family using the *S. cerevisiae* Haa1 and Ace1 homologous

complete sequences. The study of the Haa1 and Ace1 homologues considering different number of amino acids residues is related to the different levels of homology among the Haa1 and Ace1 homologues from phylogenetic groups Saccharomycetaceae and other Hemiascomycete taxonomic families.

The number of amino acid sequences retrieved by the BLASTP network traversal approach at different levels of constraining are plotted in **figure 13**. This figure shows different e-values plateaus corresponding to the different levels of constraining of sequence similarity imposed to the BLASTP network. The different e-value thresholds applied indicate four distinct BLASTP clustering ranges: E60-E40, which returned 105 sequences corresponding to the Haa1 homologues; E30-E13, which added to the previous sequences the Ace1 homologues, in a total of 250 sequences; E12-E2, which merged the Mac1 homologues to the previous ones, increasing the number of sequences to 433; and finally the E0-E1, which retrieved all the previous sequences and also false positive sequences in a total number that exceeded one million of sequences. The extensive amount of sequences incorporated by the lowest e-values is a consequence of the low level of constraining imposed by the BLASTP algorithm which cannot discriminate the amino acid sequences in the database, returning all of them.

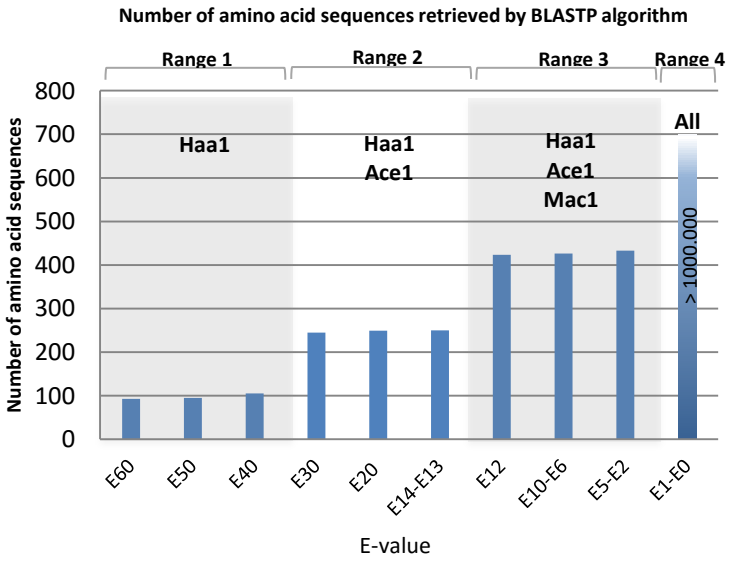

Figure 13. Representation of the number of amino acid sequences retrieved after constraining and traversing the pairwise similarity network at different e-values using the Haa1 sequence from *S. cerevisiae* S288c as 'starting node'. This analysis retrieved three successive protein sequence plateaus, depending on the e-value threshold applied, correspond to the homologues of the *S. cerevisiae* Haa1, Ace1 and Mac1 proteins.

Given that it was previously demonstrated that the Haa1 and Ace1 share a conserved domain (linked to the N-terminal domain) the identification of the Ace1 homologues in the pairwise similarity analysis is not surprising. The Mac1 identification during the BLASTP networking

traversing also resulted from the similarity existing between its N-terminal domain and the corresponding domain of the Haa1 protein. Since the amino acid identity between the Mac1 and the Haa1 is lower than that shared between the Ace1 and Haa1 homologues, the Mac1 homologues emerge at a less restrictive e-value in the pairwise analysis when compared to the Ace1 homologues. Additionally, it was observed a higher amino acid identity between the homologues pair Mac1/Ace1 in comparison to the pair Mac1/Haa1. The amino acid sequence similarity among the Haa1, Ace1 and Mac1 proteins from *S. cerevisiae* S288c is shown in **table 8**. The gap percentage observed among the sequences analysed is also consistent with protein sequences similarity data.

Table 8. Results obtained from the alignment of Haa1, Ace1, and Mac1 protein sequences from *S. cerevisiae* S288c. In this analysis, different levels of cut-off were used in order to assess the differential conservation among the amino acid sequences under study. The DBD refers to the first 124 amino acid residues. Values were obtained using the Needleman-Wunsch algorithm (EMBL-EBI).

Protein sequences aligned	Identity (%)	Similarity (%)	Gaps (%)
Haa1/Ace1 (complete sequences)	35.8	49.8	12.8
Haa1/Ace1 (Ace1 length)	36.6	48.9	16.7
Haa1/Ace1 (DBD)	47.6	58.1	13.7
Haa1/Ace1 (first 65 amino acids)	68.3	77.8	0.0
Haa1/Mac1 (complete sequences)	26.8	40.8	29.8
Haa1/Mac1 (Mac1 length)	26.8	40.8	29.8
Haa1/Mac1 (DBD)	25.0	38.9	20.8
Haa1/Mac1 (first 65 amino acids)	52.5	75.0	0.0
Ace1/Mac1 (complete sequences)	21.5	30.8	43.1
Ace1/Mac1 (Ace1 length)	21.5	30.8	43.1
Ace1/Mac1 (DBD)	52.5	72.5	0.0
Ace1/Mac1 (first 65 amino acids)	52.5	72.5	0.0

Given that the *S. cerevisiae* Mac1 homologues are not on the focus of this work, the E-13 was selected and the resulting protein sequences, a total of 250, were retrieved and used for further analyses.

In order to get some clues about the amino acid sequences homology the multiple alignment generated by MUSCLE algorithm was visualized in Jalview software (**figure 14**). This analysis confirmed the strong amino acid similarity shared among the N-terminal domain of the *S. cerevisiae* Haa1 and Ace1 homologous proteins in Hemiascomycetes (blue region in **figure 14 A**) essentially maintained for the first 65 amino acid residues (**figure 14 B**). The few similarities observed in the C-terminal domain must be considered artifacts since they result from the algorithm objective of finding amino acid residues conserved. The conclusion that the C-

terminal domain of these proteins is poorly conserved can be observed from the absence of amino acid residues aligned in **figure 14 A** (white region). Interestingly, the differential pattern of amino acid conservation over the Haa1 and Ace1 homologous sequences is conserved in Hemiascomycetes throughout the evolutionary history.

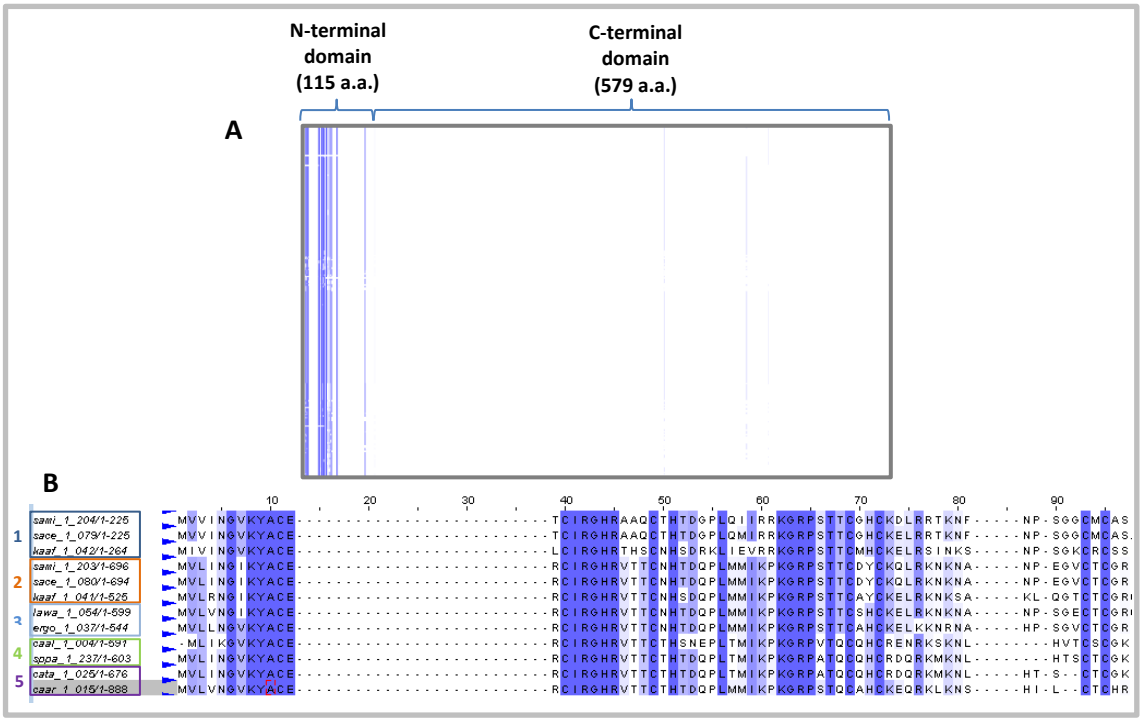

Figure 14. Alignment of the Haa1 and Ace1 homologous sequences from Hemiascomycetes using the E-13 threshold. (A) Overview of the multiple alignment obtained for the 250 sequences representing the Haa1/Ace1 protein family from Hemiascomycetes. This schematic representation was retrieved from Jalview software through the chosen of the following options: 'Colour', 'Percentage Identity', 'View' and 'Overview Window' (by this order). The number of amino acid residues corresponding to N- and C- terminal domains are reported to the Haa1 protein from *S. cerevisiae* S288c **(B)** Detail of the multiple alignment showing sequences belonging to the major phylogenetic families in study: the Haa1 and Ace1 homologues belonging to the post-WGD Saccharomycetaceae species are represented by the sequence members of groups 1 and 2, respectively; the Haa1/Ace1 homologues belonging to the pre-WGD Saccharomycetaceae, CTG complex and early-divergent hemiascomycetous species are represented by the sequence members of groups 3, 4 and 5, respectively. Protein names are listed in **table 17** of **annex III**. Blue and white regions represent conserved and non-conserved amino acid residues, respectively.

This analysis focusing the hemiascomycetous species comprises 165 genomes from 60 different species, 28 of which belong to the Saccharomycetaceae family, 14 to the CTG complex and 18 to the early-divergent hemiascomycetous species (**annex III**, **table 17**). The final sequences analysed were reduced to 244 sequences due to the existence of 5 Haa1/Ace1 homologues carrying frameshift mutations and a sequence which has lost its initial portion (corresponding to the ZBAI_B ORF from hybrid strain ISA1307). Additionally, 4 sequences were

recognized as protein fragments with less than 165 amino acids, with 2 of them corresponding to frameshift sequences. Fragmented sequences are possibly due to DNA sequencing problems which could include the introduction of a premature and non-real STOP codon. In order to clarify these cases, the DNA from the corresponding strains must be re-sequenced.

Concerning the pre-WGD species, in general it was only recognized one Haa1/Ace1 homologous sequence (by similarity) per strain genome (**table 9**). The few exceptions include two Haa1/Ace1 homologous sequences found in the genome of the hybrid strain *Pichia sorbitophila* CBS 7064 (piso_1_a06138g and piso_1_b06205g, both with 388 amino acids) and in the genome of *Nadsonia fulvescens* var. elongate DSM 6958 (nafu_1_10_j01430 and nafu_1_2_b04520 with 1248 and 285 amino acids, respectively). No obvious explanation was found for the two Haa1/Ace1 homologues found in the genome of *N. fulvescens* species which could be related to a horizontal transfer (probably linked to the larger copy) although this phenomenon does not frequently occur in hemiascomycetous species[63,106].

In general, post-WGD species do have one Haa1 and one Ace1 homologue, as it would be expected since its genomes derive from a WGD event despite the strong genetic erosion that affected this group of microorganisms. An interesting finding is that the genomes from post-WGD *Saccharomyces paradoxus* species, with few exceptions, only have one Haa1/Ace1 homologous sequence corresponding to the Ace1 homologues demonstrated by subsequent analyses. The exceptions to this behaviour include *S. paradoxus* NRRL Y-17217 and Sanger reference (consensus) genomes which carry a second Haa1/Ace1 homologous sequence with 694 (sapa_26_87_ci00100) and 702 (sapa_1_5_e03020) amino acids, respectively. At this phase it cannot be excluded the hypothesis of this additional copy that has a similar length to the Haa1 homologous sequences could encodes a functional Haa1 protein that was not identified in the other *S. paradoxus* strains maybe due to sequencing problems. On the other hand, the Haa1/Ace1 extra copy could have been acquired by horizontal gene transfer of foreign DNA.

Table 9. Number of Haa1 and Ace1 homologous sequences in the hemiascomycetous strains of the different species examined in this work. The Haa1 and Ace1 orthologues were identified according to phylogenetic and synteny analyses. The number of strains per species and the orthologues per strains is also indicated. Frameshift sequences are not included.

Phylogenetic complex	Genus	Species	*ACE1* ortologues	*HAA1* ortologues	Strain/species	Ortologues/strain	Total number of orthologues
'*Saccharomyces* complex' (Post-WGD)	*Saccharomyces*	*S. arboricola*	1	1	1	2.00	100|79
		S. bayanus	2	2	2	2.00	
		S. cerevisiae	60	61	61	1.98	
		S. mikatae	1	1	1	2.00	
		S. paradoxus	25	2	26	1.40	
		S. uvarum	1	1	1	2.00	
		S. kudriavzevii	1	1	1	2.00	
	Kazachstania	*K. Africana*	1	1	1	2.00	
		K. naganishii	1	1	1	2.00	
	Naumovozyma	*N. castellii*	2	2	2	2.00	
		N. dairenensis	1	1	1	2.00	
	Candida	*C. glabrata*	2	2	2	2.00	
	Tetrapisispora	*T. blattae*	0	1	1	1.00	
		T. phaffii	1	1	1	2.00	
	Vanderwaltozyma	*V. polyspora*	1	1	1	2.00	
'*Saccharomyces* complex' (Pre-WGD)	*Zygosaccharomyces*	*Z. bailii*	3		3	1.00	15
		Z. rouxii	1		1	1.00	
	Torulaspora	*T. delbrueckii*	1		1	1.00	
	Lachancea	*L. waltii*	1		1	1.00	
		L. thermotolerans	1		1	1.00	
		L. kluyvery	1		1	1.00	
	Kluyveromyces	*K. wickerhamii*	1		1	1.00	
		K. marxianus var. marxianus	1		1	1.00	
		Kl. Lactis	1		1	1.00	
		K. aestuarii	1		1	1.00	
	Eremothecium	*E. cymbalariae*	1		1	1.00	
		E. gossypii	1		1	1.00	
		A. acer	1		1	1.00	
CTG complex	*Candida*	*C. albicans*	11		11	1.00	26
		C. dubliniensis	1		1	1.00	
		C. maltose	1		1	1.00	
		C. orthopsilosis	1		1	1.00	
		C. parapsilosis	1		1	1.00	
		C. tenuis	1		1	1.00	
		C. tropicalis	1		1	1.00	

	Debaryomyces	*D. hansenii*	2	2	1.00	
	Lodderomyces	*L. elongisporus*	1	1	1.00	
	Meyerozyma	*M. guilliermondii*	1	1	1.00	
	Pichia	*P. sorbitophila*[#]	2	1	2.00	
	Scheffersomyces	*S. stipites*	1	1	1.00	
	Spathaspora	*S. arborariae*	1	1	1.00	
		S. passalidarum	1	1	1.00	
Early-divergent	*Candida*	*C. arabinofermentans*	1	1	1.00	24
		C. caseinolytica	1	1	1.00	
		C. tanzawaensis	1	1	1.00	
	Clavispora	*C. lusitaniae*	2	2	1.00	
	Cyberlindnera	*C. jadinii*	2	2	1.00	
	Dekkera	*D. bruxellensis**	2	2	1.00	
	Hansenula	*H. polymorpha*	1	1	1.00	
	Hanseniaspora	*H. valbyensis*	1	1	1.00	
	Hyphopichia	*H. burtonii*	1	1	1.00	
	Komagataella	*K. pastoris*	3	3	1.00	
	Metschnikowia	*M. bicuspidate*	1	1	1.00	
	Nadsonia	*N. fulvescens var. elongata*	2	1	2.00	
	Ogataea	*O. parapolymorpha*	1	1	1.00	
	Pachysolen	*P. tannophilus*	1	1	1.00	
	Pichia	*P. membranifaciens*	1	1	1.00	
		*P. kudriavzevii**	1	1	1.00	
	Wickerhamomyces	*W. anomalus*	1	1	1.00	
	Yarrowia	*Y. lipolytica* (CLIB122)	1	1	1.00	

Yeast species marked with an asterisk (*) represent those that contribute with a fragment to the analysis.

Once the phylogenetic analysis of the *S. cerevisiae* Haa1 and Ace1 homologues was limited to genomes of yeast species belonging to the Saccharomycetaceae, the final number of sequences studied herein decreased to a total of 54 sequences, 19 sequences corresponding to Haa1 homologues and 19 sequences to Ace1 homologues from post-WGD species and 16 sequences of the Haa1/Ace1 homologues from pre-WGD species (**table 10**). This analysis includes 36 genomes from 28 different species, 13 of which belong to the pre-WGD and 15 to the post-WGD species.

Table 10. Saccharomycetaceae strains examined in this work and corresponding *S. cerevisiae* Haa1 and Ace1 homologues. The presented groups and (sub)lineages were defined according to the phylogenetic (**figure 17**) and synteny (**figure 18**) analyses performed concerning this yeast family.

Genus	Species	Strain	Gene name	Length	Group	(Sub)lineage	Speciation in relation to WGD
Saccharomyces[(1)]	*S. arboricola*	H-6	saar_1_3_c00950	227	B	*ACE1*	Post
			saar_1_6_f02790	739	A	*HAA1*	
	S. bayanus	623-6C	saba_1_3_c00480	232	B	*ACE1*	
			saba_1_11_k00230	722	A	*HAA1*	
		MCYC 623	saba_2_521_ta00120	232	B	*ACE1*	
			saba_2_351_mm00100	722	A	*HAA1*	
	S. cerevisiae	S288c	sace_1_ygl166w	225	B	*ACE1*	
			sace_1_ypr008w	694	A	*HAA1*	
		cen.pk113-7d	sace_3_3_c00990	225	B	*ACE1*	
			sace_3_5_e02840	694	A	*HAA1*	
	S. kudriavzevii	IFO 1802	saku_1_7.105	221	B	*ACE1*	
			saku_1_16.292	724	A	*HAA1*	
	S. paradoxus	Sanger reference	sapa_1_2_b00960	225	B	*ACE1*	
			sapa_1_5_e03020	702	A	*HAA1*	
		CBS432	sapa_4_2_b00970	225	B	*ACE1*	
	S. uvarum	CBS 7001	sauv_1_7.97	232	B	*ACE1*	
			sauv_1_16.328	722	A	*HAA1*	
	S. mikatae	IFO 1815	sami_1_7.105	225	B	*ACE1*	
			sami_1_16.247	696	A	*HAA1*	
Kazachstania	*K. africana*	CBS 2517	kaaf_1_c03260	264	B	*ACE1*	
			kaaf_1_b01950	525	A	*HAA1*	
	K. naganishii	CBS 8797	kana_1_e03480	263	B	*ACE1*	
			kana_1_d02340	611	A	*HAA1*	
Naumovozyma	*N. castellii*	CBS 4309	naca_2_51_ay00250	291	B	*ACE1*	
			naca_2_9_i00230	743	A	*HAA1*	
		NRRL Y-12630	naca_1_a05130	291	B	*ACE1*	
			naca_1_b05730	743	A	*HAA1*	
	N. dairenensis	CBS 421	nada_1_k02530	294	B	*ACE1*	

			nada_1_b03070	893	A	*HAA1*	Post	
Candida[2]	*C. glabrata*	CBS138	cagl_1_l09339g	264	B	*ACE1*		
			cagl_1_l09339g	876	A	*HAA1*		
		CCTCC M202019	cagl_2_4_d01770	265	B	*ACE1*		
			cagl_2_2_b04230	877	A	*HAA1*		
Tetrapisispora[3]	*T. blattae*	CBS 6284	tebl_1_c00170	604	D	*HAA1*		
	T. phaffii	CBS 4417	teph_1_h00580	356	B	*ACE1*		
			teph_1_d00690	644	A	*HAA1*		
Vanderwaltozyma	*V. polyspora*	DSM 70294	vapo_1_467.5	341	B	*ACE1*		
			vapo_1_541.30	586	A	*HAA1*		
Zygosaccharomyces[4]	*Z. bailii*	ISA1307_05761	zyba_1_05761	664	C	*HAA1/ACE1*	Pre	
		ISA1307_01494	zyba_1_246	698	C	*HAA1/ACE1*		
		UTAD63	zyba_1_01494	694	C	*HAA1/ACE1*		
		CLIB 213	zyba_3_9_i00670	694	C	*HAA1/ACE1*		
	Z. rouxii	CBS 732	zyro_1_f04862g	718	C	*HAA1/ACE1*		
Torulaspora[5]	*T. delbrueckii*	CBS 1146	tode_1_c01350	627	C	*HAA1/ACE1*		
Lachancea[6]	*L. thermotolerans*	CBS 6340	lath_1_g16984g	614	C	*HAA1/ACE1*		
	L. waltii	NCYC 2644	lawa_1_14.1543	599	C	*HAA1/ACE1*		
	L. kluyvery	CBS 3082	lakl_1_b03344g	618	C	*HAA1/ACE1*		
Kluyveromyces[7]	*K. aestuarii*	ATCC 18862	klae_1_13_m00540	561	C	*HAA1/ACE1*		
	K. wickerhamii	UCD 54-210	klwi_1_120_dp00120	557	C	*HAA1/ACE1*		
	K. lactis	CLIB210	klla_1_a03047g	567	C	*HAA1/ACE1*		
	K. marxianus var. m.	KCTC 17555	klma_1_3_c06150	601	C	*HAA1/ACE1*		
Eremothecium[8]	*E. gossypii*	ATCC 10895	ergo_1_ael295c	544	C	*HAA1/ACE1*		
	E. cymbalariae	DBVPG 7215	ercy_1_6116	668	C	*HAA1/ACE1*		
	Ashbya acer	(not available)	asac_1_7_g04120	540	C	*HAA1/ACE1*		

Information regarding genus relevance described for some genera here presented:

(1) Vigorous sugar fermentation; biotechnological and industrial applications; laboratory model; food spoilage agents.
(2) Industrial applications; medical interest; human pathogens.
(3) Biotechnological applications.
(4) Biotechnological applications; fermentation importance; food spoilage agents.
(5) Industrial interest.
(6) Biotechnological and food industry applications; fermentation interest.
(7) Biotechnological application; industrial and environmental interests.
(8) Biotechnological applications; agricultural interest; plant pathogens.

4.2.2. Phylogenetic tree of the initial portion of N-terminal domain of *S. cerevisiae* Haa1 and Ace1 homologues in Hemiascomycetes

The phylogenetic analysis comprising the initial portion of the *S. cerevisiae* Haa1 and Ace1 homologous sequences corresponded to the first 65 amino acids of *S. cerevisiae* Haa1 allowed the identification of 4 different groups in hemiascomycetous species based on the amino acid sequence similarity (**figure 15** and **16**). The identified sequence groups are represented in different colours: the pink and purple comprise the Haa1 (group 1) and Ace1 (group 2) homologues from Saccharomycetaceae post-WGD species, respectively; the green corresponds to Haa1/Ace1 homologues that belong to the Saccharomycetaceae pre-WGD species (group 3); and the blue represents the Haa1/Ace1 homologues from yeast species belonging to the CTG complex and the remaining early divergent Hemiascomycetes (group 4). In general, the distribution of the *S. cerevisiae* Haa1 and Ace1 homologous sequences by groups is in agreement with the main taxonomic groups of the Hemiascomycetes[67,107]. Nevertheless, the *S. cerevisiae* Haa1 and Ace1 homologous of the CTG complex and the early divergent hemiascomycetous species are grouped together in a larger group, thus being impossible to differentiate them, although there is a tendency for the sequences of the same families to remain in the same sub-group (given by a common node). Despite the biological diversity associated to this group, the observed divergence is not so high as would be expected.

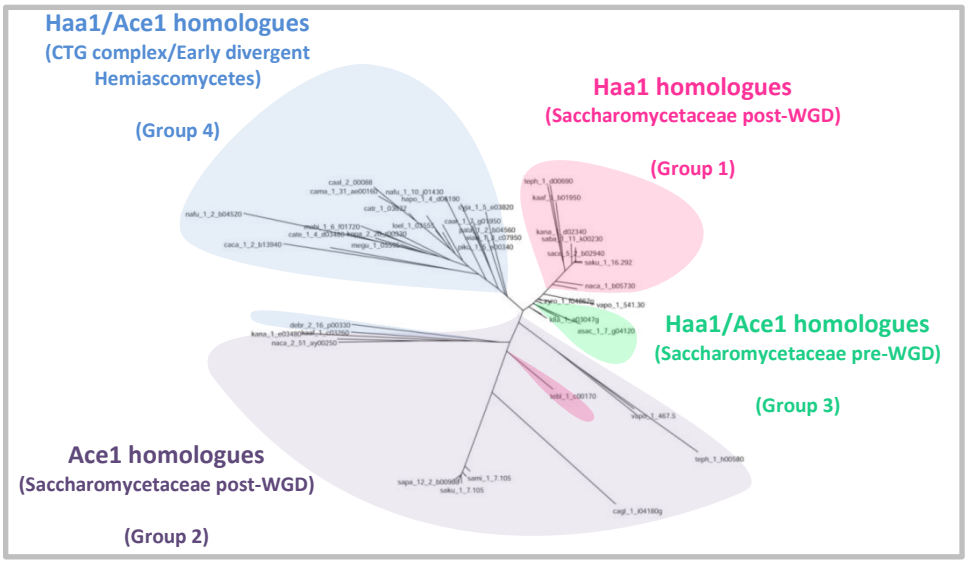

Figure 15. Phylogenetic analysis of the initial portion of N-terminal sequence of the *S. cerevisiae* Haa1 and Ace1 homologues in Hemiascomycetes. Radial phylogram shows the amino acid divergence among the Haa1 and Ace1 homologous proteins considering the first 65 amino acid residues of the Haa1 from *S. cerevisiae* S288c as the cut-off mark in the multiple alignment.

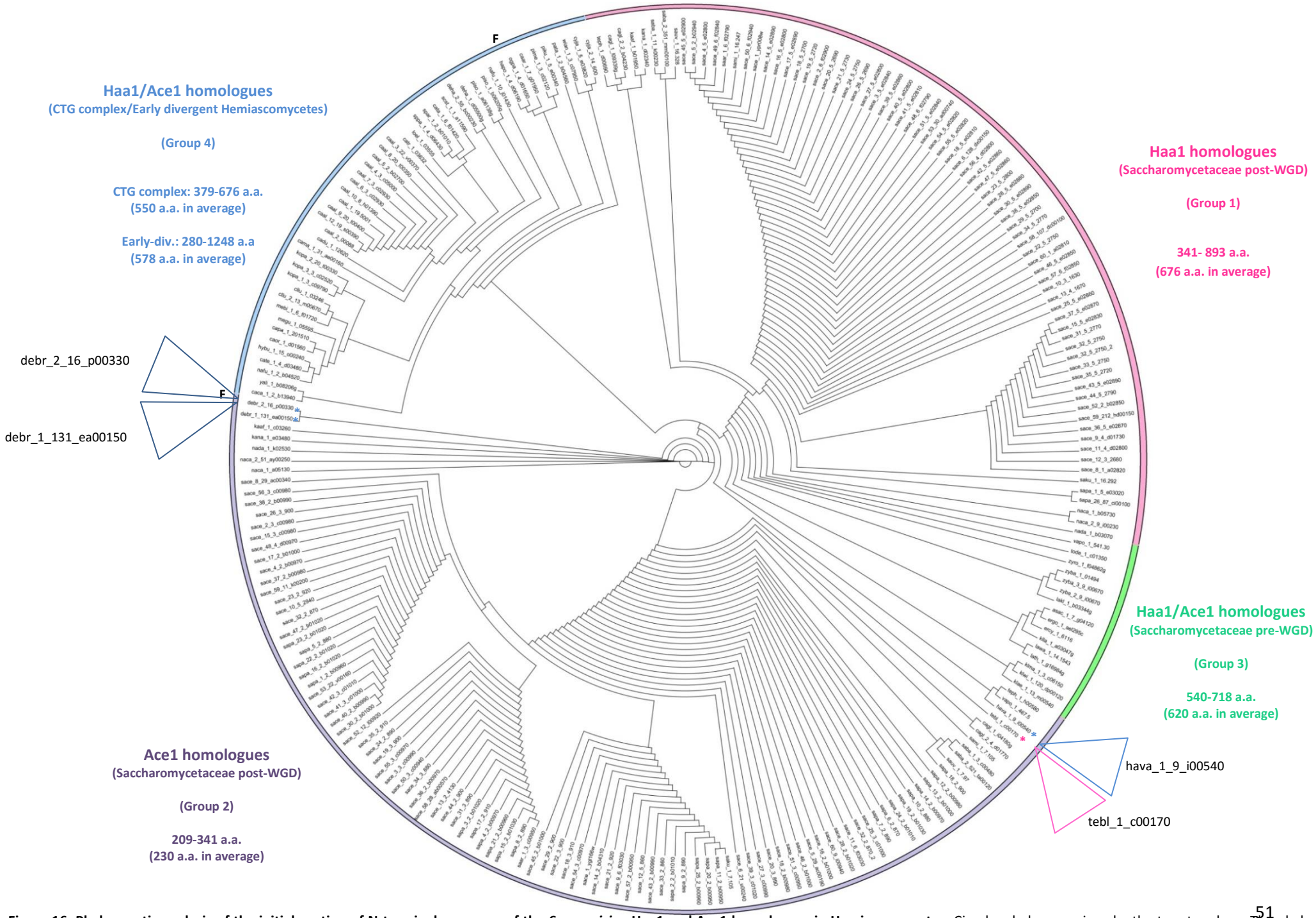

Figure 16. Phylogenetic analysis of the initial portion of N-terminal sequence of the *S. cerevisiae* Haa1 and Ace1 homologues in Hemiascomycetes. Circular cladogram given by the tree topology. The phylogenetic tree was generated based on PROTDIST/NEIGHBOUR program of PHYLIP package and retrieved results explored in Dendroscope software. The naca_1_a05130 sequence from *Naumovozyma castellii* represents the tree outgroup. Asterisks (*) represent protein sequences which according to the species phylogeny[122] must belong to another tree group (given by the asterisk colour); (F) represents sequence fragments.

A global analysis of the phylogenetic tree indicates that the *S. cerevisiae* Haa1 and Ace1 homologous sequences are organized in a radial structure, suggesting a strong divergence among the sequences under study. The longest and most dispersed branches, mostly associated to Ace1 (from post-WGD species), indicate a high degree of divergence within this group as a result of the low level of amino acid conservation.

In an ambiguous position, placed in the group 2 that includes the Ace1 homologues, emerges the tebl_1_c00170 sequence (with 664 amino acids) from *Tetrapisispora blattae*. According to the species phylogenetic relationships previously proposed and the size of this protein sequence it was expected that it falls into the group 1 that includes the Haa1 homologues[67]. This result could be related to specific mutations that affected the *HAA1*/*ACE1* homologous gene sequence from *T. blattae* in a divergent process from *HAA1* homologues.

Additionally, the Ace1 and Haa1 homologous sequences hava_1_9_i00540, debr_1_131_ea00150 and debr_2_16_p00330 (the first one from *Hanseniaspora valbyensis* and the last two from *Dekkera bruxellensis* species) were also identified in the group 2. Given that these sequences belong to the CTG complex we were expecting to find it in group 4 that encloses the Haa1/Ace1 homologous sequences belonging to the CTG complex and the early divergent Hemiascomycetes. The phylogenetic position of the Haa1/Ace1 homologues from *D. bruxellensis* and *H. valbyensis* is not completely clear and could be the result of an artificial attraction to Ace1 group influenced by the low conservation found between these protein sequences and those belonging to the CTG complex/early divergent Hemiascomycetes group. All together, these results reinforce the strong divergence associated to the Ace1 group that can support a high sequence diversity.

In addition to this analysis, the 244 sequences corresponding to Haa1/Ace1 protein family from hemiascomycetous species were also submitted to a phylogenetic analysis using the first 103 amino acids of the Haa1 from *S. cerevisiae* S288c as the cut-off mark in the multiple alignment (**annex IV**, **figures 25** and **26**). Given that the phylogenetic groups obtained with these Haa1 and Ace1 homologous sequences are similar to those observed in **figure 15**, this second analysis is not shown herein.

4.2.3. Phylogenetic tree of the *S. cerevisiae* Haa1 and Ace1 homologues in Saccharomycetaceae

The phylogenetic tree based on the 54 *S. cerevisiae* Haa1 and Ace1 homologous sequences belonging to the Saccharomycetaceae shows 3 main groups represented in pink (group A), purple (group B) and green (group C), corresponding to the Haa1 and Ace1 homologues from post-WGD species and Haa1/Ace1 homologous sequences from pre-WGD species, respectively (**figure 17**). The phylogenetic tree obtained indicates a strong divergence among the *S. cerevisiae* Haa1 and Ace1 homologous sequences, widely dispersed through the long branches of the tree, evidencing the low degree of amino acid similarity (although the main phylogenetic relationships previously defined to yeast species from this phylogenetic group are maintained[67]).

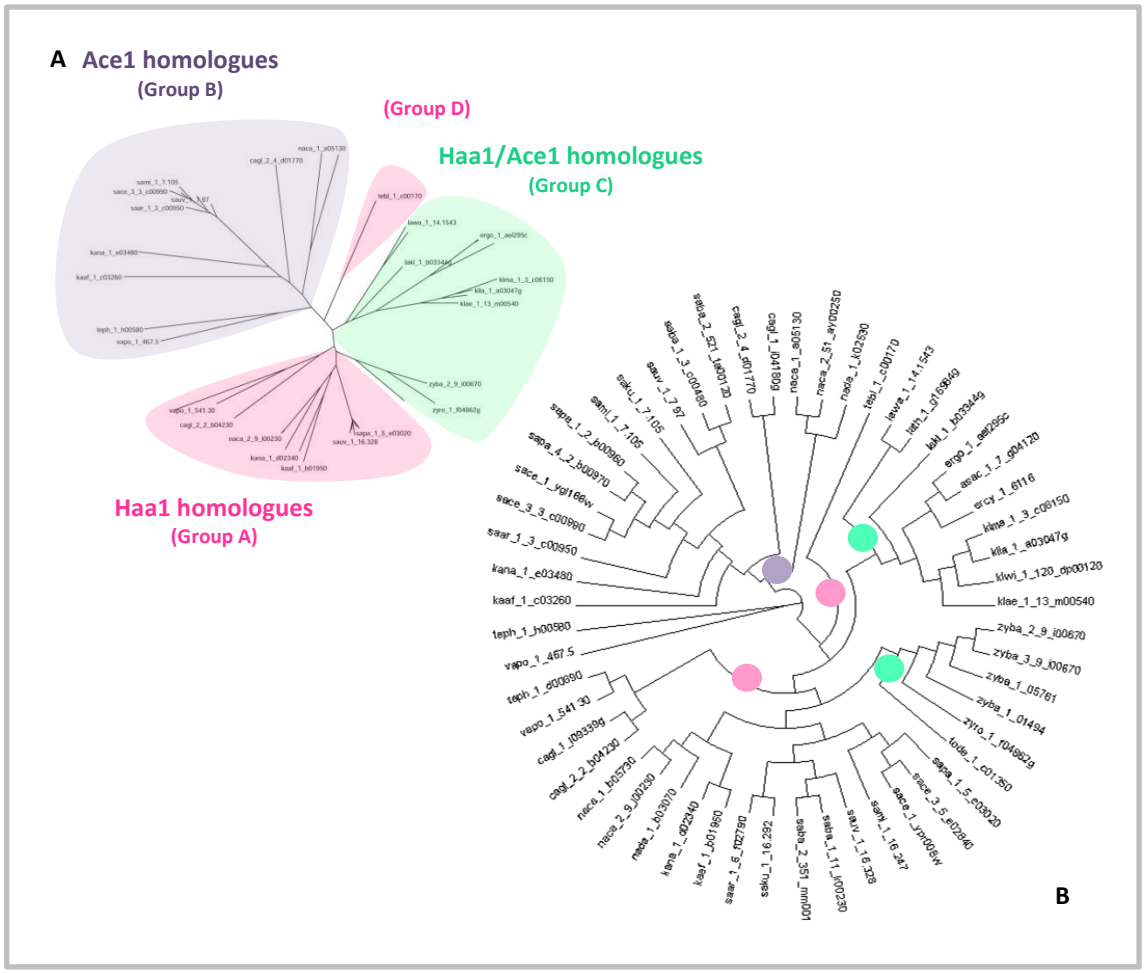

Figure 17. Phylogenetic analysis of the homologous of the *S. cerevisiae* Haa1 and Ace1 proteins (complete sequences) from yeast species belonging to the Saccharomycetaceae family. (A) Radial phylogram showing the amino acid similarity distances among the Haa1 and Ace1 homologues; **(B)** Circular cladogram showing by the tree topology. PROTML program of PHYLIP package was used in the analysis for the phylogenetic tree calculation and the Dendroscope software to subsequent visualization. The vapo_1_467.5 sequence from *Vanderwaltozyma polyspora* is the tree outgroup.

The analysis of the phylogenetic tree evidences the unexpected position of some *S. cerevisiae* Haa1 and Ace1 homologous sequences belonging to the *Zygosaccharomyces* (zyba_3_9_i00670, zyba_2_9_i00670, zyba_1_05761, zyba_1_01494 and zyro_1_f04862g) and *Torulospora* (tode_1_c01350) pre-WGD genera which share the last common node with the *S. cerevisiae* Haa1 homologues (group A) (**figure 17**) instead of with the Haa1 and Ace1 homologues from the remaining pre-WGD yeast species considered in this study (group C), as it would be expected according to the species phylogeny[67]. This finding highlights an inconsistence between the protein and species trees regarding the Haa1/Ace1 homologues from *Zygosaccharomyces rouxii*, *Z. bailii* and *Torulospora delbrueckii* species. Furthermore, it was also observed a strong divergence of the Haa1 and Ace1 homologues belonging to the *V. polyspora*, (vapo_1_467.5 and vapo_1_541.30), *Tetrapisispora phaffii* (teph_1_h00580 and teph_1_d00690) and *Candida glabrata* (cagl_1_l09339g, cagl_2_2_b04230, cagl_1_i04180g and cagl_2_4_d01770) species (by the long and dispersed branches in the phylogenetic tree) that was also extended to some protein sequences belonging to *Naumovozyma* genus (nada_1_k02530 and naca_2_51_ay00250) (**figure 17**). These observations suggest that the *S. cerevisiae* Haa1 and Ace1 homologues have suffered a large divergence in amino acid sequences during the WGD. Once again, the tebl_1_c00170 sequence from *T. blattae* appears in an unexpected tree branch (group D), located between the group B, which includes the Ace1 homologues from post-WGD species, and the group C, which contains the Haa1/Ace1 homologues from pre-WGD species.

The Ace1 homologous sequences exhibited a higher divergence compared to the Haa1 homologous sequences (which can be seen by the longer and more dispersed tree branches). This phenomenon was observed among the Ace1 homologous sequences belonging to the *Saccharomyces sensu stricto*, as well as for the sequences belonging to other post-WGD genera. For instance, the dispersion of Haa1/Ace1 homologous sequences from *Candida* (cagl_1_i04180g and cagl_2_4_d01770), *Naumovozyma* (nada_1_k02530, naca_1_a05130 and naca_2_51_ay00250) and *Kazachstania* (kana_1_e03480 and kaaf_1_c03260) genera evidence this observation very well.

Phylogenetic analysis accomplished for the Haa1/Ace1 homologues considering the first 65 amino acids of N-terminal domain and the complete ('full size') proteins reveal a common pattern regarding the amino acid sequences groups composition and, with few exceptions, with the phylogenetic relationships described for species belonging to Saccharomycetaceae family[67,107]. The differences between the present analysis and the species phylogenetic

resolution are in agreement with the extensive divergence that marks the evolution of Haa1/Ace1 protein family.

The comparison of Ace1 amino acid sequences evidences a progressive decrease in its length. This phenomenon is not observed in the Haa1 homologous sequences and in the Haa1/Ace1 homologues from pre-WGD species. On the other hand, the Haa1 homologous proteins seem to have been subjected to a series of amino acid number variations (either by the increase or decrease of amino acid residues number) (**table 10**) and, therefore, are strongly marked by fluctuations in protein length. For instance, two large Haa1 sequences were identified in both *C. glabrata* and *Naumovozyma dairenensis* species with 877 and 893, respectively, being this the maximum length observed for the Haa1 homologous sequences in Saccharomycetaceae (**table 10**). These horizontal expansions seem to be the result of internal duplications (by the analysis of amino acid sequences and the results obtained in the phylogenetic analysis).

4.2.4. Synteny analysis of *HAA1/ACE1* gene family lineage in Saccharomycetaceae

At first sight the synteny reconstruction evidences the existence of two independent sub-lineages, the *HAA1* and *ACE1* sub-lineages, with origin in a WGD event (**figure 18**).

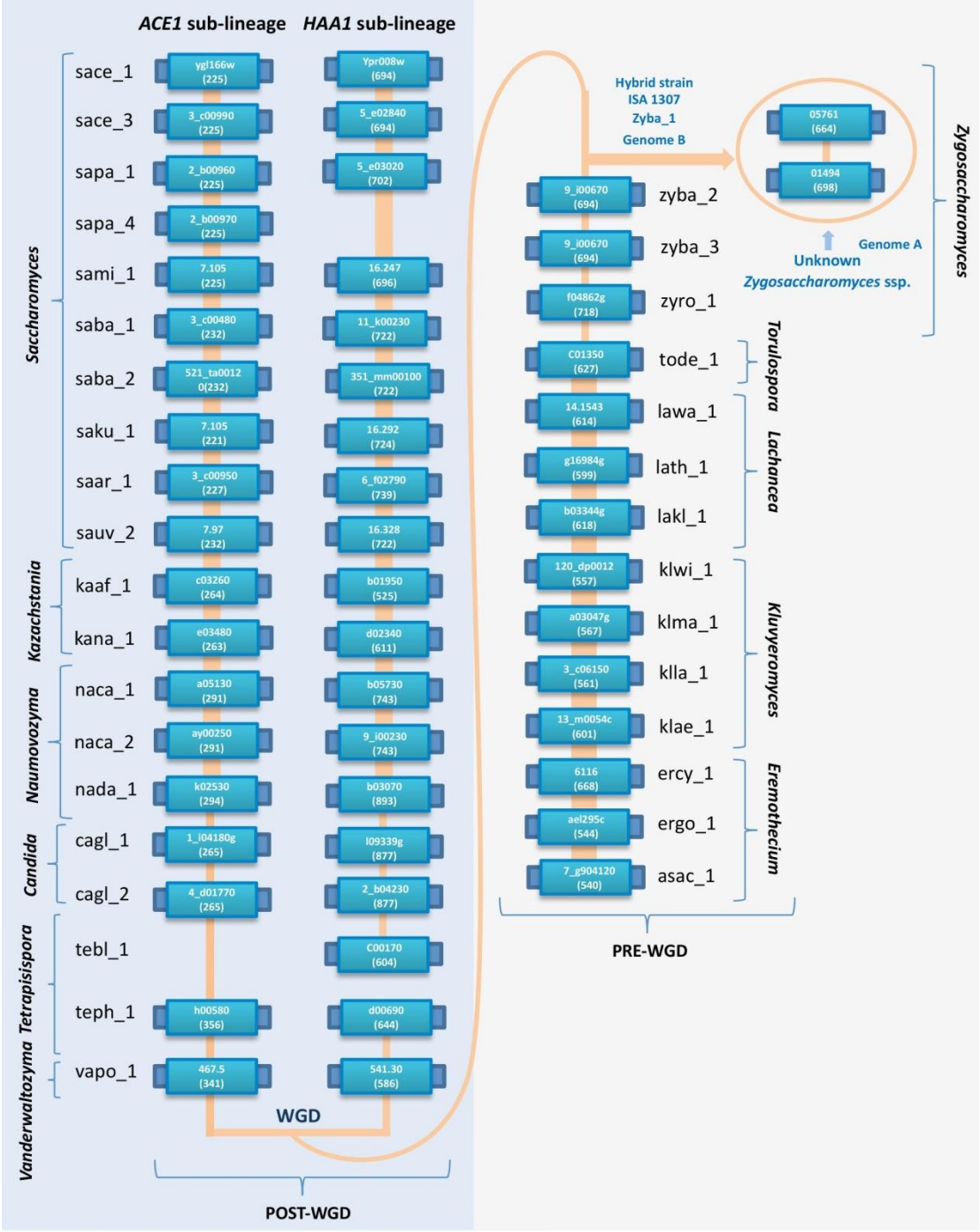

Figure 18. *HAA1/ACE1* lineage in Saccharomycetaceae. Blue boxes represent the *HAA1/ACE1* homologous genes Synteny strength is represented by the lines connecting homologous genes, thick (strong synteny) or thin (weal synteny), according to the number of sharing neighbours and other parameters as mentioned in Materials anc Methods section. The present scheme was accomplished according to the analysis of the *HAA1/ACE1* homologou genes neighbourhood performed in the Cytoscape software environment.

A strong synteny was observed among the genes encoded in the genomes of the yeast species of the *Eremothecium*, *Kluvyeromyces* and *Lachancea* genera (**figure 18**), consistent with the high number of neighbour genes in common connecting the *HAA1/ACE1* orthologues, the conservation of neighbour genes position over the chromosome environment, the protein sequence clusters (comprising a small number of members well conserved over the lineages) and also the high percentage of identity among neighbour sequences. However, this strong synteny is not observed in *Torulospora* and *Zygosaccharomyces* genera and in a more extensive way in *Vanderwaltozyma* and *Tetrapisispora* genera which have lost a high number of neighbour connections. Also, it was registered a weak synteny linking the *HAA1* and *ACE1* orthologues from *Candida* and *Naumovozyma* genera to the corresponding genes from genera positioned above and below in *HAA1/ACE1* gene lineage (given by the few number of common neighbour genes with low protein identity). Despite of the loss of the synteny observed to *HAA1* and *ACE1* homologues genes in *Naumovozyma* genus, it was observed a re-establishment of synteny between the *HAA1* and *ACE1* orthologues from *Naumovozyma* and *Kazachstania* and *Saccharomyces* genera. Altogether, these results indicate that during the WGD transition the chromosome environment of the *HAA1* and *ACE1* genes was poorly conserved. Indeed, subsequently to the WGD event the *HAA1* and *ACE1* orthologues evolution was strongly marked by the loss of its neighbour genes. In which concerns close phylogenetically related species, it was observed a strong synteny among the *HAA1* and *ACE1* genes belonging to the same genera that is remarkably conserved in *Saccharomyces sensu stricto* group. Despite the weak synteny associated to Saccharomycetaceae, a global analysis of the synteny network revealed the conservation of some neighbour protein families throughout the evolutionary history of this yeast group. Shared neighbours include the protein families represented by the number 153 (linked to chromatin remodeling), 7142 (associated to sugar metabolism and meiosis regulation) and 15140 (involved in energy production), which comprise a total members of 3173, 423 and 172 protein, respectively. These protein families have contributed to the resolution of *HAA1/ACE1* lineage since they allowed the reconstruction of the evolutionary history of the *HAA1/ACE1* gene family.

The weak synteny pattern found to *HAA1/ACE1* homologue from *T. blattae* (tebl_1_c00170g) in relation to *HAA1* and *ACE1* homologues does not allow to discriminate the position of this gene in synteny lineage. Indeed, the ambiguous position of this sequence was also anticipated from phylogenetic analysis. The option of placing this *HAA1/ACE1* homologue in the *HAA1* sub-lineage seems the most parsimonious choice. The absence of synteny found to this protein can be a consequence of a recombination event responsible for its shift position and incorporation

in another chromosome region. This hypothesis is reinforced because this *HAA1*/*ACE1* homologue resides in a subtelomeric region.

Additionally, the *S. cerevisiae HAA1* and *ACE1* homologues from Saccharomycetaceae species in study and its corresponding neighbour genes were disposed in a vertical scheme that evidences the neighbour genes chromosomal order (**figure 19**).

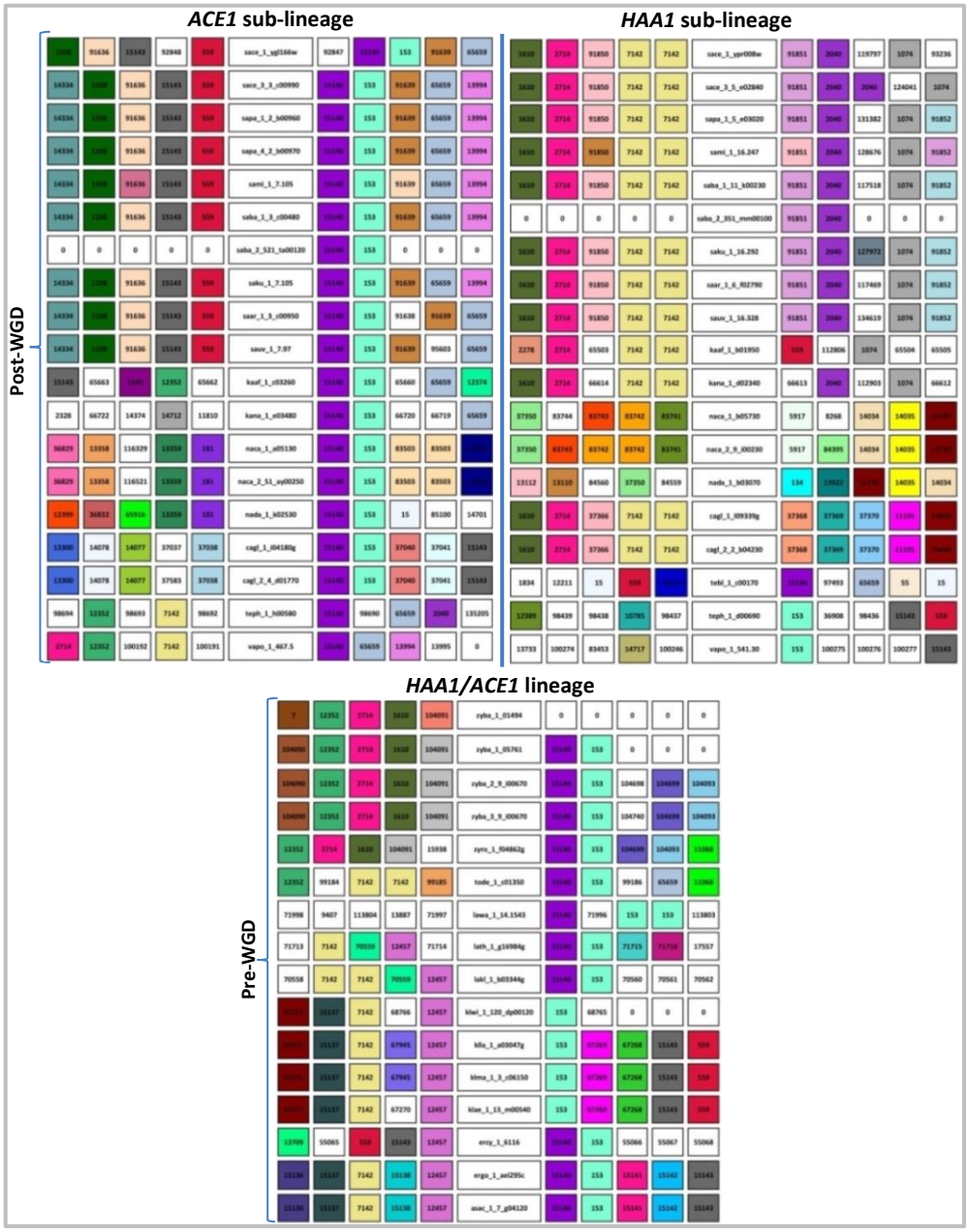

Figure 19. Gene neighbourhood comparison of the *HAA1/ACE1* homologous genes in the genome of yeast species belonging to the Saccharomycetaceae. The *HAA1* and *ACE1* homologous genes are represented in central boxes flanked by the corresponding neighbour genes highlighted at different colours according to the protein sequence cluster. White boxes represent genes without homologous neighbours in the represented chromosome region. In order to simplify the gene neighbourhood comparison the retrieved 15 neighbour genes for each adjacent region of query genes were reduced to 5 neighbour genes here shown and the obtained scheme divided according to the (sub)lineages identified. The neighbours' order in each chromosome was scrutinized by exploring the synteny network in the Cytoscape software.

In this scheme, the neighbour genes flanking the *HAA1* and *ACE1* homologous genes were positioned according to their synteny and represented by different colours consistent with their sequence cluster (neighbour genes belonging to the same sequence cluster were represented by the same colour). Horizontal elements represent the chromosomal segments and vertical elements the homologous genes across the species.

The lineages defined by the *HAA1*/*ACE1* neighbour genes distribution in Saccharomycetaceae family are consistent with the phylogenetic groups established by the phylogenetic resolution of Haa1/Ace1 protein family.

4.2.5 Haa1 amino acid variations in *S. cerevisiae*

In what concerns the analysis performed among *S. cerevisiae* strains, the 54 Haa1 homologous sequences obtained from different genomes were grouped into 16 different sub-groups as a consequence of the high protein sequence similarity observed in this species. The pairwise comparison between the 16 different Haa1 sequences found in *S. cerevisiae* illustrates the existence of 17 amino acid variations always involving point substitutions between two amino acid alternative forms in protein sequence (**table 11**). These variations are distributed throughout both N- and C-terminal protein domains and represent a tiny portion of the complete sequence (17/694 ~2.45%). The combination of two or more substitution mutations was found in many Haa1 sequences.

Table 11. Amino acid variations observed in the Haa1 transcription factor from different *S. cerevisiae* strains. 'X' symbol represent amino acid residues variation in sub-group sequences analysed in relation to the A sub-group that includes the Haa1 sequence from *S. cerevisiae* S288c. Underlined variations express synonymous mutations (both amino acids have the same biochemical proprieties).

		N-terminal domain				C- terminal domain												
Group	Sub-group	A81V	P127S	L133F	L186F	I198V	G232S	N342S	S347N	N364I	G369S	S377F	Q420D	D495N	I604M	A633V	S649R	T671A	Nº of strains
1	A																		21
	B													X					1
	C									X									2
	D					X				X									1
	L			X						X									1
2	E					X	X			X	X	X	X						1
	F					X	X			X	X	X	X		X				1
	H					X	X	X		X	X	X	X		X				3
	M		X			X	X		X	X	X	X	X		X	X		X	4
	N	X				X	X		X	X	X	X	X		X				1
3	G					X	X	X		X	X	X	X						9
	I				X	X	X	X		X	X	X	X						2
	J				X	X	X	X		X	X	X	X				X		1
	O	X				X	X	X	X	X	X	X	X				X		1
	P	X				X	X	X	X	X	X	X	X						4
	Q	X			X	X	X	X	X	X	X	X	X						1

The phylogenetic analysis of the 54 Haa1 homologues sequences of the *S. cerevisiae* strains considered in this study (**table 12**) allowed the identification of three well defined phylogenetic groups (**figure 20**). All of the 12 Haa1 homologous sequences of *S. cerevisiae* strains previously described as being involved in wine fermentation (well adapted to high ethanol concentrations and low pH) (underlined in **table 12**) are included in the group 1 of the phylogenetic tree (sub-groups A, B, C, and L). Noteworthy, the 7 Haa1 homologous sequences belonging to industrial strains already recognized as being used in industrial biorefineries or distilleries (bolded in **table 12**) are spread among the three phylogenetic groups here defined (sub-groups D, I, N, P and Q). In the first case, there seems to be a correlation between the phenotypic traits, the vigorous ability of ferment sugars to ethanol, and the Haa1 amino acid sequence conservation. However this relationship is negligible for the Haa1 from *S. cerevisiae* strains associated to industrial processes. The overall analysis was still limited to the biological information available for the phenotypic traits and ecological niches linked to each *S. cerevisiae* strain which in some cases is still poorly known. An interesting fact is that the different *S. cerevisiae* strains, characterized by different metabolic traits associated to its biological relevance (baking,

clinical, fermentation and wild) are widely dispersed through the three phylogenetic groups identified in this analysis.

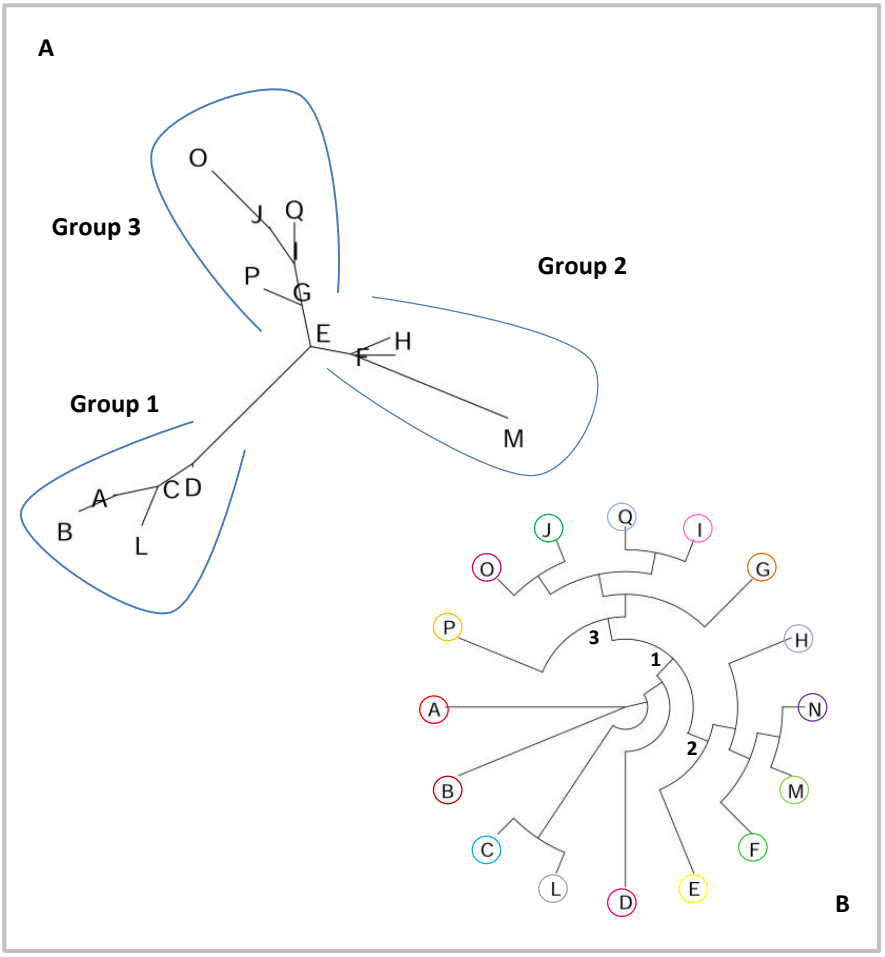

Figure 20. Phylogenetic analysis of the Haa1 amino acid sequence in *S. cerevisiae*. (A) Radial phylogram showing the amino acid sequences similarity distances between the Haa1 homologues from *S. cerevisiae* strains analysed; **(B)** Circular cladogram illustrating the tree topology. PROTDIST/NEIGHBOUR program from PHYLIP package was used to calculate the phylogenetic tree and retrieved results examined in Dendroscope software. Sequence codes were attributed according to the **table 12**. The sequence from sub-group B sequence is the outgroup. With the exception of Haa1 homologous sequences represented by the D, L, N, I, O and Q sub-group codes, all the other Haa1 sequences are represented (at strains level) at least once in a previous study[105].

Table 12. *S. cerevisiae* strains examined in this work. Underlined strains (well adapted to high ethanol concentrations and low pH) are characterized by a high fermentative performance; strains indicating at bold comprise the industrial ones. This analysis was performed using 60 *S. cerevisiae* genomes available in 2013 in the SGD, Genomes Online Database (GOLD) from Joint Genome Institute (JGI) (U.S. Department of Energy (DOE)), Genbank and Sanger Institute. The number of Haa1 sequences analysed was reduced to 54 sequences due to the existence of frameshift mutations. Currently, the number of genomes of *S. cerevisiae* available is around 400, although some of them are not completed sequenced.

Group	Sub-group	Strain	Gene name	Length	Source and geographic origin	Sub-group frequency (%)
1	A	RM11-1a	sace_14_5_e02890	660	Fermentation (wine European lineage)[105]	38.89
		S288c	sace_1_ypr008w	627	Laboratory (Americas)[105]	
		322134S	sace_16_5_e02800	687	Clinical (Europe)[105]	
		378604X	sace_17_5_e02890	694	Clinical (Europe)[105]	
		DBVPG1106	sace_18_5_2700	694	Fermentation (wine European lineage; Americas)[105]	
		DBVPG6765	sace_24_5_2750	694	Unknown (wine European lineage)[105]	
		DBVPG1373	sace_19_5_2720	694	Wild (wine European lineage; Europe)[105]	
		AWRI796	sace_2_6_f02900	694	Wine Fermentation[108]	
		DBVPG1788	sace_20_5_2690	694	Wild (wine European lineage; Europe)[105]	
		DBVPG1853	sace_21_5_2730	694	Fermentation (Africa)[105]	
		L 1374	sace_26_5_2690	694	Fermentation (wine European lineage; Americas)[105]	
		L 1528	sace_27_5_e02800	694	Fermentation (wine European lineage; Americas)[105]	
		cen.pk113-7d	sace_3_5_e02840	694	Laboratory[109]	
		YIIc17 E5	sace_39_5_e02860	694	Fermentation (Europe)[105]	
		YJM975	sace_40_5_e02800	694	Clinical (wine European lineage; Europe)[105]	
		YJM978	sace_41_5_e02810	694	Clinical (wine European lineage; Europe)[105]	
		Sigma1278b	sace_48_6_f02790	694	Laboratory[110]	
		w303	sace_51_5_e02840	694	Laboratory (unknown)[105]	
		boulardii EDRL	sace_53_30_ad00740	694	Fermentation; Therapeutic[111]	
		BY4741	sace_54_5_e02820	694	Laboratory[112]	
		BY4742	sace_55_5_e02820	694	Laboratory[112]	
	B	BC187	sace_18_5_e02810	694	Fermentation (wine European lineage; Asia)[105]	1.85
	C	YS9	sace_47_5_e02860	694	Baking (Asia)[105]	3.70
		YJM981	sace_42_5_e02860	694	Clinical (wine European lineage; Europe)[105]	
2	D	**JAY291**	sace_6_128_dx00150	694	Ethanol industrial production[113]	1.85

	L	EC1118	sace_56_4_d02800	694	Fermentation[114]	1.85
	E	DBVPG6040	sace_22_5_2750	694	Fermentation (Europe)[105]	1.85
	F	YS4	sace_46_5_e02850	694	Baking (Europe)[105]	1.85
	H	T7	sace_58_107_dc00100	694	Baking; Clinical*	5.56
		NCYC361	sace_29_5_2700	694	Fermentation (Europe)[105]	
		UWOPS83 787 3	sace_34_5_2770	694	Wild (Americas)[105]	
	M	Y55	sace_38_5_e02850	694	Laboratory (Europe)[105]	7.41
		SK1	sace_30_5_e02890	694	Laboratory (Americas)[105]	
		NCYC110	sace_28_5_e02880	694	Fermentation (West Africa lineage; Africa)[105]	
		DBVPG6044	sace_23_5_2800	694	Fermentation (West Africa lineage; Africa)[105]	
	N	**ZTW1**	sace_60_1_a02810	694	Fermentation (bioethanol production)[115]	1.85
3	**G**	273614N	sace_15_5_e02830	668	Clinical (Europe)[105]	16.67
		YJM789	sace_52_2_b02850	694	Clinical (Europe)[105]	
		YB210	sace_59_212_hd00150	694	Fermentation[116]	
		YPS606	sace_44_5_2790	694	Wild (North America lineage; Americas)[105]	
		UWOPS03 461 4	sace_31_5_2770	694	Wild (Malaysian lineage; Asia)/ Industrial distillers yeast[105]	
		UWOPS05 217 3	sace_32_5_2750_2	694	Wild (Malaysian lineage; Asia)/ Biofuel production[105]	
		UWOPS05 227 2	sace_33_5_2750	694	Wild (Malaysian lineage; Asia)/ Fermentation; Commercial wine yeast[105]	
		UWOPS87 2421	sace_35_5_2720	694	Wild (Americas)[105]	
		YPS128	sace_43_5_e02890	694	Wild (North American lineage; Americas)[105]	
	I	**M3837**	sace_11_4_d02800	644	Fermentation (West Africa lineage)[77]	3.70
		M3838	sace_12_3_2680	646	Fermentation (West Africa lineage)[77]	
	J	Y9	sace_36_5_e02870	694	Fermentation (Sake lineage; Asia)[105]	1.85
	O	Kyokai7	sace_57_6_f02850	694	Industrial sake production[117]	1.85
	P	**M3836**	sace_10_3_1630	642	Industrial distillers yeast[77]	7.41
		M3839	sace_13_4_1670	657	Industrial distillers yeast[77]	
		Y12	sace_37_5_e02870	694	Fermentation (Sake lineage; Africa)[105]	
		K11	sace_25_5_e02860	694	Fermentation (Sake lineage; Asia)[105]	
	Q	**M3707**	sace_9_4_d01730	694	Industrial distillers yeast[77]	1.85

(*) http://www.ncbi.nlm.nih.gov/bioproject/60387

The Haa1 amino acid variations revealed herein were also compared with the Haa1 mutated sequence obtained in a patent[17] that were demonstrated to improve the acetic acid tolerance in yeast cells exposed to this inhibitor (results not shown) . Nevertheless, none of reported mutations obtained from environmental efficient wine and industrial *S. cerevisiae* fermentative strains analysed in our study were found in the Haa1 sequences from this patent.

5. Discussion

In this Masters' thesis we aimed at elucidating the role in acetic acid stress of genes homologous to *S. cerevisiae HAA1* in two strains of the remarkably high weak acid tolerant species *Zygosaccharomyces bailii* and in the interspecies hybrid strain ISA1307. The heterologous expression of *HAA1* ORFs from *Z. bailii* strains CLIB213^T and IST1302 in *S. cerevisiae* BY4741 *haa1*Δ mutant, under the control of *S. cerevisiae HAA1* promoter and cloned in the same expression vector, has provided strong indications that *HAA1* is a determinant of acetic acid tolerance in *Z. bailii* species and a functional homologue of *S. cerevisiae HAA1*. Indeed, the expression of *HAA1* ORFs from *Z. bailii* strains CLIB213^T and IST1302 was found to complement the acetic acid susceptibility phenotype of *S. cerevisiae haa1*Δ mutant and also to improve the acetic acid tolerance in *S. cerevisiae* BY4741 by reducing the duration of yeast latency phase. Moreover, the expression of *ZbHAA1* was demonstrated to confer a higher tolerance to acetic acid when compared to the expression of the *ScHAA1* given by the lower latency phase duration observed in host cells expressing the first one in acetic acid conditions. Yeast cells expressing the *ZbHAA1* under study and the *ScHAA1* growing in acetic acid conditions reached a lower final biomass when compared to those cultivated in absence of acetic acid. This finding is consistent to the idea that lower biomass yields are a consequence of the energy requirements related with energy required to stress response mechanisms[2] given by a functional Haa1. The slightly different growth profiles exhibited by the recombinant strains expressing *ZbHAA1* from CLIB213^T or IST302 in comparison to the observed by the *ScHAA1* expression might be the result of specific amino acid point changes in the DBD and/or TAD domains that affect the Haa1 phosphorylatable state. Indeed the phosphorylation sate of the Haa1 from *S. cerevisiae* was previous described as being crucial to this TF activation[60]. However, due to the higher variation among *S. cerevisiae* and *Z. bailii* strains' TAD sequences it was not possible to attribute these differences to specific amino acid residues. The similar growth profiles observed by the yeast cells harbouring the *ZbHAA1* from CLIB213^T and IST302 in acetic acid culture was anticipated given that both genes code for proteins that share 99.6% of identity and all of predicted phosphorylatable amino acids are conserved in number and position.

The expression of ZBAI_B ORF from ISA1307 showed a deficient complementation of *haa1*Δ susceptibility phenotype to acetic acid stress and did not improve the acetic acid tolerance in *S. cerevisiae* BY4741 strain, being impossible to relate this ORF with a functional *S. cerevisiae* Haa1 homologue. This absence of functionality must be related to the ZBAI_B shorter length in

comparison to the *ZbHAA1* genes and the *ScHAA1* which can either be attributed to an error in ISA1307 genome sequencing, and consequently to gene annotation, or even to a point mutation that replaced the nucleotide thymine by a cytosine in the hypothesized START codon (ATG → ACG). Overall, the results suggest that the first 30 amino acids from N-terminal domain which compose the DBD of *HAA1* are essential for an efficient response against acetic acid negative effects, even though are not crucial for yeast cells survival at 80 mM of acetic acid. According to the prominent results gathered for the expression of *ZbHAA1* from CLIB213^{T} and IST302 strains, the relative conservation of the Haa1 length in Saccharomycetaceae, and the intrinsic acetic acid resistance already described for the hybrid ISA1307, it is likely that the ISA1307 has at least one functional Haa1 that might confer resistance to acetic acid. Being true that the ZBAI_B ORF has a mutation in its presumable 'ancestral START codon', upstream to the currently considered, this fact could indicate the loss of a redundant function maybe maintained through the ZBAI_A ORF. Regarding all of these evidences and the information available at this point, it would be interesting to sequence the upstream region of the annotated ZBAI_B ORF, as well as, to test the effect of the expression of the ZBAI_A ORF from ISA1307 in yeast cells response to acetic acid.

The bioinformatics analysis performed in this work was dedicated to the reconstruction of the evolutionary history of the Haa1 and its paralogue Ace1 in the Saccharomycetaceae family that include the pre-WGD species *Z. bailii* and post-WGD *S. cerevisiae* species. The phylogenetic analysis, based on protein sequence similarity, was also extended to other hemiascomycetous species. The conservation pattern favouring the N-terminal domain (that includes the copper regulatory domain)[53] of Haa1 and Ace1 proteins reflects the functional constraints associated with this protein region (proved in the experimental part of this work to Haa1) which seems to be subjected to a negative selection in Hemiascomycetes. The confirmation of this last insight requires further studies at the level of Haa1 and Ace1 homologous protein sequences, including the search of amino acid conserved motifs in the N-terminal domain.

The two phylogenetic analyses performed in this work, focused on the Haa1 and Ace1 homologous in Saccharomycetaceae for the complete sequences or the initial portion corresponding to the N-terminal domain using the first 65 amino acids of Haa1 from *S. cerevisiae* as the cut-off mark in the multiple alignment in hemiascomycetous species, led to similar protein groups, consistent with the main clusters of the phylogenetic tree of the yeast species already proposed in previous studies[67,107]. The consistency of the results here gathered reinforces the extremely high divergence associated to the C-terminal domain of the

Haa1/Ace1 protein family. Indeed, the C-terminal domain did not add any relevant information to the definition of each phylogenetic group and the N-terminal domain was sufficient to ensure the Haa1 and Ace1 phylogenetic resolution. The radial structure (with long and dispersed branches) which characterizes both phylogenetic trees evidences the high divergence among the protein sequences under analysis. This divergence is particularly strong in the phylogenetic group of the *S. cerevisiae* Ace1 transcription factor homologues. The phylogenetic analysis also indicates a strong divergence between the Haa1 and Ace1 proteins since the corresponding homologous sequences remain in different groups located at opposite sides in the phylogenetic tree. Moreover, the synteny observed between the *HAA1* and *ACE1* homologous genes is very poor. This last observation reinforces the idea that the Haa1/Ace1 protein family must be under different evolutionary pressures. Many hypotheses have been addressed to explain this issue and to the mechanisms leading to propose the Haa1 and Ace1 as paralogous proteins (according to the explanation addressed to **figure 6** of the Introduction section). An interesting discovery is that the ancient *Schizosaccharomyces pombe* presents a well conserved Mac1/Ace1/Haa1 homologue (Cuf1) linked to nutritional copper sensing[118]. The confirmation of a copper domain associated to the Cuf1 (previously described and confirmed by a BLASTP analysis, results not shown) suggests that this protein is an Haa1/Ace1 orthologue. Taking into account this insight and the Haa1 action in acetic acid tolerance demonstrated in the experimental part of this work for the pre-WGD species *Z. bailii* it is likely that the Haa1/Ace1 homologues from pre-WGD species could simultaneously possess both the Haa1 and Ace1 functions currently observed in post-WGD lineages (split off during the WGD by subfunctionalization). Nevertheless, it should not be discarded the hypothesis that Ace1 function was only acquired after WGD by neofunctionalization. In order to clarify this issue it would be interesting to investigate the functional activity of the *Z. bailii* Haa1/Ace1 homologous proteins to copper ion stress.

One possible explanation for the early divergence observed for the Haa1 and Ace1 paralogues and thus to the potential split of functions between the Haa1 and Ace1 proteins could be related with a presumable large deletion in the Ace1 C-terminal domain subsequent to WGD. In addition, successive mutations accumulated in both Haa1 and Ace1 proteins could have contributed to the functional divergence within each gene sub-lineage. The continuous small deletions over the evolutionary time that have affected the Ace1 protein reinforce the existence of a strong selective pressure to shorten this protein.

Both phylogenetic and synteny studies have demonstrated the strong divergence linked to the emergence of the *HAA1* and *ACE1* sub-lineages from a WGD event. In fact the divergence

observed for the Haa1/Ace1 homologues evidences the strong separation forces that forge the functional divergence of the Haa1 and Ace1 transcriptions factors and that could have provided an opportunity to retained new genes and create new regulatory networks[19,73]. An interesting observation is that it seems to have occurred a re-establishment of chromosomal environment conservation from *Naumovozyma* to *Saccharomyces* genera. This finding could be related to the higher evolutionary distance that separates this species from WGD event (in comparison to the other post-WGD species) and which could have promoted new and more stable genomic rearrangements.

Since we are particularly interested in yeast species with a high fermentative performance (frequently tolerant to acetic acid environments) we performed a preliminary selection in which the Haa1 amino acid sequences from several *S. cerevisiae* isolates, already described as robust fermentative strains, were compared. In this study we attempted to identify specific patterns of amino acid conservation in the Haa1 protein sequence that could be associated with a higher tolerance by yeast cells to acetic acid conditions. This analysis is based on the assumption that phenotypic variation has a good correlation with the genome traits in *S. cerevisiae*[105]. Therefore, we focused our attention in *S. cerevisiae* strains previously reported as being involved in wine fermentation and able to support high ethanol concentrations at low pH, as well as in *S. cerevisiae* industrial strains involved in biorefineries[77,105,113–115]. In this context, we submitted several Haa1 amino acid sequences to a phylogenetic analysis. The results revealed the short evolutionary time separating the *S. cerevisiae* strains given the small number of changes observed in amino acid sequences of this transcription factor in this species. Additionally, it was verified a tendency for the Haa1 from wine fermentative strains highly resistant to high ethanol concentrations and low pH to exhibit a similar amino acid sequence pattern (all the sequences fall into the same tree group) which could be related to a domestication process for the improvement of fermentation properties or with the promotion of out-breeds by the geographic movement of strains[105]. This behaviour was not observed in the industrial strains which were found widely dispersed through the three phylogenetic groups identified. In order to analyse the evolutionary history of the TF Haa1 and the *S. cerevisiae*, the phylogenetic tree obtained in this work for the Haa1 homologues from *S. cerevisiae* (**figure 20**, in Results section) was compared to a phylogenetic tree of *S. cerevisiae*[105] (**figure 21**), although some genomes included in our analysis are not present in this study.

.

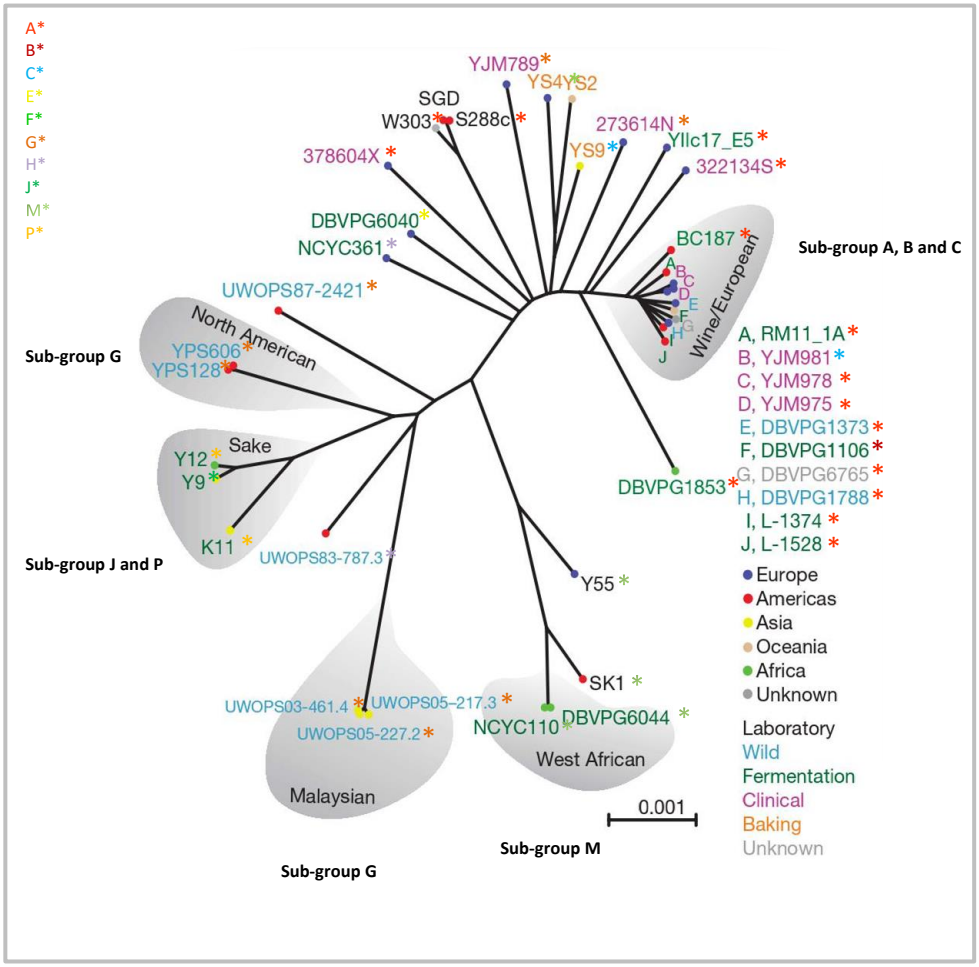

Figure 21. Phylogenetic reconstruction of *S. cerevisiae* according to the Liti and collaborators work[105]. *S. cerevisiae* strains belonging to clean lineages are highlighted in grey. Colours indicate source (name), geographic origin (dots) or its biological importance. Scale bars illustrate frequencies of base-pair differences. The letters positioned in the left side of the figure represent the sub-groups identified to the *S. cerevisiae* Haa1 phylogenetic analysis performed in this work.

At first sight, the two phylogenetic trees evidence marked high different topology, although there is a tendency for some Haa1 sequences (from a given sub-group) to fall into the corresponding and well-defined cluster linked to the clear lineages defined in this study (Wine/European, West Africa, Malaysian, Sake and North America)[105]. Nevertheless, the phylogenetic relationships among the different clean lineages clusters do not match with those observed for the Haa1 homologues. For the remaining mosaic lineages this comparison becomes more difficult and it is difficult the establishment of phylogenetic relationships regarding the Haa1 proteins and the corresponding *S. cerevisiae* strains.

Overall, the results obtained from this last analysis evidence the duality associated to genomic/phenotypic traits and environment distribution in *S. cerevisiae*. On one hand, the global population structure of *S. cerevisiae* attributes specific genotypes to specific environments; on the other hand, identical ecological niches enclose a broad range of strains which considerable differ in genomic and phenotypic characteristics[105]. This issue highlights the complexity linked to *S. cerevisiae* environment distribution that was recently entitled as a 'nomad' species[119]. The most promising results retrieved from this analysis suggest that the Haa1 sequences found in the phylogenetic tree group 1 and related to efficient wine fermentative strains (high resistance to low pH and high ethanol concentrations) are prominent *S. cerevisiae* candidates to follow up in the improvement of yeast strains for biotechnological industrial purposes. Taking into account these results and that in the experimental part of this work it was only tested the effect of the expression of *HAA1* from *S. cerevisiae* BY4741 strain in yeast cells under acetic acid conditions, it could be interesting to test other *HAA1 S. cerevisiae* candidates like those belonging to the efficient *S. cerevisiae* strains identified in this last phylogenetic analysis to acetic acid stress and compare the results with the expression of *ZbHAA1* in the same conditions.

6. Conclusion and future perspectives

The preferential approach to access the role of the Haa1 from *Z. bailii* in acetic acid response would be to evaluate the Haa1 action in this species. This process has been limited by the few molecular procedures optimized for the *Z. bailii* species. Regarding all the efforts, the first steps towards the construction the *Z. bailii* deletion mutant *haa1*Δ are actually on course in our laboratory and the preliminary results have already proven to be very promising. Indeed, some studies have also reported successful molecular experiments involving the *Z. bailii* engineering[120,121], which could be in the origin of the usage of this microorganism as a novel genetic background for bioethanol production. However there is still a long way ahead and the future approaches involving *Z. bailii* species must include the improvement of ethanol yield production by this microorganism. Another strategy to develop more robust microorganisms, strongly resistant to acetic acid, and able to be applied at industrial large scale, as cellular factories, could include the heterologous expression of *Z. bailii* proteins in genomic alternative backgrounds that are better adapted to fermentation conditions and industrial scale-up processes. In this context, this work showed that the Haa1 transcription factor is an attractive protein to follow up. Industrial applications, like those found in alcoholic fermentations, frequently enclose multiple stresses, requiring microorganisms well adapted to multiple toxic compounds and other non-physiological conditions like extreme temperature or pH, without these factors compromising its production efficiency. Beyond the industrial application linked to *Z. bailii* species as a microbial cell factory, the analysis of molecular mechanisms underlying the remarkably resistance to acetic acid present by this species will allow to get further insights about *Z. bailii* tolerance to food and beverages preservatives and, who knows, to design more efficient strategies to counteract one of the most problematic food spoilage agent. In a near future it is expected that the amount of bioinformatics tools available for *Z. bailii* and other non-*cerevisiae* strains will increase, allowing the development of metabolic engineering approaches for this species, which will be undoubtedly an important mark to enhance our understood about this microorganism. One of the main objectives of this work was to obtained new clues that could elucidate the functional analysis of the Haa1 from *Z. bailii* and also other (non-)*cerevisiae* strains as promising candidates to improve acetic acid tolerance. This kind of analysis gave us the first insights about the potential of specific Haa1 sequences to design more reliable strategies in order to explore yeast cells laboratory and industrial applications.

7. References

1. Dos Santos, S. C. & Sá-Correia, I. Yeast toxicogenomics: lessons from a eukaryotic cell model and cell factory. *Curr. Opin. Biotechnol.* **33,** 183–191 (2015).
2. 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).
3. Stratford, M. *et al.* Extreme resistance to weak-acid preservatives in the spoilage yeast Zygosaccharomyces bailii. *Int. J. Food Microbiol.* **166,** 126–134 (2013).
4. Sá-Correia, I., Guerreiro, J. & Loureiro-Dias, M. in *Encyclopedia of Food Microbiology - Zygosaccharomyces* (ed. Elsevier) **3,** 178–189 (Academic Press, 2014).
5. 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).
6. Fleet, G. H. Yeasts in foods and beverages: impact on product quality and safety. *Curr. Opin. Biotechnol.* **18,** 170–175 (2007).
7. Sousa, M. J., Ludovico, P., Rodrigues, F., Leão, C. & Côrte-Real, M. in *Cell Metabolism - Cell Homeostasis and Stress Response* 73–100 (2012).
8. Palmqvist, E. & Hahn-Hagerdal, B. Fermentation of lignocellulosic hydrolyzates. II: inhibitors and mechanisms of inhibition. *Bioresour. Technol.* **74,** 25–33 (2000).
9. Almeida, J. R. *et al.* Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae. *J. Chem. Technol. Biotechnol.* **82,** 340–349 (2007).
10. Erickson, B., Nelson & Winters, P. Perspective on opportunities in industrial biotechnology in renewable chemicals. *Biotechnol. J.* **7,** 176–85 (2012).
11. Martín, C., Marcet, M., Almazán, O. & Jönsson, L. J. Adaptation of a recombinant xylose-utilizing Saccharomyces cerevisiae strain to a sugarcane bagasse hydrolysate with high content of fermentation inhibitors. *Bioresour. Technol.* **98,** 1767–1773 (2007).
12. 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).
13. Goffeau, A. Four years of post-genomic life with 6000 yeast genes. *FEBS Lett.* **480,** 37–41 (2000).
14. 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).
15. 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).
16. 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).
17. Zahn, K. & Jacobson, S. Acetate resistance in yeast based on introduction of a mutant haa1 allele. *United States Pat. Appl. Publ.* (2014).
18. Li, H. & Johnson, A. D. Evolution of transcription networks-lessons from yeasts. *Curr. Biol.* **20,** R746–R753 (2010).

19. Pougach, K. *et al.* Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nat. Commun.* **5,** 1–11 (2014).
20. Christie, K. R., Hong, E. L. & Cherry, J. M. Functional annotations for the Saccharomyces cerevisiae genome: the knowns and the known unknowns. *Trends Microbiol.* **17,** 286–294 (2009).
21. Loureiro, V. & Querol, a. The prevalence and control of spoilage yeasts in foods and beverages. *Trends Food Sci. Technol.* **10,** 356–365 (1999).
22. Fernandes, L., Côrte-Real, M., Loureiro, V., Loureiro-Dias, M. C. & Leão, C. Glucose respiration and fermentation in Zygosaccharomyces bailii and Saccharomyces cerevisiae express different sensitivity patterns to ethanol and acetic acid. *Lett. Appl. Microbiol.* **25,** 249–253 (1997).
23. 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).
24. Van Maris, A. J. a *et al.* Alcoholic fermentation of carbon sources in biomass hydrolysates by Saccharomyces cerevisiae: Current status. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **90,** 391–418 (2006).
25. Larsson, S. *et al.* The generation of fermentation inhibitors during dilute acid hydrolysis of softwood. *Enzyme Microb. Technol.* **24,** 151–159 (1999).
26. Martín, C. & Jönsson, L. J. Comparison of the resistance of industrial and laboratory strains of Saccharomyces and Zygosaccharomyces to lignocellulose-derived fermentation inhibitors. *Enzyme Microb. Technol.* **32,** 386–395 (2003).
27. Endo, A., Nakamura, T., Ando, A., Tokuyasu, K. & Shima, J. Genome-wide screening of the genes required for tolerance to vanillin, which is a potential inhibitor of bioethanol fermentation, in Saccharomyces cerevisiae. *Biotechnol. Biofuels* **1,** 1–6 (2008).
28. Teixeira, M. C., Raposo, L. R., Mira, N. P., Lourenço, A. B. & Sá-Correia, I. Genome-wide identification of Saccharomyces cerevisiae genes required for maximal tolerance to ethanol. *Appl. Environ. Microbiol.* **75,** 5761–5772 (2009).
29. Doğan, A., Demirci, S., Aytekin, A. Ö. & Şahin, F. Improvements of Tolerance to Stress Conditions by Genetic Engineering in Saccharomyces Cerevisiae during Ethanol Production. *Appl. Biochem. Biotechnol.* **s/v,** 1–15 (2014).
30. Rodrigues, F., Ludovico, P. & Leão, C. in *Biodiversity and Ecophysiology of Yeasts - Sugar Metabolism in Yeasts: an Overview of Aerobic and Anaerobic Glucose Catabolism* 101–121 (2006).
31. Merico, A., Capitanio, D., Vigentini, I., Ranzi, B. M. & Compagno, C. Aerobic sugar metabolism in the spoilage yeast Zygosaccharomyces bailii. *FEMS Yeast Res.* **4,** 277–283 (2003).
32. Vemuri, G. N., Eiteman, M. a, McEwen, J. E., Olsson, L. & Nielsen, J. Increasing NADH oxidation reduces overflow metabolism in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 2402–2407 (2007).
33. Mollapour, M. & Piper, P. W. The ZbYME2 gene from the food spoilage yeast Zygosaccharomyces bailii confers not only YME2 functions in Saccharomyces cerevisiae, but also the capacity for catabolism of sorbate and benzoate, two major weak organic acid preservatives. *Mol. Microbiol.* **42,** 919–930 (2001).

34. 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).
35. 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).
36. Fernandes, L., Côrte-Real, M. & Leao, C. A peculiar behaviour for cell death induced by weak carboxylic acids in the wine spoilage yeast Zygosaccharomyces bailii. *Lett. Appl. Microbiol.* **28,** 345–349 (1999).
37. Ludovico, P., Sousa, M. J., Silva, M. T., Leão, C. & Côrte-Real, M. Saccharomyces cerevisiae commits to a programmed cell death process in response to acetic acid. *Microbiology* **147,** 2409–2415 (2001).
38. Ludovico, P., Sansonetty, F., Silva, M. T. & Côrte-Real, M. Acetic acid induces a programmed cell death process in the food spoilage yeast Zygosaccharomyces bailii. *FEMS Yeast Res.* **3,** 91–96 (2003).
39. Simões, T., Mira, N. P., Fernandes, a. R. & Sá-Correia, I. The SPI1 gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. *Appl. Environ. Microbiol.* **72,** 7168–7175 (2006).
40. Ullah, A., Orij, R., Brul, S. & Smits, G. J. Quantitative analysis of the modes of growth inhibition by weak organic acids in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **78,** 8377–8387 (2012).
41. Mira, N. P., Palma, M., Guerreiro, J. F. & Sá-Correia, I. Genome-wide identification of Saccharomyces cerevisiae genes required for tolerance to acetic acid. *Microb. Cell Fact.* **9,** 1–13 (2010).
42. Sousa, M. J., Miranda, L., Côrte-Real, M. & Leão, C. Transport of acetic acid in Zygosaccharomyces bailii: Effects of ethanol and their implications on the resistance of the yeast to acidic environments. *Appl. Environ. Microbiol.* **62,** 3152–3157 (1996).
43. Mollapour, M. & Piper, P. W. Hog1 mitogen-activated protein kinase phosphorylation targets the yeast Fps1 aquaglyceroporin for endocytosis, thereby rendering cells resistant to acetic acid. *Mol. Cell. Biol.* **27,** 6446–6456 (2007).
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. 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,** 1–15 (2014).
46. Tenreiro, S., Rosa, P. C., Viegas, C. a. & Sá-Correia, I. Expression of the AZR1 gene (ORF YGR224w), encoding a plasma membrane transporter of the major facilitator superfamily, is required for adaptation to acetic acid and resistance to azoles in Saccharomyces cerevisiae. *Yeast* **16,** 1469–1481 (2000).
47. Tenreiro, S. *et al.* AQR1 gene (ORF YNL065w) encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine in Saccharomyces cerevisiae. *Biochem. Biophys. Res. Commun.* **292,** 741–748 (2002).

48. 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).
49. Stratford, M. *et al.* Population heterogeneity and dynamics in starter culture and lag phase adaptation of the spoilage yeast Zygosaccharomyces bailii to weak acid preservatives. *Int. J. Food Microbiol.* **181,** 40–47 (2014).
50. Pampulha, M. E. & Loureiro-Dias, M. C. Combined effect of acetic acid, pH and ethanol on intracellular pH of fermenting yeast. *Appl. Microbiol. Biotechnol.* **31,** 547–550 (1989).
51. Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. a. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* **14,** 283–291 (2004).
52. Latchman, D. S. Transcription Factors: An Overview. *Int. J. Biochem. Cell Biol.* **29,** 1305–1312 (1997).
53. Keller, G., Ray, E., Brown, P. O. & Winge, D. R. Haa1, a Protein Homologous to the Copper-regulated Transcription Factor Ace1, is a Novel Transcriptional Activator. *J. Biol. Chem.* **276,** 38697–38702 (2001).
54. Rutherford, J. C. & Bird, A. J. Metal-Responsive Transcription Factors That Regulate Iron , Zinc , and Copper Homeostasis in Eukaryotic Cells MINIREVIEW Metal-Responsive Transcription Factors That Regulate Iron , Zinc , and Copper Homeostasis in Eukaryotic Cells. *Eurkaryotic cell* **3,** 1–13 (2004).
55. 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).
56. Teixeira, M. C. *et al.* The YEASTRACT database: An upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae. *Nucleic Acids Res.* **42,** 1–6 (2014).
57. Abbott, D. a., Suir, E., Van Maris, A. J. a & Pronk, J. T. Physiological and transcriptional responses to high concentrations of lactic acid in anaerobic chemostat cultures of Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **74,** 5759–5768 (2008).
58. Goossens, a, de La Fuente, N., Forment, J., Serrano, R. & Portillo, F. Regulation of yeast H(+)-ATPase by protein kinases belonging to a family dedicated to activation of plasma membrane transporters. *Mol. Cell. Biol.* **20,** 7654–7661 (2000).
59. Henriques, S. The Yeast Transcription Factor Haa1p : Identification of amino acid residues and nucleotides essential for functional. *IST-UL* 1–10 (2010).
60. Sugiyama, M. *et al.* Nuclear localization of Haa1, which is linked to its phosphorylation status, mediates lactic acid tolerance in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **80,** 3488–3495 (2014).
61. 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).
62. Inaba, T. *et al.* An organic acid-tolerant HAA1-overexpression mutant of an industrial bioethanol strain of Saccharomyces cerevisiae and its application to the production of bioethanol from sugarcane molasses. *AMB Express* **3,** 1–7 (2013).

63. Dujon, B. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet.* **22,** 375–387 (2006).
64. Seret, M. L. & Baret, P. V. IONS: Identification of orthologs by neighborhood and similarity-an automated method to identify orthologs in chromosomal regions of common evolutionary ancestry and its application to hemiascomycetous yeasts. *Evol. Bioinforma.* **7,** 123–133 (2011).
65. Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3: Genes|Genomes|Genetics* **1,** 11–25 (2011).
66. Kurtzman, C. P. & Robnett, C. J. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. *FEMS Yeast Res.* **3,** 417–432 (2003).
67. Kurtzman, C. P. Phylogenetic circumscription of Saccharomyces, Kluyveromyces and other members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces, Naumovia, Vanderwaltozyma and Zygotorulaspora. *FEMS Yeast Res.* **4,** 233–245 (2003).
68. Wolfe, K. H. Comparative genomics and genome evolution in yeasts. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **361,** 403–412 (2006).
69. Baldauf, S. L. Phylogeny for the faint of heart: A tutorial. *Trends Genet.* **19,** 345–351 (2003).
70. Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* **13,** 303–314 (2012).
71. Kumar, S. & Filipski, A. in *Encyclopedia of life sciences - Molecular Phylogeny Reconstruction* (ed. Sons, J. W. &) 1–6 (2009).
72. Koonin, E. V. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **39,** 309–338 (2005).
73. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10,** 725–732 (2009).
74. 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. *Omi. A J. Integr. Biol.* **14,** 701–710 (2010).
75. Dias, P. J. & Sá-Correia, I. The drug:H^+ antiporters of family 2 (DHA2), siderophore transporters (ARN) and glutathione:H^+ antiporters (GEX) have a common evolutionary origin in hemiascomycete yeasts. *BMC Genomics* **14,** 1–22 (2013).
76. Dias, P. J. & Sá-Correia, I. Phylogenetic and syntenic analyses of the 12-spanner drug:H+ antiporter family 1 (DHA1) in pathogenic Candida species: evolution of MDR1 and FLU1 genes. *Genomics* **104,** 45–57 (2014).
77. Brown, S. D. *et al.* Genome Sequences of Industrially Relevant Saccharomyces cerevisiae Strain M3707, Isolated from a Sample of Distillers Yeast and Four Haploid Derivatives. *Genome Announc.* **1,** 1–2 (2013).
78. Seret, M.-L., Diffels, J. F., Goffeau, A. & Baret, P. V. Combined phylogeny and neighborhood analysis of the evolution of the ABC transporters conferring multiple drug resistance in hemiascomycete yeasts. *BMC Genomics* **10,** 1–11 (2009).

79. Martin, N., Ruedi, E. a, LeDuc, R., Sun, F.-J. & Caetano-Anollés, G. Gene-interleaving patterns of synteny in the Saccharomyces cerevisiae genome: are they proof of an ancient genome duplication event? *Biol. Direct* **2,** 1–22 (2007).
80. Drillon, G. & Fischer, G. Comparative study on synteny between yeasts and vertebrates. *Comptes Rendus - Biol.* **334,** 629–638 (2011).
81. Wolfe, K. H. Comparative genomics and genome evolution in yeasts. *Philos. Trans. R. Soc. B Biol. Sci.* **361,** 403–412 (2006).
82. Souciet, J.-L., Dujon, B. & Gaillardin, C. Comparative genomics of protoploid Saccharomycetaceae. *Genome Res.* **19,** 1696–1709 (2009).
83. Drillon, G. & Fischer, G. Comparative study on synteny between yeasts and vertebrates. *C. R. Biol.* **334,** 629–638 (2011).
84. Rolland, T. & Dujon, B. Yeasty clocks: Dating genomic changes in yeasts. *Comptes Rendus - Biol.* **334,** 620–628 (2011).
85. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387,** 708–713 (1997).
86. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature* **428,** 617–624 (2004).
87. Zimmer, C. & Emlen, D. J. *Evolution: Making Sense of Life*. *Journal of Chemical Information and Modeling* (CO: Roberts and Company Publishers, 2013).
88. Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* **9,** 938–950 (2008).
89. Comai, L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6,** 836–846 (2005).
90. Podlaha, O. & Zhang, J. in *Encyclopedia of Life Sciences - Pseudogenes and Their Evolution* (ed. John Wiley & Sons, L.) 1–8 (2010).
91. Zhang, J. Evolution by gene duplication: An update. *Trends Ecol. Evol.* **18,** 292–298 (2003).
92. Coulombe-Huntington, J. & Xia, Y. Regulatory Network Structure as a Dominant Determinant of Transcription Factor Evolutionary Rate. *PLoS Comput. Biol.* **8,** 1–11 (2012).
93. Raijman, D., Shamir, R. & Tanay, A. Evolution and selection in yeast promoters: Analyzing the combined effect of diverse transcription factor binding sites. *PLoS Comput. Biol.* **4,** 0077–0087 (2008).
94. 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).
95. 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,** 145–53 (1997).
96. Jansen, G., Wu, C., Schade, B., Thomas, D. Y. & Whiteway, M. Drag&Drop cloning in yeast. *Gene* **344,** 43–51 (2005).
97. Lööke, M., Kristjuahan, K. & Kristjuhan, A. Extraction of Genomic Dna From Yeasts for Pcr- Based Applications. *Biotechniques* **50,** 325–328 (2011).

98. Sanger, F., Nicklen, S. & Coulson, a R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467 (1977).
99. Hoffman, C. S. Preparation of yeast DNA, RNA and proteins. *Curr. Protoc. Mol. Biol.* **13,** 1–4 (1997).
100. Gietz, R. D. & Schiestl, R. H. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2,** 38–41 (2007).
101. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5,** 1–19 (2004).
102. Felsenstein, J. PHYLIP — Phylogeny Inference Package (Version 3.2). **5,** 164–166 (1989).
103. Huson, D. H. *et al.* Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8,** 1–6 (2007).
104. Zhao, X. Q. & Bai, F. W. Mechanisms of yeast stress tolerance and its manipulation for efficient fuel ethanol production. *J. Biotechnol.* **144,** 23–30 (2009).
105. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458,** 337–341 (2009).
106. Dujon, B. *et al.* Genome evolution in yeasts. *Nat. Publ.* **430,** 35–44 (2004).
107. Salichos, L. & Rokas, A. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497,** 327–331 (2013).
108. Vilanova, M. *et al.* Assimilable nitrogen utilisation and production of volatile and non-volatile compounds in chemically defined medium by Saccharomyces cerevisiae wine yeasts. *Appl. Microbiol. Biotechnol.* **77,** 145–157 (2007).
109. Van Dijken, J., Bauer, J., Brambilla, L. & Duboc, P. An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains. *Enzyme Microb. Technol.* **26,** 706 –714 (2000).
110. Dowell, R. D. *et al.* Genotype to Phenotype: A Complex Problem. *Science.* **328,** 469–469 (2010).
111. Khatri, I. *et al.* Gleaning evolutionary insights from the genome sequence of a probiotic yeast Saccharomyces boulardii. *Gut Pathog.* **5,** 1–8 (2013).
112. 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).
113. Argueso, J. J. L. *et al.* Genome structure of a Saccharomyces cerevisiae strain widely used in bioethanol production. *Genome Res.* **19,** 2258–2270 (2009).
114. Novo, M. *et al.* Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 16333–16338 (2009).
115. Zhang, K. *et al.* Genomic reconstruction to improve bioethanol and ergosterol production of industrial yeast Saccharomyces cerevisiae. *J. Ind. Microbiol. Biotechnol.* **42,** 207–218 (2014).
116. Engel, S. R. *et al.* The Reference Genome Sequence of Saccharomyces cerevisiae: Then and Now. *G3: Genes|Genomes|Genetics* **4,** 389–398 (2014).
117. Akao, T. *et al.* Whole-genome sequencing of sake yeast Saccharomyces cerevisiae Kyokai no. 7. *DNA Res.* **18,** 423–34 (2011).

118. Beaudoin, J., Mercier, A., Langlois, R. & Labbé, S. The Schizosaccharomyces pombe Cuf1 is composed of functional modules from two distinct classes of copper metalloregulatory transcription factors. *J. Biol. Chem.* **278,** 14565–77 (2003).
119. Goddard, M. R. & Greig, D. Saccharomyces cerevisiae: a nomadic yeast with no niche? *FEMS Yeast Res.* **15,** 1–6 (2015).
120. 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).
121. 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).
122. Dujon, B. Hemiascomycetous yeasts at the forefront of comparative genomics. *Curr. Opin. Genet. Dev.* **15,** 614–620 (2005).

8. Annexes

Annex I – *HAA1* homologous sequence from *Z. bailii* IST302

ATGGTGTTGATAAACGGTGTCAAGTACGCCTGTGAAAGATGTATAAGGGGCCATAGAGTTACCACTTG
CAACCACACAGATCAGCCTCTGATGATGATTAAGCCGAAGGGGAGGCCATCCACTACGTGTAACCACT
GCAAGGAGTTGAGGAAAAACAAGAACGCCAACCCGTCGGGCGTATGTACGTGCGGGCGACTCGAGAA
GAAACGGCTGGCGCAGAAGGCCAAGGAGGAAGCTCGAGCTAAGGCGAAGGAAGAGAAAAAACTACA
CGAGTGCCGTTGCGGATATGACGAGCCATGCAGATGCCATTCCAGCAGGAGGAGGCACCAGTCAAGG
AAAGTATCGGTGAATGGTAAGGGCAGCAGTTCGAGGCCAGTGTCTCAGCAGTTCGACGCAAATAACCA
CGTGATGTCACCAGTGTCGATGGATTCGAGCTCCAGTATGCATAACGTCAGTACTTCAGGCAACGGGA
AGTCCACCGAAAGCTCGGGTATGCTAGCTTCTGGTTTTCTAGATACAGAGGTTGGAAACGGCGGCAAA
GTGTCCAAAGAGTTTCATCAGGTGCCTTCACTCGCCTCAATATCATCACTGCATTCAGGGCAATCATTCG
ACCAAAAGGTTAATCTTCCACAATCTCCTCTTCTAGGAGGAGGTGGTTCACTTGGTTTTTTGAATGGCGT
GTGGTGGAAATCGCAACAGTGGAAATTTTACAAATTGGGGGGACGAGGCGTCACTGTATTCTGTCAGA
CTGATTCTAAAGTTAACCTGACGGAGCATAGTGGAGTTCTCACGGGAATGCAGCCTCCCTCAAGTAGTG
TTCCAAAGAGAACCTCTTCCACAAAGGCACAGCTGGGTAATGTCAAGGTACCGTTGGAGGAGTATCTG
TCCCCTTCAGACAACTTTAATTCTAAGATAAATGAGACGAGTTCTCCTATGCAGGACTGGTATTTTGAGA
GGACTCCTAATGAAGACCCCAATATCTCTGATCAGTTACTTGGAGTGGACGGCAGTGGCGGCGATGCT
GGCGCTAGCCTTCAAGCAGCGCAAAATAGTGGGCTTTTAGACATGTTTATGGACTCTTCGACGATTCCA
ATTCTGTCAAAAAGCAGTTTACTCATGCAGGACAAATTAGGCCCGGCCGTAGGTAATAACACTAACAAC
TGTACTACTAGCCCTTCAAATGGTCCCAAAGGGAGAACTTGGAGTAGCCCTATGGCTGCTGATAAATCG
GATACTCTTAGTGTAGATGGAGAAAGCGTCAGGAGCGTTGAAGCATTGTCATTAATACCGAGTTACAT
GGATATCCCAGACAGGGCTCCCAGCCTACACAACGTCCAATCGGTGCTGCATTCTCATCAACACCAACC
AGGACATTTCAATCAACAAAAGCAAGGCAGGAGGTCAGGCAGTGTCAGCAGGAGTCACAGACCTCCCC
CACCTACGTCCAGGAGCGGCGTTCCCGTAACCATCAATCCCTCCATGGTTAGTAGCATCGACGACACAA
TTAGCGTAAATTCGCTGCAAAGTCCCACAGGTTCCGTCCTCGATAACCACAGTTTGTCTACTCCCCTCAG
CAATGGGGCAGACTATGGATCTTGCAGACCGGAAAGAGTGAGATCACCGCTCTTGGGATTGTCAGAGA
ACAACCCAACGTCTCCACAATTTTATCAACAACAACCTCAAGTTTTGCAATCAGCGACTGGCCCATCAGA
CTCGGAACTGGATCAGTTACTAGGATTCGACACTGGTAATAATAGTGGCATCATGGACAATACAAACTA
TTCAGATAAAAATTATGGCAATTTTAATCCGAGTGCTAACATTAACCTCAACGGCAACCGCAATGGAAA
CGGTAATGTAAGTAGCAACAGCAACGGGAGCAGGAACGGGAGCGGGAGCGGAAACGGCAACAGTAA
TGGTAACCTGTTATTCCAAAACGACATGAATACTTTATCAAGCCCAGCAATGAAGGATGTTCTCCAAGA
CACGTCACCTATGAGTAGCAACCAGACTGTCTCGCCACCGAGTCAGTTACTCACGGAAAAAGGTTTTGC
CGATTTGGACGATTTCATGTCCACATTGTGA

Figure 22. Nucleotide sequence of *HAA1* homologue from *Z. bailii* IST302 obtained by the sequencing and annotation of the genome of *Z. bailii* strain IST302 accomplished by BSRG (IST).

Annex II – Haa1 amino acid variations in Haa1 from CLIB213^{T}, IST302, ISA1307 and BY4741

Table 13. Gap percentage among the Haa1 protein from *Z. bailii* CLIB213^{T} and IST302, hybrid ISA1307 (ZBI_A and ZBI_B) and *S. cerevisiae* BY4741 strains species.

Specie	*Z. bailii* CLIB213^{T}	*Z. bailii* IST302	Hybrid ISA1307_A	Hybrid ISA1307_B	*S. cerevisiae* BY4741
***Z. bailii* CLIB213^{T}**		0.0	0.6	4.3	20.9
***Z. bailii* IST302**			0.6	4.3	20.9
ISA1307_A				4.9	21.5
ISA1307_B					24.8
***S. cerevisiae* BY4741**					

Table 14. Amino acid identity and similarity among the Haa1 protein from *Z. bailii* CLIB213^{T} and IST302, hybrid ISA1307 (ZBI_A and ZBI_B) and *S. cerevisiae* BY4741 strains. Protein identity and similarity percentages are represented at orange and white colours, respectively.

Specie	*Z. bailii* CLIB213^{T}	*Z. bailii* IST302	Hybrid ISA1307_A	Hybrid ISA1307_B	*S. cerevisiae* BY4741	ISA1307_A (-30 nt)
***Z. bailii* CLIB213^{T}**	100.0	99.6	96.7	95.2	36.9	92.4
***Z. bailii* IST302**	99.7	100.0	97.0	95.4	37.0	92.7
ISA1307_A	97.4	97.7	100.0	92.6	37.4	95.7
ISA1307_B	95.4	95.7	93.4	100.0	33.2	96.7
***S. cerevisiae* BY4741**	52.4	52.5	51.7	48.6	100.0	33.7
ISA1307_A (-30 a.a.)	93.1	93.4	95.7	97.6	47.8	100.0

Figure 23. Alignment of Haa1 homologous complete sequences from *Z. bailii* CLIB213^{T} and IST302 and hybrid ISA1307 and the Haa1 from *S. cerevisiae* S288c obtained with Geneious software. The different colours represent amino acid identity.

Table 15. Amino acid variations reported to the Haa1 protein from *Z. bailii* CLIB213^{T} and IST302 and hybrid ISA1307 (ZBI_A and ZBI_B) strains. Underlined variations indicate synonymous amino acid changes; amino acids marked with an (^{p}) represent phosphorylated residues

ISA1307(A)/ CLIB213^{T}	CLIB213^{T}/ ISA1307(A)	ISA1307(A)/ IST302	IST302/ ISA1307(A)	ISA1307(B)/ ISA1307(A)	ISA1307(A)/ ISA1307(B)
E67V	V67E	G151V	V151G	I121G	G151I
G151I	I151G	D161E	E161D	E131D	D161E
D161E	E161D	A257 T^{p}	T^{p}257A	M164I	I194M
A257T	T257A	P267Q	Q267P	T227A	A257T^{p}
P267Q	Q267P	Q275K	K275Q	Q237P	P267Q
Q275K	K275Q	G334V	V334G	K245Q	Q275K
G334V	V334G	E335D	D335E	V304G	G334V
E335D	D335E	N337 S^{p}	S^{p}337N	D305E	E335D
N337 S^{p}	S ^{p}337N	G344S^{p}	S^{p}344G	S^{p}307N	N337S ^{p}
G344 S^{p}	S^{p}344G	F377L	L377F	S314G	G344S^{p}
F377L	L377F	D378G	G378D	L347F	F377L
D378G	G378D	A381V	V381A	G348D	D378G
A381V	V381A	P407A	A407P	V351A	A381V
P407A	A407P	A469G	G469A	A377P	P407A
A469G	G469A	T^{p}623 N	N623T^{p}	G439A	A469G
H518R	R518H	625 NGS^{p}R	625Gap(4nt)	N593T	T^{p}623 N
T623N	N623T^{p}	G662D	D658G	594Gap(4nt)	625NGS^{p}R
625NGSR	625Gap(4nt)	I6223V	V659I	D628G	G662D
G662D	D659G	-	-	V629I	I663V
I663V	V660I	-	-	-	-
CLIB213^{T}/ IST302	**IST302/ CLIB213^{T}**	**CLIB213^{T}/ ISA1307(B)**	**ISA1307(B)/ CLIB213^{T}**	**IST302/ ISA1307(B)**	**ISA1307(B)/ IST302**
V67E	E67V	V67E	E67V	V151I	I121V
I151V	V151I	I194M	V151I	I194M	M164I
R518H	H518R	R518H	H518R	-	-

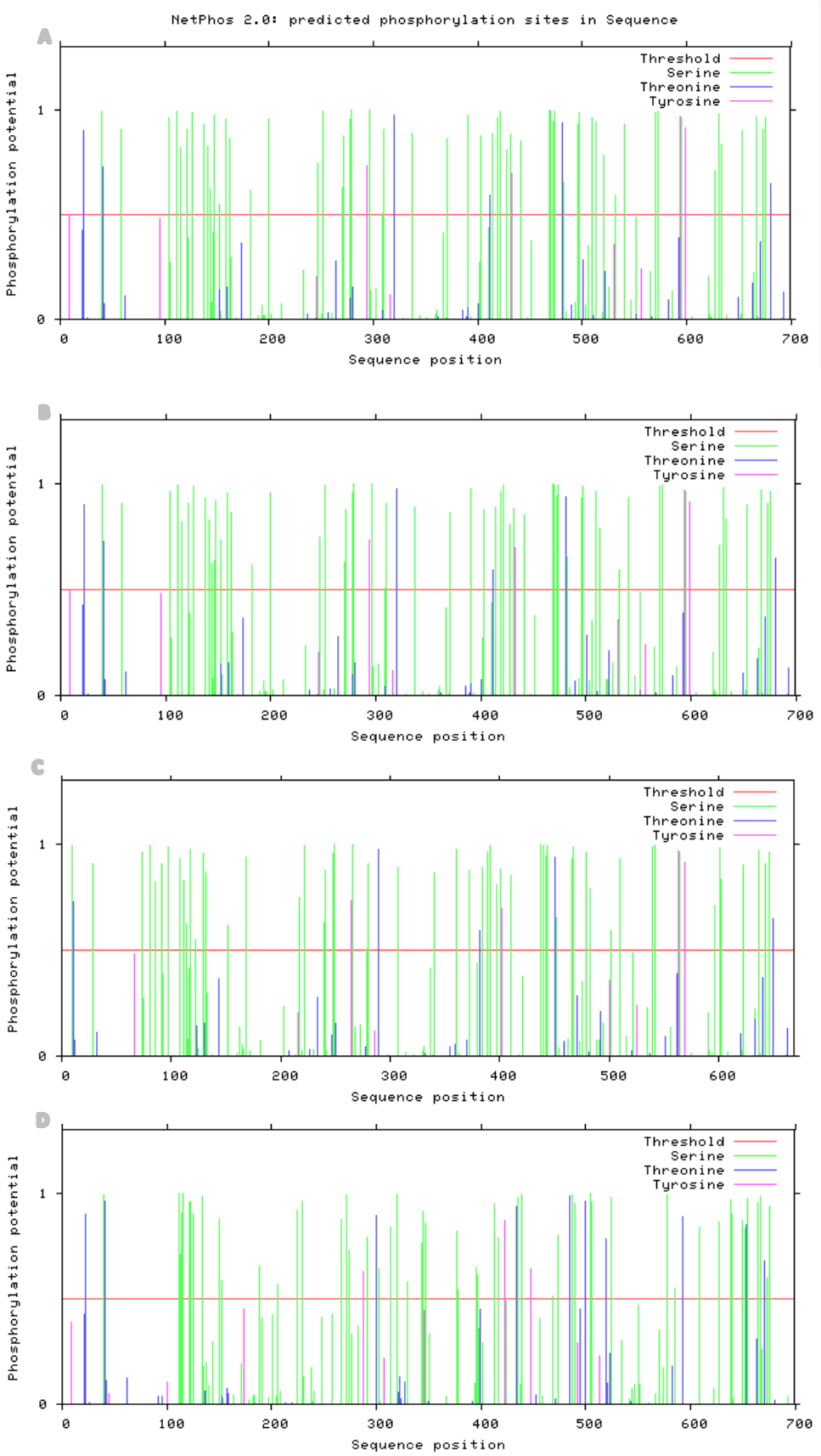

Figure 24. Amino acid predicted phosphorylated residues in the Haa1 protein from (A) *Z. bailii* CLIB213^{T}, (B) *Z. bailii* IST302, (C) hybrid ISA1307 (ZBAI_B) and *S. cerevisiae* BY4741 species. Diagrams obtained in the NetPhos 2.0 Server, Technical University of Denmark.

Table 16. Number of amino acid predicted phosphorylated residues number in the Haa1 protein from *Z. bailii* CLIB213^T and IST302, hybrid ISA1307 (ZBAI_B) and *S. cerevisiae* BY4741 species.

Strain	Serine	Threonine	Tyrosine	Total
***Z. bailii* CLIB213^T**	57	6	4	67
***Z. bailii* IST302**	57	6	4	67
Hybrid ISA1307_B	56	5	4	65
***S. cerevisiae* BY4741**	56	10	3	69

Annex III– Hemiascomycetes analysed in the Haa1 phylogenetic analysis

Table 17. Hemiascomycetous species examined in this work. The groups herein indicated are reported to the phylogenetic trees from **figures 13** and **14**. The absence of certain groups observed for some protein sequences is linked to frameshift mutations or a possible erroneous annotation (for the Haa1 homologue from ISA1397 (ZBAI_B)).

Phylogenetic complex	Genus	Specie	Strain	Gene name	Length	Cluster
'Saccharomyces complex' (Post-WGD)	*Saccharomyces*	*S. arboricola*	H-6	saar_1_3_c00950	227	2
				saar_1_6_f02790	694	1
		S. bayanus	623-6C	saba_1_3_c00480	232	2
				saba_1_11_k00230	694	1
			MCYC 623	saba_2_521_ta00120	232	2
				saba_2_351_mm00100	694	1
		S. cerevisiae	S288c	sace_1_ygl166w	225	2
				sace_1_ypr008w	627	1
			M3836	sace_10_5_2940	225	2
				sace_10_3_1630	642	1
			M3837	sace_11_6_f03030	225	2
				sace_11_4_d02800	644	1
			M3838	sace_12_5_860	225	2
				sace_12_3_2680	646	1
			M3839	sace_13_2_4130	225	2
				sace_13_4_1670	657	1
			RM11-1a	sace_14_2_b04310	225	2
				sace_14_5_e02890	660	1
			273614N	sace_15_3_c00980	225	2
				sace_15_5_e02830	668	1
			322134S	sace_16_2_b01000	225	2
				sace_16_5_e02800	687	1
			378604X	sace_17_2_b01000	225	2
				sace_17_5_e02890	694	1
			BC187 (equivalent to sace_8)	sace_18_2_b00980	225	2
				sace_18_5_e02810	694	1
			DBVPG1106	sace_18_3_910	225	2
				sace_18_5_2700	694	1
			DBVPG1373	sace_19_3_900	225	2
				sace_19_5_2720	694	1
			AWRI796	sace_2_3_c00980	225	2
				sace_2_6_f02900	694	1
			DBVPG1788	sace_20_3_890	225	2
				sace_20_5_2690	694	1
			DBVPG1853	sace_21_2_920	225	2
				sace_21_5_2730	694	1
			DBVPG6040	sace_22_3_900	225	2
				sace_22_5_2750	694	1
			DBVPG6044	sace_23_2_920	225	2
				sace_23_5_2800	694	1
			DBVPG6765	sace_24_2_890	228	2
				sace_24_5_2750	694	1
			K11	sace_25_3_c01000	225	2
				sace_25_5_e02860	694	1
			L 1374	sace_26_3_900	225	2
				sace_26_5_2690	694	1

			L 1528	sace_27_3_c00990	225	2
				sace_27_5_e02800	694	1
			NCYC110	sace_28_2_b01020	225	2
				sace_28_5_e02880	694	1
			NCYC361	sace_29_2_900	225	2
				sace_29_5_2700	694	1
			cen.pk113-7d	sace_3_3_c00990	209	2
				sace_3_5_e02840	694	1
			SK1	sace_30_2_b01000	228	2
				sace_30_5_e02890	694	1
			UWOPS03 461 4	sace_31_3_890	228	2
				sace_31_5_2770	694	1
			UWOPS05 217 3 (same sequences)	sace_32_2_870	228	2
				sace_32_5_2750	694	1
				sace_32_2_870_2	227	2
				sace_32_5_2750_2	694	1
			UWOPS05 227 2	sace_33_2_860	225	2
				sace_33_5_2750	694	1
			UWOPS83 787 3	sace_34_3_880	225	2
				sace_34_5_2770	694	1
			UWOPS87 2421	sace_35_2_910	225	2
				sace_35_5_2720	694	1
			Y9	sace_36_2_b00970	225	2
				sace_36_5_e02870	694	1
			Y12	sace_37_2_b00980	225	2
				sace_37_5_e02870	694	1
			Y55	sace_38_2_b00990	225	2
				sace_38_5_e02850	694	1
			YIIc17 E5	sace_39_3_c01020	225	2
				sace_39_5_e02860	694	1
			FostersB	sace_4_2_b00970	225	2
				sace_4_5_e02800	356	1
			YJM975	sace_40_2_b00990	225	2
				sace_40_5_e02800	694	1
			YJM978	sace_41_3_c01000	225	2
				sace_41_5_e02810	694	1
			YJM981	sace_42_3_c01010	225	2
				sace_42_5_e02860	694	1
			YPS128	sace_43_2_b00990	225	2
				sace_43_5_e02890	694	1
			YPS606	sace_44_2_900	225	2
				sace_44_5_2790	694	1
			YS2	sace_45_2_b01000	225	2
				sace_45_5_e02900	694	1
			YS4	sace_46_2_b01000	209	2
				sace_46_5_e02850	694	1
			YS9	sace_47_2_b01020	225	2
				sace_47_5_e02860	694	1
			L 1528	sace_27_3_c00990	225	2
				sace_27_5_e02800	694	1
			NCYC110	sace_28_2_b01020	225	2
				sace_28_5_e02880	694	1
			NCYC361	sace_29_2_900	225	2
				sace_29_5_2700	694	1
			cen.pk113-7d	sace_3_3_c00990	209	2
				sace_3_5_e02840	694	1
			L 1528	sace_27_3_c00990	225	2
				sace_27_5_e02800	694	1
			NCYC110	sace_28_2_b01020	225	2
				sace_28_5_e02880	694	1

			NCYC361	sace_29_2_900	225	2
				sace_29_5_2700	694	1
			cen.pk113-7d	sace_3_3_c00990	209	2
				sace_3_5_e02840	694	1
			SK1	sace_30_2_b01000	228	2
				sace_30_5_e02890	694	1
			UWOPS03 461 4	sace_31_3_890	228	2
				sace_31_5_2770	694	1
			UWOPS05 217 3 (same sequences)	sace_32_2_870	228	2
				sace_32_5_2750	694	1
				sace_32_2_870_2	227	2
				sace_32_5_2750_2	694	1
			UWOPS05 227 2	sace_33_2_860	225	2
				sace_33_5_2750	694	1
			UWOPS83 787 3	sace_34_3_880	225	2
				sace_34_5_2770	694	1
			UWOPS87 2421	sace_35_2_910	225	2
				sace_35_5_2720	694	1
			Y9	sace_36_2_b00970	225	2
				sace_36_5_e02870	694	1
			Y12	sace_37_2_b00980	225	2
				sace_37_5_e02870	694	1
			Y55	sace_38_2_b00990	225	2
				sace_38_5_e02850	694	1
			YIIc17 E5	sace_39_3_c01020	225	2
				sace_39_5_e02860	694	1
			FostersB	sace_4_2_b00970	225	2
				sace_4_5_e02800	356	1
			YJM975	sace_40_2_b00990	225	2
				sace_40_5_e02800	694	1
			YJM978	sace_41_3_c01000	225	2
				sace_41_5_e02810	694	1
			YJM981	sace_42_3_c01010	225	2
				sace_42_5_e02860	694	1
			YPS128	sace_43_2_b00990	225	2
				sace_43_5_e02890	694	1
			YPS606	sace_44_2_900	225	2
				sace_44_5_2790	694	1
			YS2	sace_45_2_b01000	225	2
				sace_45_5_e02900	694	1
			YS4	sace_46_2_b01000	209	2
				sace_46_5_e02850	694	1
			YS9	sace_47_2_b01020	225	2
				sace_47_5_e02860	694	1
			Sigma1278b	sace_48_4_d00970	217	2
				sace_48_6_f02790	694	1
			vin13	sace_49_2_b01000	225	2
				sace_49_6_f02840	544	1
			FostersO	sace_5_29_ac00190	225	2
				sace_5_2_b02940	642	1
			VL3	sace_50_3_c00940	225	2
				sace_50_6_f02940	604	1
			w303	sace_51_3_c00950	225	2
				sace_51_5_e02840	694	1
			YJM789	sace_52_12_l00920	225	2
				sace_52_2_b02850	694	1
			boulardii EDRL	sace_53_22_v00160	225	2
				sace_53_30_ad00740	694	1
			BY4741	sace_54_3_c00970	225	2
				sace_54_5_e02820	694	1

			BY4742	sace_55_3_c00970	225	2
				sace_55_5_e02820	694	1
			EC1118	sace_56_3_c00980	225	2
				sace_56_4_d02800	694	1
			Kyokai7	sace_57_2_b00950	225	2
				sace_57_6_f02850	694	1
			T7	sace_58_28_ab00570	225	2
				sace_58_107_dc00100	694	1
			YB210	sace_59_11_k00200	225	2
				sace_59_212_hd00150	694	1
			JAY291	sace_6_21_u00240	225	2
				sace_6_128_dx00150	694	1
			ZTW1	sace_60_9_i00940	225	2
				sace_60_1_a02810	694	1
			Lalvin QA23	sace_8_29_ac00340	225	2
				sace_8_1_a02820	341	1
			M3707	sace_9_6_f03030	225	2
				sace_9_4_d01730	694	1
		S. mikatae	IFO 1815	sami_1_7.105	225	2
				sami_1_16.247	694	1
		S. paradoxus	Sanger reference	sapa_1_2_b00960	225	2
				sapa_1_5_e03020	694	1
			KPN3829	sapa_10_2_880	225	2
			N 17	sapa_11_2_b00950	225	2
			N 43	sapa_12_2_b00980	225	2
			N 44	sapa_13_2_b01000	225	2
			N 45	sapa_14_2_b00970	225	2
			Q31 4	sapa_15_2_b01030	225	2
			UFRJ50791	sapa_16_2_b01020	225	2
			UFRJ50816	sapa_17_2_910	225	2
			UWOPS91 917 1	sapa_18_2_900	225	2
			W7	sapa_19_2_b01030	225	2
			A4	sapa_2_2_b01010	225	2
			Y6 5	sapa_20_2_b00950	225	2
			Y7	sapa_21_2_b00960	225	2
			Y8 5	sapa_22_2_b01020	225	2
			YPS138	sapa_23_2_b01020	225	2
			Z1	sapa_24_2_b01010	225	2
			Z1 1	sapa_25_2_b00960	225	2
			NRRL Y-17217	sapa_26_326_ln00120	164	-
				sapa_26_87_ci00100	694	1
			A12	sapa_3_2_b01020	225	2
			CBS432	sapa_4_2_b00970	225	2
			CBS5829	sapa_5_2_880	225	2
			DBVPG4650	sapa_6_2_870	225	2
			DBVPG6304	sapa_7_2_890	225	2
			IFO1804	sapa_8_2_890	225	2
			KPN3828	sapa_9_2_890	232	2
		S. uvarum	CBS 7001	sauv_1_7.97	221	2
				sauv_1_16.328	694	1
		S. kudriavzevii	IFO 1802	saku_1_7.105	264	2
				saku_1_16.292	694	1
	Kazachstania	*K. africana*	CBS 2517	kaaf_1_c03260	263	2
				kaaf_1_b01950	525	1
		K. naganishii	CBS 8797	kana_1_e03480	291	2
				kana_1_d02340	611	1
	Naumovozyma	*N. castellii*	CBS 4309	naca_1_a05130	291	2
				naca_1_b05730	743	1
			NRRL Y-12630	naca_2_51_ay00250	294	2
				naca_2_9_i00230	743	1

		N. dairenensis	CBS 421	nada_1_k02530	265	1
				nada_1_b03070	893	2
	Candida	*C. glabrata*	CBS138	cagl_1_i04180g	265	1
				cagl_1_l09339g	877	2
			CCTCC M202019	cagl_2_4_d01770	265	1
				cagl_2_2_b04230	877	2
	Tetrapisispora	*T. blattae*	CBS 6284	tebl_1_c00170	604	1
		T. phaffii	CBS 4417	teph_1_d00690	644	1
				teph_1_h00580	356	2
'*Saccharomyces* complex' (Pre-WGD)	*Vanderwaltozyma*	*V. polyspora*	DSM 70294	vapo_1_467.5	341	2
				vapo_1_541.30	586	1
	Zygosaccharomyces	*Z. bailii*	ISA1307_05761	zyba_1_01494	698	3
			ISA1307_01494	zyba_1_05761	664	-
			UTAD63	zyba_2_9_i00670	694	3
			CLIB 213	zyba_3_9_i00670	694	3
		Z. rouxii	CBS 732	zyro_1_f04862g	718	3
	Torulaspora	*T. delbrueckii*	CBS 1146	tode_1_c01350	627	3
	Lachancea	*L. waltii*	NCYC 2644	lawa_1_14.1543	599	3
		L. thermotolerans	CBS 6340	lath_1_g16984g	614	3
		L. kluyvery	CBS 3082	lakl_1_b03344g	618	3
	Kluyveromyces	*K. wickerhamii*	UCD 54-210	klwi_1_120_dp00120	557	3
		K. marxianus var. marxianus	KCTC 17555	klma_1_3_c06150	601	3
		Kl. lactis	CLIB210	klla_1_a03047g	567	3
		K. aestuarii	ATCC 18862	klae_1_13_m00540	561	3
	Eremothecium	*E. cymbalariae*	DBVPG 7215	ercy_1_6116	668	3
		E. gossypii	ATCC 10895	ergo_1_ael295c	544	3
		A. acer	(not available)	asac_1_7_g04120*	540	3
CTG complex	*Candida*	*C. albicans*	SC5314	caal_1_19.5001	591	4
			A155	caal_10_8_h01390	591	4
			CHN1	caal_12_19_s00390	598	4
			WO-1	caal_2_00088	593	4
			3153A	caal_3_22_v00370	591	4
			A20	caal_4_3_c05000	591	4
			A48	caal_5_2_b02700	591	4
			A67	caal_6_3_c02930	591	4
			A84	caal_7_3_c02930	591	4
			A92	caal_8_20_t00350	591	4
			A123	caal_9_20_t00400	591	4
		C. dubliniensis	CD36	cadu_1_12620	570	4
		C. maltosa	Xu316	cama_1_31_ae00160	537	4
		C. orthopsilosis	Co 90-125	caor_1_d01560	566	4
		C. parapsilosis	CDC317	capa_1_201510	596	4
		C. tenuis	ATCC 10573	cate_1_4_d03480	676	4
		C. tropicalis	MYA-3404	catr_1_03632	379	4
	Debaryomyces	D. hansenii	CBS76	deha_1_d05500g	567	4
			MTCC 234	deha_2_55_bc00230	409	4
	Lodderomyces	L. elongisporus	NRLL YB-4239	loel_1_03555	408	4
	Meyerozyma	M. guilliermondii	ATCC 6260	megu_1_05595	439	4
	Pichia	P. sorbitophila (H)	CBS 7064	piso_1_a06138g	388	4
			CBS 7064	piso_1_b06205g	388	4
	Scheffersomyces	S. stipitis	CBS 6054	scst_1_1_a11590	645	4
	Spathaspora	S. arborariae	UFMG-19.1A	spar_1_2_b01010	611	4
		S. passalidarum	NRRL Y-27907	sppa_1_4_d06430	603	4
Early-div.	*Ascoidea*	A. rubescens	NRRL Y17699	asru_1_15_o01530*	128	-
	Babjeviella	B. inositovora	NRRLY-12698	bain_1_7_g01580	699	-
	Candida	C. arabinofermentans	NRRL YB-2248	caar_1_7_g01950	888	4
		C. caseinolytica	Y-17796	caca_1_2_b13940	280	4
		C. tanzawaensis	NRRL Y-17324	cata_1_6_f01420	676	4
	Clavispora	C. lusitaniae	ATCC 42720	cllu_1_03246	547	4

			MTCC_1001	cllu_2_13_m00670	546	4
	Cyberlindnera	*C. jadinii*	NBRC 0988	cyja_1_5_e03820	514	4
			NRRL Y-1542	cyja_2_14_600	520	4
	Dekkera	*D. bruxellensis*	CBS 2499	debr_2_16_p00330	310	1
			AWRI1499	debr_1_131_ea00150*	46	1
	Hansenula	*H. polymorpha*	NCYC 495	hapo_1_4_d06190	571	4
	Hanseniaspora	*H. valbyensis*	NRRL Y-1626	hava_1_9_i00540	707	1
	Hyphopichia	*H. burtonii*	NRRL Y-1933	hybu_1_15_o00240	379	4
	Komagataella	*K. pastoris*	CBS 7435	kopa_1_3_c09790	523	4
			DSMZ 70382	kopa_2_20_t00330	518	4
			GS115	kopa_3_3_c02520	523	4
	Lipomyces	*L. starkeyi*	(not available)	list_1_2_b03960	386	-
	Metschnikowia	*M. bicuspidata*	NRRL YB-4993	mebi_1_6_f01720	600	4
	Nadsonia	*N. fulvescens var. elongata* (^{H})	DSM 6958	nafu_1_10_j01430	1248	4
			DSM 6958	nafu_1_2_b04520	285	4
	Ogataea	*O. parapolymorpha*	DL-1	ogpa_1_4_d01650	571	4
	Pachysolen	*P. tannophilus*	NRRL Y-2460	pata_1_2_b04560	702	4
	Pichia	*P. membranifaciens*	NRRL Y-2026	pime_1_3_c02120	863	4
		P. kudriavzevii	M12	piku_1_5_e00340*	63	4
	Wickerhamomyces	*W. anomalus*	NRRL Y-366	wian_1_3_c07950	540	4
	Yarrowia	*Y. lipolytica* CLIB122	CLIB122	yali_1_b08206g	411	4

Protein sequences marked with an asterisk (*) represent fragments. The (H) represents hybrid species.

Annex IV – Phylogenetic tree to the initial portion of the N-terminal domain of the *S. cerevisiae* Haa1 and Ace1 homologous from Hemiascomycetes

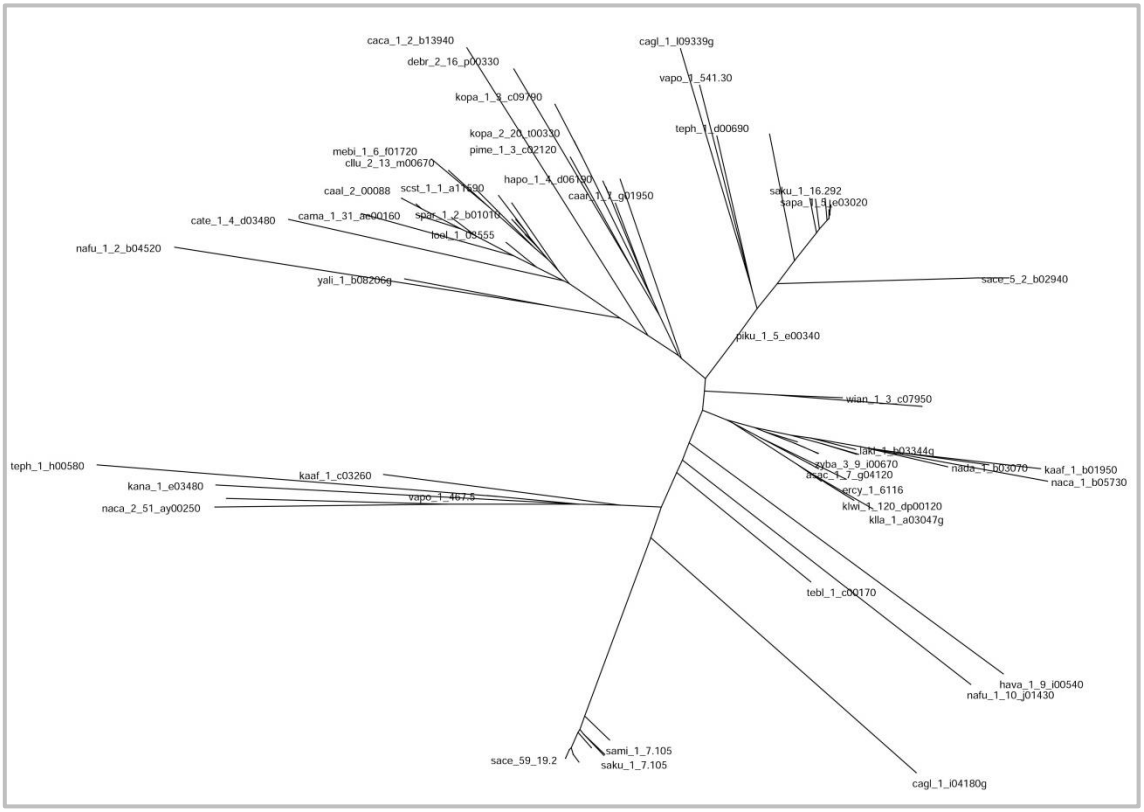

Figure 25. Phylogenetic analysis of the initial portion of N-terminal sequence of the *S. cerevisiae* Haa1/Ace1 homologues in Hemiascomycetes. Radial phylogram shows the amino acid divergence among the Haa1/Ace1 homologous proteins considering the first 103 amino acid residues of the Haa1 from *S. cerevisiae* S288c as the cut-off mark in the multiple alignment.

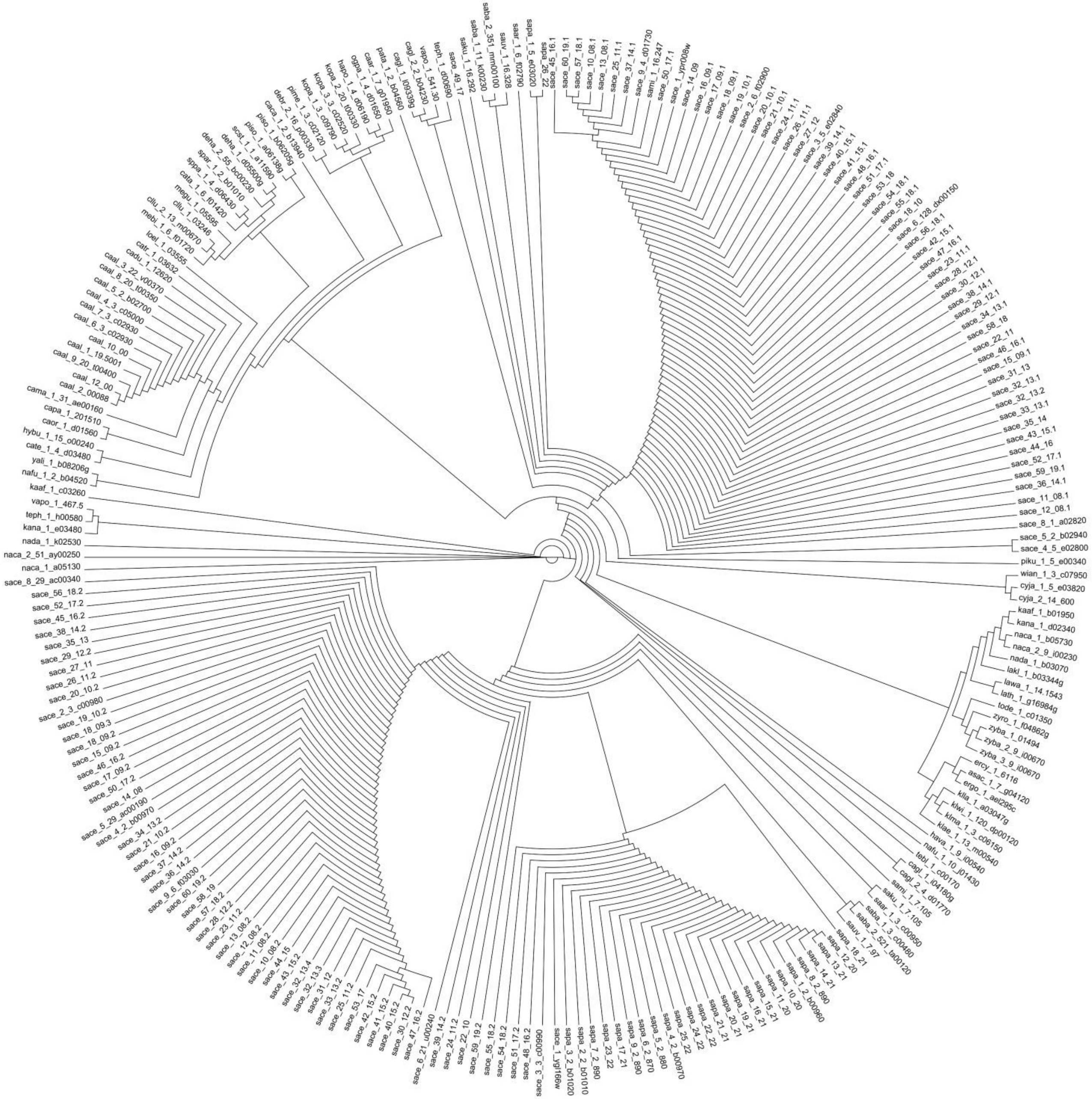

Figure 26. Phylogenetic analysis of the initial portion of N-terminal sequence of the *S. cerevisiae* Haa1/Ace1 homologues in Hemiascomycetes. Circular cladogram given by the tree topology. The phylogenetic tree was generated based on PROTDIST/NEIGHBOUR program of PHYLIP package and retrieved results explored in Dendroscope software. The naca_1_a05130 sequence from *N.castellii* represents the tree outgroup.