UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Exploiting natural and artificial diversity of plasma membrane efflux pumps towards improved cell factories for biorefineries

Madalena Isabel Nunes de Matos

Mestrado em Microbiologia Aplicada

Dissertação orientada por:
Doutora Cláudia Sofia Pires Godinho
Professora Doutora Mónica Sofia Vieira Cunha

2022

This Dissertation was fully performed at Sá-Correia's Lab, Biological Sciences Group, Institute of Bioengineering and Biosciences, Instituto Superior Técnico, University of Lisbon under the direct supervision of Post-doctoral Researcher Cláudia Godinho

Professor Mónica Cunha was the internal supervisor designated in the scope of the Master in Applied Microbiology of the Faculty of Sciences of the University of Lisbon

Acknowledgments

Firstly, I would like to thank my supervisor Doctor Cláudia Godinho for the opportunity to conduct my Master thesis under her supervision at Sá-Correia's Lab, Biological Sciences Group, Institute of Bioengineering and Biosciences, Instituto Superior Técnico, University of Lisbon. I would also like to thank the kind support, motivation, and exceptional guidance she provided me, always with good humor.

I thank Professor Doctor Isabel Sá-Correia for alloying me to perform my work at her lab, as well as for all the discussions and advice regarding the obtained results, experimental delineation and the development of the written thesis.

I thank my Internal Supervisors Professor Doctor Rogério Tenreiro and Professor Doctor Mónica Cunha for the availability to help and their contribution in the revision of this document.

I would like the acknowledge the funding that allowed the experimental work here presented. This MSc thesis research work was funded by Move2LowC project (POCI-01-0247-FEDER-046117), cofinanced by Programa Operacional Competitividade e Internacionalização (POCI), Programa Operacional Regional de Lisboa, Portugal 2020 and the European Union, through the European Regional Development Fund (ERDF), and by Fundação para a Ciência e a Tecnologia' (FCT) through ERA-IB-2/0003/2015 project. Funding received from FCT by iBB—Institute for Bioengineering and Biosciences (UIDB/04565/2020 and UIDP/04565/2020) and the Associate Laboratory Institute for Health and Bioeconomy i4HB (LA/P/0140/2020), is also acknowledged. This project work is within the framework of the FCT project 2022.01501.PTDC "Mechanistic insights into adaptation and increased robustness to acetic acid and other weak acids toxicity in yeasts", recently recommended for funding.

To all the members of the Sá-Correia's Lab, I give my profound appreciation. Thank you, Margarida for all the guidance and help. Thank you, Marta for the laughs and talks at odd hours. Thank you, Miguel for the patience in teaching me to navigate the world of bioinformatics. Thank you, Paula for the close support and for all the discussions about language diversity and nationalities. And finally, thank you Ricardo, for all the lunch discussions and all the unending book/movie/series talks.

I thank the poor yeasts that suffered all the experiments. Without you there would be nothing.

I thank my parents and close friends, you know who you are, all the support now and in all the years before.

Finally, thank you Guilherme for your constant love, support and motivation. You helped me so much. I hope you know how much that means to me.

Abstract

The industrialization, growth of global population and emergence of consumerism-based lifestyles led to an unsustainable global energy demand. Consequently, there is the need for environmentally-friendly solutions for the production of energy and chemicals, such as biorefineries. One example is the use of yeasts to catalyze the production of ethanol from lignocellulosic feedstocks, obtained from non-food crops and wastes from agroforestry activities. However, this process requires the hydrolysis of the lignocellulosic material for the release of fermentable sugars, which results in the formation of compounds that inhibit yeast metabolism and limit the process yield. Inhibitor tolerance is thus a desirable trait in cell factories, which has motivated extensive research on the mechanisms of yeast stress tolerance and the search for superior gene variants for the rational engineering of robustness.

In the present work, a set of Multidrug/Multixenobiotic Resistance (MDR/MXR) transporters was screened for a role in *S. cerevisiae* tolerance to stress induced by acetic and formic acids, furfural and 5-hydroxymethylfurfural (HMF), commonly found in lignocellulosic hydrolysates. The Major Facilitator Superfamily (MFS) transporters Tpo2 and Tpo3, and the ABC transporter Pdr18 were identified as reducing the latency phase during cultivation of unadapted yeast populations in the presence of weak acid and/or furfural stress. Artificial *TPO3* gene variants were generated by error-prone PCR and tested for improved tolerance to acetic acid. Of the 294 clones tested, 2 were found to exhibit a slightly higher maximum specific growth rate in the presence of acetic acid, when compared to the corresponding parental strain, and were sequenced for identification of putative causal mutations. Finally, the natural variability of the *TPO3* gene sequence was investigated among 1011 *S. cerevisiae* strains from different environments. Several strains harboring different *TPO3* natural variants were preliminarily screened for their tolerance to acetic acid.

Keywords: *Saccharomyces cerevisiae*, Lignocellulosic-derived inhibitors, MDR/MXR transporters, Weak acids, Gene variants

Resumo

A revolução industrial resultou num grande desenvolvimento económico, e num aumento significativo da qualidade e esperança média de vida. No entanto, levou também a um aumento exponencial da população mundial, ao surgimento de estilos de vida baseados no consumismo, à utilização intensiva de combustíveis fósseis e, consequentemente, a requisitos energéticos insustentáveis à escala global, com drásticas consequências para o meio ambiente. Assim, tornou-se necessário desenvolver formas de obtenção de energia mais sustentáveis, como é o caso da energia produzida em biorefinarias.

De forma geral, o conceito de biorefinaria refere-se a centrais onde se procede à decomposição sustentável de matéria orgânica variada, visando a produção de biocombustíveis e de outros compostos químicos de interesse económico, como por exemplo as resinas. Desta forma, as biorefinarias são uma aposta para o combate à utilização de energias fósseis de forma a reduzir a libertação de gases com efeito estufa e a impedir o agravamento da preocupante situação ambiental em que o planeta se encontra. De entre os tipos de matérias-primas utilizadas em biorefinarias, o material lenho-celulósico é dos que potenciam um melhor aproveitamento dos recursos, englobando-se nas políticas de economia circular. Estas matérias orgânicas são ricas em polímeros de açúcares (hemicelulose e celulose) e em lenhina e, idealmente, são obtidas de culturas de plantas que não competem com a alimentação humana ou animal, proveniente de atividades de limpeza florestal e do aproveitamento de resíduos agrícolas. Para que microrganismos como a levedura *Saccharomyces cerevisiae* possam catalisar a produção de bioetanol através da fermentação de açúcares presentes em biomassa lenho-celulósica, é fundamental implementar processos de degradação da matriz orgânica de forma a libertar os monossacáridos. Não obstante, as várias etapas do tratamento conduzem à produção de variados compostos com ação inibitória no metabolismo das leveduras, nomeadamente ácido acético, ácido fórmico, furfural e 5-hidroximetilfurfural (HMF), que causam inibição de crescimento e consequentemente perda de eficiência na produção de etanol e outros compostos. Posto isto, torna-se crucial perceber o processo inibitório dos compostos formados e especialmente investigar e melhorar os mecanismos implementados pelos microrganismos que lhes conferem uma maior resistência a inibidores presentes em hidrolisados lenho-celulósicos e, consequentemente, uma maior robustez enquanto fábrica celular microbiana.

Ao longo de milénios de evolução, os seres vivos e especialmente os seres unicelulares desenvolveram mecanismos que lhes permitem sobreviver, evitar e ultrapassar os danos adventícios da exposição a agentes tóxicos. O presente trabalho focou-se no estudo de transportadores de membrana plasmática envolvidos na resistência de *S. cerevisiae* a múltiplas drogas/xenobióticos (MDR/MXR), dado o extensivo estudo de que esta levedura tem sido alvo, a sua grande aplicabilidade em inúmeras áreas biotecnológicas, e tendo em conta a grande quantidade de fenótipos de resistência que têm vindo a ser associados com transportadores MDR/MXR.

Neste âmbito, este projeto teve como objetivos a identificação de transportadores MDR/MXR que têm um papel na resistência de *S. cerevisiae* a inibidores gerados no tratamento de matéria-prima lenho-celulósica e de variantes genéticas desses transportadores, geradas artificialmente ou existentes em estirpes isoladas de vários ambientes da natureza, que resultem numa maior resistência aos referidos inibidores. O objetivo final seria a aplicação dessas variantes genéticas em estipes industriais de *S. cerevisiae*, aumentando a sua resistência e, simultaneamente, a sua produtividade em processos de biorefinarias.

Desta forma, foi avaliado o crescimento de sete mutantes de eliminação para genes que codificam para transportadores MDR/MXR na presença de ácido acético, ácido fórmico, furfural ou HMF, por comparação com o comportamento da estirpe parental (BY4741). Os resultados obtidos neste estudo leva-

ram a que os transportadores MDR/MXR Pdr18, Tpo3 e Tpo2 fossem selecionados para dar seguimento ao projeto, pela sua importância na resistência a ácido acético, ácido fórmico e/ou a furfural. A caracterização das curvas de crescimento dos vários mutantes de eliminação, quando expostos aos diversos inibidores, permitiu verificar que nenhum destes transportadores é importante na tolerância de *S. cerevisiae* ao stress induzido pela presença de HMF, enquanto que em presença de furfural os mutantes *tpo2*Δ e *tpo3*Δ apresentavam um maior período de adaptação. Na presença de ácido acético, a eliminação dos transportadores Pdr18, Tpo2 e Tpo3 mostrou ser prejudicial e, em presença de ácido fórmico, apenas o mutante *pdr18*Δ mostrou ser mais suscetível do que a estirpe parental. Complementarmente, foi testado o papel dos transportadores Tpo2, Tpo3 e Pdr18 no crescimento de células adaptadas e/ou não adaptadas em presença de suplementação do meio com uma concentração sub-letal de ácido acético. Observou-se que o papel destes transportadores reside maioritariamente na redução do tempo de adaptação em células não previamente adaptadas.

Após seleção dos transportadores de interesse, foi feita uma amplificação dos genes *PDR18*, *TPO2* e *TPO3* usando uma polimerase propensa a erro, de forma a gerar formas alternativas destes genes. Os fragmentos mutados foram clonados num plasmídeo e transformados em células do respetivo mutante de eliminação. As estipes obtidas, que expressavam variantes artificiais de *TPO3*, foram cultivadas na presença de ácido acético e procurou-se identificar aquelas em que, não só houvesse uma recuperação do fenótipo de suscetibilidade exibido pelos mutantes de eliminação *tpo3*Δ, mas também um aumento de tolerância em relação ao mutante de eliminação complementado com o gene *TPO3* de BY4741. Das 294 estirpes obtidas, 2 foram identificadas como apresentando uma ligeira melhoria de tolerância à concentração de ácido acético testada, quando comparado com a respetiva estirpe parental. O gene *TPO3* codificado no DNA plasmídico destas estirpes foi sequenciado, culminando na identificação e análise de impacto na expressão e/ou estrutura da proteína, de oito mutações causais putativas, localizadas tanto na região codificante do gene, como nas respetivas regiões regulatórias a jusante e a montante.

Adicionalmente, foi feita uma análise bioinformática da variabilidade natural da região codificante do gene *TPO3* de 1011 estirpes de *S. cerevisiae*. O estudo bioinformático incluiu a construção de um dendrograma representativo das semelhanças da região codificante do gene *TPO3* nas várias estirpes e a compilação e caracterização de uma lista de SNPs (*Single Nucleotide Polymorphisms*) encontrados. Os impactos dos vários SNPs identificados na proteína Tpo3 foram estimados, tendo por base o uso de plataformas online de análise de proteínas. Neste âmbito foi também testado o crescimento de 12 das 1011 estirpes na presença de ácido acético. Este trabalho resultou na identificação de vários SNPs cujo impacto na resistência virá a ser explorado experimentalmente em estudos futuros. Da mesma forma, este trabalho permitiu a identificação de estirpes de *S. cerevisiae* com elevada tolerância a ácido acético que podem conter ferramentas genéticas a serem exploradas no futuro para guiar a construção de estirpes industriais mais robustas.

Em suma, os resultados obtidos no decorrer deste projeto e aqui apresentados contribuíram para o objetivo geral de identificação de variantes genéticas de transportadores MDR/MXR que resultem numa maior tolerância de *S. cerevisiae* face aos mais comuns inibidores resultantes do pré-tratamento de matéria orgânica lenho-celulósica. Nesse âmbito, este projeto resultou na identificação de oito mutações no gene *TPO3* que podem ser determinantes para a tolerância a ácido acético. De forma semelhante, a procura de variantes naturais do gene *TPO3* em estirpes de *S. cerevisiae* isoladas de uma gama diversificada de ambientes e a avaliação do crescimento de algumas dessas estirpes na presença de ácido acético resultou na identificação de algumas variantes genéticas e algumas estirpes que poderão constituir ferramentas genéticas para melhoramento genético de estirpes industriais de *S. cerevisiae*.

Palavras-chave: *Saccharomyces cerevisiae*, Inibidores derivados de lenho-celulose, Transportadores MDR/MXR, Ácidos fracos, Variantes genéticas

Index

Figure Index

Table Index

Abbreviations

ABC	ATP-Binding Cassette
Adh	Alcohol Dehydrogenase
ATP	Adenosine Triphosphate
bp	base pairs
CDS	Coding Sequence
CIAP	Calf Intestinal Alkaline Phosphatase
ddH_2O	Distilled and deionized water
DNA	Desoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic acid
GTP	Guanosine Triphosphate
HMF	5-hydroxymethylfurfural
LB	Luria Broth
LiAc	Lithium Acetate
MDR/MXR	Multidrug/Multixenobiotic Resistance
MFS	Major Facilitator Superfamily
MM4	Minimal Medium
MM4-ura	MM4 medium to which uracil was not added
NAD(P)	Nicotinamide Adenine Dinucleotide (Phosphate)
NBD	Nucleotide Binding Domains
OD600nm	Optic Density at 600 nm
PCR	Polymerase Chain Reaction
PDR	Pleiotropic Drug Resistance
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
SNP	Single Nucleotide Polymorphism
TAE	Tris-Acetate-EDTA
TCM	Tris-HCl Calcium Magnesium
TF	Transcription Factor
TMD	Transmembrane Domains

UTR	Untranslated Region
WGD	Whole Genome Duplication
YPD	Yeast Peptone Dextrose
YPD40	Yeast Peptone Dextrose with 40 g/l glucose

1. Introduction

1.1. *Saccharomyces cerevisiae* – an overview

Saccharomyces cerevisiae, commonly known as brewer's yeast or baker's yeast, is a unicellular fungus of the Phylum Ascomycota with great importance, both as a key experimental organism and as a powerful biotechnological tool in food production and other industrial processes. *S. cerevisiae* has been part of Human history for more than 10 thousand years; however, its study only began in the nineteenth century as a result of several scientific and technological advances [1, 2]. Since then, *S. cerevisiae* has been the focus of research in different scientific areas and in 1996 became the first eucaryotic organism with a completely sequenced genome [3].

Analysis of *S. cerevisiae*'s genome showed that it has a large number of duplicated gene blocks, which suggests that *S. cerevisiae*'s phylogenetic branch originated after a Whole Genome Duplication (WGD) event, thought to have occurred 100 million years ago [4, 5]. The WGD resulted in the duplication of the chromosome number from 8 to 16 [4]. Figure 1.1. depicts the yeast phylogeny with indication of the WGD event [6]. Natural selection then acted upon the duplicated genes: some remained as extra copies, others disappeared, and others suffered mutations that resulted in two different proteins with close resemblance (paralogs) [5].

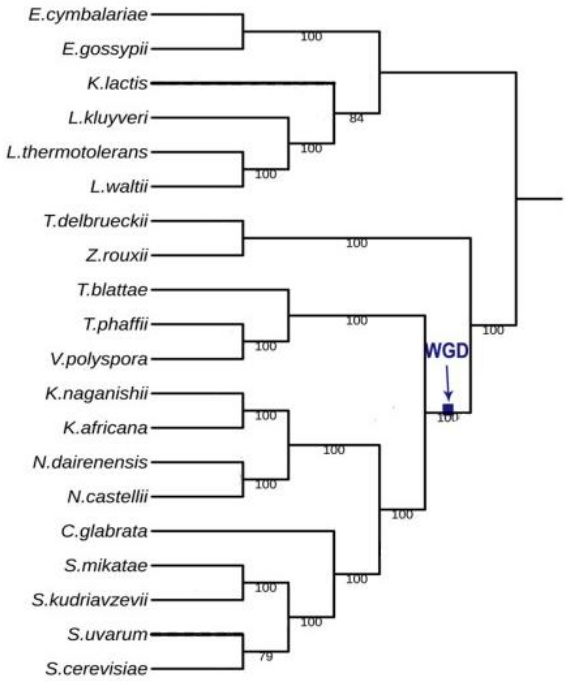

Figure 1.1. Yeast simplified phylogenetic tree
Phylogenetic tree depicting the distances of 20 yeast species. In blue is indicated the WGD event. Tree adapted from Feng *et al.* 2017.

S. cerevisiae can be found in haploid and diploid states. Both states can undergo mitotic reproduction (in a process called budding) while only the diploid state is able to reproduce meiotically. Haploid cells are divided into two different mating types, a and α, that are determined by specific loci in chromosome III. In wild strains, the mating types can be switched by homothallic switching endonuclease, according to environmental cues. In most laboratory strains, all cells have the same mating type, and the homothallic switching endonuclease is usually inactive (or not present) which leads to an inability to undergo sexual reproduction. In other laboratory strains, cells can switch mating types and it is pos-

sible to control the life cycle allowing its study. In contrast to laboratory strains, little is known regarding the factors that influence the life cycle of *S. cerevisiae* in natural habitats [2, 7].

Processes such as reproduction require enormous amounts of energy in the form of Adenosine Triphosphate (ATP). Yeasts, like most eukaryotes, can obtain ATP through two metabolic pathways, fermentation and aerobic respiration. In a fermentative process, glucose enters the glycolysis pathway whose final product is two pyruvate molecules. The pyruvate then undergoes a fermentative pathway, usually alcoholic, which culminates in the production of ethanol with a total yield (glycolysis plus fermentation) of two ATP molecules formed with phosphorylation at substrate level and without the need for oxygen. On the other hand, in aerobic respiration after the formation of pyruvate in glycolysis, this compound is transformed into acetyl-Coenzyme A by pyruvate dehydrogenase. Acetyl-Coenzyme A then enters the Krebs' cycle, where is oxidized to carbon dioxide (CO_2) with the formation of guanosine triphosphate (GTP) and nicotinamide adenine dinucleotide (NADH) that are oxidized in the mitochondrial membranes in presence of oxygen with a final total yield, in *S. cerevisiae,* of about 18 ATP molecules for each molecule of glucose consumed [4].

The transition between respiration and fermentation is subjected to rigorous controls that change according to environmental conditions and the age of the cell [8, 9]. In absence of oxygen, yeasts typically produce energy through fermentation. In presence of oxygen, the preferred pathway should be respiration since has a greater yield. However, if glucose concentration is high, the respiration in *S. cerevisiae* is downregulated and fermentation becomes the primary pathway for ATP production [10]. The preference for a fermentative pathway in presence of high glucose concentrations and oxygen is known as Crabtree effect [10]. *S. cerevisiae* displays the Crabtree effect and as such is called Crabtree-positive [4].

From an evolutionary standpoint, the appearance of yeast with a great predisposition to aerobic fermentation arose after the WGD at the same moment that plants with flowers and fruit were diversifying [4]. The exact mechanisms responsible for the Crabtree effect are still being debated, with two main theories proposed.

One theory, known as "make-accumulate-consume" strategy, is closely related to WGD and states that there was a duplication of the ancestral alcohol dehydrogenase gene (*ADH*) - responsible for the interconversion of ethanol and acetaldehyde - in *ADH1* and *ADH2*. The encoded protein, Adh1, is thought to be the most similar to the ancestral Adh and favors the reaction towards the production of ethanol leading to its excretion. Meanwhile, Adh2 tends to favor the inverse reaction, using ethanol to ultimately produce acetyl-coenzyme A and more ATP through aerobic respiration. This duology would, in theory, confer an advantage to the cell because it could accumulate ethanol, to which is resistant up to a certain concentration, while following a quicker route to energy production and then use the ethanol to obtain more energy and at the same time the toxicity of ethanol would deplete the environment of competitors [4, 11]. The other theory, called "rate/yield trade-off hypothesis", declares that the Crabtree effect is advantageous, not because of an advantage based on ethanol toxicity, but because it is more important for cells to obtain energy at a fast rate regardless of the final yield [4, 12].

In the wild, *S. cerevisiae* is considered a ubiquitous fungus, having been found in a wide range of habitats, mostly habitats with poor nutritional conditions where this yeast might be dormant for most of its life [13, 14]. From an evolutionary standpoint, studies suggest that primeval forests of the Far East Asia, are likely to have been the homeland of *Saccharomyces* yeasts [15]. *S. cerevisiae* is also found in habitats where it withstands pressure from several environmental stresses, such as acidic environments or high osmotic pressure cause by high sugar concentrations (e.g. fruits), leading to the

notion that these cells are not adapted to any habitat in particular but have the capacity to thrive in a broad range of environments [13, 16].

Despite its abundance in the wild, *S. cerevisiae* is also present in man-made environments and is a key microbe in several industrial applications (Figure 1.2.).

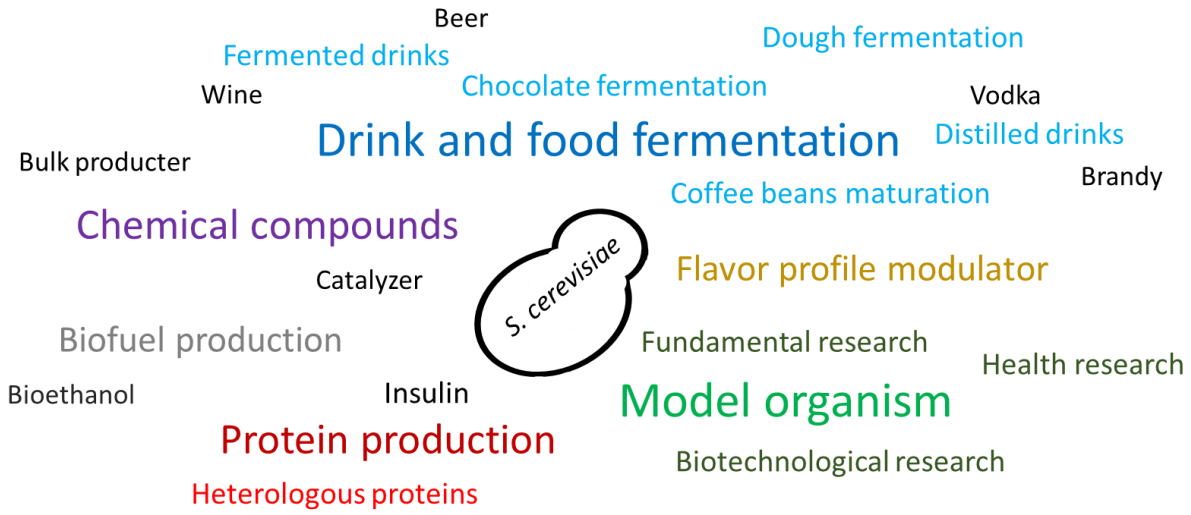

Figure 1.2. Examples of *S. cerevisiae*'s applications as described in the text

S. cerevisiae, as a single-celled eucaryote organism, is used in research as a model organism allowing the understanding of the physiological processes of more complex eukaryotic cells such as plant and animal cells. The selection of this yeast species as a model organism was potentiated by the integral sequencing of its genome, its relatively easy manipulation and the simplicity with which these cells can be cultured, especially compared to animal cells. The referred characteristics also make *S. cerevisiae* the perfect organism for bulk production of chemicals (pharmacological or not), and of proteins, either its own or heterologous proteins. For example, insulin is currently produced by *S. cerevisiae*, presenting fewer side effects and increased pharmacological efficacy, compared to the insulin harvested from pigs [1, 17].

S. cerevisiae is of great significance to the food and beverage industry. The ability of these yeast cells to resist high sugar concentrations and aerobically ferment those sugars to produce ethanol makes it a valuable organism in the production of fermented drinks, namely wine and beer, distilled drinks such as vodka and brandy, and beverages produced by fermentation of fruits, honey, among others [1, 16, 17].

The anaerobic fermentation of *S. cerevisiae* is exploited in the bread and chocolate industries. In the bread industry, the fermentation of starch present in the dough leads to the production of CO_2, which causes dough expansion, and of gluten that is responsible for the plasticity and gas retention capacity of the dough. [16]. In the chocolate industry, *S. cerevisiae* in conjugation with other yeasts and lactic and acetic acid bacteria, ferments the acidic and sugar-rich cocoa pulp present inside the cocoa rods. The fermentation and high pectinolytic activity of this yeast allow the conversion of the bitter flavor into the desired chocolate flavor. A similar process is involved in the maturation of coffee beans [16, 17].

Besides being involved in the production of several foods and drinks, *S. cerevisiae* is also implicated in their flavor profile due to the production of secondary metabolites [16]. For example, the addition of

specific strains of *S. cerevisiae* to unfermented cocoa pulp modifies the flavor of the final product, chocolate [18, 19]. Additionally, it is also possible to modulate the production of secondary metabolites in order to achieve the desired organoleptic characteristics. For instance, genetic engineering of *S. cerevisiae* led to the production of the monoterpenes, linalool and geraniol, that confer the hoppy flavor, much appreciated in beer [20].

Although essential for the production of many food products, *S. cerevisiae* may also be involved in its spoilage. The most affected foods are rich in sugar and/or alcohol such as processed fruits (juices and purees), and alcoholic beverages which can become contaminated with spoilage yeasts, altering their flavors [1].

Finally, as reviewed in the next sections, *S. cerevisiae* is of great interest for the production of bioethanol and other biofuels.

1.2. Biofuel production – process and challenges

1.2.1. Types of biofuels

In recent decades, the industrialization of many processes, the rapid growth of human population and the emergence of lifestyles based on consumerism, have led to higher requirements for energy, food and fuel [16, 17, 21]. Food and energy are obtained, direct or indirectly, from fuel. The most used fuel is non-renewable in its source, generates high levels of pollution, and is implicated in the worrying problem of global warming [17, 21]. Therefore, it becomes crucial to encounter other fuel and energy sources. Since 1826, it is possible to fuel a combustion engine using ethanol mixed with turpentine, an oil obtained from the resin of pine trees, and the technology has been evolving ever since [16].

Biofuels are defined as materials of plant nature that are used to obtain energy [22]. In the context of this work, the focus will be on fuels obtained through biological processes, mainly fermentation of organic feedstocks by microorganisms, that can substitute petroleum fuel. Bioethanol, biodiesel, propanol, butanol, and isobutanol are all considered biofuels as long as obtained by processes involving microorganisms [16, 17, 22]. Biodiesel fuels are composed of monalkyl esters of fatty acids obtained primarily from vegetable oils and it has been investigated the possibility of using microalgae or lipid-producing yeast to obtain such compounds [17].

Compared to bioethanol, butanol and propanol are less corrosive, less hygroscopic, and have higher octane values and energy density. However, contrary to ethanol, butanol, propanol, and isobutanol are not natural by-products of *S. cerevisiae*'s metabolism and their production requires the use of genetically modified *S. cerevisiae* strains [16].

Bioethanol can, nowadays, be used in engines with or without the addition of petroleum fuels. This biofuel, compared to petrol-based fuels, has higher flame speeds, broader flammability, is biodegradable, is less toxic and its use releases fewer pollutants into the atmosphere [16]. Bioethanol production results from the fermentation of naturally occurring sugars and is predominantly achieved using *S. cerevisiae*. This species produces ethanol in high amounts, is resistant to the toxicity of ethanol, up to a certain concentration, as well as to other stresses associated with the process and is easily manipulated, allowing large-scale production of ethanol. Other species capable of producing bioethanol are *Kluyveromyces marxianus, Dekkera bruxellensis* and *Scheffersomyces stipites*, although with some limitations when compared to *S. cerevisiae* [16, 23, 24].

The biomass used for bioethanol production can have different sources. Bioethanol can be produced by fermentation of plants rich in sucrose, such as beet, fruits, and other food crops directly harvested from the fields. This so-called first-generation bioethanol is considered a renewable energy source,

although it competes with land use for food [16, 23]. Alternatively, bioethanol may be obtained from lignocellulosic biomass, namely, food wastes, forest residues, non-usable parts of food crops, and from marine biomass such as micro and macro algae [16, 23]. Although lignocellulose utilization might represent future energy sources that are economically competitive, at the present there are some inherent challenges - the presence of fermentation inhibitors in the pre-treated lignocellulosic biomass - that require further research and optimization [16, 22, 23].

1.2.2. Biorefineries – a general concept

Similarly to petrol refineries, where crude is processed into fuels, a biorefinery is where the sustainable breakdown of biomass (feedstock) with the production of bioethanol and other products of economic value (for example, fertilizers, resins, organic acids and animal food) takes place [21, 25]. Depending on the feedstock used and processing techniques implemented in a biorefinery, there will be different end-products.

In a general way, any plant/algae material can be used as feedstock, and the technological processes used can be thermochemical, biochemical, mechanical, and chemical. Anaerobic digestion and fermentation are the most common biochemical processes. Fermentation uses fermentative organisms, namely *S. cerevisiae*, to produce diverse alcohols and organic acids, with ethanol being the most desired end-product. Anaerobic digestion involves the anaerobic breakdown of biomass by bacteria resulting in gas mixture [21]. Microorganisms used for biochemical processes are also known as cell factories.

1.2.3. Biofuel production from lignocellulosic feedstocks

Lignocellulosic feedstocks are obtained from non-food crops consisting of hard fibers. Such biomass is obtained as waste from lumber, agricultural and municipal wastes, and forest-cleaning activities, among others. The main components of lignocellulosic material - hemicellulose, cellulose, and lignin - can be found in different concentrations depending on the original biomass and are tightly bounded by covalent and non-covalent bonds [21, 24, 26]. A simplified scheme of lignocellulose and its main components is depicted in Figure 1.3.[27].

Cellulose represents 40 to 50% of the total lignocellulosic biomass and is a linear polymer of β-glucopyranose linked by β-(1,4) glycosidic bonds [23, 24]. In addition to being a homopolymer insoluble in the majority of solvents, the chains of cellulose are incorporated within the lignocellulosic matrix interfering with enzymatic or acidic hydrolysis [23, 24]. Hemicellulose, which represents 20 to 40 % of lignocellulosic feedstocks, is an amorphous mass of several polymers of different monosaccharides, such as, glucose, arabinose, galactose, xylose and mannose, with xylose being, the main constituent in most hardwoods [23, 24]. Hemicellulose is soluble in alkali and, although a weak polymer, is relevant to the structural integrity of the plant matrix, binding non-covalently to cellulose and covalently to lignin [23, 24]. Moreover, hemicellulose hampers the activity of hydrolases on cellulose by creating a physical barrier [23, 24]. Lignin (20-30% of total biomass) is a hydrophobic, amorphous, and rigid heteropolymer constituted by phenylpropanoid units, which are not fermentable [21, 24]. Besides being a surface for irreversible adsorption of hydrolases, the bonds between lignin, hemicellulose, and cellulose affect the general hydrolysis of the biomass by reducing the exposed area of (hemi)cellulose [24].

As a result of the intricate structure of lignocellulosic biomass and the inability of most microorganisms to degrade the lignocellulosic polymers quickly and efficiently, the production of bioethanol through fermentation of the biomass by microbial cell factories requires two initial steps: pre-treatment, and hydrolysis. Pre-treatment consists of the breakdown of the plant matrix in order to in-

crease the area/volume ratio rendering it accessible for posterior hydrolysis. On the other hand, hydrolysis involves the chemical depolymerization of the polymers into smaller and degradable molecules, that in the case of hemi(cellulose) are sugars [21, 23, 24, 26].

.

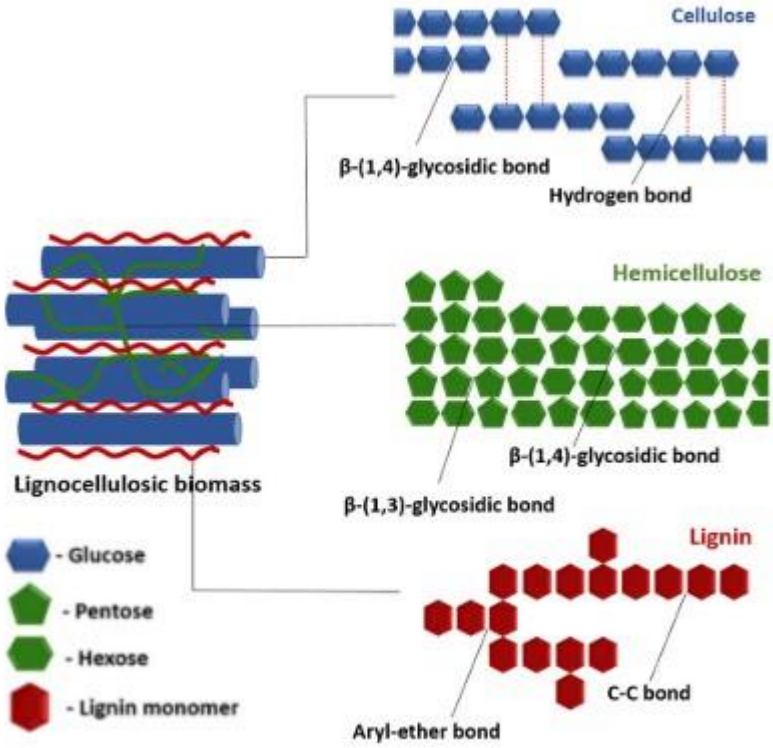

Figure 1.3. Representation of lignocellulosic biomass
Adapted from Baruah *et al.* 2018 [27].

Pre-treatment processes can be physical, chemical, physicochemical, or biological. Selection of the methods varies according to the nature of the biomass and its purpose, but must ensure increased efficiency of subsequent hydrolysis and avoid loss of carbohydrates and formation of inhibitory compounds [23, 27].

Physical treatments are the first used since they allow for a size reduction of the biomass particles and usually do not lead to the production of inhibitors. Physical methods include milling, microwave, ultrasonication, and extrusion. Milling allows the reduction of the particles, increasing the yield and rate of hydrolysis, however, needs high-energy inputs and the equipment is expensive. Microwave reduction is an energy-efficient procedure that consists of lignocellulosic structure's breakdown caused by high temperatures generated by electromagnetic waves. Extrusion, despite its high-energy requirements, is usually the preferred method for lignocellulosic breakdown. The breakdown occurs when biomass is subjected to high-temperatures and crushing forces caused by pressing the biomass into a barrel with aid of screws. Finally, ultrasonication uses ultrasonic waves to cause shear forces of biomass cleavage through cavitation [27, 28].

Chemical pre-treatments include acidic and alkali pre-treatment, ionic liquids, organosolv processes among others [27, 29]. Acidic and alkali pre-treatments are the most used not only because are well-established methods but also because lead to the production of less toxic compounds compared with

treatment with organic solvents or ion solutions [27, 29]. The alkali treatment uses basic solutions, preferably sodium hydroxide, to solubilize the lignin and parts of the hemicellulose, exposing cellulose to hydrolysis. Furthermore, alkali solutions also cause cellulose swelling reducing its structural integrity [27, 29]. This technique is advised for feedstocks with minimal lignin content and has the downside of being necessary to remove the base before fermentation and/or hydrolysis can occur due to its toxicity and/or possible interference with acidic hydrolysis [27, 28]. On the other hand, treatment with acid leads to the breakage of glycosidic bonds between hemicellulose and cellulose or between sugar monomers of the referred macromolecules, which makes acid pre-treatment also a form of hydrolysis. This method can be achieved with inorganic or organic acids and with concentrated solutions of the acid and low temperatures or low acid concentrations and temperatures up to 250 °C. The use of high concentrations of acid leads to greater sugar yield, however, can be corrosive, increasing the maintenance costs and may lead to degradation of cellulose into phenolic acids, aldehydes, and furans such as furfural and 5-hydroxy methyl furfural (HMF). Methods with diluted acids are extensively employed, especially diluted sulfuric acid, since allow a high recovery yield of hemicellulose sugars such as xylose. Nonetheless also lead to the formation of several inhibitory compounds, such as acetic acid, and the pH has to be neutralized to allow the following steps [22, 27–29].

The steam explosion method, the most frequently used pre-treatment method, is a form of physico-chemical pre-treatment that combines mechanical and chemical methods. The feedstock is first exposed to high-pressure steam at 160–260 °C which causes water to penetrate the plant matrix and the pressure is quickly released causing expansion of water and explosion of the biomass. Additionally, the high temperatures and pressure favor the breakage of glycosidic bonds and the hydrolysis of cellulose and hemicellulose releases acetic acid, that through acid treatment boosts hydrolysis [27, 28].

It is also possible to employ biological methods for the breakdown of lignocellulosic biomass. These include breakdown using microorganisms, such as white-rot fungi, or alternatively, using enzymes produced by such organisms. Biological methods imply low costs, are environmentally friendly, do not lead to the formation of inhibitors but are slow and some fungi can also consume the sugar monomers, decreasing the final yield [27–29].

Hydrolysis of the macromolecules into their structural units can be achieved by two main processes, acid hydrolysis and enzymatic hydrolysis. The process of acid hydrolysis was already described as a form of chemical pre-treatment. Despite being a fast method that can occur continuously, has a relatively low yield since the conditions lead to the transformation of the recovered sugars into furan aldehydes which may then be converted into weak acids - levulinic and formic acids, for example [22, 24, 26, 28]. Since pentoses are more easily degradable it is possible to recover 5-carbon sugars with milder treatments and then use more intense methods for hexose recovery [22, 23]. However, the production of inhibitors still exists, as well as the production of acetic acid and the need to neutralize the pH, in order for the sugars to be fermented by cell factories.

On the other hand, enzymatic hydrolysis does not lead to the formation of undesired compounds and occurs at moderate conditions [28]. Nonetheless, the final yield is influenced by a multitude of factors such as temperature, pH, enzyme activity and concentration, pre-treatment methods, and composition of the biomass among others [24, 28, 29]. Enzymatic degradation of cellulose is usually achieved with a cocktail of endoglucanases, responsible for cleaving the cellulose chains originating free extremities, exoglucanases, that remove dimers of glucose from free ends and β-glucosidases which are responsible for separating the glucose dimers [24, 28]. Since hemicellulose is a more complex and variable polymer, the enzymes used are usually adapted to the original biomass [24, 28]. Some of the more commonly used enzymes are xylanases, mannases, galactosidases, acetyl xylan esterases among others [28]. Several bacteria and fungi produce enzymes capable of degrading cellulose and hemicellulose.

For example, *Trichoderma reesei* is a fungus that produces more than 10 enzymes involved in the hydrolysis of lignocelluloses and has a complex regulatory network involved in its synthesis [22].

After the obtention of monomeric sugars, bioethanol can be produced through microbial fermentation. *S. cerevisiae* is one of the preferred organisms for industrial fermentation of biomass into ethanol, however, it is incapable of fermenting efficiently pentoses. Due to that, there are several studies that aim to engineer this yeast and transform it into a pentose-degrader [30–33].

The application of the so-called non-conventional yeasts in biorefinery environments such as pectin-ich and lignocellulosic hydrolysates has also been under investigation [34–36]. Several of these yeast species are able to metabolize complex substrates, reducing the pre-treatment requirements and are highly tolerant to temperature, acetic acid, osmotic stress, and other stresses found in lignocellulosic hydrolysates, all of which are traits with great industrial relevance [34]. For example, *Zygosaccharomyces rouxii* has a high osmotolerance, *Kluyveromyces marxianus* and *Ogataea polymorpha* are able to grow at considerably higher temperatures, *Zygosaccharomyces bailii* is tolerant to high acetic acid concentrations and *Pichia kudriavzevii* has higher ethanol yields that *S. cerevisiae* even in high temperature and acetic acid conditions [34, 35, 37].

1.3. Mechanisms of *S. cerevisiae's* response towards the presence of lignocellulose-derived inhibitors

1.3.1. Toxicity of furan aldehydes and weak acids towards *S. cerevisiae*

During pre-treatment and/or hydrolysis of the lignocellulosic feedstocks several compounds accumulate that are toxic to *S. cerevisiae* and to other microbial cell factories. Figure 1.4. displays a simplified representation of the effect of weak acids and furans on *S. cerevisiae* cells.

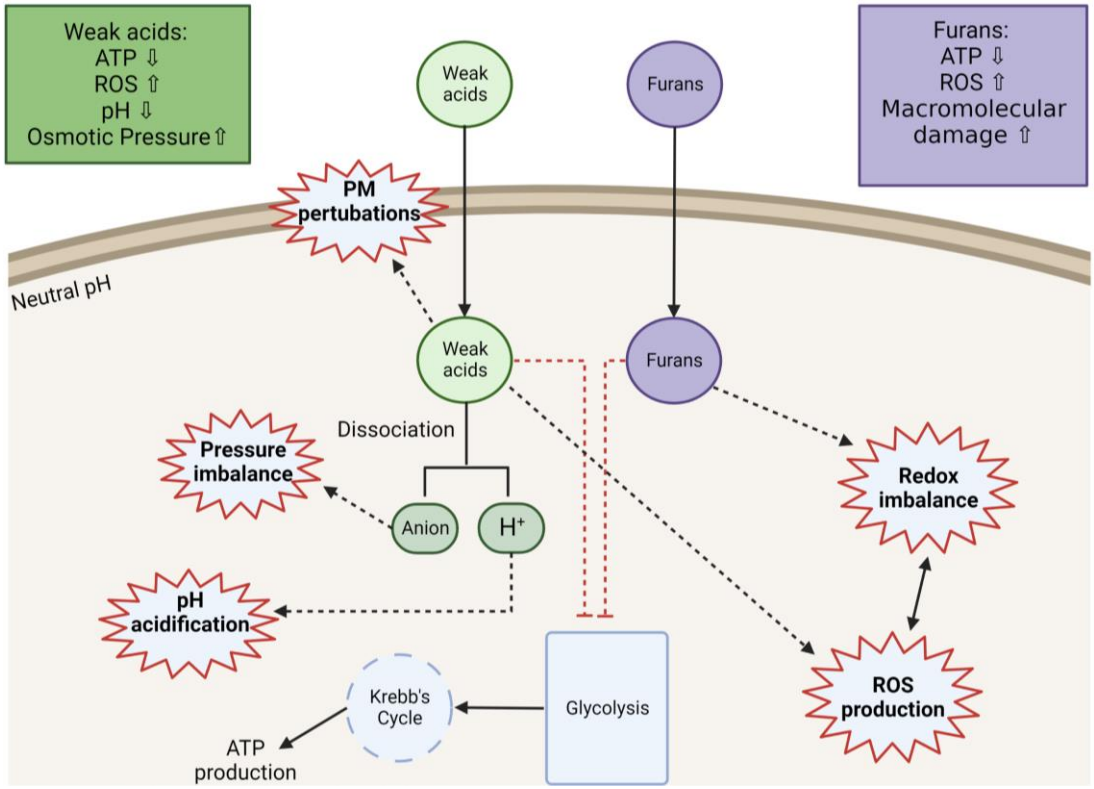

Figure 1.4. Main inhibitory effects of weak acids and furans in *S. cerevisiae* cells
Original design prepared using BioRender. PM – Plasma membrane; ROS – Reactive oxygen species.

Furfural and HMF are furan aldehydes that result from dehydration of monomeric sugars during acid hydrolysis or steam explosion. Breakdown of HMF can then originate formic acid which along with acetic acid, can also be formed by de-acetylation of lignocellulosic polymers especially hemicellulose [38, 39]. The concentration of formic acid in lignocellulosic hydrolysates varies between 9 and 63 mM, with a typical value of 30 mM while the acetic acid concentration varies between 17 and 250 mM [39–41]. The typical concentrations of HMF usually vary between 0.2 and 17 mM but on spruce hydrolysate can reach 47 mM [41, 42]. As for furfural, its typical concentration on lignocellulose hydrolysates varies between 5 and 31 mM and can reach 115 mM in hydrolysates derived from corn [41, 42].

Aldehydes such as furfural and HMF are molecules with a carbonyl group in an extremity. The double bond between carbon and oxygen creates partial positive and negative charges on the carbon and oxygen atoms, respectively. This propriety renders this group extremely reactive, being able to react with proteins and DNA by nucleophilic addition [43]. As a result, furfural and HMF are responsible for damaging proteins and DNA. Furthermore, it has been described that furfural and HMF inhibit RNA and protein synthesis, reducing the metabolic rate of the cell and delaying growth [43]. In particular, these compounds have been linked to the inhibition of aldehyde dehydrogenase, alcohol dehydrogenase, and pyruvate dehydrogenase [44]. Additionally, furfural appears to be involved in a destabilization of the internal structures of the cell linked to the induction of production of reactive oxygen species (ROS) [45]. Some structures affected are the vacuolar and mitochondrial membranes, actin cytoskeleton and nuclear chromatin [45].

Other compounds typically produced in the treatment of lignocellulosic materials are acetic and formic acids. Acids can be found in two states, undissociated and dissociated. Dissociation of the acid corresponds to the separation of the anion and the proton and occurs mostly when the pH of the medium is superior to the pKa of the acid, being Ka the dissociation constant of the acid and pKa the negative logarithm of Ka.

The pKa of acetic and formic acids is 4.75 and 3.75, respectively. The medium's pH during lignocellulosic fermentation is usually found between 4 and 4.5. At such pH levels, acetic acid is primarily undissociated and formic acid is mostly dissociated. These weak acids are composed of a small fatty acid chain and a carboxylic group and, for that reason, are fairly hydrophobic when undissociated, entering the cell through the plasma membrane. Within the cell, the cytosol is nearly neutral, and the acids dissociate. Although the dissociation constant of formic acid is 10 times higher than acetic acid's, drastically decreasing the concentration of the undissociated form found in the medium, this molecule is also smaller which facilitates its entry into the cell where it readily dissociates [38, 46].

After dissociation of the acid, the proton contributes to the acidification of the cytosol, dissipating the cell's proton gradient, impairing its energy production and consequently all metabolic pathways, affecting the chemical environment of all structural proteins and enzymes, hampering their function and perturbing vesicular transport which is highly dependent on pH differences [47–49]. On the other hand, the anion, a charged molecule that cannot diffuse out of the cell through the plasma membrane, accumulates intracellularly, being able to cause cell death, or at least plasma membrane damage, through the entry of water that generates high pressures. Additionally, the anion may cause oxidative stress by inducing the formation of ROS which have the potential to affect and destroy almost all macromolecules in the cell [47–49]. Weak acids, especially acetic acid, has also been shown to reduce the amino acid uptake of the cell, and cause a depletion of the ergosterol content of the cell, being that the later affects the resistance to higher temperatures that can be present in bioreactors and increases the permeability of the membrane [48, 50]. It is thought that formate might be intrinsically more toxic than acetate. In fact, besides being an acid, formic acid is also an aldehyde and for that reason is re-

sponsible for inhibiting enzymes involved in the central metabolism of the cell (glycolysis and Krebs cycle) [39, 51]. As such exposure to weak acids, through the formation of ROS, acidification of the cytosol and increments in the cell internal pressure is deleterious to growth and has impactful effects on the plasma membrane rendering the cell more susceptible to external offenses [38, 47–49].

All these factors lead to growth impairment and/or cell death, slowing the transformation of substrate into biofuels. As a result, it is imperative to understand the mechanisms of resistance and detoxification of the cell.

1.3.2. General mechanisms of *S. cerevisiae* response to inhibitors

In the presence of toxic compounds, yeast cells respond towards reducing the intracellular concentration of the stress agent by exporting it to the extracellular medium, by metabolizing it into a compound that is neutral or even beneficial to the cell, and by restricting the diffusional entry of the liposoluble acid form into the cell. It is also imperative to activate mechanisms that allow counteracting the inhibitor's consequences, especially those with fast and deadly action.

Weak acid dissociation results in a drop in the intracellular pH, dissipating the proton gradient across the plasma membrane and causing the inhibition of major intracellular enzymes involved in cell metabolism. In order to reestablish a neutral pH, cells activate plasma and vacuolar membrane H^+-ATPases that extrude protons at expense of ATP molecules [49, 52].

Weak acid resistance is also dependent upon the modification of the plasma membrane and cell wall composition, which leads to a decrease in the permeability of the cell envelope [53, 54]. This action reduces the diffusional entry of non-dissociated weak acids and avoids re-entry of the acid form after the expulsion of the anion from the cytosol, limiting the associated futile cycle [53, 54]. Extrusion of the anion is thought to be perpetrated by efflux pumps such as Tpo2 and Tpo3 [49].

Once in the cytosol, acetate can be activated into acetyl-Coenzyme A which can be used for lipid and sterol biosynthesis [55]. Formate can be converted into CO_2 by formate dehydrogenase and furfural and HMF can be modified, at the expense of NAD(P)H, into less toxic compounds [39]. Furfural is converted into furfuryl alcohol or furoic acid and HMF in 5-hydroxymethylfurfuryl alcohols like furan dimethanol [41, 56, 57].

Unregulated production of ROS and deregulation of antioxidant balance is one of the most threatening effects of weak acid and furan-induced stresses and requires quick activation of several response mechanisms by the cell. Yap1 is one of the transcription factors (TF) activated during oxidative stress. Its activation is induced by furans and phenolic compounds and results in the expression of proteins involved in the H_2O_2 and thiol reduction, elimination of superoxide anions, and the glutathione system. Moreover, Yap1 is also involved in the activation of other TFs, such as Stb5, which plays an important role in the regeneration of NADPH [41, 58].

Other responses induced by lignocellulosic inhibitors include regeneration of ATP and NADH, regulation of osmotic stress, and induction of autophagy and apoptosis. Stress responses in the model yeast are incredibly complex processes, and their full examination goes beyond the purpose of this work. Nonetheless there are several reviews and omics studies focused on identifying all the players involved in these responses [41, 46, 63, 64, 47, 49, 51, 57, 59–62].

The next sections will be dedicated to exploring the role of the cell envelope and multidrug/multixenobiotic resistance (MDR/MXR) transporters on the cell's ability to withstand stressful conditions.

1.4. Yeast cell envelope: a key player in multi-stress resistance in *S. cerevisiae*

S. cerevisiae's cell envelope is composed of the cell wall and the plasma membrane.

The cell wall is responsible for the cell's shape and its structural integrity, which is especially important when the cell is subjected to changes in osmotic pressure, mechanical aggressions and during the cell replicative process [65]. *S cerevisiae*'s cell wall is composed primarily of β-glucans, mannoproteins, and chitin. Mannoproteins and β-glucans represent about 95% of the cell wall mass. Mannoproteins are in direct contact with the exterior and β-glucans are its structural foundation, while also being responsible for the high elasticity of the cell wall. Chitin is usually found in association with β-glucans and is crucial during the budding process [65]. The yeast cell wall is permeable to any solute with a molecular weight less than 600 Da [66].

The plasma membrane is a selectively permeable, asymmetric, and fluidic barrier composed of an approximately 7.5 nm phospholipid bilayer in which are embedded proteins, and other lipids. The lipid bilayer is impermeable to charged molecules [65, 66].

The plasma membrane provides a selective barrier between the intra and extracellular spaces and its main functions are to establish and regulate its electrochemical gradients, to regulate the import and export of molecules (toxic and beneficial), and to be a support matrix for catalysis of several reactions, for binding of signaling molecules and for the biosynthesis of cell wall components [65, 66].

Over the years, studies have demonstrated that, contrary to original belief, the yeast's plasma membrane is not uniform in all its extension. For example, there is a physical separation between the plasma membranes of the bud and the mother cell that ensures that all essential plasma membrane components are present in the daughter cell. There are dynamic zones, with an abundance of plasma membrane H^+-ATPase, where diffusion, endocytosis and cell wall synthesis occur at a fast pace. Other domains are smaller, punctuate patches, and immobile where there is a specific accumulation of proteins in response to different stimulus. These are abundant, have distinct lipid composition and may be associated with cytoplasmatic protein complexes denominated eisosomes, especially in the growing bud or in the mother cell under abnormal environmental conditions The association between punctuate patches and eisosomes is highly regulated, creates invaginations of the membrane and is thought to protect proteins from endocytosis and to be involved in stress responses by agglomeration of required proteins. Other type of punctuate patches are regions with accumulation of TORC2 kinase that are thought to be involved in the cell polarity and sphingolipid biosynthesis [67].

1.4.1. Lipids of the plasma membrane

The lipids that compose the plasma membrane belong to three major groups, glycerophospholipids (70%), sphingolipids (15%), and sterols (15%) [68].

Glycerophospholipids are the most abundant lipids in the plasma membrane. These are amphipathic molecules composed of two fatty acids chains, linked by an ester bond to glycerol-3-phosphate. The hydrophobic regions of the two glycerophospholipids are facing each other, and the hydrophilic regions interact with the intracellular and extracellular media. The fatty acid chains are usually saturated or monounsaturated with 16 or 18 carbons and the phosphoryl group of the glycerol-3-phosphate can be bonded to several molecules, for instance, glycerol, choline, serine and inositol. The distribution of phospholipids between the inner and outer layers of the lipid bilayer is asymmetrical. In *S. cerevisiae* the inner layer has phosphatidylethanolamine, phosphatidylinositol, and phosphatidylserine, while the outer leaflet is enriched in phosphatidylcholine and sphingolipids. This dual distribution allows to

modulate the environment surrounding the plasma membrane, creates sites for specific interactions with proteins and signaling molecules and can physically distort the membrane [66, 69].

Sphingolipids are constituted by two long hydrocarbon chains and a head group. One of the chains has a hydroxyl group to which the head group is bonded and an amino group that is attached to the other long chain by an amide bond. The two hydrocarbon chains form a molecule known as ceramide. In *S. cerevisiae* the most common sphingolipids are inositol-phosphate ceramide, mannosyl-inositolphosphate ceramide, and mannosyl-diinositolphosphate ceramide. Due to their complex structure, the existence of several biologically relevant stereoisomers and the compartmentalization of its metabolism, the functions of sphingolipids remain unclear. Nonetheless they appear to be important in response to some stress conditions and to be implicated in the transduction of signals across the plasma membrane [66, 69, 70].

Sterols are also amphipathic molecules with a robust hydrophobic region and a hydrophilic region composed by a hydroxyl group. In *S. cerevisiae* the primary sterols of the plasma membrane are ergosterol and, to a smaller extent, zymosterol. Ergosterol is identical to cholesterol, the major sterol in animal cells' plasma membrane, except for two double bonds and its synthesis requires nearly 30 enzymes denominated Erg proteins [66].

In the plasma membrane, ergosterol acts as an elastic glue, preventing disruption of the membrane due to large vibrations caused by high temperatures, and loss of fluidity in response to constriction of phospholipids at low temperatures. In sum, ergosterol is involved in the regulation of plasma membrane's fluidity, including protein's movement and activity [66].

The permeability and fluidity of the plasma membrane are dependent on the composition of the membrane and influence the cell's ability to withstand deleterious environmental conditions [66, 71]. For example, the composition of the membrane in terms of length and saturation of the glycerophospholipids' fatty acids can be modulated in response to temperature variations. A membrane constituted by longer fatty acids with few double bonds is more rigid and thus more able to withstand higher temperatures without loss of structure [66]. Modulation of the sterol content of the membrane can also modify the membrane's capacity to remain intact when exposed to temperature variations, since, as already mentioned, ergosterol acts as a binding agent between other lipids at high temperatures and as a spacer, impeding severe constriction of the membrane at low temperatures [71, 72]. In addition, deletion of Erg proteins leads to accumulation of precursors and depletion of ergosterol and consequently, the cell's membranes become hyperpolarized, the cell is more sensitive to osmotic stresses, metal cations, cationic drugs, and the function of MDR transporters is impaired [71].

The plasma membrane has also been proven to play crucial roles in adaptation to acetic acid stress (Figure 1.4.). The non-dissociated acetic acid enters the cell by simple diffusion through the lipid bilayer, as such its composition and consequent permeability are key factors in acetic acid stress response. The fact that insertion of alcohols in the plasma membrane can increase acetic acid susceptibility by increasing the space between lipids, and consequently increasing permeability, is a clear indicator that the plasma membrane's composition needs to be tightly controlled and has severe impacts on resistance [73]. The role of sphingolipids in increasing the density and thickness of the plasma membrane, and decreasing permeability to acetic acid has been described for *Zygosaccharomyces bailii*, a yeast species with high tolerance to acetic acid, and a similar function can be hypothesized for *S. cerevisiae* [74]. In fact, it has already been verified that *S. cerevisiae* cell's under acetic acid-induced stress presented sphingolipids with longer carbon chains and double the normal concentration of ceramide, sphingolipid precursor, in the membrane [75]. The same study also identified an increase

in phosphatidylinositol concentration [75]. The ergosterol content on the membrane is also impactful in permeability. Acetic acid has been shown to induce a decrease in the ergosterol content of the cell, augmenting the permeability and has been shown that an important factor in acetic acid resistance is the ability to counteract this effect maintaining the ergosterol homeostasis [50]. Additionally, cell wall composition and structure have also been shown to be modified in acetic acid-induced stress responses. In such conditions the cell wall becomes stiffer and more robust decreasing the ability to re-associated acetic acid to reenter the cell and helping to put an end to the energetically expensive diffusion/expulsion cycle [54, 76].

Overall, the plasma membrane's lipid composition and structure are crucial for yeast resistance to environmental insults, determining passive diffusion of compounds and maintaining cellular integrity.

1.4.2. MDR/MXR transporters

Proteins interacting with the plasma membrane can be classified into transmembrane proteins, interacting with up to three different environments, cytosolic or fully extracellular proteins, exposed only to one side of the membrane. As such, they usually contain at least one hydrophobic region where interaction (anchoring) with the membrane occurs. The plasma membrane proteins account for approximately 50 % of its mass and the plasma membrane H^+-ATPase is by far the most abundant protein. Like H^+-ATPase, the majority of plasma membrane proteins are transporters, allowing a regulated exchange of molecules between the interior and exterior of the cell. Plasma membrane proteins can also participate in signal transduction, cell wall synthesis, interaction with cytoskeleton and endo- or exocytosis processes [66].

The simultaneous acquisition of multidrug/multixenobiotic resistance (MDR/MXR) is a generalized phenomenon with clinical, agricultural, and biotechnological repercussions that, in *S. cerevisiae*, is frequently associated with the action of a group of plasma membrane transporters. These transporters belong mainly to the ATP-binding cassette (ABC) superfamily [77–80], and to the Major Facilitator Superfamily (MFS) [81–83]. Such transporters have been traditionally considered to pump drugs/xenobiotics out of the cell, therefore lowering its concentration within the cell and diminishing its deleterious effects. However, studies have shown that they also play physiological roles in unstressed conditions, having natural substrates and that their mechanisms of protection might be more complex and go far beyond drug efflux [84].

1.4.2.1. Major Facilitator Superfamily transporters

MFS transporters carry small molecules across the plasma membrane, either by uniport, symport or antiport systems, being powered by ion gradients. However, their role in MDR/MXR is thought to be associated with the antiport of protons and drugs/xenobiotics [81, 85].

At a structural level, MFS transporters associated with MDR/MXR are constituted by one polypeptide chain of 500 to 600 amino acids [85]. The polypeptide chain is assembled in two sets of six or seven transmembrane α-helices linked by a cytoplasmatic loop [86].

According to phylogenetic studies, MFS superfamily is divided into two families: 12-spanner drug:H+ antiporter family 1 (DHA1) (scheme in Figure 1.5.- green) and DAG (DHA2/ARN/GEX) family. The DAG family includes three formerly independent MFS transporter families, the siderophore transporters (ARN) family, the glutathione exchangers (GEX) family, and the 14-spanner drug:H+ antiporter family 2 (DHA2) [81, 82, 85, 87].

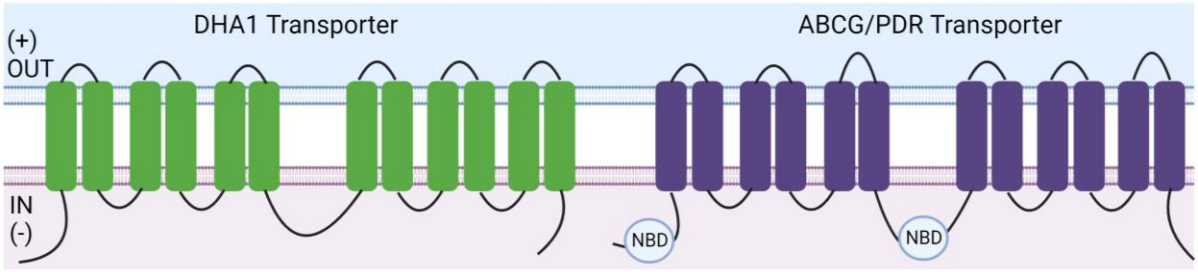

Figure 1.5. Schematic representation of the structural organization of DHA1 and ABCG transporters
Scheme illustrating the organization of the transmembrane domains in the membrane of DHA1 (green, left) and ABCG/PDR (purple, right) transporters. IN and OUT indicate the interior and exterior of the cell, respectively. Original design prepared using BioRender. NBD – nucleotide binding domain.

S. cerevisiae's genome encodes 12 transporters belonging to the DHA1 family - Aqr1, Dtr1, Flr1, Qdr1, Qdr2, Qdr3, Tpo1, Tpo2, Tpo3, Tpo4, Yhk8, and Hol1. Qdr1, Qdr2 and Aqr1 are paralogous proteins. *QDR1/QDR2* and *AQR1* diverged from an original gene in consequence of the WGD and *QDR1* and *QDR2* genes comprise a tandem repeat originated post-WGD. *TPO2* and *TPO3* are also paralogous genes that originated during WGD [88]. All of these transporters are generally found only in the plasma membrane, and all except Yhk8 and Hol1 have described roles regarding MDR/MXR. Nevertheless, cells with low susceptibility to azole drugs appear to have higher *YHK8* transcription levels, and Hol1 seems to be involved in the importation of histidine precursors and cations [85, 89, 90]. The physiological and MDR/MXR roles of DHA1 transporters are listed in Table 1.1.

In *S. cerevisiae*, the DAG family includes 16 transporters - Atr1, Azr1, Sge1, Vba1, Vba2, Vba3, Vba4, Vba5, Atr2, Amf1, Arn1, Arn2, Arn3, Arn4, Gex1 and Gex2 – with variable sub-cellular locations, physiological and MDR/MXR roles [87]. For example, the Vba transporters are localized in the vacuolar membrane, being involved in the transport of basic amino acids and all but Vba3 are involved in resistance against azole fungicides and anti-arrhythmic drugs [91]. The Gex transporters are involved in the transport of glutathione both out of the cell and into the vacuolar lumen [92]. The Arn transporters have a role in the transport of the complex siderophore/iron [93]. In the context of this thesis is also noteworthy to mention that Atr1, Azr1 and Atr2 were described to have a role in resistance to weak acids [85].

Table 1.1. Physiological role and MDR/MXR phenotype of DHA1 transporters
Table constructed with information reviewed in Sá-Correia and Godinho, 2022 and Godinho and Sá-Correia, 2019 [84, 85].

Transporter	Physiological role	MDR/MXR phenotype
Aqr1	Excretion of excess amino acids; Possible role in DNA replication during stress	Anti-arrhythmic and antimalarial drugs; Azole fungicides; Cationic dyes; Osmotic pressure; Acetic, propionic and butyric acids
Dtr1	Spore wall maturation (bisformyl dityrosine translocation)	Propionic acid; Anti-arrhythmic and antimalarial drugs
Flr1	Possible role in DNA replication during stress	Agriculture fungicides; Aniline analgesics; Immunosuppressants; Mutagens; Oxidizing and alkylating agents; Antibiotics
Qdr1	Possible role in cell wall assembly	Propionic acid; Anti-arrhythmic drugs; Azole fungicides; Herbicides

Qdr2	Divalent cation and potassium homeostasis; Oxidative stress response	Anti-arrhythmic drugs; Chemotherapy agents; Herbicides; Azole fungicides
Qdr3	Export of polyamines (spermine and spermidine); Possible role in yeast respiration; Cell wall assembly	Anti-arrhythmic drugs; Chemotherapy agents; Herbicides; Cations; Glutaric, adipic, pimelic, muconic and glutaconic acids
Tpo1	Polyamine homeostasis (spermine, putrescine, and spermidine) in response to pH levels and oxidative stress; Squalene synthase activation	Herbicides; Fungicides; Antimalarial and anti-inflammatory drugs; Chemotherapy agents; Metal ions; Immunosuppressants; Benzoic, octanoic, decanoic and propionic acids
Tpo2	Polyamine homeostasis (spermine)	Acetic and propionic acids
Tpo3	Polyamine homeostasis (spermine)	Acetic, formic, and propionic acids
Tpo4	Polyamine homeostasis (spermine, putrescine, and spermidine)	Acetic acid; Antibiotics; Anti-arrhythmic drugs

1.4.2.2. ATP-binding cassette transporters

The ABC superfamily is one of the largest families of proteins, with at least 3000 members widespread among all kingdoms of life. ABC proteins are active transporters that couple hydrolysis of ATP to the influx of nutrients, essentially in plants and procaryotes, or efflux, in fungi and animals, of different classes of molecules, either endogenous or exogenous [80, 85, 94]. In *S. cerevisiae* 30 ABC transporters have been identified, but only 22 are considered true ABC transporters since the remaining do not have predicted TMDs. The 22 ABC transporters, except two (CAF16 and YDR061w), belong to six subfamilies named ABCB, ABCC, ABCD, ABCE, ABCF, and ABCG [85, 95].

Structurally, the ABC transporters encoded in the genome of *S. cerevisiae* are comprised of two hydrophobic regions with six transmembrane domains (TMD) each and two hydrophilic segments in which are included two nucleotide binding domains (NBD), where ATP binds. The architecture of the transporters usually is 6TMD-NBD-6TMD-NBD. Nonetheless, the transporters with a better-described role in MDR/MXR (ABCG) have an inverted topology, NBD-6TMD-NBD-6TMD (scheme in Figure 1.5. – purple). Nowadays, is believed, that in efflux pumps, binding of the substrate to the TMD open to the intracellular medium triggers a change in the NBD and consequently leads to a conformation change that culminates in the release of the substrate in the extracellular medium. The resting configuration is then restored by ATP hydrolysis [85, 96, 97].

Transporters from the ABCG family are called PDRs for their role in pleiotropic drug resistance. These transporters exhibit a reverse topology and comprise between 1350 and 1550 amino acids in length. This subfamily includes Pdr5, Pdr12, Pdr15, Pdr10, Pdr11, Aus1, YOL075c, Snq2, and its paralog Pdr18. However, Pdr11, Aus1 and YOL075c do not have described roles in MDR/MXR and although present the characteristic reverse topology, do not have cysteine in the place of lysine in N-terminal Walker A motifs like the remainder. From a physiological standpoint, Pdr11 and Aus1 are involved in sterol uptake while the function of YOL075c is unknown [80, 85, 98]. The physiological and MDR/MXR roles of ABCG transporters are listed in Table 1.2.

The ABCC transporter Yor1, contrary to most members of its subfamily, and similarly to ABCG transporters, is localized in the plasma membrane instead of the vacuolar membrane. Structurally, Yor1 has a forward topology with an N-terminal extension characteristic of its subfamily [80, 85]. Physiologically, Yor1 is involved in the translocation of glycerophospholipids between plasma mem-

brane layers and regarding its MDR/MXR phenotype, is involved in resistance to reveromycin A, tautomycin, aleptomycin B, oligomycin, rhodamine B, cadmium, rhodamine 6G, doxorubicin and organic anions with carboxyl groups[99–104].

Table 1.2. Physiological role and MDR/MXR phenotype of ABCG transporters
Table constructed with information reviewed in Sá-Correia and Godinho, 2022 and Godinho and Sá-Correia, 2019 [84, 85].

Transporter	Physiological role	MDR/MXR phenotype
Pdr5	Translocation of glycerophospholipids between plasma membrane layers; Possible role in quorum sensing	Agricultural and clinical fungicides (vanillin and dodecanol); Antibiotics; Cations; Chemotherapy drugs; Herbicides; Human steroid hormones
Pdr10	Microdomain formation; Possible role in modulating the environment of Pdr12	Anionic dyes (congo red and calcofluor white); Organic solvents
Pdr12	Unknown	Weak acids of medium to long chains (benzoic, propionic, sorbic, levulinic and octanoic acid); Short chain alkanols
Pdr15	Unknown	Antibiotics (chloramphenicol); Detergents (polyoxyethylene-9-lauryl ether)
Pdr18	Ergosterol homeostasis	Alcohols (ethanol and 2-butanol); Anti-arrhythmic and antimalarial drugs; Metal cations (Zn^{2+}, Mn^{2+}, Cu^{2+}, Cd^{2+}); Clinical and agricultural fungicides (mancozeb); Herbicides (barban and 2-methyl-4-chlorophonoxyaceticacid), Polyamines; Acetic, formic and benzoic acids; High temperatures
Snq2	Putative role in lipid translocation and quorum sensing.	Antibiotics; Decanoic and oxalic acids; Cations; Chemotherapy agents; Detergents, Steroid hormones; Ionophores; Mutagens (triaziquone and 4-nitroquinoline 1-oxid)

1.4.2.3. MDR/MXR transporters' transcriptional regulation

Regulation of MDR/MXR transporters is achieved at many levels, among which is the regulation of their transcription by transcription factors (TFs) that, by binding to specific DNA regions, inhibit or activate it.

The main TFs involved in transcriptional regulation of *PDR5*, *PDR10*, *PDR15*, *PDR18*, *SNQ2*, *PDR12*, *YOR1*, *AQR1*, *DTR1*, *FLR1*, *QDR1-3* and *TPO1-4* (transporters from the DHA1 and ABCG families – plus Yor1 – groups studied in this work) in conditions of general stress, not exclusively MDR/MXR related, are depicted in Figure 1.6. The data used to construct the regulation matrix was obtained from the YEASTRACT+ database [105].

Of the 10 TFs included in Figure1.6., four (Rpn4, Pdr1, Pdr3 and Yap1) act upon almost all the genes considered. Rpn4 and Yap1 regulate the 17 genes while Pdr1 and Pdr3 regulate expression of all the ABC transporters and only 8 and 9, respectively, of the MFS transporters.

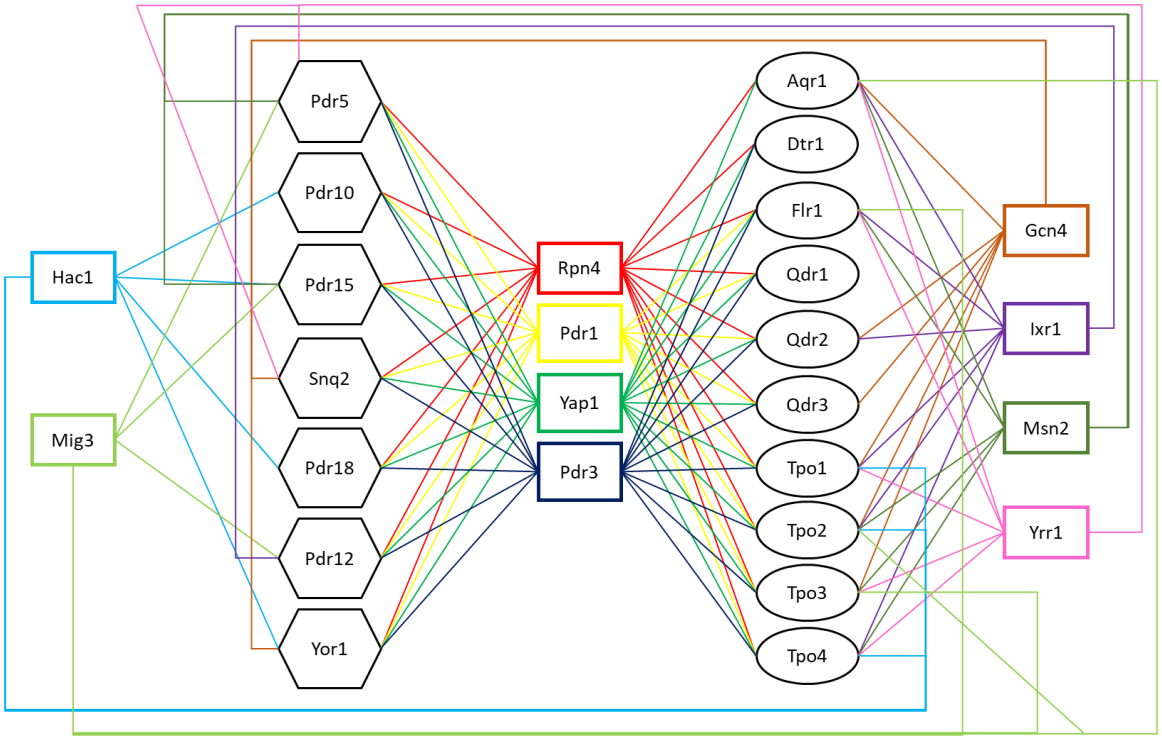

Figure 1.6. Simplified scheme of the transcriptional regulatory network controlling MDR/MXR transporters under stress conditions.
The data used to construct the regulatory network was extracted from YEASTRACT+ database (consulted in July 2022). The environmental condition selected was stress and were considered all the documented regulations based on DNA binding or expression evidence. To simplify, interactions between transcription factors are not shown and the transcription factors selected regulate at least 50 % of the transporters of one superfamily. Transcription factors are represented by rectangles, ABC transporters by hexagons and MFS transporters by ellipses.

Pdr1 and Pdr3 are two paralogue transcription factors with a zinc cluster ($Zn(II)_2Cys_6$) that play a central role in the regulation of MDR/MXR, including the regulation of the MDR/MXR transporters [85]. These TFs form homo and heterodimers and recognize the same sequence of nucleotides in the gene's promoters activating the expression of the transporters [106]. Rdr1 and Stb5 act as repressors by forming dimers with Pdr1 and/or Pdr3 and competing for the binding sequences [107]. Deleterious mutations on Pdr1 or Pdr3 result in decrease in resistance levels, indicating a crucial role in MDR/MXR. However, this role is not dependent on binding of the TFs since they are constitutively bound to DNA, independently of the presence or absence of toxic compounds. The concrete functions of Pdr1 and Pdr3 individually and as heterodimers are not clear. Nonetheless Pdr1 is linked to the activation of genes in early response to drugs and Pdr3 appears to intervene in regulation of stress regarding protein import in the mitochondria [106].

Yap1 TF is a zinc-finger TF that, as referred before, is the main player in the activation of oxidative stress response and is known to directly activate transcription of MDR/MXR transporters [58, 85]. On the other hand, Rpn4 is an unstable TF that activates a negative feedback loop by activating genes involved in the ubiquitin-proteosome system that will then destroy it. Rpn4 is also involved in the activation of base/nucleotide excision DNA repair. Both these functions are essential to cells under stress, since the stresses most commonly associated with industrial processes, usually result in DNA and protein damage, and also because stress resistance require quick activation of new pathways and deactivation of others which implies increased proteosome activity [108, 109].

Another TF of interest in the regulation of MDR/MXR transporters in the presence of weak acids, although with a smaller regulatory network among the considered transporters, is the transcription factor Haa1. The YEASTRACT+ analysis revealed that Haa1 is responsible for regulating the transcription of *PDR12, TPO2* and *TPO3*. Haa1 although closely related to copper-regulated TFs, has no identified role in copper homeostasis. Haa1 is a key TF in response to weak acid stress especially acids with a small lipophilic chain and most of all is known for its role in acetic acid resistance. Activation of Haa1 directly leads to transcriptional activation of *TPO2* and *TPO3* and to activation of several TF and genes, among which, two of the most important, are *SAP30* and *HRK1* genes. *SAP30* encodes a subunit of Rpd3L and HRK1 encodes a kinase involved in the post-translational regulation of plasma membrane transporters. Rpd3L is a histone deacetylase complex involved in the regulation of gene expression by chromatin remodeling and in transcriptional response to stress associated with Msn2 and Msn4. Haa1 also leads to cell wall remodeling by promoting the expression of Ygp1, a cell wall-related secretory glycoprotein [60, 110].

1.5. Project objectives and thesis outline

With ever-increasing energy demands and the worsening of environmental health after less than two centuries of industrial activities, becomes crucial to find ways to mitigate pollution while maintaining energy production.

Biorefineries, especially those involved in the production of biofuels from lignocellulosic substrates using microbial cell factories, are a promising cleaner and more environment-friendly solution. Lignocellulosic substrates have multiple natural sources, and their use allows us to give a profitable final destination to plant materials, contributing to the promotion of a circular economy. However, as detailed in the sections above, pre-treatment of lignocellulosic feedstocks generates toxic compounds that impose restrictions on microbial growth and consequently decrease the production yield [24].

Several microorganisms can be used in biorefineries and in fact, there is an ongoing attempt to diversify the microorganism's array [34–36]. Nonetheless, *S. cerevisiae* continues to be the preferred organism due to the in-depth knowledge that exists of its genome and physiological processes.

The yeast's plasma membrane and its embedded MDR/MXR transporters are one of the primary lines of yeast defense against toxic compounds and in particular those derived from lignocellulosic substrates. As such, the study of MDR/MXR transporters and their natural diversity allows us to better understand *S. cerevisiae*'s resistance mechanisms and ultimately design more robust industrial strains [16].

The main purpose of this work was to contribute to the characterization of MDR/MXR transporters' role in yeast tolerance to the fermentation inhibitors that are transversal to lignocellulosic hydrolysates from different waste materials: acetic and formic acids, furfural and HMF. Also, this work intended to identify artificial and natural gene variants of those transporters that can contribute to the resistance to the referred inhibitors. In that regard, the first step was defined to be the screening of a set of previously selected transporters, the ABC and MFS transporters with greater impact on *S. cerevisiae* resistance to furfural, HMF and acetic and formic acids. The most promising genes were randomly mutated by error-prone PCR and the resulting gene variants were characterized for their tolerance to acetic acid, when comparing to the expression of the BY4741 proteins. In parallel, a bioinformatic analysis of the natural variability of the sequence of MDR/MXR transporters genes across the genome of 1011 strains were conducted and strains harboring different gene variants were screened for high susceptibility toward acetic acid-induced stress.

2. Materials and Methods

2.1. Strains and growth conditions

2.1.1. Yeast strains and yeast cultivation conditions

Saccharomyces cerevisiae BY4741 (MATa *his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0) and the derived deletion mutant strains *pdr18*Δ, *pdr10*Δ, *pdr15*Δ, *yor1*Δ, *qdr1*Δ, *tpo2*Δ and *tpo3*Δ used in this work were obtained from the EUROSCARF collection (http://www.euroscarf.de/). These deletion mutants were selected for this work based on preliminary results obtained in Sá-Correia's Lab.

Saccharomyces cerevisiae strains RM11-1a (AAA), DBVPG3591_1b (ABS), CBS1385 (AGE), CBS2247 (AHK), NPA02-1 (AKH), A-18 (ANM), CLIB227 (ASH), DBVPG5760 (BRF), 8-3 (CAN), K14 (CLN), SM.8.7.BR1 (CNG) and MTF2549 (CQF) are part of the evolutionary study published by Peter *et al.* [117] and were kindly provided by the authors to Sá-Correia's Lab.

For yeast cultivation, Yeast Peptone Dextrose medium (YPD) and minimal medium (MM4) were used. YPD is composed of 10 g/l yeast extract (Gibco, Thermo Fisher Scientific, USA), 20 g/l peptone (Gibco), 20 g/l glucose (Scharlau, Spain), and distilled and deionized water (ddH_2O). In lignocellulosic hydrolysates, the amount of glucose is usually equal to that of xylose. Since the *S. cerevisiae* strains used do not metabolize xylose, the concentration of glucose was doubled. Therefore, for testing yeast susceptibility in the presence of fermentation inhibitors YPD40 was used, containing 10 g/l yeast extract, 20 g/l peptone, and 40 g/l glucose.

Minimal medium (MM4) is a yeast medium composed of 20 g/l glucose, 2.67 g/l $(NH_4)_2SO_4$ (PanReac AppliChem, Germany), 1.7 g/l yeast nitrogen base without amino acids nor ammonium sulphate (BD Difco, Thermo Fisher), 60 mg/l leucine (Sigma, Germany), 20 mg/l histidine (Sigma), 20 mg/l methionine (Sigma), 20 mg/l uracil (Sigma) and ddH_2O. To ensure selective pressure in yeast cells transformed with pGREG506 and pGREG506-derived plasmids, MM4 medium without uracil supplementation (MM4-ura) was used since the parental strain is incapable of synthesizing it and *URA3* is a selection maker encoded in these plasmids.

YPD, YPD40, MM4-ura and MM4 media's pH was adjusted to 4.5 with HCl and NaOH.

Unless stated otherwise, yeast cells were cultivated at 30 ºC in a climatized room, with orbital agitation (250 rpm; Agitorb 200, ARALAB, Portugal).

2.1.2. Bacterial strains and bacterial cultivation conditions

The *Escherichia coli* DH5α (Thermo Fisher Scientific) strain was used for the transformation and rapid replication of plasmids.

For *E. coli* cultivation, Luria Broth medium (LB) was used. This medium is composed of 25 g/l LB broth (NZYTech) and ddH_2O. Bacterial cells were cultivated at 37ºC with orbital agitation (250 rpm; Agitorb 200).

For both yeast and bacterial cultivation media, solid media were obtained by addition of 20 g/l agar (NZYTech, Portugal).

2.2. Preparation of exponentially-growing yeast cells

Yeast cells were stored as a 30% glycerol stock at -80ºC. Cells were defrosted on an YPD solid media plate and incubated for 48 h at 30 ºC. A loop of cells was used to inoculate 25 ml YPD in a 50 ml Er-

lenmeyer flask and cultivated overnight (between 12 and 16 h). The overnight cell culture was used to inoculate fresh YPD medium in the same media:flask volume proportion, to an OD600nm of 0.1 and cultivated at 30 ºC to a final OD600nm of 1. This cell suspension was used to prepare the susceptibility assays performed in the present work.

2.3. Assessment of yeast strains' susceptibility to fermentation inhibitors

To test the susceptibility of BY4741 and its derived deletion mutants for ABC or MFS transporters (*pdr18Δ*, *pdr10*Δ, *pdr15*Δ, *yor1*Δ and *qdr1*Δ, *tpo2*Δ and *tpo3*Δ) towards fermentation inhibitors, we analyzed their growth in Elenmeyer flasks and in microplates, in presence or absence of supplemented stress. For that, exponentially-growing cells were prepared as described in 2.2 and used to inoculate either 100 µl (microplates; Cell Culture Microplate, 96 Well, Ps, U-Bottom, Greiner, Austria) or 25 ml (Erlenmeyer flasks) fresh YPD40 media. Yeast growth was measured by following OD600nm every 15 minutes (microplates) or 2 hours (Erlenmeyer flasks).

Microplates were read in FilterMax F5 (Molecular Devices, USA) which was programmed to read OD595nm every 15 minutes with 5s orbital agitation before and between reads.

The inhibitors tested were acetic acid (Fluka, Thermo Fisher Scientific), formic acid (Fluka), furfural (Sigma, USA), and 5(Hydroxymethyl)furfural (HMF; Sigma). For testing yeast susceptibility to each individual inhibitor, stock solutions were prepared and set to pH 4.5. A range of concentrations was tested in BY4741 growth, to determine equivalent sub-lethal concentrations: acetic and formic acids (80 – 120 mM), furfural (10 – 80 mM), and HMF (10 – 60 mM).

Maximum specific growth rates were obtained by identifying the most linear portion of the exponential growth curve and calculating the slope of the line that best fits the points in the identified region.

2.4. Sub-cultivation of acid-adapted yeast cells

S. cerevisiae BY4741 and the derived deletion mutant cells (*pdr18Δ*, *tpo2*Δ, and *tpo3*Δ), prepared as described in 2.2 were used to inoculate liquid YPD40 medium (at pH 4.5) supplemented with 80 mM acetic acid or 100 mM formic acid. When cultures resumed growth and attained the exponential phase, they were used to inoculate fresh YPD40 medium supplemented with the same acetic or formic acid concentration. Growth was followed by measuring OD595nm in a microplate reader (section 2.3.).

Maximum specific growth rates were calculated as described in section 2.3.

2.5. Generation of artificial variants of MDR/MXR transporters genes

BY4741's genome was extracted and used to amplify genes *PDR18*, *TPO2*, and *TPO3* using error-proof and error-prone polymerases. The obtained fragments, together with the digested and purified pGREG506 plasmids, were transformed in the respective deletion mutant strains for homologous recombination-mediated cloning. To confirm the correct insertion of the fragments in the vector, the recombined vector was extracted from the yeast cells, inserted in competent *E. coli* cells to generate multiple copies, and tested by PCR and Sanger sequencing. Simplified scheme of the workflow in Figure 2.1.

2.5.1. Yeast genomic DNA extraction

Genomic DNA of parental strain BY4741 was extracted using a modified version of the phenol:chloroform method [111]. Briefly, yeast cells, collected from a loop of yeast cell biomass cultivated for 48h in a solid YPD media plate, were combined, in a microcentrifuge tube, with 150 µl of Tris-

EDTA buffer (pH 8) (Sigma), 150 µl of a phenol:chloroform:isoamyl alcohol solution (25:24:1, pH 8.0; Amresco, USA) and 0.5 mm diameter glass beads (BioSpec Products, USA). The solution was then vortexed for 5 min at maximum speed in the hotte, and centrifuged for 10 min at 13000 rpm (Scanspeed 1730R, Labogene, Denmark). The supernatant was recovered and precipitated by adding 300 µl cold 100% ethanol and 90 µl sodium acetate (3 M; Merck, Germay). The mixture was kept at -80 °C for 30 min. After the incubation time, DNA was collected by centrifugation (15 min, 1300 rpm), washed with 300 µl 70% ethanol, and centrifuged (10 min, 13000 rpm). The pellet was dried using a SpeedVac (Eppendorf Concentrator Plus, Eppendorf, Germany) and the DNA was resuspended in 50 µl of nuclease-free water (VWR, USA). The final concentration was quantified by absorbance measurement at 260 nm using an Eppendorf Nanodrop (Spectrophotometer ND-1000, Nanodrop, USA).

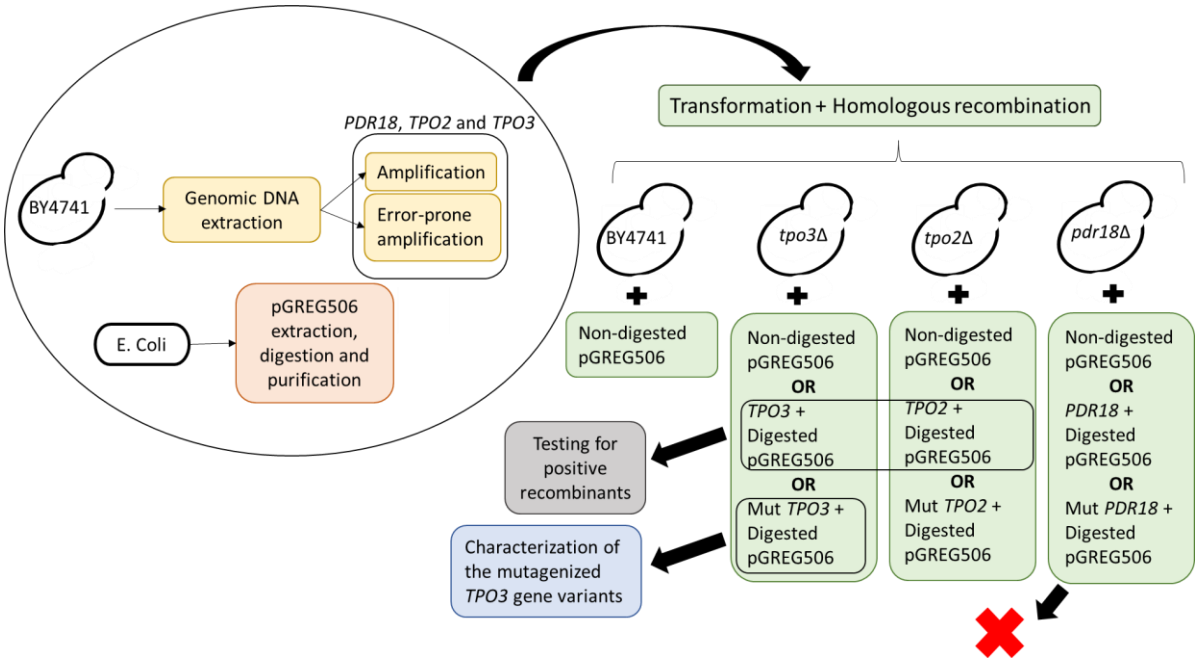

Figure 2.1. Workflow adopted for the generation and testing of artificial gene variants

2.5.2. PCR and error-prone PCR amplification of selected transporter's encoding genes

Amplification of *PDR18*, *TPO2*, and *TPO3* was achieved by polymerase chain reaction (PCR) using Invitrogen Platinum SuperFi II DNA polymerase (Invitrogen, Thermo Fisher), following the manufacturer's specifications. The primers' pairs used were TPO2_Rec, TPO3_Rec, and PDR18_Rec (Table 2.1.), ordered from STABVida, Portugal. These primers allow the amplification of both the promotor [considered the 1000 base pairs (bp) upstream of the start codon], the gene coding sequence (CDS), and a downstream region (about 400 bp) in the genome of BY4741. These primers also include tails that are homologous to regions Rec2 and Rec5 (for reverse and forward primers, respectively) of plasmid pGREG506 (EUROSCARF)[112]. Sequences of amplified fragments are included in Appendix A.

Error-prone PCR amplification of *PDR18*, *TPO2*, and *TPO3* was conducted using GeneMorph II Random Mutagenesis kit (Agilent Technologies, USA), following the low mutation frequency protocol. The primer pairs used were the same as for amplification.

All PCR reactions were conducted in a thermocycler GTC96S (Cleaver Scientific, UK). The amplification conditions for PCR and error-prone PCR are described in Appendix B.

Amplification was confirmed by gel electrophoresis in 1x TAE buffer (BioRad, USA) with 1% agarose (NZYTech, Portugal) using NZY DNA ladder III (NZYTech), 1x loading buffer (TaKaRa, Japan) and intercalator (RedSafe CHEMBIO, USA). A current of 100 V was applied for 30 min.

Table 2.1. Sequences of primer pairs used in this work
Capital letters represent the sequence homologous to the target region and lower-case letters indicate a tail homologous with region Rec5 (forward primer) and Rec2 (reverse primer), from pGREG506 vector. RR – regulatory regions. Primers were designed using Benchling and ordered from STABVida, Portugal.

Name	Forward sequence (5' → 3')	Reverse sequence (5' → 3')	Target /Amplified region	Amplicon size (bp)
TPO2_Rec	aacaaaagctggagctcgtttaaac-ggcgcgccGCTGCTTACATAC-CACTAGCTG	gcgtgacataactaattacatgac-tcgaggtcgacTTCA-GGAAGAGTAGCAAGAG C	*TPO2* plus RR	3333
TPO3_Rec	aacaaaagctggagctcgtttaaac-ggcgcgccGATTTCTCGAGAA-TCCGG	gcgtgacataactaattacatgac-tcgaggtcgacTCAAATT-CGTATTCCTCGG	*TPO3* plus RR	3348
PDR18_Rec	aacaaaagctggagctcgtttaaac-ggcgcgccGAGAGCTCATTGGA TGGAAG	gcgtgacataactaattacatgac-tcgaggtcgacATAC-GCTCTGTAAGTAGGTCC C	*PDR18* plus RR	5763
pGREG506_ CONF	GCGCAATTAACCCTCAC	CGAAGTTAAGTGCGCA-GA	Between Rec2 and Rec5 re-gions	Variable
TPO3_Fw1	CCTGTTACAACCTGTTCTCC	NONE	*TPO3*/RR	Variable
TPO3_Fw2	GTTATTGGGTTCTCGCTG	NONE	*TPO3*/RR	Variable
TPO3_Fw3	GGTTTCGGTATGGTGTTG	NONE	*TPO3*/RR	Variable

2.5.3. pGREG506 plasmid extraction, digestion, and purification

The pGREG506 and pGREG506-derived plasmids (simplified map in Appendix C) were extracted from *E. coli*, cultivated in liquid LB supplemented with (0.1 mg/ml) ampicillin (Sigma), using the NZYTech Miniprep kit, following the manufacturer instructions.

The extracted plasmid was digested with AscI and XhoI (both from NZYTech) using CutSmart as buffer. Calf Intestinal Alkaline Phosphatase (CIAP; NEB) was added during the enzymatic digestion. The reaction was incubated overnight at 37 ℃ in a water bath.

The digested plasmid-containing solution was applied in an agarose gel (1%) along with NZY DNA ladder III, 1x loading buffer, and intercalator. A current of 100 V was applied for 40 min and the band corresponding to the open plasmid backbone – 7000 bp – was purified using NZYTech GelPure Kit, following the manufacturer's instructions.

2.5.4. Yeast Transformation

The linearized pGREG506 (AscI/XhoI – digested) vector and the *PDR18*, *TPO2*, and *TPO3* PCR fragments, either mutagenized or not, were used to transform BY4741 and the derived deletion mutants *pdr18Δ, tpo2*Δ and *tpo3*Δ, respectively. As a control, each strain was also transformed with the pGREG506 (both digested or non-linearized) or with no DNA fragment (vector or amplicon). The protocol for transformation and cloning followed is an adaptation of the Gietz method [113]. The schematic representation of the recombination process after transformation is included in Appendix C.

Exponentially-growing yeast cells were obtained as described in section 2.2. Fifty milliliters of exponentially-growing cells were harvested by centrifugation, resuspended in 25 ml of sterile water, harvested by centrifugation, and resuspended in 1 ml of 0.1 M Lithium Acetate (LiAc; Sigma). The solution was centrifuged at top speed for 30 s in a microcentrifuge tube, resuspended in 1 ml of 0.1 M LiAc solution and 100 µl were distributed in new microcentrifuge tubes and pelleted by centrifugation at maximum speed for 1 min.

To each tube with pelleted cells, 240 µl of 50% (m/v) Polyethylene glycol MW3350 (Merck), 36 µl of 1 M LiAc, 10 µl of 10 mg/ml ss-carrier DNA (Thermo Fisher), 100 ng of AscI/XhoI-digested pGREG506, 500 ng of DNA fragment and sterile water were added to a final volume of 360 µl. After mixing, the solution was incubated at 42 °C (QBD1, Grant Instruments, UK) for 40 min. Finally, the supernatant was discarded after 30 s of centrifugation at top speed and cells were plated in a solid MM4-ura medium (pH 4.5).

The obtained colonies were looped to fresh MM4-ura media, grown over-night and stored as a 30% glycerol stock at -80 °C. Strains with non-mutagenized gene fragments were then tested to verify that the homologous recombination was successful. Section 2.5.5 describes that process. Colonies of cells with mutagenized gene fragments were screened for their resistance to acetic acid-induced stress as described in section 2.6.

2.5.5. Testing for positive recombinants

2.5.5.1 Yeast colony PCR

Colonies from the MM4-ura plate were recovered in PCR tubes containing commercial water for molecular Biology. The mixture was boiled for 4 min in the microwave, in presence of a glass of water to avoid sample evaporation.

Speedy NZYTaq 2× Green Master Mix (NZYTech) was used and the specifications from its manufacturer were followed. The annealing temperature of 47 °C was found to be optimal. The pair of primers used was pGREG506_CONF (STABVida) which bind to regions of the vector pGREG506 outside the insertion site of the gene (sequences in Table 2.1).

2.5.5.2. Extraction of plasmids from transformed yeasts

For the most promising candidates from yeast colony PCR results, we extracted the plasmids for cloning in *E. coli* and further testing.

Frozen stocks obtained from the protocol described in section 2.5.4 were cultivated in liquid MM4-ura overnight at 30 °C with orbital agitation. The culture was collected in a microcentrifuge tube by centrifugation and resuspended in A1 buffer of NZYTech Miniprep kit. Glass beads (0.5 mm diameter) were added, and cells were vortexed for 10 min followed by a period of rest of 10 min. The supernatant was recovered after 5 min centrifugation and buffer A2 was added. After this step extraction was achieved following the regular protocol from NZYTech Miniprep kit.

2.5.5.3. Preparation of *E. coli* competent cells

E. coli cells from strain DH5α were cultivated overnight in 50 ml of LB at 37 °C and then used to inoculate 100 ml of fresh LB to a starting OD600nm of 0.05. When the culture reached an OD600nm of 0.5, cells were collected by centrifugation at 4 °C, washed twice: first in 100 ml of a 0.1 M $MgCl_2$ (Sigma) solution and then in 100 ml of a 0.1 M $CaCl_2$ solution and incubated in ice for 25 min. Cells were recollected by centrifugation, resuspended in 22 ml of 0.1M $CaCl_2$ and 3.5 ml of 86% (v/v) glycerol (Himedia, India), and distributed in 150 µl aliquots for future use.

2.5.5.4. Transformation of competent *E. coli* cells by heat-shock and plasmid extraction

Extracted plasmid DNA (20 µl) was added to 50 µl of TCM solution [10 mM $CaCl_2$,10 mM $MgCl_2$ and 10 mM Tris-Hcl (pH7.5; Sigma)] and 150 µl of competent *E. coli* DH5α cells. The mixture was incubated in ice for 20 min, followed by heat sock for 3 min at 42 °C and 5 min in ice. To finalize, 800 µl of LB were added, and after 1 h of recovery at 37 °C with orbital agitation, cells were plated in solid LB medium supplemented with (0.1 mg/ml) ampicillin.

2.5.5.5. Confirmation of positive transformants

Colonies of transformed *E. coli* were cultivated overnight in liquid LB (supplemented with 0.1 mg/ml ampicillin) at 37 °C and the vector was extracted using NZYTech Miniprep kit. Confirmation of homologous recombination between non-mutagenized gene fragments and digested vector was achieved by amplification of the cloned DNA (pair of primers pGREG506_CONF, as described before), sanger sequencing (STABVida, pair of primers pGREG506_CONF) and by comparison of growth of these strains in microplate with MM4-ura supplemented with 70 mM acetic acid (section 2.3.) with the growth of BY4741 and deletion mutants with non-digested plasmid.

2.6. Characterization of the mutagenized *TPO3* gene variants

For clarity purposes, *tpo3*Δ and parental strain transformed with a non-linearized vector (pGREG506) will be called *tpo3*Δ_υ and BY4741_υ, respectively. The *tpo3*Δ successfully transformed with amplified *TPO3* gene, respectively, and open pGREG506 will be denominated *tpo3*Δ_υ/Tpo3, corresponding to a case where the inserted gene complements the deletion mutant, restoring the phenotype. Schematic representation of strains' nomenclature in Figure 2.2. The nomenclature for Tpo2 strains is the same as for Tpo3. Mutants corresponding to *tpo3*Δ transformed with open vector and mutant versions of *TPO3* will be denominated Mut_(number of the mutant).

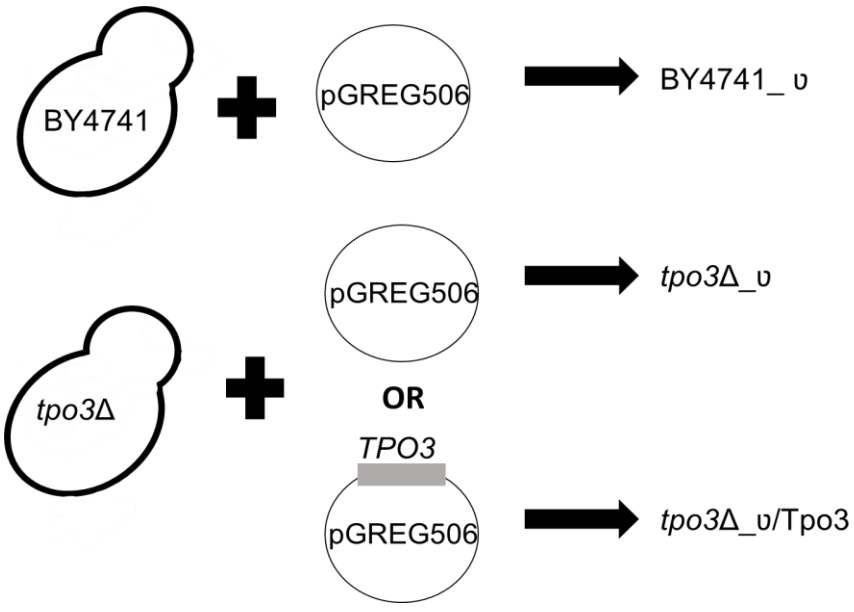

Figure 2.2. Schematic representation of the nomenclature of the strains obtained by transformation

2.6.1. Acetic acid susceptibility assay

The susceptibility of the *tpo3*Δ strain harboring each of the mutagenized variants of *TPO3* gene obtained was tested by growth in MM4-ura medium supplemented with 80 - 90 mM acetic acid. The

procedure was done as described before for yeast growth in microplates (section 2.3). The growth of each strain was compared to that of the *tpo3*Δ strain harboring the BY4741 *TPO3* gene variant.

2.6.2. Genetic characterization of mutants

Plasmids of the best-performing strains harboring mutagenized variants of *TPO3*, [Mut_44, Mut_47, and Mut_71 (one of the worst strains)], were extracted, used to transform *E. coli* and extracted from *E. coli*. Protocols used for confirmation were the same as described (sections 2.5.5.2 to 2.5.5.5). For sanger sequencing, primers used were TPO3_Fw1, TPO3_Fw2, TPO3_Fw3, and pGREG506_CONF Fw (Table 2.1.). Primer pGREG506_CONF forward binds to the plasmid, directly upstream of the Rec5 region while primers TPO3_Fw1-3 bind to the inserted fragment (Tpo3 CDS, upstream and downstream regulatory regions), allowing full coverage of the inserted fragment. Plasmid from *tpo3*Δ_υ/Tpo3 was also fully sequenced at this point.

Visualization of the chromatogram/alignments was achieved using the Benchling platform. Prediction of the impact of mutations on the proteins was achieved through analysis of the Tpo3 structure - retrieved from UniPort; code Q06451 [114] and predicted using AlphaFold software [115]. The retrieved protein structure was visualized with Coot [116]. The YEASTRACT+ database [105] was explored to identify possible impacts of mutations outside of the CDS.

2.7. Search for natural variants of *TPO3* gene across 1011 *S. cerevisiae* strains

2.7.1. Bioinformatic extraction of *TPO3* gene variants' sequence

The sequence variants for the gene *TPO3* (YPR156C) for each of the 1011 *S. cerevisiae* isolates, whose genome sequences were recently published [117] were retrieved from the respective single nucleotide polymorphism (SNP) matrix. Extraction of the *TPO3* sequence variants for each isolate was made using the bioinformatic tool Tabix [118]. Variant annotation to the protein sequences was made using the R/Bioconductor packages, BSgenome, and VariantAnnotation [119, 120].

2.7.2. Construction of *TPO3* sequence variants rooted tree

Prior to the tree construction, *S. cerevisiae* isolates with identical genomic sequences were grouped to get only isolates displaying unique sequences. The maximum likelihood tree was constructed using RAxML-HPC v.8 on XSEDE, available at CIPRES [121, 122], based on the GTRCAT model with 100 rapid bootstraps. All parameters were set as default for nucleotide sequences, and the *S. cerevisiae* S288c (BY4741) genomic sequence was set as the tree outgroup.

2.7.3. Prediction of the impact of SNPs on Tpo3 function

The identified SNPs were analyzed with mutfunc [123] – a software for prediction of the impact of mutations on the proteins - and through analysis of the Tpo3 structure - retrieved from UniPort; code Q06451 [114] and predicted using AlphaFold software [115]. The retrieved protein structure was visualized with Coot [116].

2.7.4. Screening of several *S. cerevisiae* strains' resistance to stress induced by acetic acid

Twelve *S. cerevisiae* strains already available at the lab and included in a recent paper, [117] and BY4741 were cultivated in microplates containing YPD supplemented with 0, 70, and 90 mM of acetic acid. The protocol was performed as described in section 2.3.

3. Results

3.1. MDR/MXR transporters are determinants of yeast tolerance to common fermentation inhibitors

3.1.1. Pdr18, Tpo2, and Tpo3 are determinants for yeast tolerance to weak acid stress

In order to search for the role of selected MDR/MXR transporters in yeast tolerance to the inhibitors usually found in lignocellulosic hydrolysates and other biotechnologically-relevant industrial processes, the deletion mutants *pdr18*Δ, *pdr10*Δ, *pdr15*Δ, *yor1*Δ, *qdr1*Δ, *tpo2*Δ, and *tpo3*Δ were cultivated in YPD40 media (pH 4.5) supplemented with different concentrations of either acetic acid, formic acids, furfural or 5-hydroxymethylfurfural (HMF).

The susceptibility of the referred individual deletion mutants towards stress induced by sub-lethal concentrations of each of the fermentation inhibitors was tested by assessing their growth in a microplate reader and compared to that of the corresponding parental strain. For that, a range of concentrations of each inhibitor was tested and the concentrations used in this work were selected for producing equivalent inhibitory effects in the growth of BY4741. The results are plotted in Figure 3.1. and the corresponding maximum specific growth rates calculated are shown in Table 3.1.

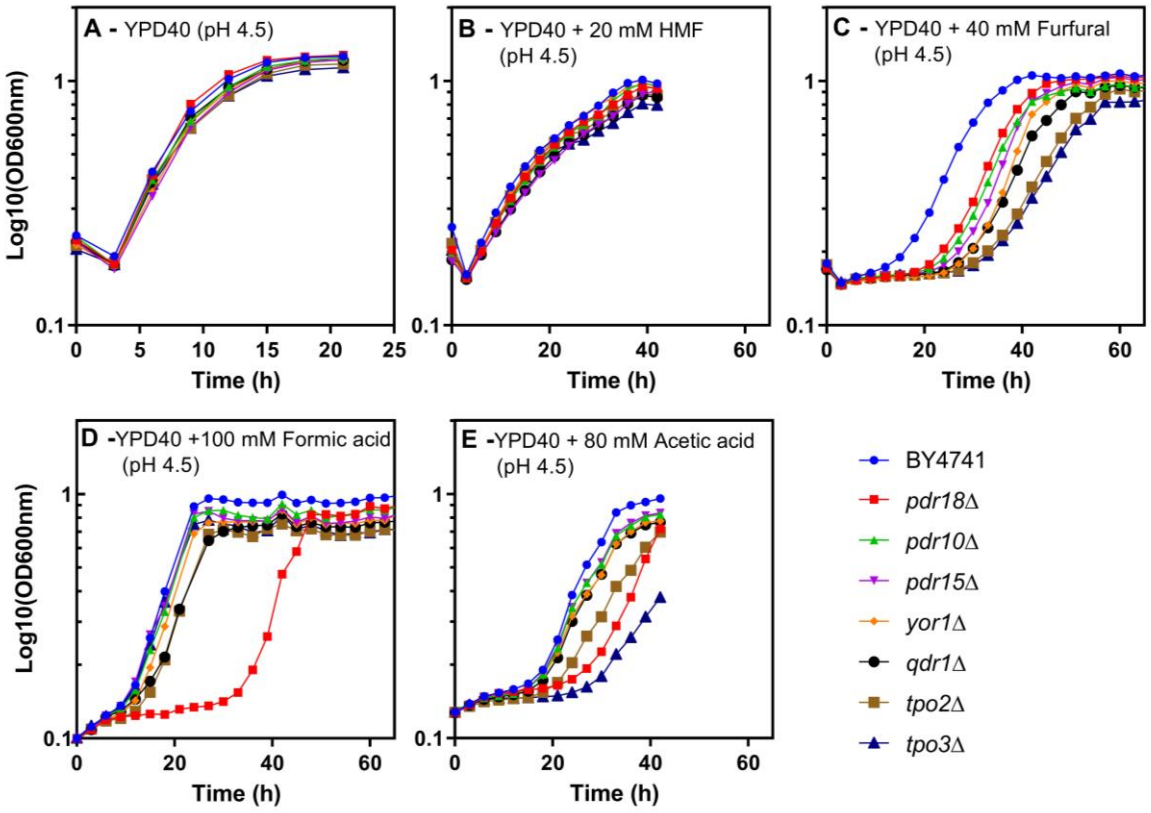

Figure 3.1. Susceptibility of individual deletion mutants for selected MDR/MXR transporters to the presence of fermentation inhibitors.
Growth curves of the parental and *pdr18Δ*, *pdr10*Δ, *pdr15*Δ, *yor1*Δ, *qdr1*Δ, *tpo2*Δ and *tpo3*Δ strains in liquid YPD40 medium (pH 4.5) in the presence (B - E) or absence (A) of 20 mM HMF (B), 40 mM furfural (C), 100 mM of formic acid (D), or 80 mM acetic acid (E), based on culture OD600nm reads in a microplate reader. Results of all panels of the figure are representative of at least two independent growth experiments

Table 3.1. Maximum specific growth rates of deletion mutants for selected MDR/MXR transporter encoding genes in the presence of fermentation inhibitors

Maximum specific growth rate (h^{-1}) of BY4741 and derived deletion mutants *pdr18Δ*, *pdr10*Δ, *pdr15*Δ, *yor1*Δ, *qdr1*Δ, *tpo2*Δ and *tpo3*Δ strains when exposed to sub-lethal concentrations of each fermentation inhibitor tested. Maximum specific growth rates were calculated from the growth curves presented in Figure 3.1.

	Maximum specific growth rate (h^{-1})							
	BY4741	*pdr18*Δ	*pdr10*Δ	*pdr15*Δ	*yor1*Δ	*qdr1*Δ	*tpo2*Δ	*tpo3*Δ
YPD40	0.094	0.099	0.088	0.090	0.091	0.088	0.079	0.079
YPD40 + 20 mM HMF	0.024	0.021	0.020	0.016	0.022	0.017	0.022	0.020
YPD40 + 40 mM Furfural	0.043	0.051	0.050	0.064	0.063	0.041	0.039	0.034
YPD40 + 100 mM Formic acid	0.071	0.062	0.041	0.077	0.062	0.050	0.054	0.061
YPD40 + 80 mM Acetic acid	0.048	0.049	0.038	0.039	0.035	0.036	0.031	0.019

The results show that in absence of supplementation with the inhibitors (Figure 3.1.A), the strains' maximum specific growth rate varies around 0.090 h^{-1}. The deletion mutant *pdr18*Δ has the highest maximum specific growth rate (0.099 h^{-1}) while *tpo2*Δ and *tpo3*Δ have the lowest (0.079 h^{-1}).

In the presence of 20 mM HMF (Figure 3.1.B), the BY4741 maximum specific growth rate was reduced from 0.094 h^{-1} to 0.024 h^{-1}, whereas no lag phase was detected. Supplementation with this inhibitor was found to result in lower final biomass than in unstressed conditions. Overall, the maximum specific growth rates and growth curves of deletion mutants, in presence of HMF, were found to be similar to the ones calculated for BY4741. The similarity between BY4741 behavior and that of the deletion mutants suggests that the impact of the deleted transporters on HMF resistance is probably negligible in the conditions tested.

When the medium was supplemented with 40 mM furfural (Figure 3.1.C), parental strain cell culture exhibited a lag-phase of approximately 19 h and a maximum specific growth rate of 0.043 h^{-1}. The derived deletion mutants displayed longer lag-phases ranging from 20 (*pdr18*Δ) to 34 h (*tpo2*Δ and *tpo3*Δ). In terms of maximum specific growth rate, deletion mutants for MFS transporters exhibited lower maximum specific growth rates than deletion mutants for ABC transporters. The deletion mutants *tpo2*Δ and *tpo3*Δ showed lower maximum specific growth rates and longer lag-phases, clearly differentiating from the other strains in terms of susceptibility to furfural.

Supplementation of the medium with 100 mM of formic acid (Figure 3.1.D) induced an extended lag-phase of about 10 h in BY4741 and reduced its maximum specific growth rate to 0.071 h^{-1}. The lag-phase of all deletion mutants, except *pdr18*Δ, is approximately 10 h. The mutant *pdr18*Δ exhibited a dramatically extended lag-phase of about 35 h.

Supplementation of the medium with 80 mM acetic acid (Figure 3.1.E) resulted in a decrease of BY4741 maximum specific growth rate from 0.094 h^{-1} (in the absence of added stress) to 0.048 h^{-1} and in a lag-phase of about 18 h. The *pdr18*Δ, *tpo2*Δ, and *tpo3*Δ deletion mutants exhibited very extended lag-phases of about 25, 20, and 30 h, respectively. The maximum specific growth rate of *tpo3*Δ was also dramatically reduced (0.019 h^{-1}) by supplementation of the medium with this concentration of acetic acid. All the other deletion mutants displayed a phenotype similar to that of BY4741.

As observed in Figure 3.2., expression of *TPO2* and *TPO3* from a centromeric plasmid (pGREG506) was found to rescue the acetic acid susceptibility exhibited by the *tpo2*Δ and *tpo3*Δ deletion mutants, respectively, enabling these mutant strains to display a susceptibility profile similar to the one exhibit-

ed by the parental strain harboring the empty vector pGREG506. The cloning of *PDR18* was not successful and therefore the experiment could not be performed using this gene.

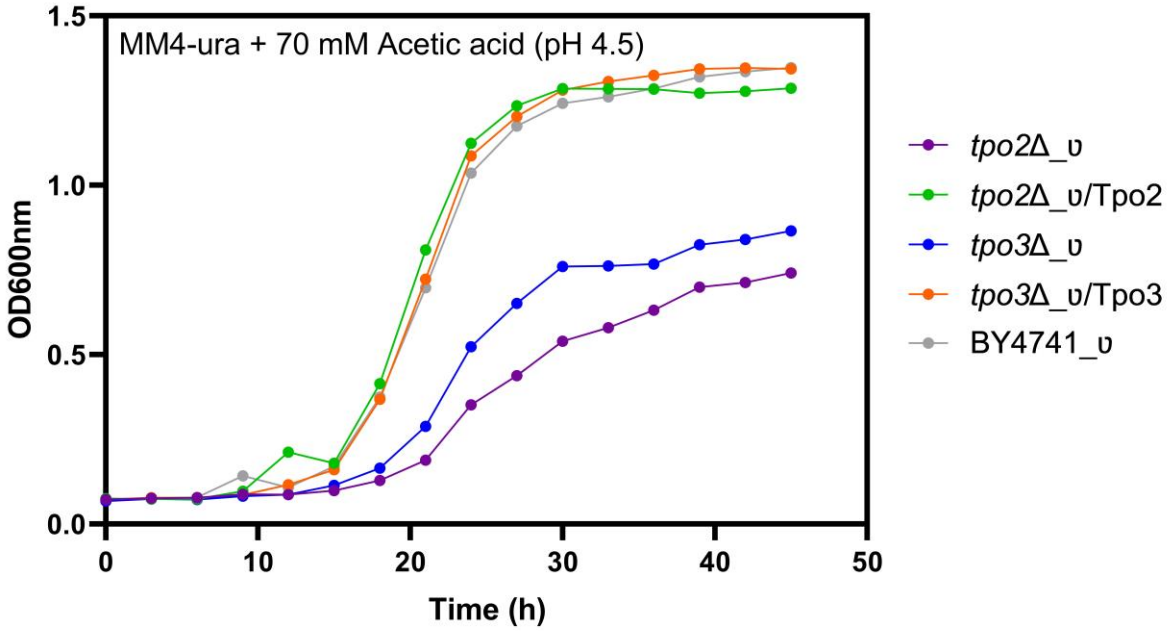

Figure 3.2. Phenotype rescue confirmation and schematic representation of strains used
Growth curve of BY4741_υ (*grey dots*), *tpo3*Δ _υ (*blue dots*), *tpo2*Δ _υ (*purple dots*), *tpo3*Δ_υ/Tpo3 (*orange dots*) and *tpo2*Δ_υ/Tpo2 (*green dots*) in 2 ml of MM4-ura supplemented with 70 mM acetic acid (pH 4.5). Curve plotted with culture OD600nm reads in a microplate reader.

3.1.2. Pdr18, Tpo2, and Tpo3 are required for yeast adaptation to weak acid stress in unadapted cell populations

In order to test the ability of the parental strain and derived deletion mutants (*pdr18*Δ, *tpo2*Δ and *tpo3*Δ) to adapt to weak acid-induced stresses, yeast cells, cultivated in YPD40 (pH 4.5) supplemented with 80 mM acetic acid (pH 4.5) or 100 mM formic acid (pH 4.5), were harvested in mid-exponential phase and reinoculated in fresh media with the same acetic/formic acid concentration.

This assay allows to compare variations in the maximum specific growth rate and duration of the lag-phase before and after adaptation to the weak acids tested, and the values of growth in absence of stress conditions. The growth curves are shown in Figure 3.3 and the maximum specific growth rates are plotted as bar graphs in Figure 3.4.

Figure 3.3. allows confirming that deletion of Pdr18, Tpo3, and at a smaller degree Tpo2 render the cell more susceptible to acetic acid, which is visible by the increase in the lag-phase of these deletion mutants in relation to BY4741 (Figure 3.3.A). Similarly, Figure 3.3.B corroborates that only *pdr18*Δ, and not *tpo3*Δ nor *tpo2*Δ, is highly susceptible to formic acid.

When each population resumed growth and attained the exponential phase, cells were sub-cultivated in fresh media supplemented with the same concentration for each stress condition. This adapted population showed no observable lag-phase or reduction in the maximum specific growth rate, when compared to the values observed in unstressed conditions. The maximum specific growth rates after sub-cultivation were also, in general, equal, or superior to the ones of BY4741 in unstressed conditions, except for *tpo2*Δ and *tpo3*Δ which exhibited the lowest maximum specific growth rates.

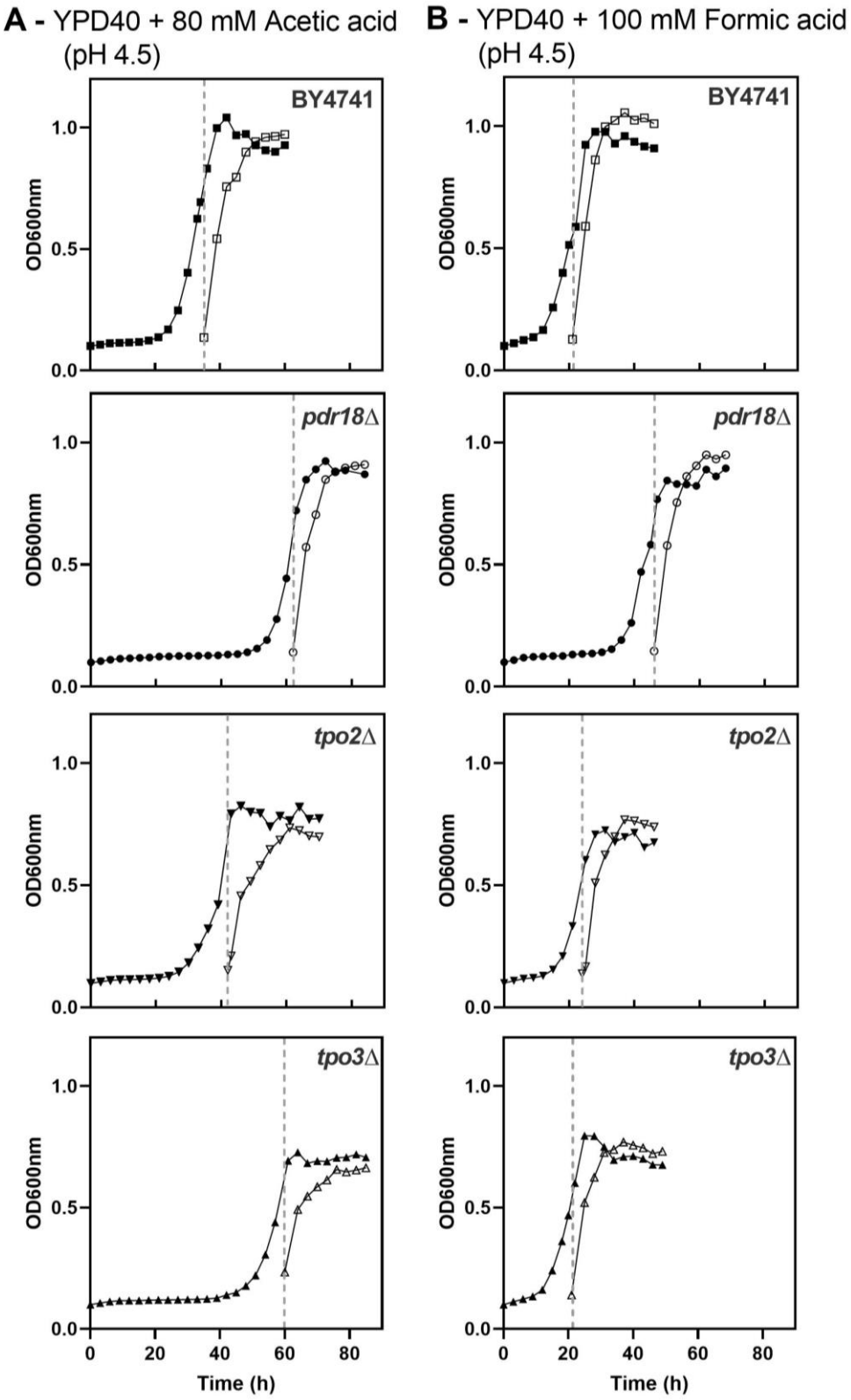

Figure 3.3. Sub-cultivation of acetic and formic acid-adapted cells shows yeast cell adaptation independently of the expression of the selected MDR/MXR transporters.

Growth curves of the parental strain (*squares*) and derived deletion mutants *pdr18*Δ (*circles*), *tpo2*Δ (*inverted triangles*) *tpo3*Δ (*triangles*) in YPD40 (pH 4.5) supplemented with 80 mM acetic acid (**A**) or 100 mM formic acid (**B**). Growth curves were plotted with culture OD600nm reads in a microplate reader. Black symbols represent growth before sub-cultivation and open symbols represent growth after sub-cultivation, in fresh YPD40 supplemented with the same acid concentrations, of cells harvested in the time-point marked with the dashed line. The results are representative of three independent experiments.

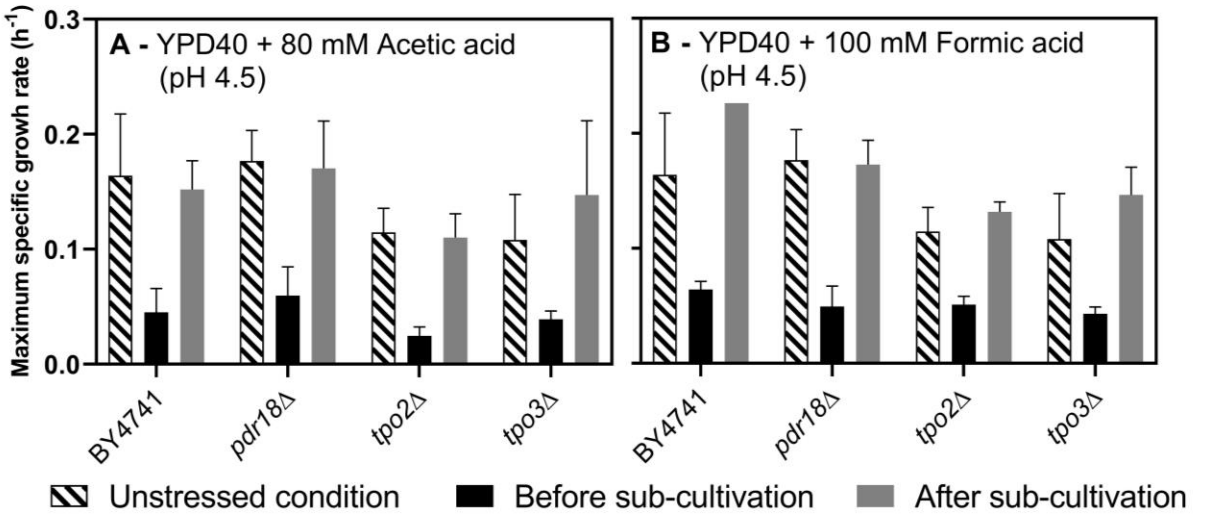

Figure 3.4. Variation of maximum specific growth rate in the sub-cultivation assay.
Strains *pdr18Δ*, *tpo2*Δ, *tpo3*Δ and BY4741 were grown in liquid YPD40 in presence (*black and grey bars*) or absence (*striped bar*) of 80 mM acetic acid (**A**) or 100 mM formic acid (**B**). Bars represent the maximum growth rate determined by observation of the growth curves plotted with culture OD600nm reads in microplate. *Striped bars* – maximum specific growth rate in absence of stress; *Black bars* - maximum specific growth rate before sub-cultivation; *Grey bars* – maximum specific growth rate after sub-cultivation of pre-adapted cells. The values showed are a mean value from three independent experiments and error bars represent standard deviation.

In conclusion, the data suggest that the tested transporters' role, is predominantly in the adaptation to weak acid stress in unadapted cells.

3.2. Screening and characterization of Tpo3 artificial variants

Tpo3, Tpo2 and Pdr18 were identified in the previous sections as the most relevant MDR/MXR transporters tested in yeast adaptation to sub-lethal concentrations of the fermentation inhibitors tested. However, cloning of *PDR18* was not successfully achieved and therefore this transporter was not selected for further studies. Due to time restrictions, it was not possible to test Tpo2 artificial variants.

We aimed at exploiting Tpo3 variants that could potentially improve yeast cells' tolerance against the fermentation inhibitors under study. For this, we generated mutagenized variants of the *TPO3* gene by amplifying it using error-prone PCR and cloned the resulting DNA fragments, as described in the Materials and Methods section.

3.2.1. Two artificial Tpo3 mutants identified as promising for improved yeast tolerance

Randomly mutated *TPO3* gene fragments (including CDS plus regulatory regions) were cloned into the pGREG506 vector and transformed into *tpo3*Δ, resulting in 294 colonies that were systematically tested for improved growth in MM4-ura supplemented with 80 mM acetic acid (pH 4.5), using a microplate reader (Appendix D). The strains *tpo3*Δ_υ and *tpo3*Δ_υ/Tpo3 were used as controls. Acetic acid was used for testing the artificially-generated Tpo3 variants, considering that this is the most noticeable phenotype described for this transporter.

Of the 294 mutants tested, 57 (about 19.4 %) seemed to at least rescue the susceptibility phenotype exhibited by *tpo3*Δ_υ in presence of acetic acid. The growth curves of these strains are depicted in Figure 3.5.

Analysis of Figure 3.5. led to the selection of the top-performing 25 mutants for more detailed growth experiments, including a higher acetic acid concentration (90 mM; Appendix D). Figure 3.6 shows the growth of five mutants chosen either because they are representative of the general behavior of the 25 mutants (Mut_56, Mut_87, Mut_292), or because they exhibited improved tolerance (Mut_44, Mut_47) to acetic acid when compared with the remaining mutants of the batch.

Figure 3.6. reveals that, although the expression of all the Tpo3 artificial variants rescued (at least partially) the phenotype exhibited by *tpo3*Δ_υ, only Mut_44, and Mut_47 exhibited improved tolerance to acetic acid when compared to *tpo3*Δ_υ/Tpo3.

3.2.2. Identification of eight putative causal mutations for acetic acid tolerance in *TPO3* gene sequence

The *TPO3* gene sequence of the selected strains Mut_44 and Mut_47 was investigated for the identification of putative causal mutations for improvement of yeast tolerance towards acetic acid stress. The *TPO3* gene sequence of Mut_71, one of the strains exhibiting a more susceptible phenotype towards acetic acid stress, was also sequenced.

Results showed that Mut_71 did not encode a *TPO3* gene. However, portions of the upstream and downstream regulatory regions were found to be present (Figure 3.7).

The mutations found in the *TPO3* gene sequence of Mut_44 and Mut_47 are listed in Table 3.2. Mut_44_*TPO3* exhibited five mutations, two in the upstream regulatory region and three in the *TPO3* CDS, while Mut_47_*TPO3* encoded three mutations in regulatory regions, two upstream and one downstream.

In Mut_44, from the three mutations identified in the CDS, one is synonym, having resulted in a change of codon but not of amino acid and two are non-synonyms (or missense). The mutation on position 2088 resulted in a change of A to C, and consequently in an isoleucine (Ile) to be substituted by a leucine (Leu) in position 351 of the protein. This region corresponds to the sixth transmembrane region of the protein, structurally composed by alpha-helices. Observation of the predicted structure of Tpo3 generated by AlphaFold [115] and available in UniProt [114] platform revealed that the side chain of the amino acid is facing the outside of the helix, inside of the plasma membrane. However, both amino acids have hydrophobic side chains and are structurally similar, thus the change is not expected to impact the structure and function of the protein. Finally, the mutation on position 2730 results in an alanine (Ala) – hydrophobic side chain - being switched for a serine (Ser) – polar, non-charged, side chain – in position 565 of the protein, which is included in the eleventh transmembrane region of the transporter. The AlphaFold predicted structure shows that the side chain of the amino acid is also facing the internal part of the membrane. In this case, a change in the polarity of the side chain could result in the deformation of the protein. Nevertheless, considering the environment surrounding the amino acid, is possible that the hydroxyl group of serine is stabilized by a hydrogen bond of nearby amino acids. Additionally, Figure 3.6. shows that Mut_44 revealed to be less susceptible to acetic acid than *tpo3*Δ_υ/Tpo3, which leads to the conclusion that the mutation, although potentially deleterious when conjugated with the other mutations, does not appear to negatively affect the protein.

YEASTRACT+ database [105] was used to investigate for putative transcription factor binding sites on the upstream regulatory region in which the mutations were identified. This analysis revealed that none of the mutations detected occurred in described TF binding sites. Moreover, analysis of the Mut_44 and Mut_47 promoter sequences did not result in new binding sites considering the consensus sequences documented for known transcription factor.

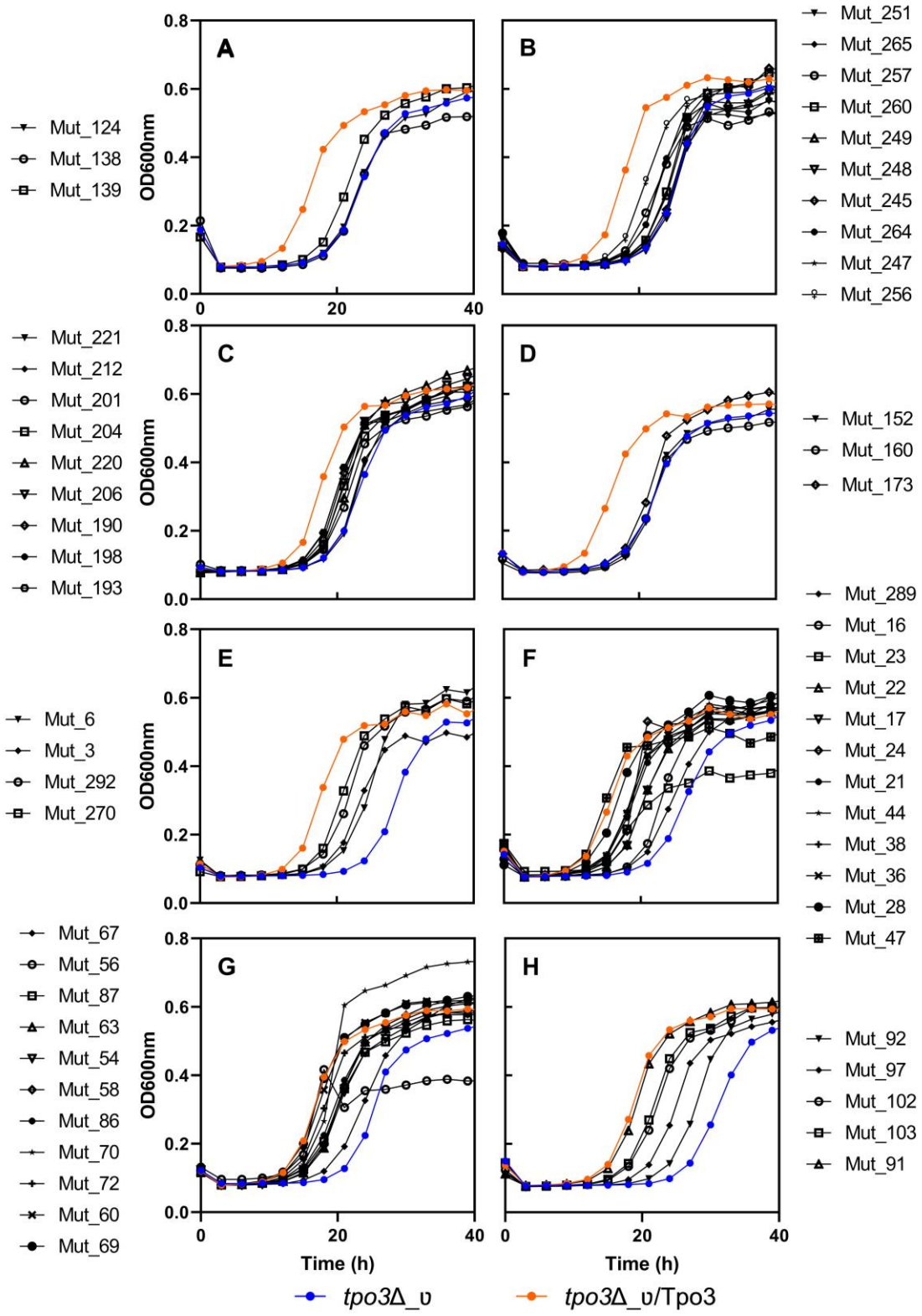

Figure 3.5. Susceptibility profile of Tpo3 artificial variants with phenotypic improvement regarding acetic acid resistance

Growth curves of strains harboring artificial variants of Tpo3 (Mut_1-294) and control strains - *tpo3Δ*_υ/Tpo3 (*orange*) and *tpo3*Δ_υ (*blue*) – in liquid MM4-ura supplemented with 80 mM acetic acid (pH 4.5). Assessment of growth was based on culture OD600nm reads in a microplate reader. Control strains were present in triplicate and the mean curve is shown. Each panel corresponds to the screening of different groups of strains each with independent controls.

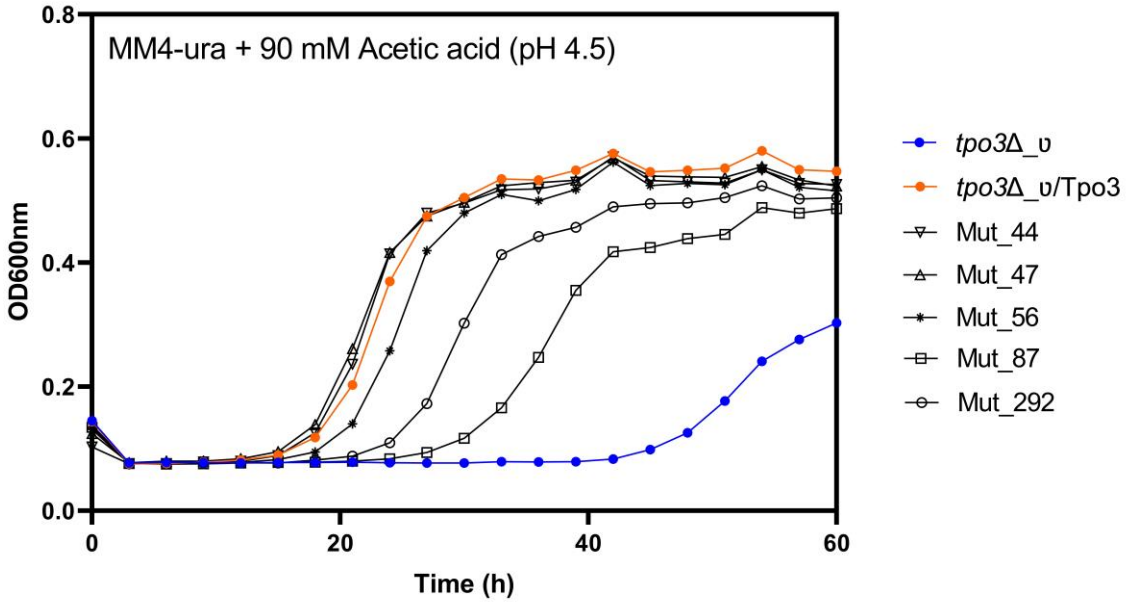

Figure 3.6. Mut_44 and Mut_47 are more tolerant to acetic acid-induced stress than the complemented deletion mutant, *tpo3*Δ_υ/Tpo3

Growth curves of five strains with artificial variants of Tpo3 - Mut_44 (*open inverted triangle*), Mut_47 (*open triangle*), Mut_56 (*asterisk*), Mut_87 (*open square*), Mut_292 (*open circle*) -and control strains - *tpo3Δ*_υ/Tpo3 (*orange*) and *tpo3*Δ_υ (*blue*) – in liquid MM4-ura with 90 mM acetic acid (pH 4.5) based on culture OD600nm reads in a microplate reader. Results are representative of two replicas.

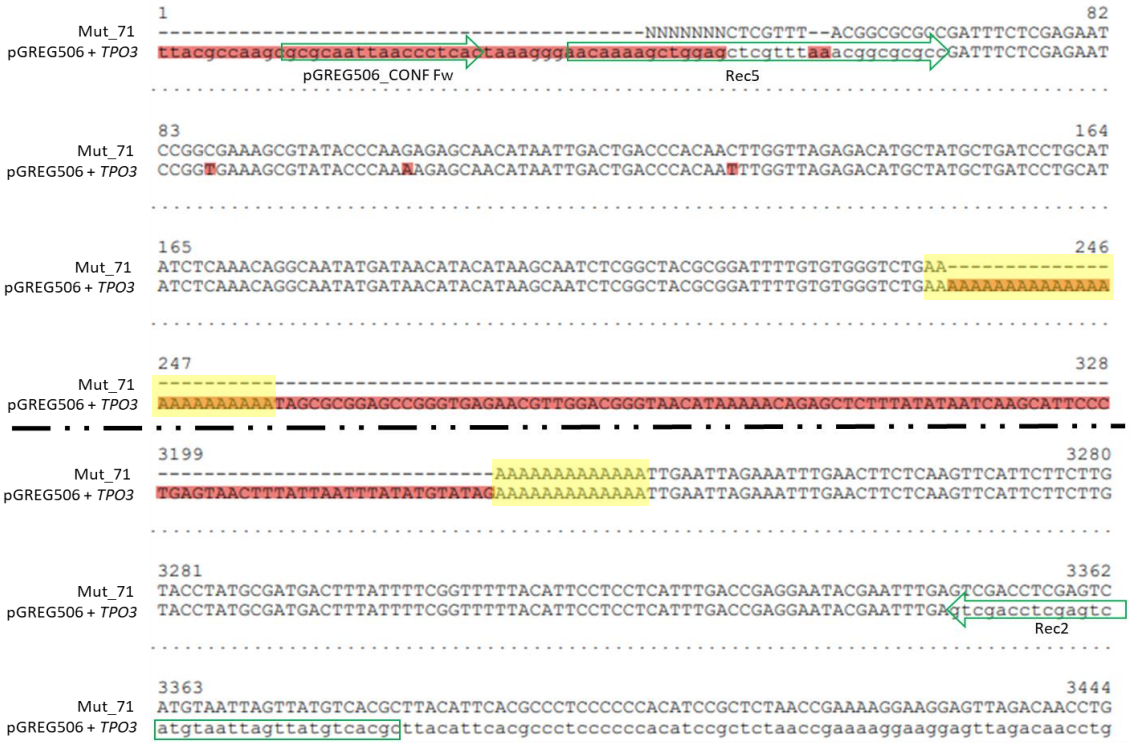

Figure 3.7. Partial sequence of the inserted fragment found in Mut_71

Plasmid of Mut_71 was sequenced with primer pGREG506_CONF forward. The inserted fragment is flanked by regions Rec5 and Rec2 that recombine with the ones found in the vector. Sequence of Mut_71 plasmid was aligned, using Benchling with sequence of pGREG506 in which region between Rec5 and Rec2 was substituted by *TPO3* plus upstream and downstream regions (sequence identified as pGREG506 +*TPO3*). Black line represents a jump in terms of position. Yellow highlight indicates the probable origin of the truncated fragment. Red highlights represent a non-perfect correspondence of nucleotides.

Table 3.2. List of mutations found in *TPO3* sequence of Mut_44 and Mut_47
Complete list of mutations found in *TPO3* sequence of inserted fragments in Mut_44 and Mut_47's plasmids. DNA position refers to position of the mutation in relation to the first nucleotide of the START codon. The gene has 1869 nucleotides. NA – Not Applicable, TM – transmembrane region, IN – intracellular region, CDS – coding sequence

	Position (DNA)	Region (DNA)	Nucleotide alteration	Position (Protein)	Original codon	Mutated codon	Amino acid alteration	Region (Protein)
Mut_44	-876	Upstream	AA/--	NA	NA	NA	NA	NA
	-272	Upstream	C/T	NA	NA	NA	NA	NA
	1050	*TPO3 CDS*	A/C	351	ATA	CTA	Ile/Leu	TM6
	1076	*TPO3 CDS*	C/T	359	TAC	TAT	Tyr/Tyr	IN4
	1692	*TPO3 CDS*	G/T	565	GCC	TCC	Ala/Ser	TM11
Mut_47	-876	Upstream	A/-	NA	NA	NA	NA	NA
	-167	Upstream	T/G	NA	NA	NA	NA	NA
	1985	Downstream	T/C	NA	NA	NA	NA	NA

In conclusion, the generation of random variants of *TPO3* by an error-prone polymerase, resulted in 294 strains, that were individually tested. From the 294, almost 20 % were able to resist acetic acid-induced stress better than *tpo3*Δ_υ and two had phenotypes that showed a decrease in susceptibility towards acetic acid when comparing to the phenotype exhibited by *tpo3*Δ_υ/Tpo3. Sequencing of those two mutants uncovered five mutations in regulatory regions with unknown impact and three mutations in the CDS, one with no predicted impact and two with theoretically neutral or detrimental impacts. As such this assay resulted in the identification of five mutations in the regulatory regions and three mutations in the CDS that can be investigated in future studies.

3.3. Exploitation of natural *TPO3* variants using bioinformatic analysis

The *TPO3* genomic sequences were extracted from the genome sequences of 1011 *S. cerevisiae* strains, together with the *TPO3* sequence variants, included in the available data published by Peter *et al.* 2018 [117]. The retrieved sequences comprised a total of 194 different variants. In other words, there were 194 positions in which at least one of the strains had a single nucleotide polymorphism (SNP) when using the *S. cerevisiae* S288c (BY4741) *TPO3* sequence as a reference.

The main goal was to assess the natural genomic variability of Tpo3 generated by different environmental pressures and identify strains with point mutations or structural variants that might be promising for further exploration towards the improvement of plasma membrane transporters' activity and increasing yeast cell resistance against fermentation inhibitors.

3.3.1. Analyzing *TPO3* sequence similarity among *S. cerevisiae* strains

A rooted tree was constructed using the *TPO3* gene sequences of variants encoded in the genomes of each of the 1011 *S. cerevisiae* strains. The used software, groups the strains with the same *TPO3* gene sequence, reducing the entropy and ensuring better visualization.

We verified that there was unexpected variability, with very few strains sharing exactly identical *TPO3* sequences. After examination of the sequences, we found that the variability was artificially generated since most sequences were full of ambiguity (N) signs. Moreover, the *TPO3* sequences of two strains (BPH_Wine_European and SACE_YCN_Other) were fully constituted by Ns and were hence removed from the analysis.

As a way to reduce the artificial variability, all the variants that were unclear and represented by N in more than 1 % of the dataset sequences were removed. This process led to the removal of 90 variants from the initial 194. Moreover, strains whose sequence contained more than 5 % of the 104 variants represented by N were also removed.

The original list of 1011 sequences composed of 194 different SNPs was reduced to 1002 sequences with 104 SNPs (79 synonym, 24 non-synonyms, and one nonsense). The cleaned dataset was used to group *TPO3* sequences according to similarity. Of the 1002 sequences, 75 were unique and the remaining were included in 68 groups. As such, the rooted tree was constructed considering 141 unique sequences, both from groups and individual isolate sequences. The reference strain's sequence is included in Group_A, which was used as the outgroup in the tree's construction.

In Appendix E is represented a schematic depiction of the clade's representation in each group. In general, Group_A represents the group of strains encoding the most common *TPO3* gene sequence (422 members). Although BY4741 is a laboratory strain with genomic manipulations and several growth restrictions, it encodes the most common *TPO3* gene sequence. Most of the members of Group_A belong to the more numerous clade, Wine European. The remaining groups are generally composed of strains from the same clade.

The final simplified dendrogram is represented in Figure 3.8. To facilitate analysis, each group was named after the clade of the majority (at least 75 %) of the members, and each strain/group's name was colored according to its attributed clade.

The general structure of the tree and clades' disposition matches other trees constructed in the laboratory for the variant analysis of other genes (unpublished data) from which can be inferred that the degree of variability is linked to the clades and not particularly with this specific gene.

Analysis of Figure 3.8. also shows that while some clades are clustered together, like French Guiana human, Wine European subclade 4, African Beer and Sake, other clades are dispersed along the dendrogram, like the clade Ale Beer. Nonetheless all clades are relatively clustered, indicating a close proximity between their members and leading to the hypothesis that the fixation of certain mutations is related to environments and selective pressures present - naturally or induced.

Additionally, the rooted tree allows to hypothesize that the strains encoding the *TPO3* genomic sequences displaying the largest dissimilarity from the reference strain are members of the clade French Guiana Human, the BX_Other group, at a smaller extent, the BAL_CHNI strain, and some members of the clade Mosaic Region 3. On the other hand, close to the outgroup are some strains of the clades Brazilian Bioethanol, Mixed Origin, and Ale Beer.

3.3.2. Analyzing naturally occurring SNPs in the *TPO3* gene

In order to circumvent the ambiguity found in the genome sequences, for each variant we assessed all the nucleotides (or ambiguity signs) present in all the strains. The 42 missense SNPs that are clearly represented by at least two distinct nucleotides and disregarding the amount and type of ambiguity signs present are listed in Table 3.3.

In Table 3.3. is also provided information regarding its position within the CDS, the position of the respective amino acid, the change in the codon and amino acid (three-letter code), the region of the protein where the affected amino acid is located (information retrieved from UniProt). In the prevalence column is indicated the number of strains that have the SNPs (nucleotide different from reference or an appropriate ambiguity code, Ns where not considered), whereas the distribution column is indicated the clade of the majority (>75%) of the strains that encode the SNP.

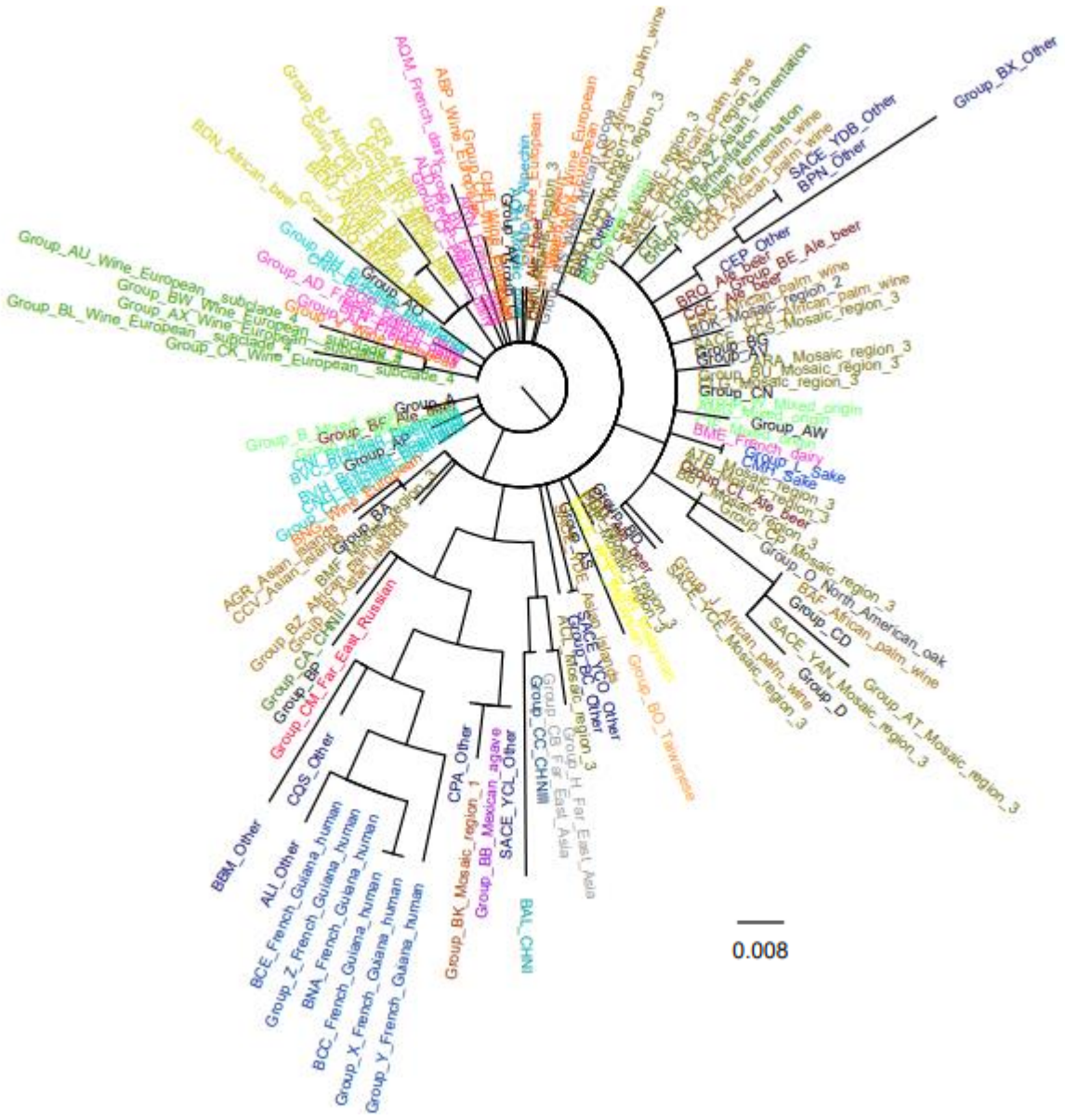

Figure 3.8. Rooted tree of the *TPO3* gene variants encoded across 1002 *S. cerevisiae* strains
Rooted tree constructed based on variant sequences (104 variants) of 141 distinct *TPO3* sequences (groups or isolates). The tree was obtained by rapid bootstrap with 100 bootstrapping iterations and corresponds to the best-scoring maximum likelihood tree. The outgroup is Group_A, group where BY4741 *TPO3* sequence is inserted. Groups and isolates are colored according to the corresponding predominant clade (≥75 % in case of groups).

Of the 42 identified SNPs, 57% are transitions (pyrimidine is substituted by a pyrimidine, or a purine by a purine) and 43% are transversions (pyrimidine changes to a purine or *vice-versa*).

Regarding the distribution of the SNPs in relation to the clade, only 10 SNPs were found to be randomly distributed within the clades, meaning that the remaining 32 SNPs are present, essentially, in a restricted population belonging to the same clade. This is an interesting finding because it reinstates the notion that the specific properties of each environment condition the evolution of these genomes. Also relevant is the fact that SNPs in positions 808 and 619 are present in most of the strains, but not

Table 3.3. List of *TPO3* gene SNPs encoded in the genome of 1011 *S. cerevisiae* strains.
The SNPs shown are represented by two nucleotides. From this group, the SNPs included in the construction of the rooted tree are highlighted in orange. The considered regions of the protein are IN – intracellular, OUT - extracellular and TM – transmembrane domain, each category is numbered independently and starting in the N-terminal. The IN1 region can be structured (S) or not structured (NS). In TM regions the amino acid sidechain can be facing the plasma membrane (PM) or the pore (P) formed by the protein. Prevalence refers to the number of strains have the SNP or an ambiguity code representing the two possibilities (Ns are not considered). The distribution indicates the clade with more than 50 % representativity in the universe of the strains considered in Prevalence. "None" means no clade representativity.

Position (CDS)	Position (Protein)	Region (Protein)	Codon alteration	Amino acid Alteration	Prevalence	Distribution (Clade)
64	22	IN1 (NS)	GAA/CAA	Glu/Gln	1	Other (100%)
68	23	IN1 (NS)	AGT/AAT	Ser/Asn	1	Wine European (100%)
101	34	IN1 (NS)	AGT/AAT	Ser/Asn	8	Alpechin (62.5%)
122	41	IN1 (NS)	GCT/GGT	Ala/Gly	1	CHNI (100%)
124	42	IN1 (NS)	CCT/TCT	Pro/Ser	1	Other (100%)
175	59	IN1 (S)	ATG/GTG	Met/Val	3	French dairy (100%)
184	62	IN1 (S)	AGC/GGC	Ser/Gly	42	Sake (100%)
185	62	IN1 (S)	AGC/AAC	Ser/Asn	1	Mosaic region 3 (100%)
190	64	IN1 (S)	AAA/CAA	Lys/Gln	6	Far East Asia (100%)
200	67	IN1 (S)	GTT/GCT	Val/Ala	1	Asian Islands (100%)
208	70	IN1 (S)	GTT/ATT	Val/Ile	3	Alpechin (100%)
298	100	IN1 (S)	CAT/TAT	His/Tyr	1	Wine European (100%)
338	113	IN1 (NS)	AGA/AAA	Arg/Lys	3	Wine European (100%)
385	129	IN1 (NS)	GCT/ACT	Ala/Thr	2	African palm wine (100%)
391	131	IN1 (NS)	GAA/AAA	Glu/Lys	1	Mosaic Region 3 (100%)
410	137	IN1 (NS)	AAG/ATG	Lys/Met	1	French dairy (100%)
448	150	IN1 (NS)	GCT/ACT	Ala/Thr	19	Brazilian bioethanol (79%)
523	175	IN1 (S)	GCA/TCA	Ala/Ser	12	None
529	177	IN1 (S)	ATC/GTC	Ile/Val	9	None
538	180	IN1 (S)	AGT/GGT	Ser/Gly	50	None
550	184	TM1 (PM)	CTA/GTA	Leu/Val	1	African beer (100%)
598	200	TM1 (P)	AGT/TGT	Ser/Cys	17	Wine European subclade 4 (88.2%)
599	200	TM1 (P)	AGT/ACT	Ser/Thr	19	None
619	207	OUT1	GAA/CAA	Glu/Gln	603	None
620	207	OUT1	GAA/GGA	Glu/Gly	1	Ale beer (100%)
629	210	OUT1	TAC/TTC	Tyr/Phe	38	None
808	270	OUT2	AGT/TGT	Ser/Cys	647	None
809	270	OUT2	AGT/ACT	Ser/Thr	1	Wine European (100%)
821	274	OUT2	TGT/TTT	Cys/Phe	2	Wine European (100%)
955	319	TM5 (PM)	GTT/ATT	Val/Ile	5	None
1174	392	IN4	GGT/AGT	Gly/Ser	11	None
1216	406	IN4	TCT/GCT	Ser/Ala	6	None
1325	442	OUT4	GGC/GAC	Gly/Asp	1	Mosaic region 3 (100%)
1366	456	TM8 (PM)	GTT/ATT	Val/Ile	1	Mosaic region 3 (100%)
1391	464	TM8 (PM)	GCC/GTC	Ala/Val	49	Mixed Origin (87,7%)
1423	475	IN5	CAA/AAA	Gln/Lys	13	Far East Asia (61.5%)
1435	479	IN5	CAG/GAG	Gln/Glu	4	Mosaic region 3 (75%)
1466	489	TM9	CTA/CCA	Leu/Pro	2	Mosaic region 2 (100%)
1672	558	TM11 (PM)	TTA/ATA	Leu/Ile	13	Wine European (84.6%)
1682	561	TM11 (PM)	GCC/GTC	Ala/Val	33	Wine European subclade 4 (78.8%)
1712	571	OUT6	ATT/ACT	Ile/Thr	19	African beer (100%)
1783	595	TM12 (PM)	GCT/TCT	Ala/Ser	32	French Guiana human (96.9%)

in the laboratory strain BY4741 used in most studies that aim at characterizing protein function.

The investigation of the impact of SNPs on the structure and function of a given protein is a laborious and time-consuming task that, among other techniques, may involve transporter activity assays and crystallization of native and mutated proteins which go far beyond the duration and scope of this MSc thesis. It is especially difficult to identify possible benefic nucleotide alterations since transporter proteins are highly structured and conserved biological systems resultant from millions of years of evolution. However, by making use of prediction software and the AlphaFold predicted structure for Tpo3 (available at Uniprot – code Q06451) some considerations were attempted, which constitute only informed speculations based on basic knowledge available on MFS transporters' structure and function.

The analysis of the SNPs listed in Table 3.3. using mutfunc [123] led to the identification of three SNPs on conserved regions and with possible impact on function. These SNPs are in positions 1466, 598, and 298 of the CDS. All three SNPs have SIFT scores (analogous to p-values) below 0.05 which may indicate that the SNPs are deleterious to function. This conclusion is only corroborated by the SNP on position 1466 of the CDS. This SNP results in the substitution of a leucine by a proline (amino acid 489) within a helical transmembrane domain, as observed using the AlphaFold prediction for Tpo3. The inclusion of prolines is known to disrupt alpha helices [124]. Hence this allele might cause perturbation in the ability of the protein to correctly insert into the plasma membrane and therefore might hamper the ability of the encoded protein to function as a transporter.

The SNP on position 598 also causes a non-synonymous substitution of amino acid 200, which is also predicted to be part of an alpha helix. Nonetheless, alteration of a serine by a cysteine, whose sidechains are facing the pore of the protein, does not necessarily result in a misfunction since both amino acids are hydrophilic and no other cysteine is present to form a possible sulfur bridge.

No prediction can be made on the effect of the SNP on amino acid 100 (position 298 of the CDS) since no information was found regarding the exchange of a histidine by a tyrosine in the first intracellular region of the protein (IN1).

The first intracellular region comprises about 180 amino acids and is likely involved in regulatory activities. In fact, according to Uniprot [114] (code Q06451), there are four sites for post-translational phosphorylation within this region (amino acids 55, 98, 101, 132). Although none of the identified SNPs affects these amino acids, there are cases, such as SNP on position 124, 298, 385, or 448 of the CDS, that result in the appearance of amino acids that can be phosphorylated (serine, threonine, and tyrosine), therefore being able to cause changes in the regulation of the transporter activity.

Additionally, IN1 is a sequence of amino acids that can be structured or not structured (intrinsically disordered regions). However, the quality scores of the AlphaFold prediction for this region are low. Nevertheless, it is important to note that even the not structured regions can have regulatory purposes. For example, interaction with a regulatory partner can induce conformational changes that result in a structured region. Thus, it is not possible to assume that mutations in not-structured regions cannot affect the structure/regulation of the protein [125].

Examples of apparently innocuous SNPs are those on positions 550, 955, and 1366. The SNPs affect amino acids located in transmembrane regions with sidechains facing the interior of the membrane. However, they result in the substitution of hydrophobic amino acid for another hydrophobic amino acid hence the structure and insertion of the protein remain apparently unchanged.

On the other hand, a change of charge of the amino acid's sidechain – for example in positions 1423, 1435, and 1325 – can implicate the generation of repulsive or attractive forces that might force conformational changes or can change the affinity of the protein with the substrate, but once again confirmation of these speculations required a more biochemical approach.

3.3.3. Evaluating the acetic acid tolerance of *S. cerevisiae* strains encoding different natural variants of TPO3

The acetic acid tolerance of 12 of the 1011 strains was assessed and compared to that of BY4741. These strains were chosen among the ones available at the laboratory. The purpose of this work was to assess the susceptibility towards acetic acid of strains harboring different natural variants of *TPO3* gene.

The selected strains were AAA, ABS, AGE, AHK, AKH, ASH, CAN, BRF, CLN, ANM, CNG, and CQF. Table 3.4. contains the clade, group, ploidy and SNPs of the selected strains.

Table 3.4. List of *S. cerevisiae* strains screened for acetic acid tolerance
Strain code, group, (as defined during dendrogram construction), ploidy of the strains and *TPO3* SNPs. The SNPs are identified by the respective position within CDS. SNPs with ambiguous codes instead of definitive nucleotide are showed with the respective code is in brackets. Repetitive SNPs were colored to facilitate visualization.

Code	Clade	Group	Ploidy	Mutations
AAA	Wine_European	A	1	619; 808
ABS	Wine_European	A	1	619; 808
AGE	Mixed_origin	AW	2	538
AHK	Wine_European__subclade_4	BL	1	1682
AKH	African_palm_wine	J	2	538; 619; 808
ANM	African_beer	BR	4	523 (M); 529 (Y); 1712
ASH	Ale_beer	BE	4	None
BRF	Wine_European	A	2	1672
CAN	Wine_European__subclade_4	A	2	598; 619; 808
CLN	Sake	L	2	184; 808
CNG	Brazilian_bioethanol	None	2	619
CQF	West_African_cocoa	AS	2	619 (Y); 629 (W); 808

Figure 3.9. shows the growth of the 13 strains in YPD in the absence or presence of supplementation with 90 mM acetic acid (pH 4.5). The name of the strain is concatenated to the respective clade.

Results show that this concentration of acetic acid induced a lag-phase of approximately 22 h for the BY4741 strain. AHK was found to be unable to grow at the acetic acid concentration tested, whereas ASH exhibited approximately the same latency phase but lower final biomass when compared to BY474. All the remaining strains tested were found to adapt and resume growth faster than BY4741, although only one strain attained significantly higher biomass (CLN). Remarkably, besides rendering approximately the same final biomass in the presence of 90 mM acetic acid as BY4741, strains AKH, ANM, CQF, and BRF show no observable latency phase.

Of the identified SNPs (listed in Table 3.4.), most of them have unknown implications since result in amino acids changes of the same class or are in extra(intra)cellular regions making its importance difficult to predict. The SNP on position 598 is the only one with an identified negative role on the pro-

tein since according to mutfunc occurs in a conserved area. In fact, the strain CAN is one of the strains with slower growth and less resistance in presence of acetic acid (Figure 3.9.B).

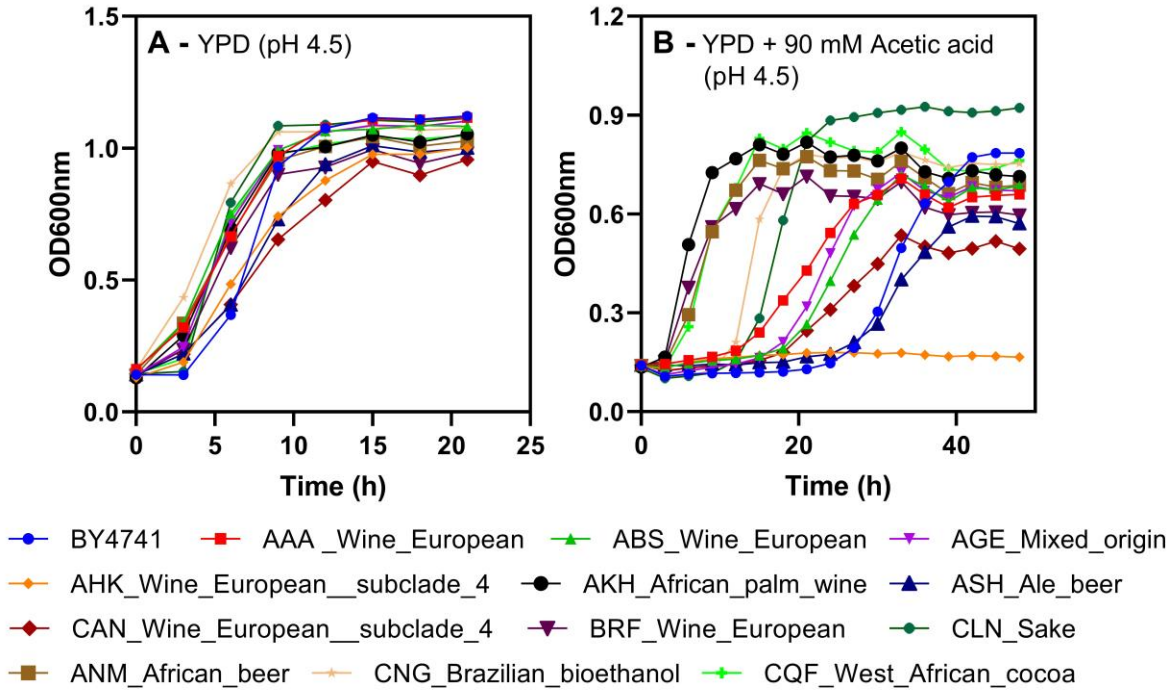

Figure 3.9. Behavior of selected strains in presence of acetic acid
Twelve strains, described in Peter, J. et al. (2018) [117] – AAA, ABS, AGE, AHK, AKH, ASH, CAN, BRF, CLN, ANM, CNG and CQF – and parental strain BY4741 were grown in microplate in YPD medium (pH 4.5) in absence (A) and presence of 90 mM acetic acid (**B**). The growth curves were plotted with OD600nm values.

On the other side, strains AKH and CQF are amongst the more tolerant strains towards acetic acid-induced stress and share two SNPs (at positions 619 and 808), with CLN also sharing the SNP at position 808. These SNPs occurred in the regions OUT1 (619) and OUT2 (808), both located on the exterior of the cell. SNP in position 619 results in the change of a negatively charged amino acid by a polar uncharged amino acid which may have implications in the interaction with the exterior layer of the membrane or with components of the medium, including possible substrates. On the other hand, SNP on position 808 results in a switch of two polar uncharged amino acids, and the new one has a Sulphur instead of oxygen, with the possibility of subsequent alteration in interaction with the membrane and medium components.

4. Discussion

Lignocellulosic materials are promising biorefinery feedstocks, potentiating a circular economy [21]. However, their use by yeasts requires physical and chemical treatments for the release of fermentable sugars, which results in formation of several growth inhibitors that hamper the ability of *S. cerevisiae*, to ferment the released sugars for the production of ethanol and other biofuels [27, 41, 42]. Among several mechanisms that are essential for yeast tolerance to these inhibitors is the expression of a variety of plasma membrane transporters involved in the simultaneous acquisition of resistance to a multitude of drug/xenobiotics (MDR/MXR), allowing *S. cerevisiae* to adapt and thrive [81, 106]. This work aimed at characterizing MDR/MXR transporters regarding their relevance in yeast tolerance to fermentation inhibitors and more specifically aimed at identifying natural and artificial gene variants that can contribute to an increased tolerance to lignocellulose-derived inhibitors.

Screenings such as the one performed in this work are crucial since they provide clues that are the basis for more detailed functional studies. This was the case of two studies focusing on characterizing the role of Pdr18 in yeast tolerance to acetic acid [50, 76]. The deletion mutant for *PDR18* gene was found to be susceptible to acetic acid stress in a genome-wide screening of *S. cerevisiae* deletion mutants performed in Sá-Correia's Lab (unpublished results). Later, our group proposed a role for Pdr18 in the ergosterol transport at the plasma membrane level, allowing the maintenance of the physiological ergosterol concentration of this membrane under acetic acid stress [50]. Furthermore, a more recent study described the influence of the expression of *PDR18* in yeast cell wall properties at the molecular and biophysical levels, under acetic acid-induced stress, revealing a crosstalk between plasma membrane ergosterol content and cell wall composition and organization [76]. Several other works describe Pdr18 as a major determinant for yeast tolerance to both chemical and physical stress, with special emphasis on weak acids [50, 72, 98, 126] which is consistent with the susceptibility phenotype observed in this work for the *pdr18*Δ strain in presence of sub-lethal concentrations of acetic and formic acids. Although this has been observed before by our group by spot assays, this is the first description of the impact of formic acid supplementation on the growth of *pdr18*Δ lag-phase and maximum specific growth rate [98]. Formic acid is a short-chain weak acid that is believed to directly penetrate the plasma membrane through simple diffusion [51]. As such, the role of Pdr18 in the modulation of plasma membrane permeability, is likely to be crucial in the cell's tolerance to formic acid, therefore justifying our results. It remains to be studied if the formic acid stress results in the depletion of cellular and plasma membrane concentrations of ergosterol, similarly to the effect of acetic acid [50]. This could be tested by quantification of ergosterol in both cell lysates and purified plasma membranes.

Tpo2 and Tpo3 are important determinants of yeast tolerance to acetic acid in *S. cerevisiae* [37, 60, 64] which was also verified in our results. The mechanism by which Tpo2 and Tpo3 exert a protective effect against acetic acid-induced stress has been hypothesized to be related to the removal of the acetate counterion from the cytosol, limiting its deleterious effects on the cell [60], namely the increase of intracellular pressure and the possible role in the inhibition of the central metabolism of the cell and in production of reactive oxygen species (ROS) [41, 127]. Tpo2 and Tpo3 were previously found to play a physiological role in the cell related to the transport of polyamines, which are positively charged molecules with more than one amino group [128, 129]. The polyamine content of the cell is tightly regulated since these molecules are essential for yeast growth and proliferation but in elevated concentrations becomes cytotoxic [129]. A closely related transporter, Tpo1, is also a polyamine transporter and, contrary to Tpo2 and Tpo3, has been extensively studied. Tpo1 performs the export of the polyamines spermidine, and spermine during oxidative stress [130]. This export is thought to act as a switch, inducing the expression of oxidative stress response proteins (e.g., Hsp70, Hsp90, and Sod1) and extending the cell cycle arrest induced by oxidative stress, probably as a way to decrease the prop-

agation of genetic mutations caused by ROS [130]. In our results, we verified that after adaptation to acetic acid, *tpo3*Δ exhibited a low maximum specific growth rate, when compared to the remaining deletion mutants and parental strain. It is possible that this phenomenon occurs because the extrusion of acetate from the cell is a role that is necessary even when occurs simultaneously to other resistance mechanisms, however *S. cerevisiae*'s adaptation to acetic acid stress includes a decrease of the permeability of the plasma membrane which results in less diffusional entry of acetic acid and less acetate in the cytosol [76]. Other explanation for the prolonged lag-phase and low maximum specific growth rate of *tpo3*Δ and *tpo2*Δ is a possible involvement in the oxidative stress response as described for Tpo1 [130] or even a role in counteracting, by polyamine export, the depolarization of the plasma membrane caused by acidification of the cytosol.

When setting up the experimental conditions for this work we aimed to select equivalent concentrations of the inhibitors tested. However, for acetic and formic acids, the results obtained for maximum specific growth of the parental strain indicate that this may not be the case. This may have occurred because the selection of the equivalent concentrations was based on the duration of the lag-phase in the parental strain cell population induced by the weak acids. We do not consider this to have influenced the conclusions of this work since our main goal was to compare the growth curves of the parental strain with the deletion mutants. In future works in which it is important to work with equivalent concentrations, a more viable approach would be to assess the cell viability of the populations either by assessing colony forming units (CFUs) or by using viability and vitality probes and measure these parameters by flow cytometry, such as described in other works [72, 131].

The mechanism of action of HMF in yeast is not fully understood but it is known to damage proteins and DNA, and decrease the protein and RNA content in yeast cells due to the inhibition of several enzymes involved in the central metabolism of the cell [38, 43]. Such information is consistent with the decrease of maximum specific growth rate observed for all the strains tested in medium supplemented with this inhibitor. The transcriptional regulation responsible for adaptation to HMF is dependent on the transcription factors (TF), Yap1, Pdr1, Pdr3, Rpn4, and Hsf1 [132]. According to information gathered from YEASTRACT+ database, the TFs Yap1, Pdr1, Pdr3 and Rpn4 are also responsible for activating the expression of all the transporters analyzed. However, in HFM-induced stress, the expression of these transporters was not significantly increased, except in the case of Yor1 and Pdr15 [132]. In our results, the effect of the deletion of *PDR15* gene in growth in presence of HMF is negligible. Overall, the work of Ma and Liu 2010 supports our results indicating that adaptation to HMF in *S. cerevisiae* seems not to be associated with the expression of the MDR/MXR transporters tested.

Furfural, like HMF, is a furan derivate and, similarly to HMF, causes inhibition of enzymes in the central metabolism of the cell, impairing cell growth and alcoholic fermentation, which supports the obtained results that show a decrease in maximum specific growth rate in all the strains, relatively with the maximum specific growth rate in unstressed conditions. Interestingly, in the presence of stress induced by furfural, deletion mutants for ABC transporters, especially *yor1*Δ and *pdr15*Δ, exhibited higher maximum specific growth rates than the parental strain and the deletion mutants for MFS transporters (mostly *tpo3*Δ). One hypothesis is that ATP hydrolysis performed by ABC transporters may constitute a disadvantage in the presence of stress for which these transporters do not confer tolerance [133]. Additionally, and to the best of our knowledge, we verified for the first time the involvement of Tpo2 and Tpo3 in tolerance to furfural. In this study we focused on tolerance to acetic acid, however in the future, is crucial to pursue this line of work, investigating the role of Tpo2 and Tpo3 on *S. cerevisiae*'s tolerance to furfural.

Despite the obvious importance of studying yeast tolerance to individual stressors, the conjugated effect of these inhibitors is also of major importance. In the present work, we aimed at assessing the role of these MDR/MXR transporters in the presence of a mix of inhibitors that mimicked the concentrations usually found in lignocellulosic hydrolysates. However, the results lacked consistency between both technical and biological replicates (results not shown). It is difficult to understand the reasons behind the discrepancy in the obtained results. One hypothesis is that the inhibitors reacted with each other over time, changing the mix's composition. In fact, over time, the mix darkened, and was possible to observe the formation of a precipitate. In future studies, we will test the supplementation of each inhibitor directly from their stock solution to the culture media.

In this work, we aimed to explore artificially-generated variants of the *TPO3* gene to find mutations with a favorable impact on yeast resistance to the inhibitors tested. A very high frequency (approximately 80%) of *TPO3* artificial variants exhibited very low tolerance to acetic acid stress, in most cases more susceptible than the deletion mutant. There are some possible reasons for this phenomenon. First of all, it is possible that the yeast transformation with the lithium acetate method resulted in DNA damage and/or induced genomic alterations caused by unintended recombination of the plasmid or fragment with semi-homologous regions of the genome [134]. Additionally, it is also possible that the randomly generated *TPO3* variants resulted in deleterious mutations in the Tpo3 protein. This could lead the encoded transporter to lose specificity, creating open pores on the membrane, importing substances detrimental to growth, or effluxing valuable cell metabolites.

In our work, of the 294 strains tested, two displayed slightly higher resistance to acetic acid stress than *tpo3*Δ_υ/Tpo3. These strains exhibited five mutations on the 3' and 5' regulatory regions and three on the coding sequence (CDS) of *TPO3*.

The upstream regulatory region includes 5'-UTR (Untranslated Region), the sequences for polymerase attachment and initiation of transcription, as well as possible binding sites for transcription factors (activators or inhibitors). The 5'-UTR contains sequences for attracting and binding the ribosome, for binding regulatory proteins involved in translational regulation, might contain sequence that by forming secondary structures, regulate translation, among others. On the other end, the downstream region contains signals for termination of transcription, and the 3'-UTR sequence, responsible for the half-life of the transcript [135]. However, these sequences and regions are highly variable, difficult to identify and it is hard to understand without further studies if a single nucleotide alteration affects its function or not. One important sequence in *S. cerevisiae*, and eucaryotes in general, is the Kozak sequence, which is located near the START codon (ATG) and is involved in translational initiation. The optimal consensus Kozak sequence in *S. cerevisiae* is 5'-(A/T)A(A/C)A(A/C)A**ATG**-3'. In the case of *TPO3*, this sequence is ACAGAA**ATG,** therefore attesting to the variability of such sequences [136]. Nevertheless, none of the identified mutations are located in this region.

Additionally, the possibility of the mutation being located within binding sites of transcription factors was investigated. For example, it is known that the binding site for transcription factor Haa1 is located within positions -790 and -590 bp (upstream of the start codon) and binds to the consensus sequence 5'-(G/C)(A/C)GG(G/C)G-3'[110]. However, no mutations were found in this region. In fact, according to information gathered in YEASTRACT+ none of the mutations occurred in documented TF binding sites. It is possible that the mutation might cause alterations in the secondary structure of the DNA, affecting TF and general protein binding [137]. Another possibility is that the altered nucleotides lead to a sequence that is a binding site for undocumented TFs or for TFs for which no consensus sequences has yet been documented. The complexity of the transcriptional and translational regulation makes it impossible to truly attest for the actual impact of these mutations, without further studies.

Considering that these two strains exhibited increased tolerance towards acetic acid stress, each individual single nucleotide polymorphism (SNP) should be tested by directed mutagenesis in BY4741 *TPO3* gene sequence to search for the causative mutation(s) that lead to improved phenotype.

One strain harboring a *TPO3* gene variant that led to a detrimental phenotype (Mut_71) was also sequenced. Analysis of the sequenced fragment led us to believe that during PCR amplification the polymerase staggered in repetitive sequences of the same nucleotide. Since the repetitive nucleotide was the same in both regions (A - adenine), the fragments paired – reverse reaction inserted T (thymine) – and amplification proceeded with a smaller fragment. These results justify the high susceptibility of Mut_71 towards acetic acid stress, since the *TPO3* gene is not present. However, this does not explain why the strain is more susceptible to acetic acid stress than *tpo3*Δ_υ. It is possible that the inserted upstream regulatory region (promoter sequence), is recruiting transcription machinery and therefore wasting resources. A more likely hypothesis is that during the transformation process the cell suffered unrepairable DNA damage.

A similar mutagenesis protocol was followed for Tpo1, resulting in the identification of two strains expressing a Tpo1 variant with increased activity, leading to higher resistance against 10-carbon fatty acids [138]. Based on that work, we identified some modifications that can be done to the experimental setup of our work in order to increase the percentage of strains of interest and decrease the amount of work on individually testing each strain in future works. Firstly, purification of the fragments amplified by the error-prone polymerase prior to transformation could decrease the interference of PCR reagents in the transformation/recombination. Especially, the removal of excess primers could reduce unwanted recombination events. On the other hand, and more importantly, Zhu *et al*. performed several competition assays with the intent to enrich the population with the best-performing strains and therefore reducing the number of strains to test. In this case, transformants could have been cultivated in sub-sequentially higher concentrations of acetic acid, plated in the presence of selective pressure, and picked for testing.

The natural variation of the *TPO3* gene sequence was also a focus of this study. The *TPO3* gene sequence of 1011 *S. cerevisiae* strains was retrieved from Peter *et al.* (2018) [117]. The sequences compiled were used to construct a rooted tree representing the variability of the *TPO3* within the 1011 strains isolated from different environments. Also each SNP was analyzed to identify the possible impact of the newly found allele on the Tpo3 structure and function, and for correlating the different SNPs with the isolation environment. This analysis had previously been performed at Sá-Correia's Lab for other transporters and transcriptions factors related to acetic acid stress with interesting results (unpublished results). However, the majority of the strains' *TPO3* sequence indicated ambiguity signs (mostly Ns) instead of specific nucleotides. Such abundance of ambiguity led to the removal of several strains and SNPs from our analysis and negatively impacted the actual representativeness of the results.

The evolution of *S. cerevisiae* and whether is conditioned by adaptation and natural selection or genetic drift of populations has been the focus of other works [13, 15, 139]. These studies indicate that the variability of wild strains is largely dependent on geographic location and has a large genetic drift component, whereas domesticated strains appear to have evolved in an environmental directed manner [13, 139]. The strains used in our analysis are both from domesticated and wild environments hence evolutionary conclusions cannot be made. Nevertheless, the rooted tree, constructed using BY4741 *TPO3* sequence as an outgroup, revealed an interesting agglomeration of the strains from the same clade, as also described in some genome-wide studies [117, 140]. Also, the majority of the SNPs identified were found in a restricted group of strains, generally, from the same clade. Such agglomeration allows that in future works, besides evaluating the impact of all individual SNPs on Tpo3 by directed

mutagenesis of BY4741 *TPO3* sequence, it will also be possible to select for SNPs of greatest interest by identification of the environments in which the stressor in study is present.

In the present work, we tested the growth of 12 *S. cerevisiae* strains included in the 1011 strains dataset, in the presence of a concentration of acetic acid that is sub-lethal for BY4741. The strains AKH, CQF, ANM and BRF were identified as having a high tolerance to acetic acid. All of these strains except BRF were isolated from African foods and beverages. These are highly acidic environments, due to the action of lactic and acetic acid bacteria [141]. The *TPO3* gene variants encoded in these strains are promising tools to be tested in the same genetic background, in the presence of acetic acid stress, similarly to the experiments performed for the artificially generated variants. Nonetheless, it is important to notice that acetic acid tolerance is not exclusively linked to *TPO3* SNPs but is instead a result of a multifactorial process of adaptation. Also, in the future, would be beneficial to select and study other strains based on the SNPs found in *TPO3*, since due to time restrictions the strains used in the present work were the ones already available at the lab, selected in the scope of other studies. Despite its limitations, the results obtained herein can be used to select strains from which the *TPO3* gene could be amplified and more thoroughly studied.

5. Conclusion

The production of bioethanol and other biofuels from lignocellulosic feedstocks by *S. cerevisiae* (or other microorganisms) is a process limited by the presence of inhibitors generated during pre-treatment of biomass. The identification of genetic determinants for yeast tolerance to these inhibitors and the search for natural and artificial gene variants conferring a superior performance in *S. cerevisiae* is of paramount importance. Altogether, the results presented in this thesis contribute to the overall goal of identifying MDR/MXR transporters that confer *S. cerevisiae* with tolerance to the most common inhibitors present at lignocellulosic hydrolysates. It also contributes to the identification of eight artificially generated mutations in the *TPO3* gene sequence that may be causative for acetic acid tolerance. Also, the search for natural gene variants across a very wide range of *S. cerevisiae* strains isolated from harsh environment conditions can provide us with genetic tools for yeast engineering. In this regard, the present work encountered time-consuming technical challenges. However, the analysis performed can guide the selection of yeast species for further studies. Overall, the search for both artificial and natural gene variants seems to be a promising strategy for guiding the rational genetic improvement of industrial yeast strains' robustness.

6. References

1. Stewart GG. *Saccharomyces*: *Saccharomyces cerevisiae*. In: Batt CA, Tortorello M Lou, editors. Encyclopedia of Food Microbiology. Second Edi. Elsevier; 2014. p. 309–15.

2. Liti G. The Natural History of Model Organisms: The fascinating and secret wild life of the budding yeast S. cerevisiae. Elife. 2015;4:1–9.

3. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, et al. Life with 6000 Genes. Science. 1996;274 October:546–67.

4. Pfeiffer T, Morley A. An evolutionary perspective on the Crabtree effect. Front Mol Biosci. 2014;1 OCT:1–6.

5. Escalera-Fanjul X, Quezada H, Riego-Ruiz L, González A. Whole-Genome Duplication and Yeast's Fruitful Way of Life. Trends Genet. 2019;35:42–54.

6. Feng B, Lin Y, Zhou L, Guo Y, Friedman R, Xia R, Hu F. Reconstructing Yeasts Phylogenies and Ancestors from Whole Genome Data. Nat Sci Reports. 2017;7:1–12.

7. Duina AA, Miller ME, Keeney JB. Budding yeast for budding geneticists: A primer on the *Saccharomyces cerevisiae* model system. Genetics. 2014;197:33–48.

8. Gasmi N, Jacques PE, Klimova N, Guo X, Ricciardi A, Robert F, Turcotte B. The switch from fermentation to respiration in *Saccharomyces cerevisiae* is regulated by the Ert1 transcriptional activator/repressor. Genetics. 2014;198:547–60.

9. Leupold S, Hubmann G, Litsios A, Meinema AC, Takhaveev V, Papagiannakis A, Niebel B, Janssens G, Siegel D, Heinemann M. *Saccharomyces cerevisiae* goes through distinct metabolic phases during its replicative lifespan. Elife. 2019;8:1–19.

10. Deken RH. The Crabtree Effect: A Regulatory System in Yeast. J Gen Microbiol. 1966;44:149–56.

11. Thomson JM, Gaucher EA, Burgan MF, De Kee DW, Li T, Aris JP, Benner SA. Resurrecting ancestral alcohol dehydrogenases from yeast. Nat Genet. 2005;37:630–5.

12. Pfeiffer T, Schuster S, Bonhoeffer S. Cooperation and competition in the evolution of ATP-producing pathways. Science. 2001;292:504–7.

13. Jouhten P, Ponomarova O, Gonzalez R, Patil KR. *Saccharomyces cerevisiae* metabolism in ecological context. FEMS Yeast Res. 2016;16:1–8.

14. Belda I, Ruiz J, Santos A, Van Wyk N, Pretorius IS. *Saccharomyces cerevisiae*. Trends Genet. 2019;35:956–7.

15. Duan SF, Han PJ, Wang QM, Liu WQ, Shi JY, Li K, Zhang XL, Bai FY. The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. Nat Commun. 2018;9:1–13.

16. Parapouli M, Vasileiadis A, Afendra AS, Hatziloukas E. *Saccharomyces cerevisiae* and its industrial applications. 2020; 6:1-31.

17. Johnson EA, Echavarri-Erasun C. Yeast biotechnology. In: Kurtzman CP, Fell JW, Boekhout T, editors. The Yeasts, a Taxonomic Study. 5th edition. Elsevier B.V.; 2011. p. 21–44.

18. Assi-Clair BJ, Koné MK, Kouamé K, Lahon MC, Berthiot L, Durand N, Lebrun M, Julien-Ortiz A, Maraval I, Boulanger R, Guéhi TS. Effect of aroma potential of *Saccharomyces cerevisiae* fermentation on the volatile profile of raw cocoa and sensory attributes of chocolate produced thereof.

Eur Food Res Technol. 2019;245:1459–71.

19. Moreira I, Costa J, Vilela L, Lima N, Santos C, Schwan R. Influence of *S. cerevisiae* and *P. kluyveri* as starters on chocolate flavour. J Sci Food Agric. 2021;101:4409–19.

20. Denby CM, Li RA, Vu VT, Costello Z, Lin W, Chan LJG, Williams J, Donaldson B, Bamforth CW, Petzold CJ, Scheller H V., et al. Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. Nat Commun. 2018;9:1–10.

21. Cherubini F. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. Energy Convers Manag. 2010;51:1412–21.

22. Guo M, Song W, Buhain J. Bioenergy and biofuels: History, status, and perspective. Renew Sustain Energy Rev. 2015;42:712–25.

23. Balat M, Balat H. Recent trends in global production and utilization of bio-ethanol fuel. Appl Energy. 2009;86:2273–82.

24. Bertacchi S, Jayaprakash P, Morrissey JP, Branduardi P. Interdependence between lignocellulosic biomasses, enzymatic hydrolysis and yeast cell factories in biorefineries. Microb Biotechnol. 2021;15:1–11.

25. Ree R van. IEA Bioenergy 29th update. In: IEA BIOENERGY Task42 BIOREFINING. 2014. p. 1–66.

26. Fernando S, Adhikari S, Chandrapal C, Murali N. Biorefineries: Current status, challenges, and future direction. Energy and Fuels. 2006;20:1727–37.

27. Baruah J, Nath BK, Sharma R, Kumar S, Deka RC, Baruah DC, Kalita E. Recent trends in the pretreatment of lignocellulosic biomass for value-added products. Front Energy Res. 2018;6 DEC:1–19.

28. Taherzadeh MJ, Karimi K. Enzyme-based hydrolysis processes for ethanol from lignocellulosic materials: A review. 2007;2:707-738.

29. Sarkar N, Ghosh SK, Bannerjee S, Aikat K. Bioethanol production from agricultural wastes: An overview. Renew Energy. 2012;37:19–27.

30. Demeke MM, Dietz H, Li Y, Foulquié-Moreno MR, Mutturi S, Deprez S, Den Abt T, Bonini BM, Liden G, Dumortier F, Verplaetse A, et al. Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. Biotechnol Biofuels. 2013;6:1–24.

31. Desfougeres T, Pignede G, Techel J. Pentose-fermenting strain with optimized propagation. 2019;1–38.

32. Nijland JG, Driessen AJM. Engineering of Pentose Transport in *Saccharomyces cerevisiae* for Biotechnological Applications. Front Bioeng Biotechnol. 2020;7 January:1–13.

33. Verhoeven MD, De Valk SC, Daran JMG, Van Maris AJA, Pronk JT. Fermentation of glucose-xylose-Arabinose mixtures by a synthetic consortium of single-sugar-fermenting *Saccharomyces cerevisiae* strains. FEMS Yeast Res. 2018;18:1–12.

34. Radecka D, Mukherjee V, Mateo RQ, Stojiljkovic M, Foulquié-Moreno MR, Thevelein JM. Looking beyond *Saccharomyces*: The potential of non-conventional yeast species for desirable traits in bioethanol fermentation. FEMS Yeast Res. 2015;15:1–13.

35. Mukherjee V, Radecka D, Aerts G, Verstrepen KJ, Lievens B, Thevelein JM. Phenotypic

landscape of non-conventional yeast species for different stress tolerance traits desirable in bioethanol fermentation. Biotechnol Biofuels. 2017;10:1–19.

36. Martins LC, Monteiro CC, Semedo PM, Sá-Correia I. Valorisation of pectin-rich agro-industrial residues by yeasts: potential and challenges. Appl Microbiol Biotechnol. 2020;104:6527–47.

37. Palma M, Guerreiro JF, Sá-Correia I. Adaptive response and tolerance to acetic acid in *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*: A physiological genomics perspective. Front Microbiol. 2018;9 FEB:1–16.

38. Parawira W, Tekere M. Biotechnological strategies to overcome inhibitors in lignocellulose hydrolysates for ethanol production: Review. Crit Rev Biotechnol. 2011;31:20–31.

39. Du C, Li Y, Xiang R, Yuan W. Formate Dehydrogenase Improves the Resistance to Formic Acid and Acetic Acid Simultaneously in *Saccharomyces cerevisiae*. Int J Mol Sci. 2022;23:1–14.

40. Oshoma CE, Greetham D, Louis EJ, Smart KA, Phister TG, Powell C, Du C. Screening of non-*Saccharomyces cerevisiae* strains for tolerance to formic acid in bioethanol fermentation. PLoS One. 2015;10:1–17.

41. Cunha JT, Romaní A, Costa CE, Sá-Correia I, Domingues L. Molecular and physiological basis of *Saccharomyces cerevisiae* tolerance to adverse lignocellulose-based process conditions. Appl Microbiol Biotechnol. 2019;103:159–75.

42. Li B, Liu N, Zhao X. Response mechanisms of *Saccharomyces cerevisiae* to the stress factors present in lignocellulose hydrolysate and strategies for constructing robust strains. Biotechnol Biofuels Bioprod. 2022;15:1–20.

43. Cao D, Tu M, Xie R, Li J, Wu Y, Adhikari S. Inhibitory Activity of Carbonyl Compounds on Alcoholic Fermentation by *Saccharomyces cerevisiae*. J Agric Food Chem. 2014;62:918–26.

44. Modig T, Lidén G, Taherzadeh MJ. Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase. Biochem J. 2002;363:769–76.

45. Allen SA, Clark W, Mccaffery JM, Cai Z, Lanctot A, Slininger PJ, Liu ZL, Gorsich SW. Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae*. Biotechnol Biofuels. 2010;3:1–10.

46. Mira NP, Palma M, Guerreiro JF, Sá-Correia I. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. Microb Cell Fact. 2010;9:1–13.

47. Piper P, Calderon CO, Hatzixanthis K. Weak acid adaptation : the stress response that confers yeasts with resistance to organic acid food preservatives. Microbiology. 2001;147:2635–42.

48. Sousa MJ, Ludovico P, Rodrigues F, Leão C, Côrte-Real M. Stress and Cell Death in Yeast Induced by Acetic Acid. In: Bubulya P, editor. Cell Metabolism – Cell Homeostasis and Stress Response. 2012. p. 73–102.

49. Guaragnella N, Bettiga M. Acetic acid stress in budding yeast: From molecular mechanisms to applications. Yeast. 2021;38:391–400.

50. Godinho CP, Prata CS, Pinto SN, Cardoso C, Bandarra NM, Fernandes F, Sá-Correia I. Pdr18 is involved in yeast response to acetic acid stress counteracting the decrease of plasma membrane ergosterol content and order. Sci Rep. 2018;8:1–13.

51. Lee S-E, Park B-S, Yoon J-J. Proteomic Evaluation of Cellular Responses of *Saccharomyces cerevisiae* to Formic Acid Stress. Mycobiology. 2010;38:302–9.

52. Fernandes AR, Durão PJ, Santos PM, Sá-Correia I. Activation and significance of vacuolar H+-ATPase in *Saccharomyces cerevisiae* adaptation and resistance to the herbicide 2,4-dichlorophenoxyacetic acid. Biochem Biophys Res Commun. 2003;312:1317–24.

53. Ullah A, Chandrasekaran G, Brul S, Smits GJ. Yeast adaptation to weak acids prevents futile energy expenditure. Front Microbiol. 2013;4 JUN:1–10.

54. Ribeiro RA, Vitorino M V., Godinho CP, Bourbon-Melo N, Robalo TT, Fernandes F, Rodrigues MS, Sá-Correia I. Yeast adaptive response to acetic acid stress involves structural alterations and increased stiffness of the cell wall. Sci Rep. 2021;11:1–9.

55. Chen Y, Siewers V, Nielsen J. Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in *Saccharomyces cerevisiae*. PLoS One. 2012;7:1–9.

56. Petersson A, Almeida JRM, Modig T, Karhumaa K, Hahn-Hägerdal B, Gorwa-Grauslund MF, Lidén G. A 5-hydroxymethyl furfural reducing enzyme encoded by the *Saccharomyces cerevisiae* ADH6 gene conveys HMF tolerance. Yeast. 2006;23:455–64.

57. Jung YH, Kim S, Yang J, Seo JH, Kim KH. Intracellular metabolite profiling of *Saccharomyces cerevisiae* evolved under furfural. Microb Biotechnol. 2017;10:395–404.

58. Herrero E, Ros J, Bellí G, Cabiscol E. Redox control and oxidative stress in yeast cells. Biochim Biophys Acta. 2008;1780:1217–35.

59. Zeng L, Huang J, Feng P, Zhao X, Si Z, Long X, Cheng Q, Yi Y. Transcriptomics Analysis of Formic Acid Stress Response in *Saccharomyces cerevisiae*. World J Microbiol Biotechnol. 2022;38:1–35.

60. Mira NP, Teixeira MC, Sá-Correia I. Adaptive Response and Tolerance to Weak Acids in *Saccharomyces cerevisiae* : A Genome-Wide View. Omi A J Integr Biol. 2010;14:525–40.

61. Guo Z, Olsson L. Physiological response of *Saccharomyces cerevisiae* to weak acids present in lignocellulosic hydrolysate. FEMS Yeast Res. 2014;14:1234–48.

62. Henriques SF, Mira NP, Sá-Correia I. Genome-wide search for candidate genes for yeast robustness improvement against formic acid reveals novel susceptibility (Trk1 and positive regulators) and resistance (Haa1-regulon) determinants. Biotechnol Biofuels. 2017;10:1–11.

63. Crawford RA, Pavitt GD. Translational regulation in response to stress in *Saccharomyces cerevisiae.* Yeast. 2019;36:5–21.

64. Mira NP, Becker JD, Sá-Correia I. Genomic expression program involving the Haa1p-regulon in *Saccharomyces cerevisiae* response to acetic acid. Omi A J Integr Biol. 2010;14:587–601.

65. Stewart GG. The Structure and Function of the Yeast Cell Wall, Plasma Membrane and Periplasm. In: Stewart GG, editor. Brewing and Distilling Yeasts. Springer International Publishing; 2017. p. 55–75.

66. Van Der Rest ME, Kamminga AH, Nakano A, Anraku Y, Poolman B, Konings WN. The plasma membrane of *Saccharomyces cerevisiae*: Structure, function, and biogenesis. Microbiol Rev. 1995;59:304–22.

67. Douglas LM, Konopka JB. Fungal Membrane Organization : The Eisosome Concept. Annu Rev Microbiol. 2014;68:377–93.

68. Klose C, Surma MA, Gerl MJ, Meyenhofer F, Shevchenko A, Simons K. Flexibility of a eukaryotic lipidome - insights from yeast lipidomics. PLoS One. 2012;7:1–11.

69. Meer G, Voelker DR, Feigenson GW. Membrane lipids, where they are and how they behave: Nat Rev Mol Cell Biol. 2008;9:112–24.

70. Ren J, Hannun YA. Metabolism and Roles of Sphingolipids in Yeast *Saccharomyces cerevisiae*. In: Geiger O, editor. Biogenesis of Fatty Acids, Lipids and Membranes. Springer; 2016.

71. Kodedová M, Sychrová H. Changes in the sterol composition of the plasma membrane affect membrane potential, salt tolerance and the activity of multidrug resistance pumps in *Saccharomyces cerevisiae*. PLoS One. 2015;10:1–19.

72. Godinho CP, Costa R, Sá-Correia I. The ABC transporter Pdr18 is required for yeast thermotolerance due to its role in ergosterol transport and plasma membrane properties. Environ Microbiol. 2021;23:69–80.

73. Lindahl L, Genheden S, Faria-Oliveira F, Allard S, Eriksson LA, Olsson L, Bettiga M. Alcohols enhance the rate of acetic acid diffusion in S. Cerevisiae: Biophysical mechanisms and implications for acetic acid tolerance. Microb Cell. 2017;5:42–55.

74. Lindahl L, Genheden S, Eriksson LA, Olsson L, Bettiga M. Sphingolipids contribute to acetic acid resistance in *Zygosaccharomyces bailii*. Biotechnol Bioeng. 2016;113:744–53.

75. Lindberg L, Santos AXS, 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. 2013;8:1–12.

76. Ribeiro RA, Godinho CP, Vitorino M V., Robalo TT, Fernandes F, Rodrigues MS, Sá-Correia I. Crosstalk between Yeast Cell Plasma Membrane Ergosterol Content and Cell Wall Stiffness under Acetic Acid Stress Involving Pdr18. J Fungi. 2022;8:1–19.

77. Balzi E, Goffeau A. Yeast Multidrug Resistance: The PDR Network. J Bioenerg Biomembr. 1995;27:71-76.

78. Gulshan K, Moye-rowley WS. Multidrug Resistance in Fungi. Eukaryot Cell. 2007;6:1933–42.

79. Higgins CF. Multiple molecular mechanisms for multidrug resistance transporters. Nature. 2007;446 April:749–57.

80. Jungwirth H, Kuchler K. Yeast ABC transporters - A tale of sex, stress, drugs and aging. FEBS Lett. 2006;580:1131–8.

81. dos Santos SC, Teixeira MC, Dias PJ, 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. 2014;5 May:1–15.

82. Sá-Correia I, dos Santos SC, Teixeira MC, Cabrito TR, Mira NP. Drug:H+ antiporters in chemical stress response in yeast. Trends Microbiol. 2009;17:22–31.

83. Gbelska Y, Krijger J, Breunig KD. Evolution of gene families : the multidrug resistance transporter genes in five related yeast species. FEMS Yeast Res. 2006;6:345–55.

84. Sá-Correia I, Godinho CP. Exploring the biological function of efflux pumps for the development of superior industrial yeasts. Curr Opin Biotechnol. 2022;74:32–41.

85. Godinho CP, Sá-Correia I. Physiological Genomics of Multistress Resistance in the Yeast Cell Model and Factory: Focus on MDR/MXR Transporters. In: Sá-Correia I, editor. Progress in molecular and subcellular biology. Springer Nature Switzerland; 2019. p. 1–35.

86. Drew D, North RA, Nagarathinam K, Tanabe M. Structures and General Transport Mechanisms

by the Major Facilitator Superfamily (MFS). Chem Rev. 2021;121:5289–335.

87. Dias PJ, 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. 2013;14:1-22.

88. Dias PJ, Seret ML, Goffeau A, Correia IS, Baret P V. Evolution of the 12-spanner drug:H+ antiporter DHA1 family in hemiascomycetous yeasts. Omi A J Integr Biol. 2010;14:701–10.

89. Wright MB, Howell EA, Gaber RF. Amino acid substitutions in membrane-spanning domains of Hol1, a member of the major facilitator superfamily of transporters, confer nonselective cation uptake in *Saccharomyces cerevisiae*. J Bacteriol. 1996;178:7197–205.

90. Barker KS, Pearson MM, Rogers PD. Identification of genes differentially expressed in association with reduced azole susceptibility in *Saccharomyces cerevisiae*. J Antimicrob Chemother. 2003;51 April:1131–40.

91. Kawano-kawada M, Pongcharoen P, Kawahara R, Yamasaki T, Akiyama K, Sekito T. Vba4p , a vacuolar membrane protein , is involved in the drug resistance and vacuolar morphology of *Saccharomyces cerevisiae*. Biosci Biotechnol Biochem. 2016;80:279–87.

92. Dhaoui M, Auchère F, Blaiseau P, Lesuisse E, Landoulsi A, Brodsky JL. Gex1 is a yeast glutathione exchanger that interferes with pH and redox homeostasis. Mol Biol Cell. 2011;22:2054–67.

93. Heymann P, Ernst JF. Identification and substrate specificity of a ferrichrome-type siderophore transporter (Arn1p) in Saccharomyces cerevisiae. FEMS Microbiol Lett. 2000;186:221–7.

94. Banerjee A, Moreno A, Pata J, Falson P, Prasad R. ABCG: a new fold of ABC exporters and a whole new bag of riddles! Adv Protein Chem Struct Biol. 2021;123:163–91.

95. Paumi CM, Chuk M, Snider J, Stagljar I, Michaelis S. ABC Transporters in *Saccharomyces cerevisiae* and Their Interactors : New Technology Advances the Biology of the ABCC (MRP) Subfamily. Microbiol Mol Biol Rev. 2009;73:577–93.

96. Higgins CF, Linton KJ. The ATP switch model for ABC transporters. Nat Struct Mol Biol. 2004;11:918–26.

97. Prasad R, Khandelwal NK, Banerjee A. Yeast ABC transporters in lipid trafficking. Fungal Genet Biol. 2016;93:25–34.

98. Godinho CP, Dias PJ, Ponçot E, Sá-Correia I. The paralogous genes *PDR18* and *SNQ2*, encoding multidrug resistance ABC transporters, derive from a recent duplication event, *PDR18* being specific to the *saccharomyces* genus. Front Genet. 2018;9 OCT:1–17.

99. Katzmann DJ, Hallstrom TC, Voet M, Wysock W, Golin J, Volckaert G, Moye-rowley WS. Expression of an ATP-Binding Cassette Transporter-Encoding Gene (*YOR1*) Is Required for Oligomycin Resistance in *Saccharomyces cerevisiae*. Mol Cell Biol. 1995;15:6875–83.

100. Katzmann DJ, Epping EA, Moye-rowley WS. Mutational Disruption of Plasma Membrane Trafficking of *Saccharomyces cerevisiae* Yor1p, a Homologue of Mammalian Multidrug Resistance Protein. Mol Cell Biol. 1999;19:2998–3009.

101. Cui Z, Hirata D, Tsuchiya E, Osada H, Miyakawa T. The multidrug resistance-associated protein (MRP) subfamily (Yrs1/Yor1) of *Saccharomyces cerevisiae* is important for the tolerance to a broad range of organic anions. J Biol Chem. 1996;271:14712–6.

102. Decottignies A, Grant AM, Nichols JW, Wet H De, Mcintosh DB, Town C, Africa S. ATPase

and Multidrug Transport Activities of the Overexpressed Yeast ABC Protein Yor1p *. J Biol Chem. 1998;273:12612–22.

103. Nagy Z, Montigny C, Leverrier P, Yeh S, Goffeau A, Garrigos M, Falson P. Role of the yeast ABC transporter Yor1p in cadmium detoxification. Biochimie. 2006;88:1665–71.

104. Grigoras I, Lazard M, Plateau P, Blanquet S. Functional characterization of the Saccharomyces cerevisiae ABC-transporter Yor1p overexpressed in plasma membranes. Biochim Biophys Acta. 2008;1778:68–78.

105. Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M, Godinho P, Martins C, Bourbon N, Mota MN, Ribeiro RA, et al. YEASTRACT + : a portal for cross-species comparative genomics of transcription regulation in yeasts. Nucleic Acids Res. 2020;48 October 2019:642–9.

106. Buechel ER, Pinkett HW. Transcription factors and ABC transporters : from pleiotropic drug resistance to cellular signaling in yeast. FEBS Lett. 2020;594:3943–64.

107. Akache B, Turcotte B. New Regulators of Drug Sensitivity in the Family of Yeast Zinc Cluster Proteins. J Biol Chem. 2002;277:21254–60.

108. Xie Y, Varshavsky A. RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: A negative feedback circuit. Proc Natl Acad Sci U S A. 2001;98:3056–61.

109. Owsianik G, Balzi E, Ghislain M. Control of 26S proteasome expression by transcription factors regulating multidrug resistance in *Saccharomyces cerevisiae*. Mol Microbiol. 2002;43:1295–308.

110. Mira NP, Henriques SF, Keller G, Teixeira MC, Matos RG, Arraiano CM, Winge DR, Sá-Correia I. Identification of a DNA-binding site for the transcription factor Haa1, required for *Saccharomyces cerevisiae* response to acetic acid stress. Nucleic Acids Res. 2011;39:6896–907.

111. Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol:chloroform. Cold Spring Harb Protoc. 2006;2006.

112. Jansen G, Wu C, Schade B, Thomas DY, Whiteway M. Drag&Drop cloning in yeast. Gene. 2005;344:43–51.

113. Gietz RD, Schiestl RH, Willems AR, Woods RA. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast. 1995;11:355–60.

114. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, et al. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49:480–9.

115. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9.

116. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr Sect D Biol Crystallogr. 2010;66:486–501.

117. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, Cruaud C, et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates Species-wide genetic and phenotypic diversity. Nature. 2018;556:339–46.

118. Li H. Tabix: Fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics. 2011;27:718–9.

119. Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M. VariantAnnotation : a

Bioconductor package for exploration and annotation of genetic variants. Bioinformatics. 2014;30:2076–8.

120. Pagès H. BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs. R package version 1.64.0. 2022.

121. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. 2010;:1–4.

122. Stamatakis A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics. 2014;30:1312–3.

123. Wagih O, Galardini M, Busby BP, Memon D, Typas A, Beltrao P. A resource of variant effect predictions of single nucleotide variants in model organisms. Mol Syst Biol. 2018;14:1–16.

124. Yun RH, Anderson A, Hermans J. Proline in α-helix: Stability and conformation studied by dynamics simulation. Proteins Struct Funct Bioinforma. 1991;10:219–28.

125. Van Der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, et al. Classification of intrinsically disordered regions and proteins. Chem Rev. 2014;114:6589–631.

126. Cabrito TR, Teixeira MC, Singh A, Prasad R, Sá-Correia I. The yeast ABC transporter Pdr18 (ORF YNR070w) controls plasma membrane sterol composition, playing a role in multidrug resistance. Biochem J. 2011;440:195–202.

127. Semchyshyn HM, Abrat OB, Miedzobrodzki J, Inoue Y, Lushchak VI. Acetate but not propionate induces oxidative stress in bakers' yeast *Saccharomyces cerevisiae*. Redox Rep. 2011;16:15–23.

128. Tomitori H, Kashiwagi K, Asakawa T, Kakinuma Y, Michael AJ, Igarashi K. Multiple polyamine transport systems on the vacuolar membrane in yeast. Biochem J. 2001;353:681–8.

129. Rocha RO, Wilson RA. Essential, deadly, enigmatic: Polyamine metabolism and roles in fungal cells. Fungal Biol Rev. 2018;33:47–57.

130. Krüger A, Vowinckel J, Mülleder M, Grote P, Capuano F, Bluemlein K, Ralser M. Tpo1-mediated spermine and spermidine export controls cell cycle delay and times antioxidant protein expression during the oxidative stress response. EMBO Rep. 2013;14:1113–9.

131. Opitz C, Schade G, Kaufmann S, Di Berardino M, Ottiger M, Grzesiek S. Rapid determination of general cell status, cell viability, and optimal harvest time in eukaryotic cell cultures by impedance flow cytometry. Appl Microbiol Biotechnol. 2019;103:8619–29.

132. Ma M, Liu ZL. Comparative transcriptome profiling analyses during the lag phase uncover *YAP1*, *PDR1*, *PDR3*, *RPN4*, and *HSF1* as key regulatory genes in genomic adaptation to the lignocellulose derived inhibitor HMF for *Saccharomyces cerevisiae*. BMC Genomics. 2010;11:660–80.

133. Krasowska A, Łukaszewicz M, Bartosiewicz D, Sigler K. Cell ATP level of *Saccharomyces cerevisiae* sensitively responds to culture growth and drug-inflicted variations in membrane integrity and PDR pump activity. Biochem Biophys Res Commun. 2010;395:51–5.

134. Stojiljković M, Claes A, Deparis Q, Demeke MM, Subotić A, Foulquié-Moreno MR, Thevelein JM. Whole-Genome Transformation of Yeast Promotes Rare Host Mutations with a Single Causative SNP Enhancing Acetic Acid Tolerance. Mol Cell Biol. 2022;42:1–25.

135. Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, Walter P. Control of Gene Expression. In: Lucas G, editor. Molecular Biology of the Cell. 6th edition. y Garland Science, Taylor & Francis Group; 2015. p. 369–438.

136. Li J, Liang Q, Song W, Marchisio MA. Nucleotides upstream of the Kozak sequence strongly influence gene expression in the yeast S. cerevisiae. J Biol Eng. 2017;11:1–14.

137. Inukai S, Kock KH, Bulyk ML. Transcription factor–DNA binding: beyond binding site motifs. Curr Opin Genet Dev. 2017;43:110–9.

138. Zhu Z, Hu Y, Teixeira PG, Pereira R, Chen Y, Siewers V, Nielsen J. Multidimensional engineering of *Saccharomyces cerevisiae* for efficient synthesis of medium-chain fatty acids. Nat Catal. 2020;3:64–74.

139. Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, Forsmark A, Durbin R, Omholt SW, Louis EJ, Liti G, et al. Trait variation in yeast is defined by population history. PLoS Genet. 2011;7:1–15.

140. Salinas F, De Boer CG, Abarca V, García V, Cuevas M, Araos S, Larrondo LF, Martínez C, Cubillos FA. Natural variation in non-coding regions underlying phenotypic diversity in budding yeast. Sci Rep. 2016;6:1–13.

141. Sawadogo-lingani H, Owusu-kwarteng J, Glover R. Sustainable Production of African Traditional Beers With Focus on Dolo , a West African Sorghum-Based Alcoholic Beverage. Front Sustain Food Syst. 2021;5 May:1–26.

Appendices

Appendix A – Sequences

Sequence A.1. - Sequence (5'- 3') of the amplified fragment containing *PDR18*. Lowercase letters represent the Rec5 (5') and Rec2 (3') sequences. The coding sequence is highlighted in blue, and the binding site of the primers used in yellow.

aacaaaagctggagctcgtttaaacggcgcgccGAGAGCTCATTGGATGGAAGAAAGCATAAACGTAAATC
AGGGTCCTGAGTGTCTGGGTGCTGTCTACGTTAGTACTATTCGGCTGCATTATATTCCCAA
TTCACAAATAAATTAGGCTCATAACCGTAATTTTATTCGAGACATTTTTGGTTACTTCAAA
ATATTGTTATTATATAAAGATCATATAAAGTTCTTGGACAAGATTGGATACATTTAGTTTT
ATTTTTGAAAATCACAAAGATGAAACAAAATAAAAAATGTCAGCGTGCCGCATTGAAAG
GTAAAAACTAAAATTAATGAACTTTTCTAATAAAGATGAAAGGGAGAAAACAGTATACA
TTGTAGCAGAGCGTTTAGAAGGAGTACTCAGAAAGTTCTAAAAAGGCCTTATTTCAAACT
CATGAAGAAAAAAAATTTCCATGAATGAATGATTGTATGATTGTATTGGCATGTAATATC
TGCAGGATCCACTCAATCTTGAAAAGATACTGTAGAGGAATACCCGAACAACAAATAAA
TCATGTTACAACTCTTTGACGGCGAAGGGTGTGGCTACCGTGCTGCAACATTTTAGTTACT
GCTGGTACATTAAAATTGGGAAGATGTCCAAGAGAAGAACATGTTCTGCTATCGAATCTC
TCACGATATATATATATATATATTTATATGACTCACCTGGTAAAACATCTATGTTTCGTTT
CATATGAGGGCTTAGGTACCATTCTTTTAGAGTTAAGATCTCTTATTGTGGAAAAAAGAA
GTTGAACGTGATTTCATTTGTTTATTGTGCTAGTGGCTGGAAATATGAAAACTTTTGCTCT
TGCAACGTACCTTTTTACTTAATGTCTTATATTTGAACACTGCTATCCCAGCTACACAACG
CATTTTCTGAAGGAAGAGAGTTTAGAAGAAACCTTTCTCTGTCCATCGCTATAATCAGCC
GTTTCTTTTTCGCTTCCTTACGAAGTAAAGGGTGTCCCTTTATAACGTAAGCGTCCAAAAT
GCAGTTATTAATATACTCCATTGCACGCAAAGAAATTTCAAGGAAAGTATCCGGGAATTC
TGATATTTTACACTTGCACTAATATACTCTTATTAGAAGTAGTGATGGTAGACCTCAAATC
AATCTAAACATAATGGATTTCCATACTGTTAAAGACGGCGATACGGAATTGAGGTGTCCT
ATTCCCGACACTTTTGATGCGTCAGCTATAATCAAATCTTATGTTAAATAAGCTTCTGAAC
AAGGGATTCATCTTCGTAAGGCAGGTGTAGCCATGGAATGCGTTTCAGTAGAAGGTTTGG
ATTCTTCTTTTTTGGAGGGCCAAACCTTTGGCGATATTTTGTGTTTACCATGGACAATTAT
CAAGGGTATCCGTGAGCGGAAGAATCGCAATAAGATGAAGATCATTTTGAAGAATGTCA
GTTTGCTGGCTAAATCAGGAGAGATGGTCCTTGTCCTAGGAAGACCAGGCGCTGGCTGTA
CATCATTTTTAAAGAGCGCTGCTGGTGAGACCAGTCAGTTTGCAGGTGGTGTAACAACAG
GACATATATCGTACGATGGTATCCCTCAGAAAGAAATGATGCAACATTACAAGCCAGATG
TAATCTATAATGGTGAGCAAGATGTTCATTTCCCACATTTGACAGTAAAACAAACTCTAG
ATTTTGCTATTTCCTGTAAGATGCCCGCAAAAAGAGTCAATAATGTAACGAAAGAAGAGT
ATATTACTGCCAATAGAGAATTCTATGCTAAAATTTTTGGTTTGACGCATACTTTCGATAC
CAAAGTTGGTAACGATTTCATCAGCGGTGTATCTGGAGGTGAGCGTAAACGCGTTTCCAT
TGCTGAAGCATTAGCAGCGAAAGGTTCAATTTACTGCTGGGATAATGCTACAAGAGGTCT
TGACTCTTCTACCGCGCTAGAATTTGCACGAGCTATTCGTACTATGACAAATCTGTTAGGT
ACAACGGCCCTTGTTACGGTTTACCAAGCCAGTGAAAACATTTATGAAACTTTTGATAAA
GTCACTGTTCTATACGCTGGAAGACAAATATTTTGCGGCAAAACTACTGAAGCAAAAGAT
TATTTTGAAAACATGGGTTACTTGTGTCCACCGAGACAATCGACTGCTGAATATTTGACC
GCAATTACTGATCCTAATGGTCTGCACGAAATAAAGCCTGGCTTTGAGTATCAAGTACCT
CATACCGCTGATGAATTCGAAAAATACTGGCTTGATTCCCCAGAATATGCCCGCCTAAAA
GGTGAAATTCAGAAGTACAAACATGAAGTGAATACTGAATGGACCAAAAAAACATACAA
TGAGTCTATGGCACAAGAAAAGTCGAAAGGTACAAGAAAAAAATCTTATTATACGGTCT
CTTATTGGGAGCAAATTAGACTTTGCACTATTAGAGGTTTTCTAAGGATTTACGGTGATAA

GTCATACACCGTTATCAACACATGCGCTGCTATAGCACAGGCTTTTATCACTGGGTCATTG
TTCTACCAAGCACCTTCTTCAACTCTAGGGGCCTTTTCTAGAAGTGGCGTTCTGTTTTTTTC
CCTCTTATATTATTCTTTGATGGGTTTAGCTAATATTAGTTTCGAGCACAGGCCAATATTG
CAAAAACACAAGGTCTATTCACTATATCATCCCTCAGCTGAAGCGTTAGCAAGTACGATT
TCTTCTTTTCCATTCAGAATGATTGGTCTAACATTTTTCATAATCATCCTGTACTTCTTAGC
CGGTTTGCATAGAAGCGCCGGTGCTTTTTTTACTATGTATTTGTTATTGACAATGTGTTCA
GAAGCTATTACAAGTTTGTTTCAGATGGTTTCATCTTTATGCGATACATTGTCCCAGGCCA
ACTCCATTGCCGGTGTTGTGATGTTATCTATTGCCATGTATTCGACGTACATGATACAATT
ACCTTCAATGCATCCATGGTTTAAGTGGATTTCGTACATTCTACCCATTAGATATGCATTT
GAATCGATGTTAAATGCAGAATTTCATGGAAGACATATGGATTGTGGTGGCACTTTGGTT
CCTTCTGGACCTGGGTTTGAAAACATCTTGCCAGAAAATCAAGTGTGTGCTTTTGTTGGTT
CAAGGCCTGGCCAATCTTGGGTCCTAGGTGATGATTATTTGAGGGCCCAATATCAATATG
AGTACAAAAATACTTGGAGAAACTTCGGCATCATGTGGTGTTTCTTAATTGGCTACATCG
TCTTGAGGGCGGTTTTCACTGAGTACAAAAGTCCTGTCAAAAGTGGTGGTGATGCTCTGG
TCGTCAAGAAGGGCACAAAGAATGCTATACAAAGATCATGGAGCAGCAAAAATGACGAA
GAGAACCTTAATGCCTCTATAGCAACACAGGATATGAAAGAGATAGCTTCAAGTAACGA
CGATAGCACAAGTGCAGACTTTGAAGGTTTGGAATCTACCGGGGTGTTTATTTGGAAAAA
TGTTTCTTTCACAATTCCTCATTCTAGCGGACAACGTAAACTTCTGGACAGCGTAAGCGGT
TACTGTGTTCCTGGTACTTTGACAGCATTAATAGGTGAGTCCGGCGCTGGTAAGACCACA
TTATTAAACACTTTGGCTCAAAGGAACGTGGGAACGATTACTGGTGATATGTTAGTTGAT
GGTCTCCCAATGGACGCTAGTTTCAAAAGGCGTACCGGTTATGTTCAACAGCAAGATCTT
CATGTTGCTGAACTTACTGTCAAAGAATCGCTACAATTTAGTGCTCGTATGCGTCGGCCAC
AGTCTATTCCTGACGCTGAAAAGATGGAATATGTTGAAAAAATTATATCTATTCTTGAAA
TGCAGGAGTTCTCAGAAGCTCTTGTCGGCGAAATTGGTTACGGCTTGAATGTTGAACAGA
GAAAGAAACTATCAATTGGCGTTGAGCTAGTTGGCAAGCCGGATCTGTTATTGTTTTTGG
ACGAACCGACCTCTGGCTTGGATTCCCAATCCGCGTGGGCCGTCGTCAAAATGTTAAAAA
GATTAGCTCTAGCAGGTCAATCAATTTTATGTACTATTCATCAACCATCAGCTACTCTTTT
TGAACAGTTTGACAGATTATTGCTTTTGGGAAAAGGCGGTCAAACAATTTATTTTGGTGA
AATAGGTAAAAATTCAAGTTCTGTCATTAAGTATTTCGAAAAGAACGGAGCCAGGAAAT
GTCAACAAAATGAAAATCCGGCAGAGTATATTTTAGAAGCCATAGGAGCTGGGGCCACC
GCTTCTGTTCAACAGAACTGGCCTGATATATGGCAAAAATCTCACGAGTATGCAAACATT
AACGAAAAAATAAATGACATGATTAAGGACTTATCTTCTACAACTTTGCATAAAACAGCT
ACAAGGGCCTCTAAGTATGCAACATCGTATTCCTACCAATTCCATCATGTGCTGAAAAGA
TCCAGTTTAACATTTTGGAGAAACTTGAACTACATCATGGCCAAAATGATGTTATTGATG
ATCAGTGGCTTGTTCATTGGTTTTACCTTTTTCCATGTGGGTGTAAATGCTATTGGATTAC
AAAATAGCTTATTTGCCTGTTTCATGGCTATCGTTATATCAGCTCCTGCAACAAACCAAAT
ACAGGAGCGTGCTACCGTTGCTAAGGAGCTATATGAAGTTCGTGAGTCCAAATCTAATAT
GTTTCATTGGTCTTTACTTTTGATTACCCATTATTTAAATGAATTGCCTTACCATTTATTGT
TTTCAACAATTTTTTTCGTTTCATCATATTTCCCTCTGGGTGTCTTTACCGAAGCCTCTAGG
TCAAGTGTTTTTTATCTGAACTATGCCATACTTTTTCAACTTTACTATATTGGTCTTGCTTT
AATGATTCTGTACATGTCTCCAAATCTACAATCTGCCAATGTTATTGTAGGTTTTATACTT
TCGTTTTTGCTCTCTTTCTGCGGTGCTGTCCAACCTGCCTCTTTAATGCCTGGTTTCTGGAC
ATTCATGTGGAAACTATCCCCTTACACGTATTTTTTGCAAAATTTAGTTGGATTATTGATG
CATGACAAACCCGTAAGATGTTCAAAGAAAGAGTTGTCTCTTTTCAACCCCCCGGTAGGC
CAAACATGTGGTGAATTTACCAAACCGTTTTTTGAATTTGGGACTGGGTATATTGCAAATC
CAGATGCAACAGCAGATTGCGCCTATTGTCAGTACAAAGTAGGTGATGAATACTTGGCGC
GCATAAATGCTAGCTTTAGTTACTTATGGAGAAACTTCGGTTTCATTTAGGCTTATATCCT
TTTCAACATTGCCGGTATGATTGTGGTTTATTACGTTGTTCAAGTTAAGCATTTTTCCCCTA

TGAAAATTGGCTTCGTAAAAAGAATAACAAGTAAATTTAAGAGAAAATGAAGAAAAAAA
TTGCAGGTATACTATGGGTGCAGGTGAGGCGGTTTTTTAACCGATTTAGCACACGATACT
GTGAATTTTGGGTTATATGATGATGTGAATATTGCAGAAATGTAATGTTATATTATTCAAA
GAAATAGGGTAACTTCTATTTATGCAAACATTGGTAATCAAGGCTATAAACACATAATCC
CAAATAAGTATTTTGGTACACAGTTGATGTTTCTTTATGAATACGACTCATTAGTTCACTG
TCAAATTGCCAACGTTCAGAGGGACCTACTTACAGAGCGTATgtcgacctcgagtcatgtaattagttatgtc
acgc

Sequence A.2. – Sequence (5'- 3') of the amplified fragment containing *TPO2*. Lowercase letters represent the Rec5 (5') and Rec2 (3') sequences. The coding sequence is highlighted in blue, and the binding site of the primers used in yellow.

aacaaaagctggagctcgtttaaacggcgcgccGCTGCTTACATACCACTAGCTGAGCAAGACAGCTGCCGT
TCAATGTTTGTACTTCGCACCATGAGAGTTGTGGATATGTACTTTATGGACGAAGCTGAA
ACGGCTAACACTATTATTACGCGAACTTTTTCAACACTTTTACTTCGCCACAAGTGCCGAA
GATCGAATCTTACCCAAAATCCGGTTATTTTGTACCTATGCAAAAACCCTTCCCCTGATCT
ATTTGCCATTTAGCCGCCCAAAGTACCGCAGAAGCTGTTACAAACACATCACGGGCCACC
GGAAAAAGTATTTTTTTGCTGCTTACCTGAATATATTTGCGCTTAAAAAAATTTTCCGGCG
CTTCCGAGTGAAATATACTGCTTAACTGTGTCCGGAGAAACCTCGTGCGGAAAACAGCGC
AAACCGCAAAAACCCCGCAATGGGCGCGCTTACTCAAATTCCCTCGGGTGGTCTCACGGG
CTGGCCGCAGCGCAGTCGCCTGCCGCACTAAACCTTAAAAATGCGAGAGCGCATGCACG
GCGGACGTCGTGGAAGCCTGACGCGTGCGCCCTTCTCCTCCCGAGGGACATAATGTCTTG
GTGCCTTACTAGAACCGATTTCTCGAGATGATTCCATAGCCGTTAAATTCATCTCAAAAG
AATAGAGAACACAGCAGCGCAGAACGAAAACAAACTGTCGCGTCGCATTTTACTGAACG
AGTCATTCGCATATTCTGCTTAGGTGTCTGAGTCTTTTAAACGCATTCTAGGTTTTTCGAA
AAAGCCACCGAAGAACACTGCCATCCGCATGCACTGTAATGCCACTGTTTATTTTTTTTTC
TTGCTATATATATACATTGCGCATACACCCTTTTGAGGTTTCGTACTAGGATTTTTTTGTAT
TTGTTTATGTTCTTTTTTCCTGAAGACATCAAGAAGAAATAAGAAAACAAATATATCATTT
TATTGTTTGTGTTTATAGAGAACAATATACAACTTCCTTCTCCTACTATATCAAGTTTAAT
ATTTAATTTTTTTTTTTTTTCGGTTTGAATTTTAATTTTAACTTTAATTTCAAAAAGTAATTG
ATACATTCTTTTATCAAAGCCCTGCTATTATTCTTTATTTCTACACCCTAATATTTTTTGTC
CATTGATATTTTTCCTTTAATTCCATCTAATAACTAATCACAAAAATAATACACAAAACAA
ATATGAGTGATCAAGAATCTGTTGTTTCATTCAACTCACAAAACACTTCCATGGTGGACG
TTGAGGGCCAACAACCTCAACAGTATGTCCCCTCAAAAACCAACTCTCGTGCAAATCAAC
TTAAGTTAACTAAGACTGAGACCGTCAAGTCTTTACAAGATTTAGGTGTCACTTCAGCCG
CCCCAGTGCCTGATATCAATGCGCCACAAACTGCTAAGAATAACATTTTCCCTGAAGAAT
ATACCATGGAAACACCATCTGGGCTGGTTCCAGTGGCTACCTTACAATCTATGGGTAGAA
CCGCCTCTGCCTTATCTCGTACTAGAACAAAGCAATTGAACCGTACCGCTACCAATTCCTC
ATCCACAGGTAAAGAAGAAATGGAAGAGGAAGAAACTGAAGAACGTGAAGACCAGAGC
GGTGAAAACGAGCTAGATCCAGAGATCGAATTCGTTACTTTTGTTACTGGTGATCCAGAA
AACCCTCACAATTGGCCCTCATGGGTTCGTTGGAGTTACACTGTCCTGTTGTCCATCTTAG
TTATTTGCGTTGCCTACGGTTCTGCTTGTATCAGTGGTGGGTTGGGAACCGTTGAAAAGAA
ATACCATGTAGGTATGGAAGCCGCTATTTTATCATGTTCTTTAATGGTTATTGGGTTCTCG
CTGGGTCCTTTGATTTGGTCTCCTGTTAGTGATCTTTACGGTAGAAGAGTTGCTTACTTTGT
TTCTATGGGTCTTTATGTCATCTTCAATATCCCTTGCGCCTTAGCTCCAAATCTAGGTTGTC
TTTTAGCTTGTAGATTTTTATGTGGTGTTTGGTCATCATCTGGTTTGTGTTTAGTTGGTGGG
TCTATTGCCGATATGTTCCCAAGTGAAACAAGAGGTAAGGCTATTGCTTTCTTCGCTTTTG
CTCCTTACGTTGGTCCCGTTGTTGGTCCACTAGTTAACGGTTTTATTTCCGTTTCTACCGGA

CGTATGGACCTGATTTTCTGGGTCAATATGGCCTTTGCAGGTGTTATGTGGATCATATCTT
CTGCCATCCCAGAAACGTACGCTCCAGTTATCTTGAAGAGAAAGGCTGCTAGATTAAGAA
AGGAAACTGGTAATCCCAAGATTATGACTGAGCAGGAAGCGCAAGGTGTCAGTATGAGT
GAAATGATGAGGGCTTGTCTGTTGAGACCTTTGTACTTCGCTGTCACTGAACCTGTTCTAG
TTGCCACTTGTTTCTACGTGTGTTTGATTTACTCTCTACTATATGCGTTCTTCTTTGCCTTCC
CTGTCATTTTCGGTGAACTATATGGCTACAAAGATAACCTTGTGGGTTTAATGTTTATTCC
TATTGTTATCGGTGCTCTTTGGGCGTTAGCCACAACTTTCTACTGTGAAAACAAATATTTA
CAAATTGTCAAACAGCGTAAACCTACTCCTGAAGATCGTTTGCTAGGTGCTAAGATCGGT
GCTCCATTTGCTGCAATTGCTCTATGGATCCTGGGTGCTACCGCTTATAAACATATTATTT
GGGTTGGTCCAGCTTCAGCTGGTTTAGCTTTTGGTTTCGGTATGGTGTTGATTTATTATTCA
TTGAATAATTACATTATTGATTGCTACGTCCAATACGCATCCAGTGCTCTGGCTACAAAGG
TTTTCTTAAGATCCGCCGGTGGTGCTGCCTTCCCCTTGTTTACCATTCAAATGTACCACAA
ATTGAATTTGCACTGGGGTTCTTGGTTGTTGGCTTTCATCTCCACTGCTATGATTGCTTTAC
CTTTTGCATTTTCTTACTGGGGTAAGGGCTTGAGACATAAGTTGTCCAAGAAGGACTATTC
CATCGACAGTGTTGAGATGTAAGCTGGTGCAAGTTTCCGGTAAAAATAATGATGTTCTAG
TCATTCATATATACGATACAAAAATAACAAATTTCCACATATATTCTGAAAAATTAATTTT
TAATATTTTATATTAGGAACTTCTCAAAAAAACGTCGGACATTTCTACATACTCTTTAACG
AAATATCAAAGTATTATTGTTTTGCAGCTTCTGAATATCAAGATCGCTCTTGCTACTCTTC
CTGAAgtcgacctcgagtcatgtaattagttatgtcacgc

Sequence A.3. – Sequence (5'- 3') of the amplified fragment containing *TPO3*. Lowercase letters represent the Rec5 (5') and Rec2 (3') sequences. The coding sequence is highlighted in blue, and the binding site of the primers used in yellow.

aacaaaagctggagctcgtttaaacggcgcgccGATTTCTCGAGAATCCGGTGAAAGCGTATACCCAAAAGA
GCAACATAATTGACTGACCCACAATTTGGTTAGAGACATGCTATGCTGATCCTGCATATC
TCAAACAGGCAATATGATAACATACATAAGCAATCTCGGCTACGCGGATTTTGTGTGGGT
CTGAAAAAAAAAAAAAAAAAAAAAAAAAAATAGCGCGGAGCCGGGTGAGAACGTTGGAC
GGGTAACATAAAAACAGAGCTCTTTATATAATCAAGCATTCCCTCCATACAAAATGTGTC
TGCAGCCACGTGCAAAGTACGGTCGTCGTGTGCTAAAGACCAGTAAGTACAGTCTAGCGG
GCCTCTGACTAGCTTCTCTGTGCTTGGCGAGGGGTTTACTGGAGCCCAATCGGACTAGCC
GGGCTTCGTCACTGCGGCAATGTCTTCGCCCAACGGATCGGAACAGCTCACTGATTTCAC
CCAAACGGGAAAAAGGAAAAACAAAAACAAAACAGAACGGCGTAGTTTCCGGCAGCCA
TGATGATCCCTGATCCATAGAGCCATTTTTTTCATTTATTTCCTAATGCATTTTGTTCATAT
TCTGGATCTGCGTTAAACGCTGTAAAACTGGCTCGTTTCTTTTATATATCTTGCAATCATT
CGCTGCTCAACCATCCTAGTCTCTTTTTACACGTACCCTTTTTCCACTTTTTCTTTTAGTTTA
TATATTCTTTTCGCCTAAGAACATTTGTCTTTGTAATCTACTTAACTTAGCAATTGTCAGCT
GTCCTGTTACAACCTGTTCTCCAAAGTGAATACAATAAGCAGTATTTGCTCTTTTTCGTCT
AGGTGCGTTGATCTAATCTTCTTTTTATTTTTCTTCTTCGTTCAGCCTTTTTTGTCCTCACCG
GTAGCGTTCACCCTAATTCTTTAATTGTCATACCCATAATTATACTACAAATTTCGTGGTC
TAGTTTTCTAATATAACTAATTACTTTCTCCCTTCTTACTTCATTATTTTAATTTTGCATTAG
TACTCCTCTAGCCAAAGATAAACAGAAATGAACAGACAGGAATCCATAAATTCGTTTAAT
TCAGACGAAACATCTTCGTTGTCTGATGTAGAAAGTCAGCAGCCGCAACAATATATCCCT
TCAGAGAGTGGATCTAAATCCAACATGGCTCCTAATCAACTGAAGTTGACCCGGACGGAA
ACCGTGAAGTCATTGCAGGACATGGGTGTGAGCTCCAAAGCCCCCGTTCCTGATGTTAAT
GCTCCTCAATCTAGCAAGAATAAGATTTTTCCTGAAGAATATACTTTAGAAACCCCTACA
GGTTTAGTTCCTGTCGCCACTCTACATTCCATAGGTAGAACTTCTACTGCGATTTCCCGTA
CGAGAACTAGACAGATCGATGGCGCTTCTTCGCCTTCTTCTAATGAAGATGCTTTAGAAA

GTGATAATAACGAAAAGGGTAAAGAAGGCGACTCTAGTGGTGCGAATGACGAAGCTCCA
GATCTAGATCCGGAAATTGAATTCGTTACCTTTGTGACTGGCGATCCAGAAAACCCCCAT
AACTGGCCTGCATGGATCCGTTGGAGTTACACTGTCCTACTGTCAATCTTAGTTATTTGTG
TCGCCTACGGGTCTGCTTGTATCAGTGGTGGGTTGGGAACCGTTGAAAAGAAATACCATG
TAGGTATGGAAGCCGCTATTTTATCGGTTTCTTTAATGGTTATTGGGTTCTCGCTGGGTCC
TTTGATTTGGTCTCCTGTTAGTGATCTTTACGGTAGAAGAGTTGCTTACTTTGTTTCTATGG
GTCTTTATGTCATCTTCAATATCCCTTGCGCCTTAGCTCCAAATCTAGGTAGTCTTTTAGCT
TGTAGATTTTTATGTGGTGTTTGGTCATCATCTGGTTTGTGTTTAGTTGGTGGGTCTATTGC
CGATATGTTCCCAAGTGAAACAAGAGGTAAGGCTATTGCTTTCTTCGCTTTTGCTCCTTAC
GTTGGTCCCGTTGTTGGTCCACTAGTTAACGGTTTTATTTCCGTTTCTACCGGACGTATGG
ACCTGATTTTCTGGGTCAATATGGCCTTTGCAGGTGTTATGTGGATCATATCTTCTGCCAT
CCCAGAAACGTACGCTCCAGTTATCTTGAAGAGAAAGGCTGCTAGATTAAGAAAGGAAA
CTGGTAATCCCAAGATTATGACTGAGCAGGAAGCGCAAGGTGTCAGTATGGGTGAAATG
ATGAGGGCTTGTCTGTTGAGACCTTTGTACTTCTCTGTCACTGAACCTGTTCTAGTTGCTA
CTTGTTTCTACGTGTGTTTGATTTACTCTCTACTATATGCGTTCTTCTTTGCCTTCCCTGTCA
TTTTCGGTGAACTATATGGCTACAAAGACAACCTTGTGGGTTTGATGTTTATTCCTATTGT
TATCGGTGCTCTTTGGGCGTTAGCCACAACTTTCTACTGTGAAAACAAGTATTTACAAATT
GTCAAACAGCGTAAACCTACTCCTGAAGATCGTTTGCTAGGTGCTAAGATCGGTGCTCCA
TTTGCTGCAATTGCTTTATGGATCCTGGGTGCTACCGCTTATAAACATATTATTTGGGTTG
GTCCAGCTTCAGCTGGTTTAGCTTTTGGTTTCGGTATGGTGTTGATTTATTATTCATTGAAT
AATTACATTATTGATTGCTACGTCCAATACGCATCCAGTGCTCTGGCTACAAAGGTTTTCT
TAAGATCCGCCGGTGGTGCTGCCTTCCCCTTGTTTACCATTCAAATGTACCACAAATTGAA
TCTGCACTGGGGTTCTTGGTTGTTGGCTTTCATCTCCACTGCTATGATTGCTTTACCTTTTG
CATTTTCTTACTGGGGCAAAGGTTTGAGACATAAGTTATCCAAGAAGGATTATTCCATAG
ATAGTATTGAATAAGAGTAACGATTTTACGAATCGAGCATAATAATGAAAAATTATAGTA
GTAGTAATGATAAAAACATAAGCAAACATAAGCAAACATAAGCAAACATATCAAACATG
CATTCAAACCATTATCCGACAGTAATGTACATATATTTTTACGTATTTTACATTCTTTCTCT
ATTGTTCTTTTATCTATTGCTATCCTAAACAAATATATTGATTGACCTTCAAACCGTGAGT
AACTTTATTAATTTATATGTATAGAAAAAAAAAAAAATTGAATTAGAAATTTGAACTTCT
CAAGTTCATTCTTCTTGTACCTATGCGATGACTTTATTTTCGGTTTTTACATTCCTCCTCAT
TTGACCGAGGAATACGAATTTGAgtcgacctcgagtcatgtaattagttatgtcacgc

Sequence A.4. – Amino acid sequence (N-terminal to C-terminal) of Tpo3.

MNRQESINSFNSDETSSLSDVESQQPQQYIPSESGSKSNMAPNQLKLTRTETVKSLQDMGVSS
KAPVPDVNAPQSSKNKIFPEEYTLETPTGLVPVATLHSIGRTSTAISRTRTRQIDGASSPSSNED
ALESDNNEKGKEGDSSGANDEAPDLDPEIEFVTFVTGDPENPHNWPAWIRWSYTVLLSILVIC
VAYGSACISGGLGTVEKKYHVGMEAAILSVSLMVIGFSLGPLIWSPVSDLYGRRVAYFVSMG
LYVIFNIPCALAPNLGSLLACRFLCGVWSSSGLCLVGGSIADMFPSETRGKAIAFFAFAPYVGP
VVGPLVNGFISVSTGRMDLIFWVNMAFAGVMWIISSAIPETYAPVILKRKAARLRKETGNPKI
MTEQEAQGVSMGEMMRACLLRPLYFSVTEPVLVATCFYVCLIYSLLYAFFFAFPVIFGELYGY
KDNLVGLMFIPIVIGALWALATTFYCENKYLQIVKQRKPTPEDRLLGAKIGAPFAAIALWILGA
TAYKHIIWVGPASAGLAFGFGMVLIYYSLNNYIIDCYVQYASSALATKVFLRSAGGAAFPLFTI
QMYHKLNLHWGSWLLAFISTAMIALPFAFSYWGKGLRHKLSKKDYSIDSIE

Appendix B – PCR protocol

Table B. 1. PCR protocol used for amplification (SuperFi II), error-prone amplification (GeneMorph II) and confirmation of positive recombinants (Speedy)

	SuperFi II (*PDR18*)		SuperFi II (*TPO*s)		GeneMorph II (*PDR18*)		GeneMorph II (*TPO*s)		Speedy (*TPO*s)	
Number of cycles	30		30		20		20		25	
	Temperature (°C)	Time	Temperature (°C)	Time	Temperature (°C)	Time	Temperature (°C)	Time	Temperature (°C)	Time
Initial denaturation	98	30''	98	30''	95	2'	95	2'	95	1’
Denaturation	98	10''	98	10''	95	30''	95	30''	94	30’’
Annealing	60	10''	60	10''	55	30''	55	30''	47	25’’
Extension	72	3'30''	72	2'	72	6'	72	4'	72	1’45’’
Final extension	72	5'	72	5'	72	10'	72	10'	72	3’

Appendix C – Cloning vector and recombination design

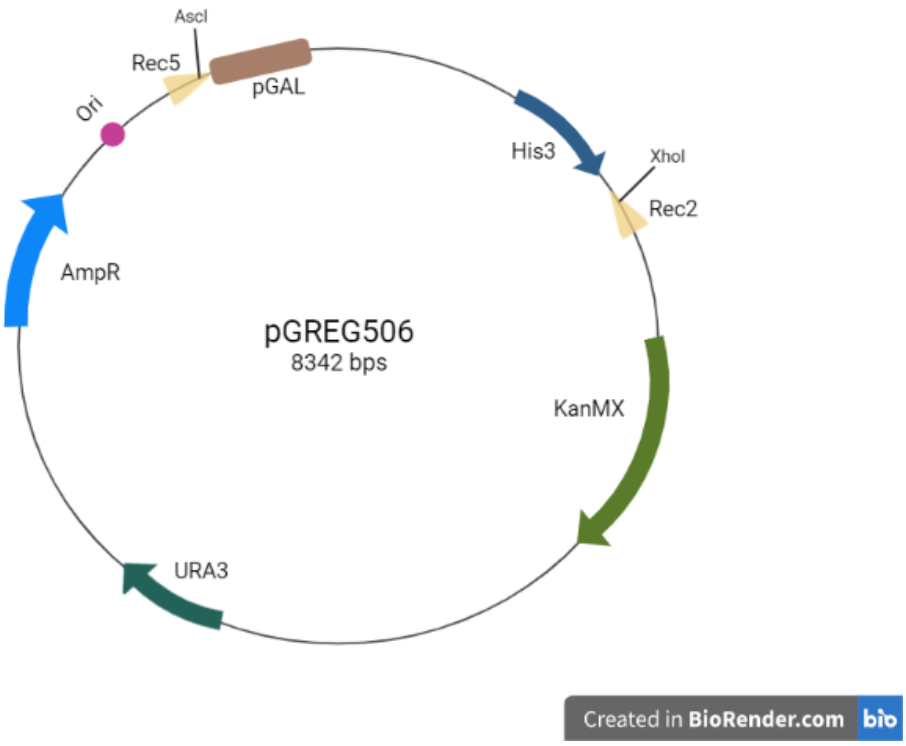

Figure C. 1. Simplified representation of the cloning vector pGREG506
The scheme represents the cloning vector used in this project. The cutting sites of the enzymes used to open the plasmid are indicated. Gene AmpR confers resistance to ampicillin in *E. coli*. KanMX is the gene required for resistance to G418 in *S. cerevisiae*. URA3 is the gene required for uracil de novo synthesis in BY4741. Size and position of gene is only indicative, nothing is at scale.

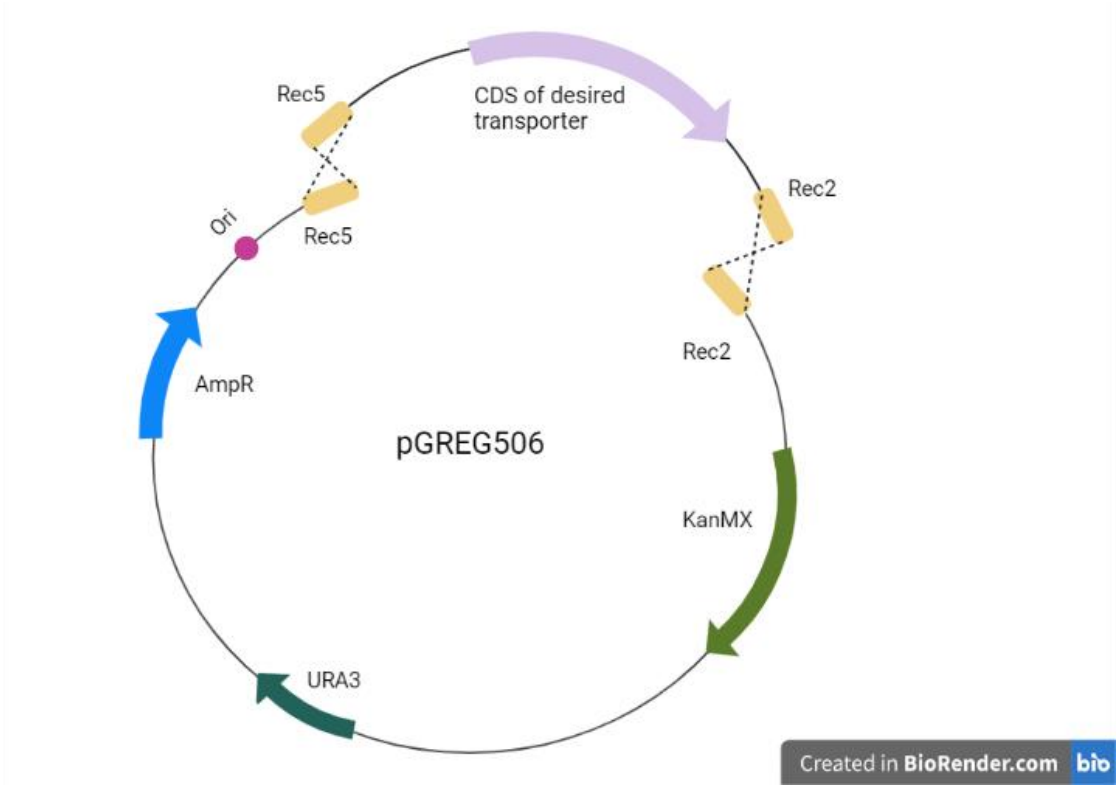

Figure C. 2. Simplified scheme of the homologous recombination between open vector and amplified fragment

The pGREG506 vector cut by AscI and XhoI, has exposed Rec5 and Rec2 that recombine with the same regions present in the amplified fragment of the CDS plus upstream and downstream regions.

Appendix D – Screening of strains with artificial *TPO3* variants

In this section are presented the screening curves of all the 294 Tpo3 variants obtained. Each graphic represents a different microplate with different strains tested. Strains were grown in 100µl MM-ura supplemented with 80 mM acetic acid (pH 4.5). The growth curves were plotted using OD600nm values. The strains *tpo3*Δ_υ/Tpo3 (orange dashed lines) and *tpo3*Δ_υ (blue dashed lines) were used as controls. Each control was present in the microplate in triplicate and all replicas are showed.

Figure D.9 represents the testing of the strains with higher tolerance to acetic acid. In this case the concentration of acetic acid used was 90 mM.

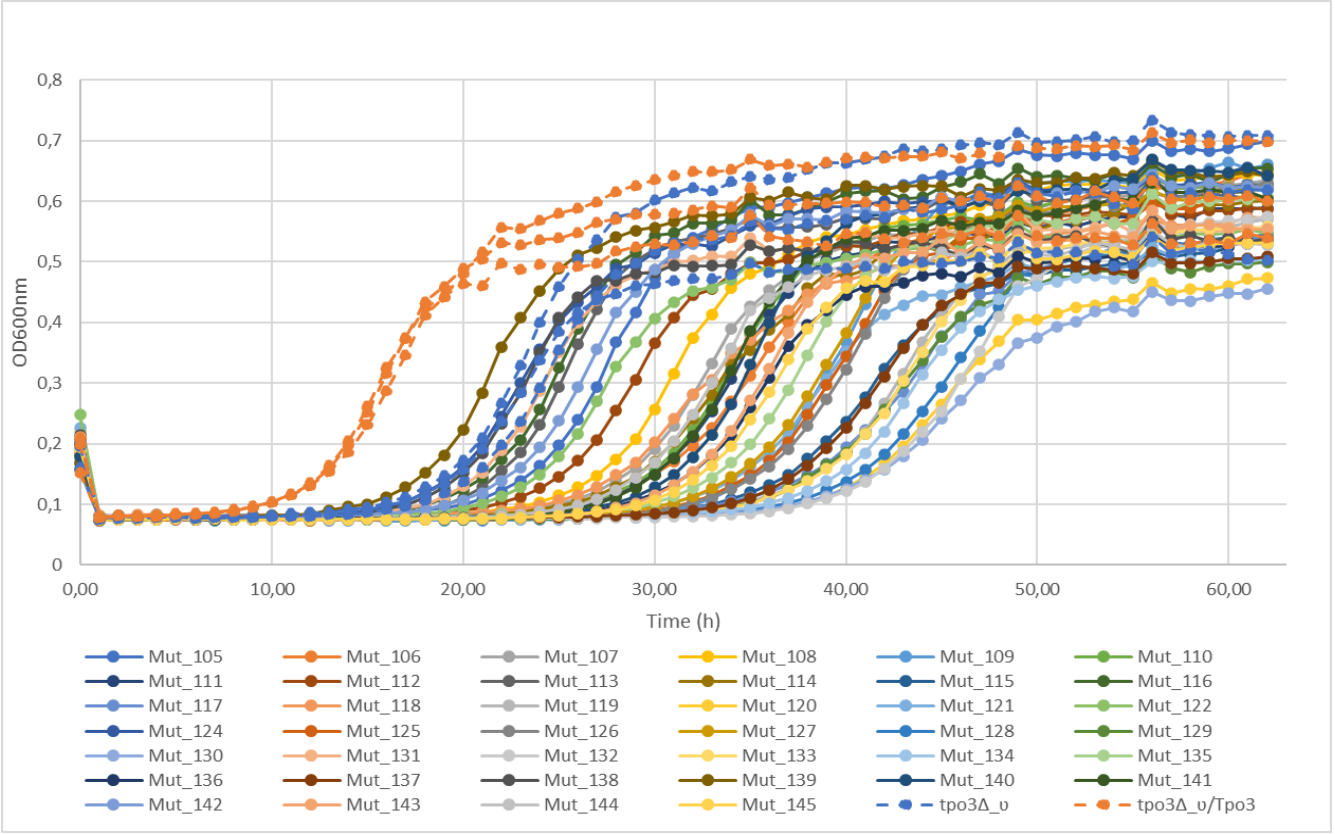

Figure D.1. Growth curves of Mut_105 to Mut_145 (except Mut_123)

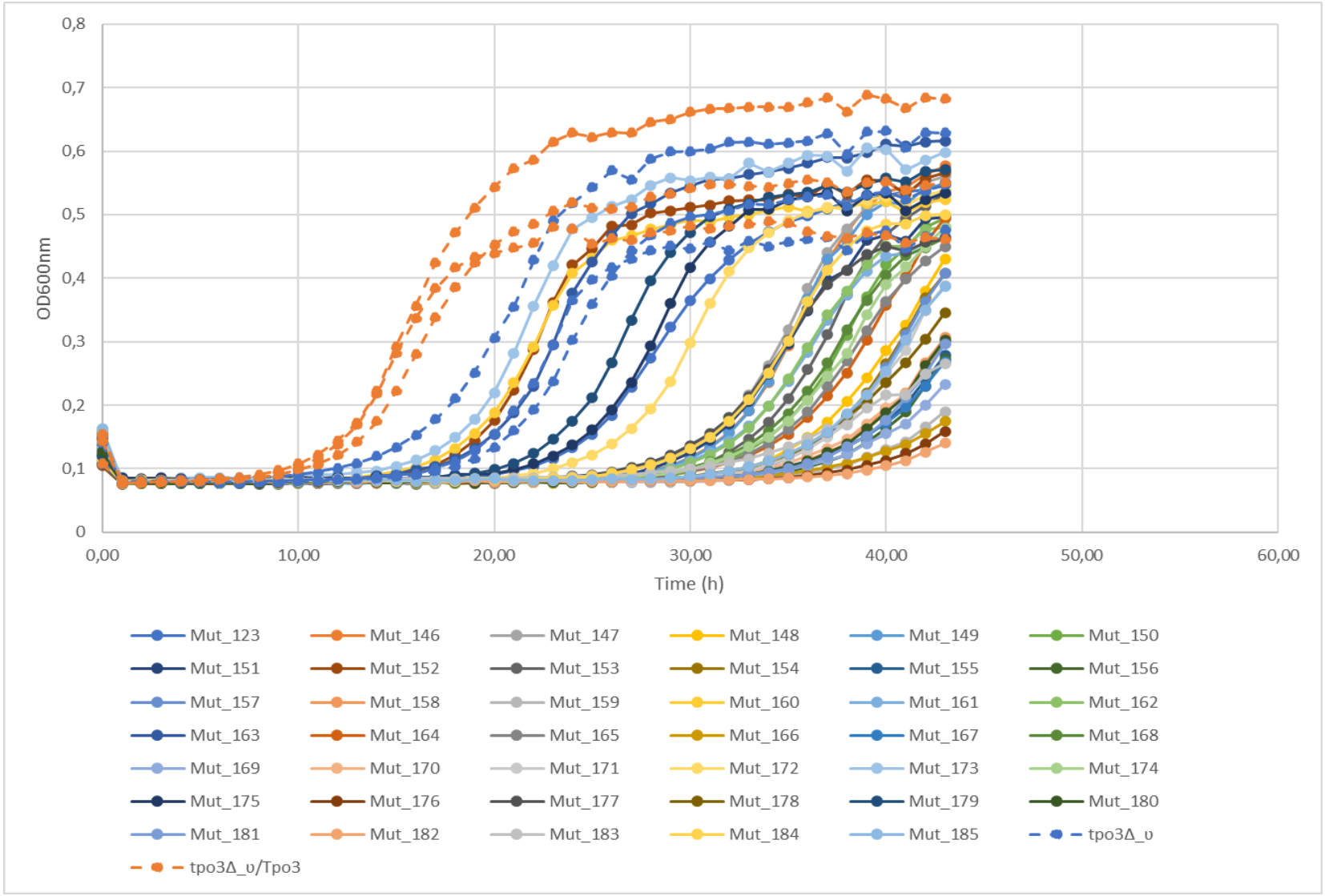

Figure D.2. Growth curves of Mut_146 to Mut_185 (plus Mut_123)

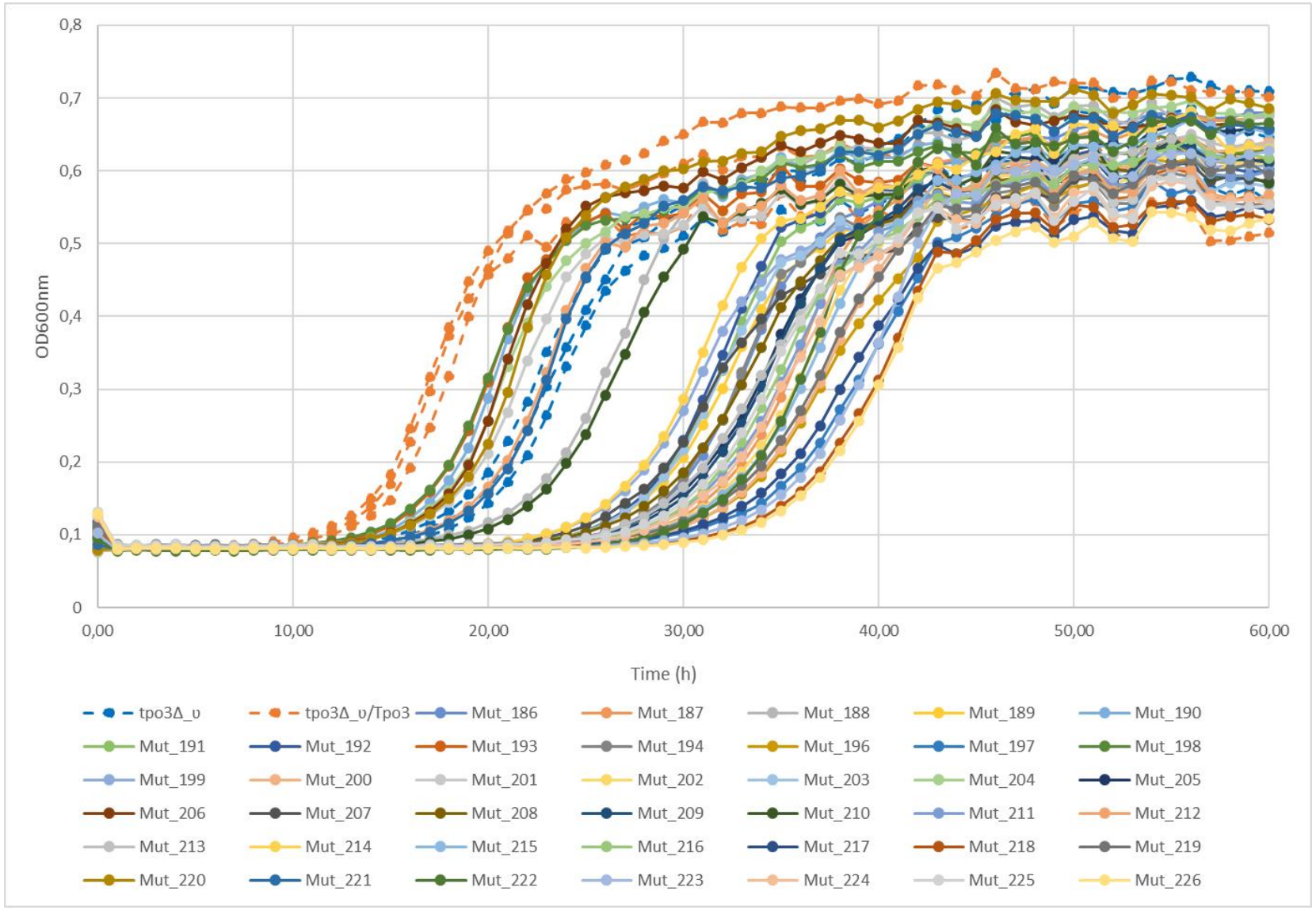

Figure D.3. Growth curves of Mut_186 to Mut_226

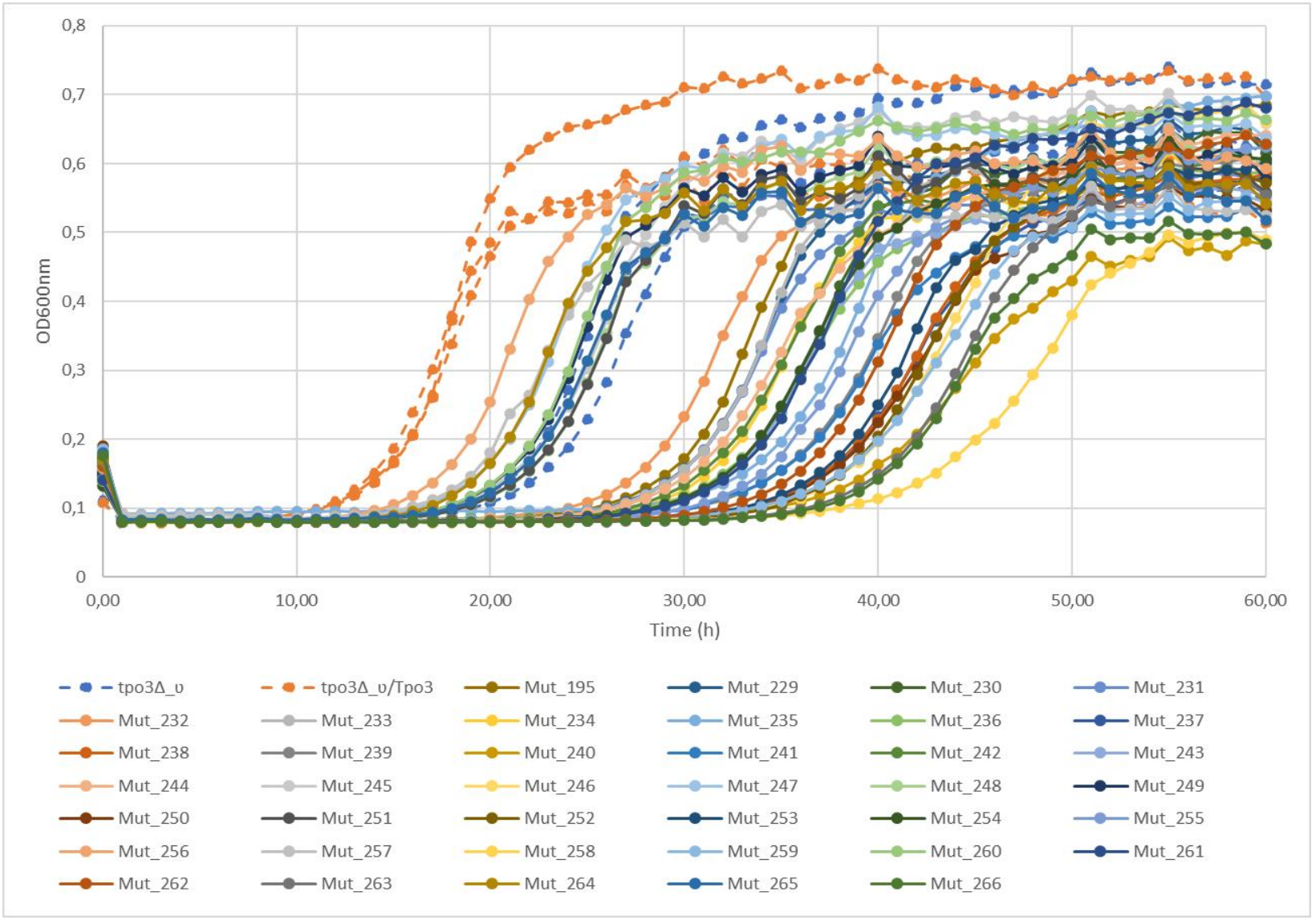

Figure D.4. Growth curves of Mut_229 to Mut_266 and Mut_195

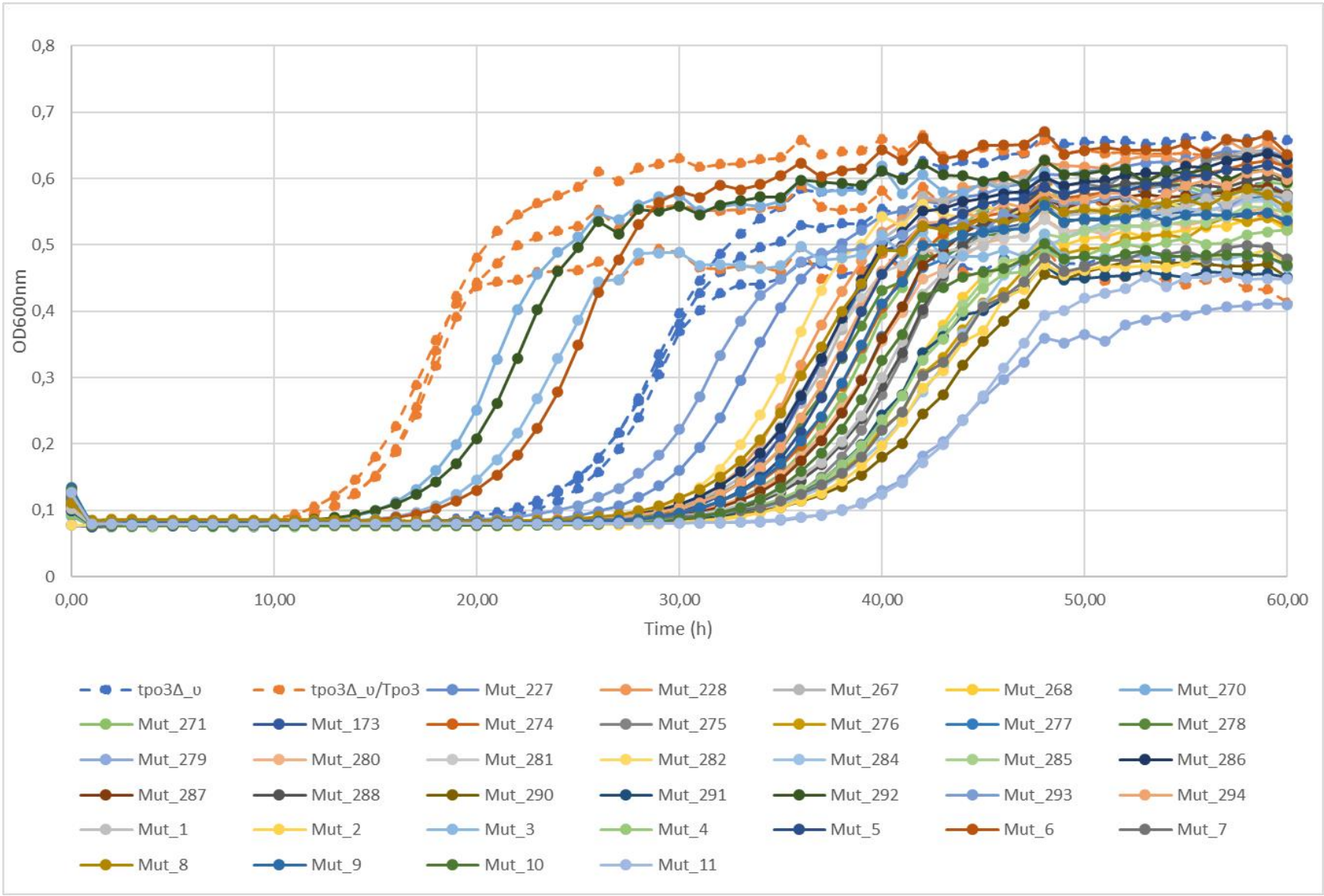

Figure D.5. Growth curves of Mut_267 to Mut_294 (except Mut_269, Mut_272, Mut_283 and Mut_289) , Mut_1 to Mut_11, Mut_227 and Mut_228

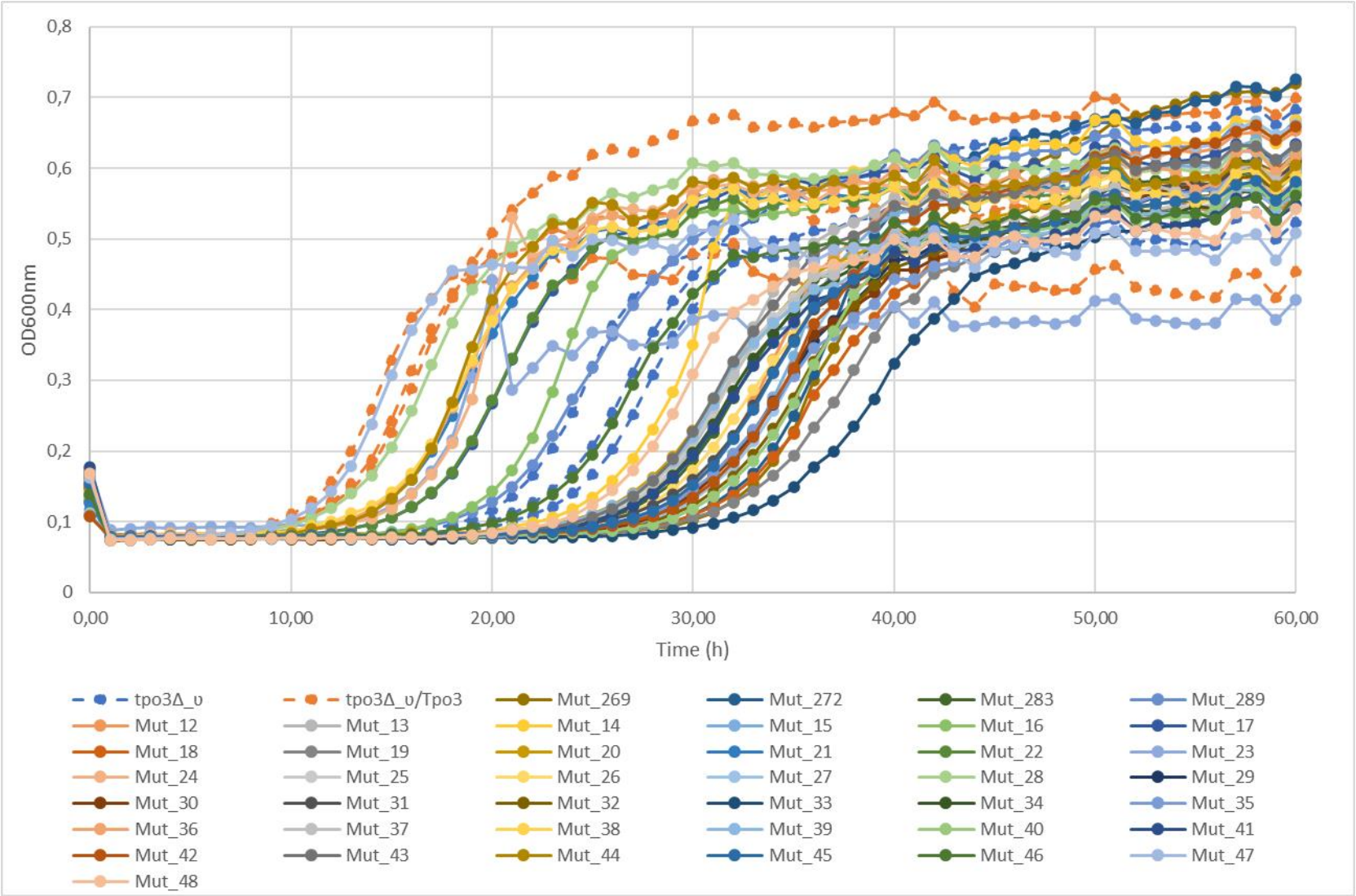

Figure D.6. Growth curves of Mut_12 to Mut_48, Mut_269, Mut_272, Mut_283 and Mut_289

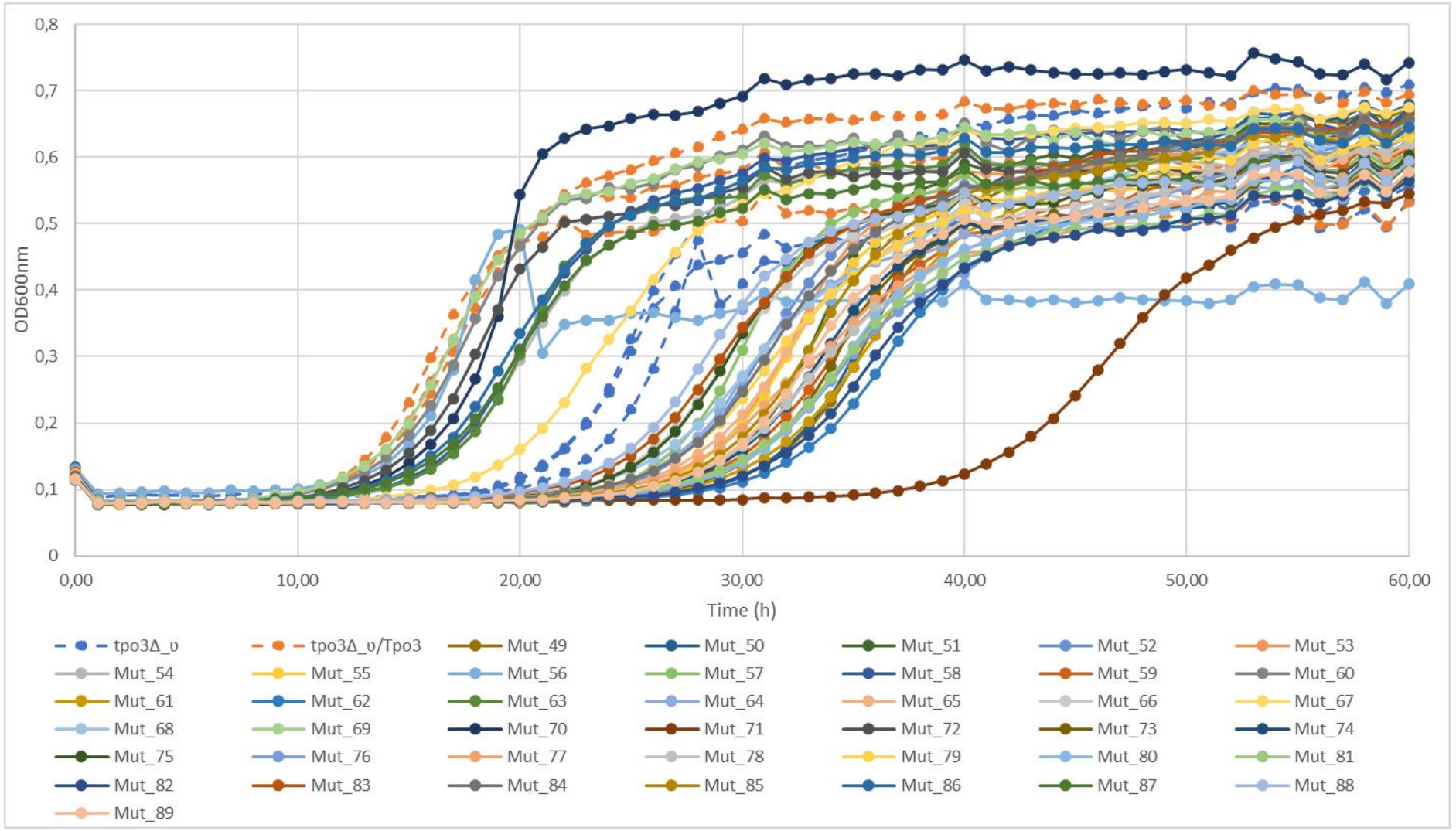

Figure D. 7. Growth curves of Mut_49 to Mut_89

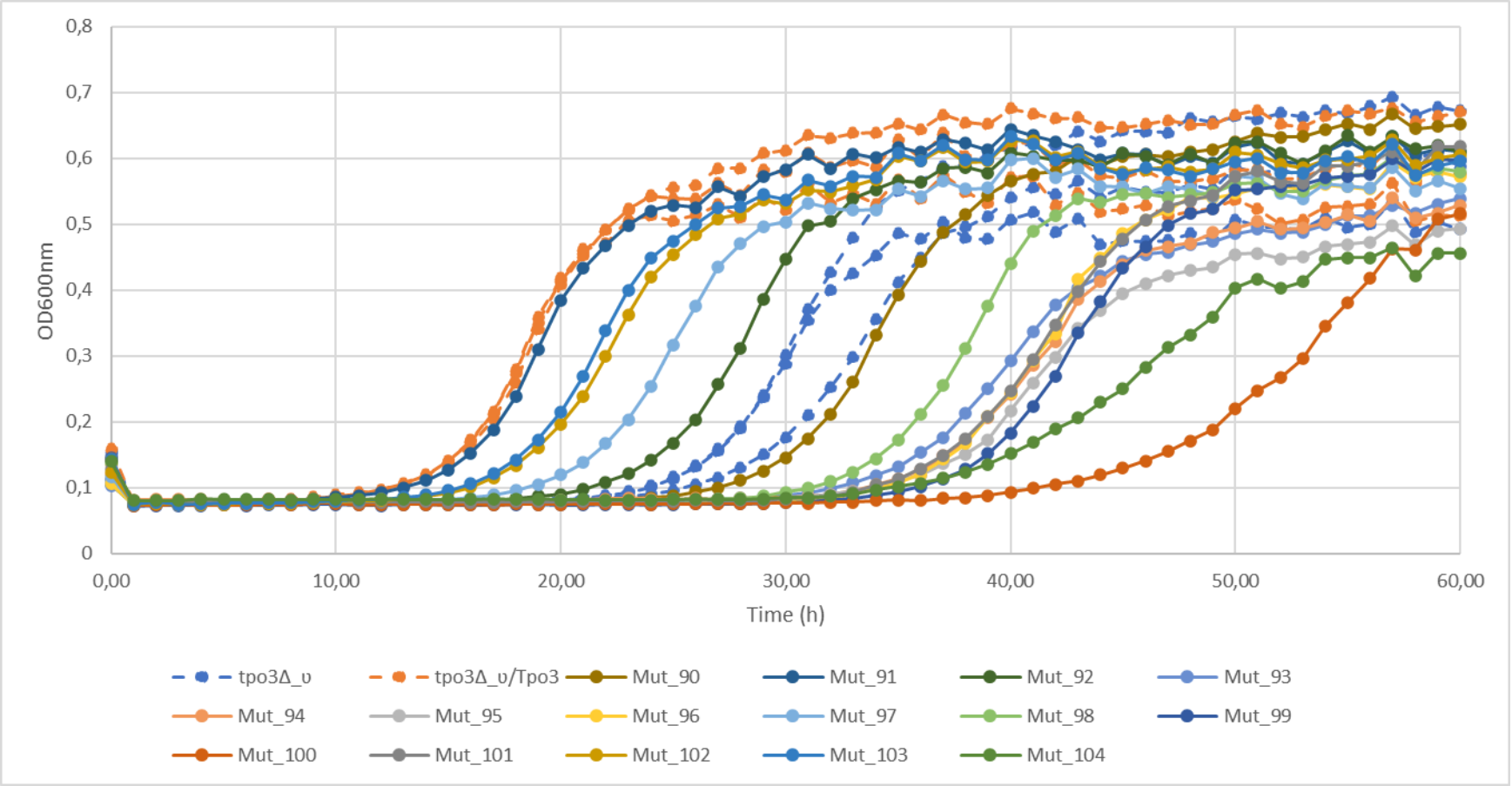

Figure D. 8. Growth curves of Mut_90 to Mut_104

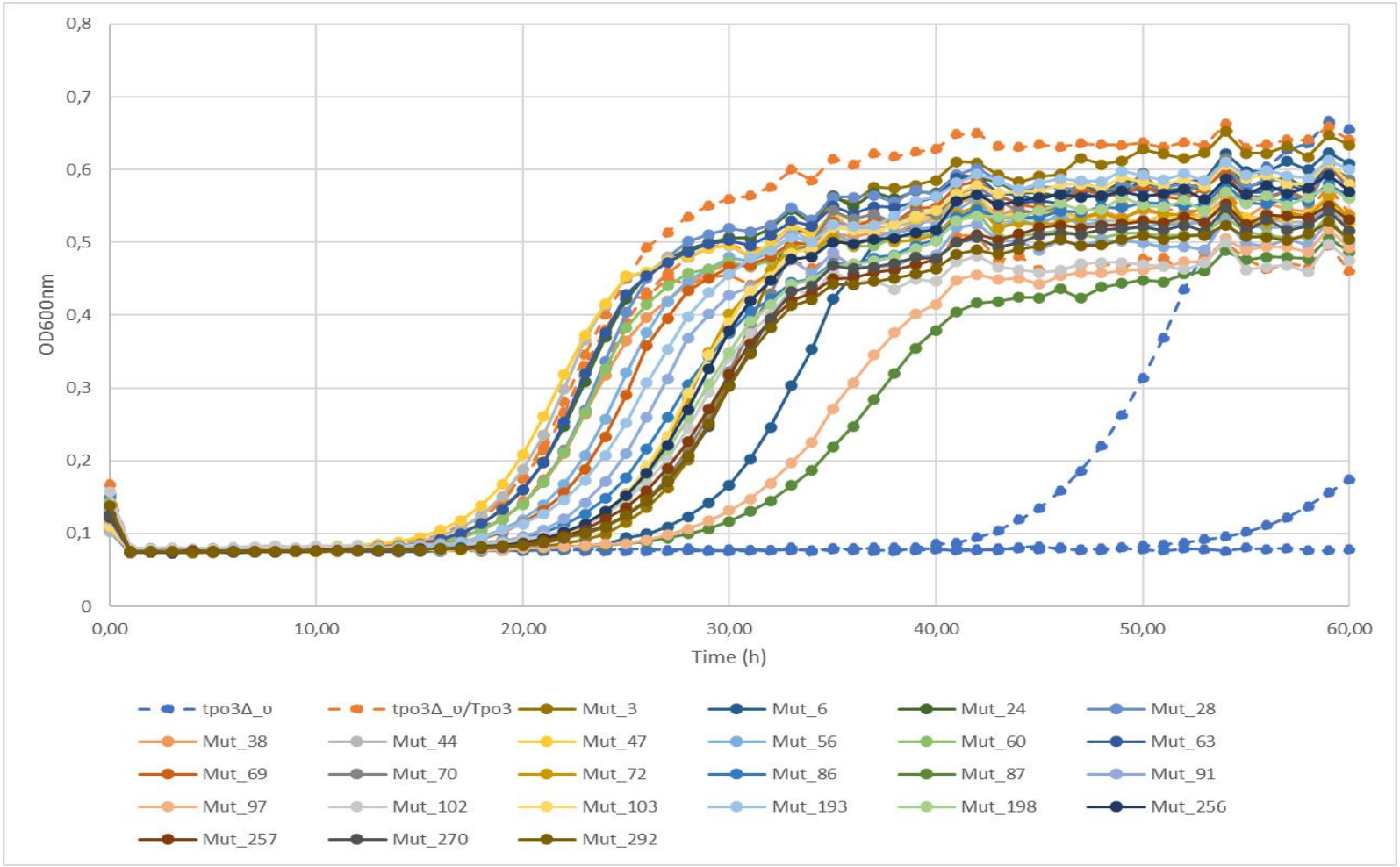

Figure D. 9. Growth curve of the strains harboring the *TPO3* artificial variants with highest tolerance to 90 mM acetic acid

Appendix E – Groups of strains used for dendrogram construction

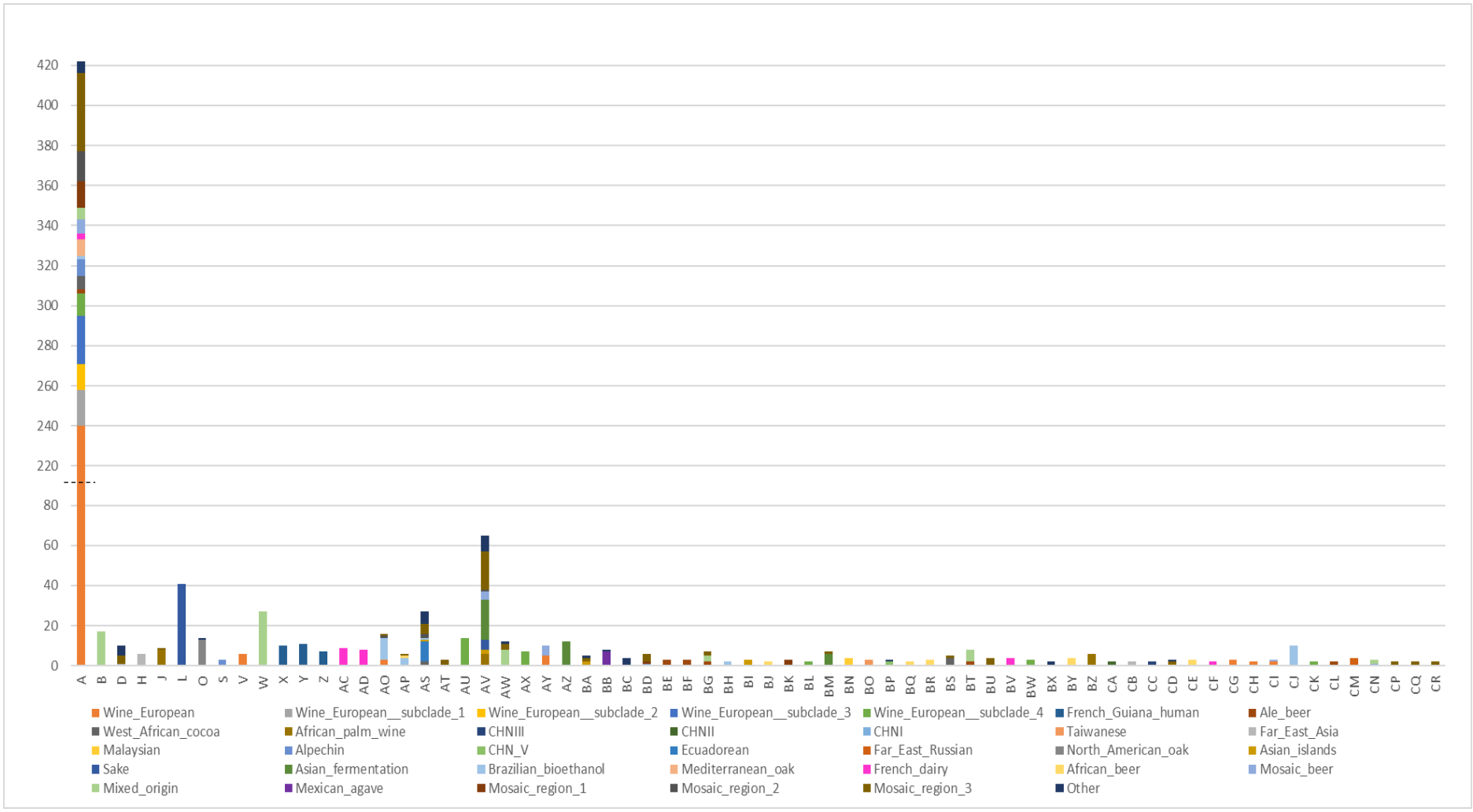

Figure E. 1. Clade composition of each group
Each group corresponds to strains with the same sequence of variants. These groups were formed after removal of variants composed by Ns in more than 1% of the strains and after the following removal of strains with Ns in more than 5 % of the remaining variants.