Characterization of transcriptional regulatory proteins in *Pichia pastoris*

Dissertation

zur Erlangung des Doktorgrades an der Universität für Bodenkultur Wien

Eingereicht von

Diane Barbay

Betreuer

Assoc. Prof. Dipl.-Ing. Dr. Brigitte Gasser

Department für Biotechnologie

Universität für Bodenkultur Wien

Wien, September 2019

ACKNOWLEDGEMENTS

First of all, I would like to thank Brigitte Gasser and Diethart Mattanovich for giving me the opportunity to do my PhD in their group as well as for their guidance and support throughout the project.

I am also grateful to my colleagues from the lab, for their help and for the nice working atmosphere. I especially thank Bojana Radoman, Sonakshi De, Domen Zavec, David Peña, Thomas Gaßler, Nora Tir, Özge Ata, for the fun discussions (and the occasional work-related conversations, of course) we had over the last years.

I also want to thank my two students: Daniela Unger who did a lot of growth and spotting assays, which helped me to focus on other parts of my project, and Monika Macakova who did a really good work during her master thesis on the further characterization of some transcription factors. Additionally, I would like to thank my BioToP colleague Leander Sützl who did the phylogenetic analysis of the Cat8/Sip4 homologs for me.

Thanks also to Ms. Burian, who was of great help for me during the past two years in overcoming various difficulties.

Finally, I would like to thank my family for supporting me throughout the years.

ABSTRACT

The methylotrophic yeast *Pichia pastoris* (syn. *Komagataella* spp.) plays a major role in the synthesis of recombinant proteins for biopharmaceutical or industrial purposes, but the production of a large number of complex secretory proteins at the desired high levels with the current industrial bioprocesses remains difficult. Cell engineering approaches using transcription factors were shown to be successful in increasing the recombinant protein productivity. However, many transcriptional regulators of *P. pastoris* are still uncharacterized, which means that their application in the design of new cell engineering strategies remains limited. In order to elucidate the function of yet uncharacterized transcription factors of *P. pastoris*, deletion and overexpression strains were generated for selected regulators. The growth characteristics as well as the colony and cell morphologies of the mutant strains were analyzed on different media representing various stress and growth conditions. Two transcription factors, Cat8-1 and Cat8-2, were further characterized. These two proteins were shown to be required for the growth of *P. pastoris* on ethanol. Cat8-1 is necessary for the activation of genes of the glyoxylate cycle, whereas Cat8-2 is necessary for the activation of genes of the carnitine shuttle. Both are required for activation of genes of the ethanol utilization pathway. Finally, the *CAT8-2* gene is repressed by the TF Mig1-2 on glucose, and autoregulated by the Cat8-2 protein on all carbon sources.

ZUSAMMENFASSUNG

Die methylotrophe Hefe *Pichia pastoris* (syn. *Komagataella* spp.) spielt eine wichtige Rolle bei der Herstellung großer Mengen rekombinanter Proteine für biopharmazeutische oder industrielle Zwecke. Viele komplexe sekretorische Proteine können jedoch bei industriellen Produktionsprozessen immer noch nicht in den gewünschten hohen Mengen hergestellt werden. Das konzertierte Engineering der zellulären Antwort durch Modulation von Transkriptionsregulationsproteinen hat sich als vielversprechender Ansatz erwiesen, um die Produktivität von *P. pastoris* zu steigern. Bisher ist jedoch eine große Anzahl der in *P. pastoris* vorhandenen Transkriptionsfaktoren (TF) uncharakterisiert, was bedeutet, dass ihre Anwendung für das gezielte Zellengineering begrenzt ist. Um die regulatorischen Ziele von noch nicht charakterisierten Transkriptionsregulatoren in *P. pastoris* zu beleuchten, wurden Transkriptionsfaktor-Überexpressions- und Knock-out-Stämme erzeugt und auf Medien kultiviert, die unterschiedliche Wachstums- und Stressbedingungen repräsentieren. Zusätzlich zur Wachstumsfähigkeit in Gegenwart oder Abwesenheit von Stressfaktoren und verschiedenen Nährstoffquellen wurde die Morphologie der Stämme auf ihrer makroskopischen und mikroskopischen Ebene berücksichtigt. Zwei Transkriptionsfaktoren, Cat8-1 und Cat8-2, wurden weiter charakterisiert. Es wurde gezeigt, dass diese beiden Proteine für das Wachstum von *P. pastoris* auf Ethanol erforderlich sind. Cat8-1 ist für die Aktivierung von Genen des Glyoxylat-Zyklus erforderlich, während Cat8-2 für die Aktivierung von Genen des Carnitin-Shuttles erforderlich ist. Beide werden zur Aktivierung von Genen des Ethanol-Verwertungswegs benötigt. Außerdem wird das *CAT8-2* Gen durch den TF Mig1-2 auf Glucose unterdrückt und durch das Cat8-2 Protein auf allen Kohlenstoffquellen autoreguliert.

CONTENTS

Introduction 5

I. Promoter architecture and promoter elements 5

II. Molecular mechanisms of transcription 7

II.1. Structure of RNA polymerase II 7

II.2. Transcription initiation with RNA polymerase II 8

II.3. Transcription elongation and termination 9

III. Structure and regulation of transcription factors 10

III.1. Classification of transcription factors 10

III.2. Structure of transcription factors 11

III.3. Regulation of transcription factor activity 12

IV. Mechanisms for transcriptional regulation 15

IV.1. Transcription factors in transcriptional regulation 15

IV.2. Transcription activation mechanisms 20

IV.3. Transcription repression mechanisms 21

V. *Pichia pastoris* in biotechnology 23

V.1. *Pichia pastoris*, a host for recombinant protein production 23

V.2. Cell engineering for improved heterologous protein production 24

VI. Aim of the Study 29

Material and methods 30

I. Strains, primers and plasmids 30

II. Sequence analyses, alignment and phylogenetic tree 30

III. Media 31

IV. Genomic DNA extraction and PCR 31

V. Construction of overexpression and knock-out strains 32

V.1. Overexpression 32

V.2. Knock-out 33

VI. Preparation of electro-competent *Pichia pastoris* cells 33

VII. Transformation of *Pichia pastoris* 34

VIII. Gene copy number determination 34
IX. Growth assay in liquid medium 35
X. Spotting assay 36
XI. RNA extraction and transcript levels analysis 36
XII. Construction of the HA-tagged Cat8-1 and Cat8-2 strains 36
XIII. Enhanced GFP reporter assay 37
XIII.1. Cloning of eGFP under the control of P_{CAT8-1} and P_{CAT8-2} 37
XIII.2. Reporter assay 37
XIII.3. Flow cytometry 38
XIV. Screening of eGFP production under P_{AOX1} in TF mutant strains 38
XIV.1. Cloning of eGFP under the control of P_{AOX1} 38
XIV.2. Screening procedure 38
XIV.3. Flow cytometry 39
XV. Microscopy 39
Chapter 1: Phenotypic characterization of transcription factors from *Pichia pastoris* 40
I. Introduction 40
II. Selection of TFs and generation of TF overexpression and knock-out mutants 40
II.1. Selection of transcription factors in *P. pastoris* 40
II.2. Generation of TF overexpression and knock-out TF mutant strains 42
III. Arraying the TF mutants on different media representing different growth conditions 43
III.1. Arraying of *P. pastoris* TF mutants on different stressors 43
III.2. Arraying of *P. pastoris* TF mutants on different carbon sources 46
III.3. Phenotype of *SWI5* mutant clones 49
IV. Influence of TF mutants on *AOX1* promoter activity 50
IV.1. *AOX1* transcript levels measurement 50
IV.2. Effect of *SWI5* overexpression on P_{AOX1} 51
V. Discussion 52
Chapter 2: The Cat8 TFs regulate carbon source utilization in *P. pastoris* 57
I. Introduction 57

II. Phylogenetic analysis of Cat8 and Sip4 homologs in *Saccharomycetes* 59
III. *CAT8-1* and *CAT8-2* are essential for ethanol assimilation in *P. pastoris* 62
IV. *CAT8-1* and *CAT8-2* expressions are higher on ethanol and methanol than on glucose and glycerol 64
V. Cat8-1 and Cat8-2 regulate the expression of genes important for ethanol assimilation 67
VI. The *CAT8-2* gene expression is autoregulated by the Cat8-2 protein 70
VII. The *CAT8-2* expression is regulated by the Mig1-2 transcription factor 72
VIII. Discussion 75
VIII.1. Cat8-2 is a Sip4 equivalent in *P. pastoris* 75
VIII.2. Role of Cat8-1 and Cat8-2 in the regulation of ethanol utilization 75
VIII.3. Cat8-1 and Cat8-2 possibly recognize different subsets of CSRE motifs 77
VIII.4. Mechanisms of Cat8 and Sip4 activation 78
Conclusion 80
Appendices 81
I. Appendix 1: List of primers used in this study 81
II. Appendix 2: List of plasmids generated and used in this study 93
III. Appendix 3: List of strains used in this study 97
IV. Appendix 4: Results of the spotting assay 99
V. Appendix 5: Results of the growth of the TF mutants on different carbon sources 105
VI. Appendix 6: List of proteins their identifiers (NCBI accession numbers) used for the phylogenetic analyses of Cat8-1 and Cat8-2 109
References 112
Curriculum vitae 124

INTRODUCTION

Production of recombinant proteins gained an increasing importance for the manufacturing of biopharmaceuticals and enzymes. The production is often performed in *Escherichia coli*, but mammalian cell lines and yeasts are also used as expression systems. Recombinant protein production in Eukaryotes is beneficial due to efficient protein folding and post-translational protein modifications. In particular, yeasts have been widely used because of their higher specific growth rate and the availability of molecular and genetic manipulation tools. Among industrially relevant yeast species, the methylotrophic yeast *Pichia pastoris* (syn. *Komagataella* spp.) has been widely employed as an expression system and is under intense scientific investigation in order to optimize its use as a recombinant protein production platform. The detailed understanding of genetic regulation and knowledge about related genes and factors in *P. pastoris* is of great value for various strain improvement strategies.

The function of gene expression is to synthesize of a functional gene product, which can be a non-coding ribonucleic acid (ncRNA) or a protein. RNAs are produced by RNA polymerases from information encoded in DNA during transcription. During translation, the messenger RNA (mRNA) code is converted to a protein by ribosomes. The regulation of transcription is an important process in all living organisms, since it enables cells to control the quantity of RNA being produced, thereby orchestrating gene activity. Overall, transcriptional control allows cells and organisms to respond to a variety of intra- and extra-cellular signals.

I. Promoter architecture and promoter elements

In Eukaryotes, transcription is a very well organized process involving an abundance of factors. One very important regulatory element of a gene is its promoter, which is usually located upstream of the coding sequence. Yeast promoters contain the core and the upstream activation sequence elements.

The core promoter is the site where the RNA polymerase II and the general transcription factors assemble in the pre-initiation complex (PIC) before the transcription initiation begins. The core promoter was first identified in mammalian promoters and was defined as “the minimum DNA element required for basal transcription” (Smale and Kadonaga, 2003).

A typical eukaryotic core promoter contains the following DNA sequence elements (**Figure 1**):

1. **The TATA element**, the recognition site for the general transcription factor TATA-binding protein (TBP)

2. **The initiator element (INR)**, the binding site of two independent proteins: a TBP associated factor (Taf) and the RNA polymerase II. INR directly overlaps the transcription start site (TSS).
3. **The downstream core promoter element (DPE)**, a recognition site for the Taf subunits of the TFIID coactivator.
4. **The TFIIB recognition element (BRE)**, located upstream of the TATA box and where the general factor TFIIB binds.

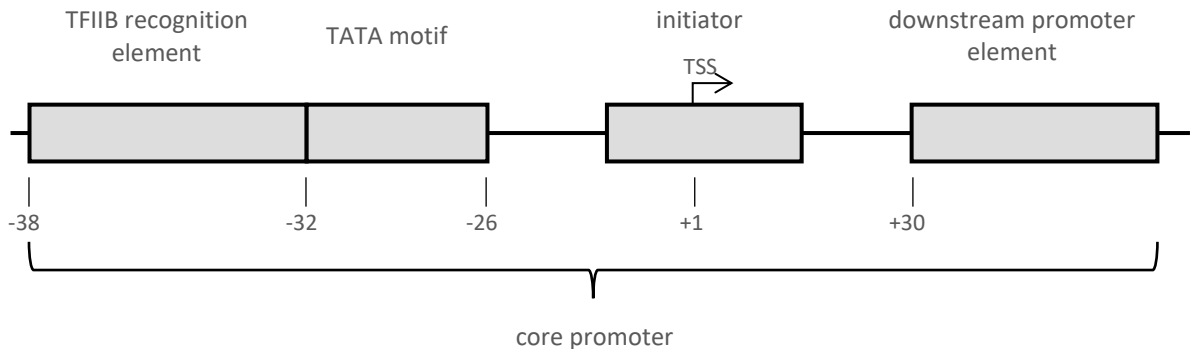


Figure 1: Sequence elements in a typical core promoter.
The different motifs presented here are not present in all core promoters: a specific core promoter can contain none, some, or all these elements. Positions of the motifs are given relative to the transcription start site (TSS, with +1). Adapted from (Carey and Smale, 2001).

All these core promoter elements are short, degenerate and low specificity elements. Their combination varies among promoters and determines activator and enhancer specificity. Out of these metazoan elements, the TATA box is the only one to be clearly conserved in yeasts. The TATA element is located between -40 and -120 bp of the transcription start site (Hampsey, 1998) and is present in only around 20% of all *Saccharomyces cerevisiae* promoters (Basehoar et al., 2004), but often with improper consensus sequence (Lubliner et al., 2013).

Other important promoter regions are enhancers and silencers, which are bound by transcription activators and transcription repressors, respectively. Metazoan enhancers can alter gene expression from positions up to 1 Mbp away from the gene, downstream or upstream from the transcription start site and independently of their orientation (Ogbourne and Antalis, 1998; Shlyueva et al., 2014). In yeast, regulatory sequences that are equivalent to enhancers and silencers are called upstream activation sequence (UAS) and upstream repression sequence (URS), respectively. UAS/URS function is generally orientation independent but must be located in 5' of the core promoter, except for few exceptions (Guarente and Hoar, 1984; Mellor et al., 1987). Most UASs are located in the nucleosome depleted region of promoters, or exposed on the surface of nucleosomes, probably because the transcription factors accessibility is negatively affected by the presence of nucleosomes (Albert et al., 2007).

In *S. cerevisiae*, promoter regions are generally nucleosome depleted because of sequence properties (G/C content, poly A/T sequences) (Bernstein et al., 2004; Lee et al., 2004). The nucleosome depleted region is found at active and inactive promoters and thus does not correlate with the transcription status. Some regulated promoters do not contain these uniformly nucleosome-depleted regions. In these promoters, the modulation of the chromatin structure contributes to the gene regulatory mechanism, and is usually dependent on the recruitment of remodeling factors by transcription activators.

II. Molecular mechanisms of transcription

Transcription, which takes place in the nucleus of eukaryotic cells, represents the first step of gene expression. It consists in the synthesis of RNA from a DNA template by the enzyme RNA polymerase, and it is performed in three sequential steps: initiation, elongation and termination.

Three different nuclear RNA polymerases are present in eukaryotic cells, and each of them is responsible for the synthesis of a distinct subgroup of RNA:

- RNA polymerase I produces large ribosomal RNAs (rRNAs)
- RNA polymerase II synthesizes messenger RNAs (mRNAs), most small nuclear RNAs (snRNAs) and microRNAs (miRNAs)
- RNA polymerase III produces transfer RNAs (tRNAs), small rRNAs and other small RNAs

II.1. Structure of RNA polymerase II

RNA polymerase II is the most studied of the three types of RNA polymerase in Eukaryotes. It consists in a 550 kDa complex containing 12 subunits, called Rpb1-12, numbered from largest to smallest (Cramer et al., 2008).

The two largest subunits of RNA polymerase II (Rbp1 and Rpb12), form the pore for the entry of the nucleotide triphosphates, the active site and the binding sites for the DNA and the DNA-RNA hybrid in the transcription elongation complex. Rpb3, Rpb11 and Rpb6 are important for the assembly and the stability of the polymerase. The other subunits are involved in interaction of the polymerase II with general factors, nucleic acids and coactivators (Werner and Grohmann, 2011).

RNA polymerase II also contains a repeated 7 residue motif (YSPTSTS) at the C-terminus of Rpb1, called the C-terminal domain (CTD) (Buratowski, 2009). This motif is targeted by different kinases phosphorylating serine residues at the positions 2, 5 and 7 of the repeat. The degree of CTD phosphorylation is strictly regulated during the different phases of transcription. These

modifications regulate the association of many factors with the RNA polymerase II, such as mRNA capping factors, chromatin modifiers, mRNA export and transcription termination factors (Akhtar et al., 2009; Feaver et al., 1994).

II.2. Transcription initiation with RNA polymerase II

Transcription initiation requires several steps. It starts by the formation of the pre-initiation complex (PIC) in the promoter. The PIC is composed of the general transcription factors and the RNA polymerase II. This complex then transitions from a closed promoter complex to an open promoter complex, in which around 15 base pairs of the promoter DNA are denaturated to form a transcription bubble. Next steps in the transcription initiation include transcription start site selection, *de novo* RNA synthesis and promoter escape (**Figure 2**) (Hahn, 2004; Liu et al., 2013; Sims et al., 2004).

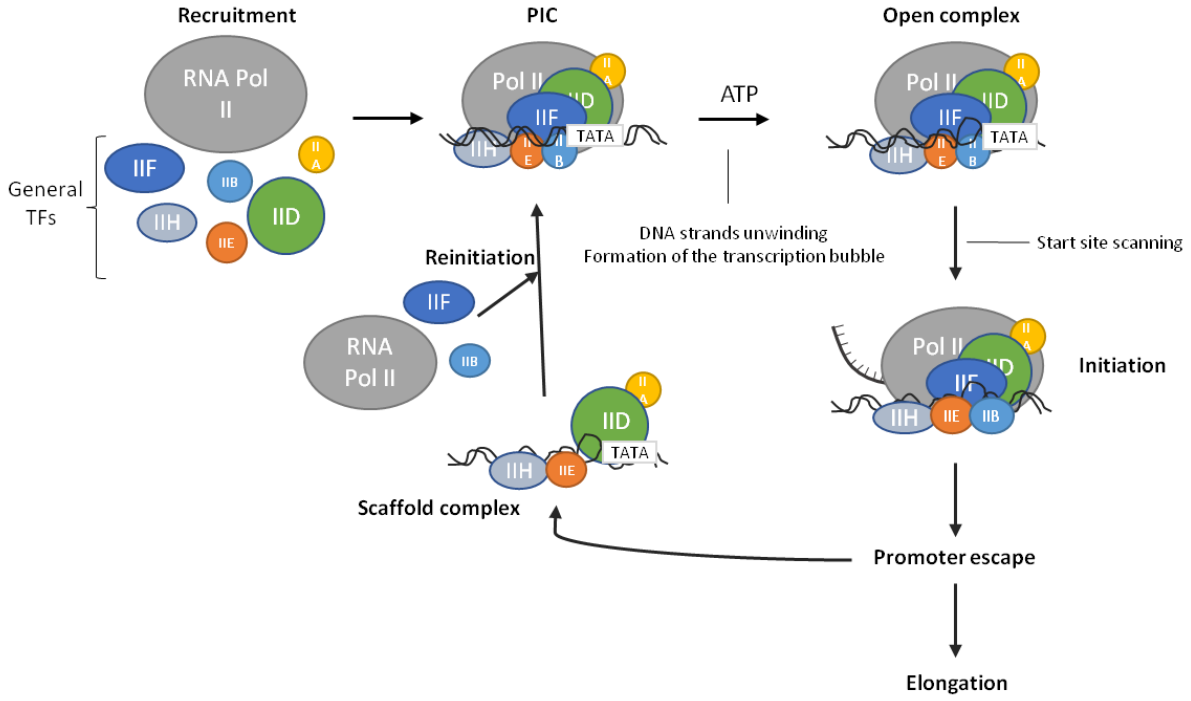


Figure 2: Steps of the transcription initiation with RNA polymerase II. Adapted from (Hahn, 2004).

FORMATION OF THE PRE-INITIATION COMPLEX

Transcription initiation starts with the recruitment of the transcriptional machinery to the core promoter. In this step, a subset of the general transcription factors recognizes the promoter DNA and forms a platform to recruit the RNA polymerase II. The first general TFs to form a complex with DNA are TBP, TFIIA and TFIIB. TBP (a subunit of TFIID) and TFIIB recognize the TATA and the BRE elements, respectively (**Figure 2**). TFIIA stabilizes the binding of TBP to DNA and

promotes binding of TFIID to the promoter. The RNA polymerase II and the other general TFs are then recruited to form the pre-initiation complex (PIC). The factors in the PIC are responsible for recognizing the core promoter, recruiting the RNA polymerase, and interacting with coactivators or repressors to modulate transcription. At this point, all the factors and the polymerase are bound to the promoter, but not in an active conformation for enabling the start of the transcription.

TRANSITION OF THE PIC TO THE OPEN COMPLEX

After the PIC is formed, it undergoes a transition from a closed promoter complex (with entirely double stranded promoter DNA) to an open promoter complex. In this open complex, the 11 to 15 bp of DNA around the transcription start site are denaturated, and the template strand for the synthesis of RNA is inserted within the active site of the RNA polymerase II. The general factors TFIIE and TFIIF play important roles in the separation and the stabilization of the promoter DNA strands during the transition to the open complex.

START SITE SCANNING

Once the open complex is formed, the RNA polymerase II scans the promoter downstream sequences to find a suitable transcription start site. The general TFs TFIIB and TFIIF as well as the polymerase are involved in the start site selection.

INITIATION OF TRANSCRIPTION

The initiation of the transcription can then start with the synthesis of the first phosphodiester bond of RNA. First, multiple short RNAs (3-10 bases) called abortive products are synthesized before the polymerase II initiates the transcription of the full length RNA product.

PROMOTER ESCAPE

After the synthesis of around 30 base pairs of RNA, the RNA polymerase II leaves the core promoter and the rest of the transcriptional machinery and enters the elongation stage. In this step, the factors promoting RNA synthesis, processing and export as well as the chromatin modifiers can be recruited to the polymerase. After the transcription initiation, many of the general TFs remain at the promoter and form the scaffold complex. This complex seems to mark the genes that have been transcribed and enables to bypass the relatively slow step of factors recruitment in further rounds of transcription.

II.3. Transcription elongation and termination

After initiation, the RNA polymerase II gets new factors for the transcription elongation. During this step, the double stranded DNA is melted by the polymerase so the template strand is available

for RNA synthesis. The DNA strands and the nascent RNA product exit the polymerase through different channels, and the two DNA strands reunite as double helix at the end of the transcription bubble.

Once elongation of the product RNA is finished, transcription enters its final phase: termination. During this step, the complete RNA transcript, the DNA and the RNA polymerase dissociate. The exact process differs for each polymerase, and is the least understood of the three transcription stages. Transcription termination is followed by capping of the 5' end of the mRNAs, splicing by the spliceosome (a RNA-protein complex) for intron-containing transcripts and maturation as well as polyadenylation of the 3' end of the transcripts. Finally, mRNAs are exported through nuclear pores and can translocate to other cellular compartments or stay in the cytoplasm and undergo translation.

All three transcription steps are subject to regulation. In particular, transcription initiation is the primary level at which gene expression is regulated. Initiation can be regulated by trans-acting elements such as transcription factors. It can also be regulated by targeting the movement of the polymerase during elongation.

III. Structure and regulation of transcription factors

Transcription factors are DNA binding proteins that are able to regulate gene expression by stimulating or diminishing transcription. Since TFs play a critical role in the regulation of gene expression, they are present in all living organisms. The amount of TFs present in a given organism increases with its genome size, so a bigger genome usually results in more TFs per gene (van Nimwegen, 2003).

III.1. Classification of transcription factors

Transcription factors are grouped into classes depending on the DNA-binding domain (DBD) they possess. The three main classes of transcription factors are:

1. **Zn^{+2}-stabilized DNA binding domain**, the most abundant class in all organisms (Krishna et al., 2003). This class can be further subdivided into:
    a. Cys_2His_2 zinc fingers (Böhm et al., 1997) (*e.g.* Adr1, Mig1), whose DNA binding domain is formed by a ~30 amino acids sequence folded around a central zinc ion. Cys_2His_2 zinc finger TFs usually bind DNA as monomers with each finger recognizing consecutive triplets of bases.
    b. Cys_6 proteins, also called zinc knuckle or Zn_2Cys_6 binuclear zinc cluster proteins (MacPherson et al., 2006) (*e.g.* Gal4), are unique to fungi. They have a DBD

containing two zinc ions bound to six cysteine residues. The DBD is in most Cys_6 TFs located in N-terminus of the protein. These TFs usually bind as dimers to symmetric DNA sites, using a dimerization domain located C-terminal to the DBD.

c. Cys_4 or GATA fingers, whose DBD is formed by a four-cysteine zinc finger and an adjacent basic region, bind to six base-pair long DNA sequences with the consensus HGATAR. In *S. cerevisiae*, GATA factors comprise proteins mainly involved in nitrogen metabolism and mating-type switching (*e.g.* Gln3, Ash1) (Scazzocchio, 2000).

2. **Zipper type**, defined by a DBD containing a dimerization motif and a basic region. This class can be further divided into:
   a. bZIP, or basic leucine zippers proteins (*e.g.* Gcn4, Yap1), defined by a basic DNA binding region followed by a leucine zipper motif (**Figure 4C**). bZIP can form hetero- or homodimers in metazoans, but they mostly function as homodimers in yeast (Reinke et al., 2013). bZIP regulators can be further divided in several subfamilies depending on their protein sequences and DNA binding characteristics (Jindrich and Degnan, 2016).
   b. bHLH, or helix-loop-helix proteins (*e.g.* Ino2, Pho4), who have a basic region preceded by a loop of variable length that separates two α-helices (Robinson and Lopes, 2000). These proteins generally bind DNA as heterodimers, generating a multitude of different complexes.
3. **Helix-turn-helix (HTH)** (*e.g.* Matα1, Matα2, Mata1) forming both homo- and hetero dimers.

III.2. Structure of transcription factors

Transcription factors usually contain multiple domains (**Figure 3**) (Latchman, 1997):

1. A **DNA-binding domain (DBD)**, which binds to specific sequences of DNA.
   Some proteins classified as TFs, such as Met4 and Swi6, lack a DBD motif and interact with DNA via a binding partner (Lee et al., 2010; Sidorova and Breeden, 1993). Some other TFs have an optional DBD, possibly because they are able to form heterodimers with another subunit that mediates the DNA binding (Bricmont et al., 1991; Scott et al., 2000; Tornow et al., 1993).
2. An **effector domain,** usually including an activation/repression domain (AD/RD) containing binding sites for other proteins such as transcriptional coregulators. Activation domains are specific regions needed for activation of transcription. Repression domains have been found in many TFs to have a negative effect on gene

expression. Some TFs have the ability, depending on the growth conditions, the chromatin context and the promoter, to both activate and repress transcription (Piña et al., 2003; Polish et al., 2005).

3. An optional **signal-sensing domain (SSD)**, which senses and transmits external signals to the rest of the transcription complex, resulting in a modification of gene expression.
4. A **nuclear localization sequence (NLS)** and an optional **nuclear export signal (NES)**, which interact with soluble receptors recognizing specific nucleoporins and enabling translocation of the protein into the nucleus or from the nucleus to the cytoplasm (Hahn et al., 2008; Turpin et al., 1999). Indeed, regulating the access of a TF to the nucleus is a common way to control transcription, as it will be discussed later.

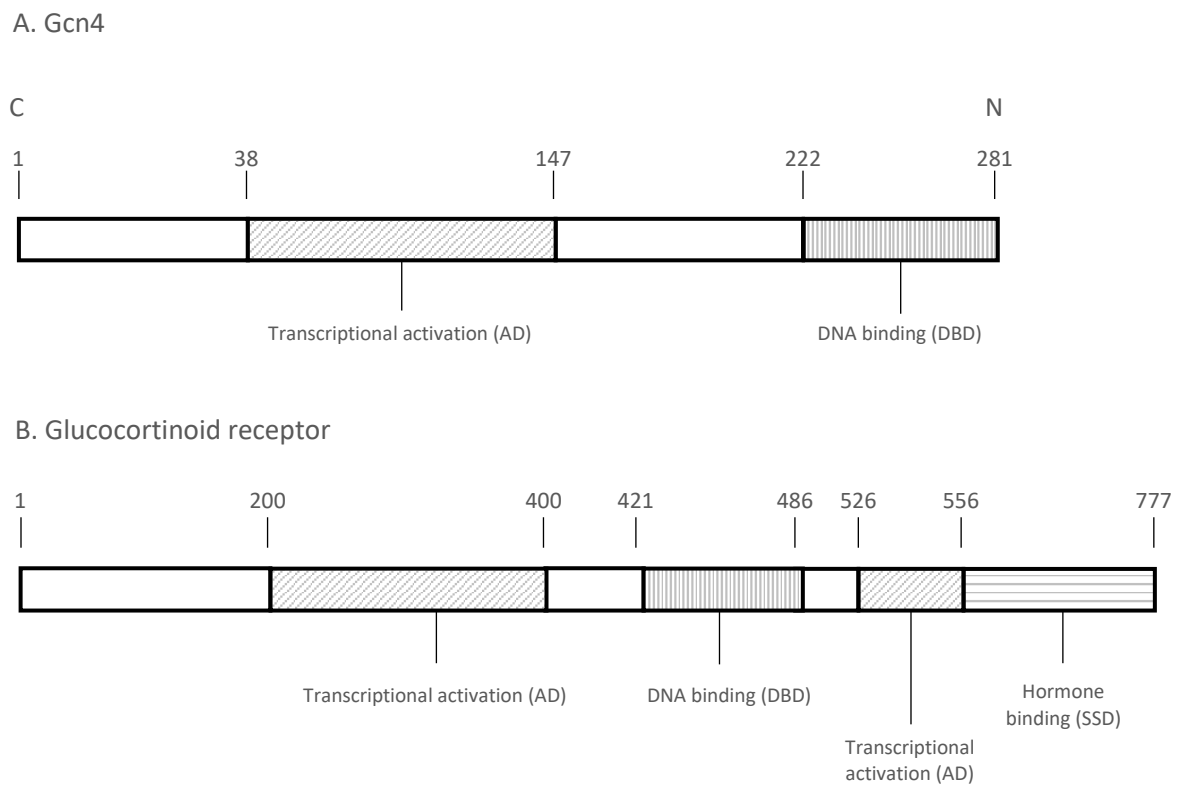


Figure 3: Structure of (A) the yeast Gcn4 factor, and (B) the mammalian glucocortinoid receptor indicating the distinct regions mediating DNA binding and transcriptional activation. Positions are given relative to amino acids sequence. Adapted from (Latchman, 1997).

III.3. Regulation of transcription factor activity

In order to express the correct sets of genes under specific conditions, transcription factor activity must be controlled by an upstream layer of gene regulation. This can be achieved by regulating TF gene expression, and by various post-translational events affecting the TF localization, conformation or activation.

REGULATION BY DIRECT BINDING OF A LIGAND

Transcription factors can be activated or inhibited by binding of a ligand. This ligand can either be a protein (as for Gal4), a small molecule such as a metabolite (in the case of *e.g.* Leu3 and Put3) or a metal ion (*e.g.* for Zap1).

When the ligand is a small molecule: Zap1 and Zn^{2+}

In *S. cerevisiae*, the TF Zap1 regulates the expression of various genes in response to zinc deficiency. Zap1 activates the expression of its target genes when cellular zinc levels are low. It is an 880 amino acids protein that contains a DNA binding domain consisting of five zinc fingers located at the C-terminal domain of the protein. Zap1 also has two activation domains, AD1 and AD2, mediating the increased transcription of target genes. When zinc levels are high, the metal ion binds to both AD1 and AD2, which inhibits the ability of the two domains to activate transcription (Bird et al., 2000; Wu et al., 2008).

When the ligand is a protein: Gal4 and Gal80

Gal4 is a transcription factor involved in the regulation of the Gal genes, which are required for the growth of *S. cerevisiae* on galactose. Under non-inducing conditions, when no galactose is present, the activity of Gal4 is blocked by the interaction with the protein Gal80 which occludes the C-terminal activation domain of Gal4. Hence, coactivators of transcription such as TBP and TFIIB cannot be recruited. Galactose is sensed by Gal3, which binds and inactivates Gal80 in the presence of this compound (Traven et al., 2006).

REGULATION BY CHAPERONES

Another mean of regulating the activity of some transcription factors is achieved by using chaperones.

Hap1 is an important oxygen sensor in *S. cerevisiae*. It has been showed that in this yeast, intracellular heme levels correlate with oxygen levels of the environment. In response to heme, and thus to oxygen levels, Hap1 activates the transcription of genes involved in respiration and control of oxidative damages. Hap1 contains two classes of elements required for its regulation by heme: repression modules (RPMs) and heme responsive motifs (HRMs) (**Figure 4**).

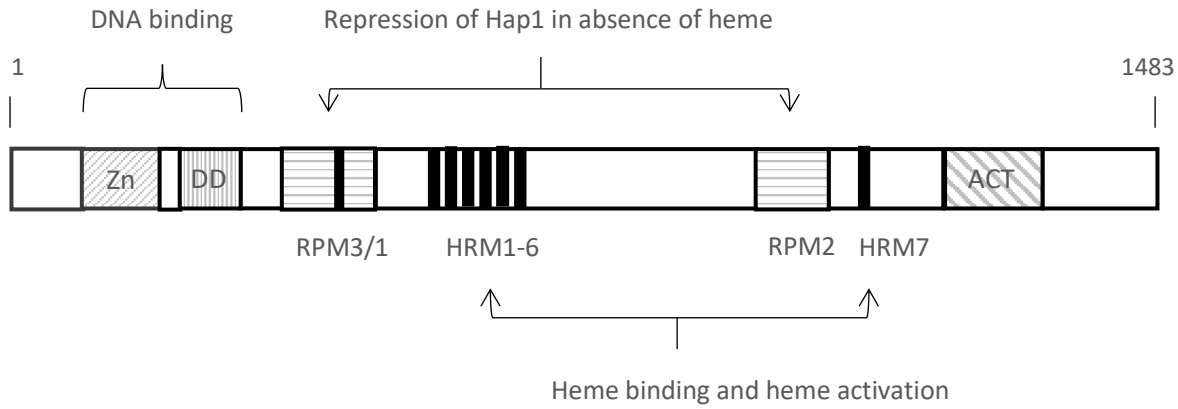


Figure 4: Structure of Hap1. Zn: Zinc cluster, DD: dimerization domain, RPM3/1, RPM2: repression modules, HRM1-6, HRM7: heme responsive motifs, ACT: activation domain. Zn and DD mediate the DNA binding. Two classes of Hap1 elements mediate heme regulation: RPMs mediate repression of Hap1 in absence of heme, HRMs mediate heme-binding and heme activation of Hap1. Positions are given relative to amino acids sequence. Adapted from (Lee and Zhang, 2009).

At low heme levels, the chaperone Hsp70 and co-chaperones Ydj1 and Sro9 bind to RPMs and keep Hap1 in an inactive conformation. At high heme levels, heme binds to HRMs, which promotes a conformation change leading to the recruitment of Hsp90 to Hap1. Interaction between Hap1 and Hsp90 causes further conformational changes in the Hap1-chaperone complex, which induces activation of Hap1 (Lee and Zhang, 2009).

REGULATION BY COVALENT MODIFICATIONS

Another way to regulate transcription factor activity involves alteration of the proteins by covalent modifications such as phosphorylation, acetylation, glycosylation or methylation.

An example of such regulation is found for the transcription factor Adr1, which activates the expression of genes essential for the growth of *S. cerevisiae* on non-fermentable carbon sources such as glycerol, ethanol and lactate. Under repressing conditions (in presence of glucose) Adr1 is phosphorylated at Ser230 by an unknown protein kinase. The 14-3-3 protein (Bmh) inhibits Adr1 activation function by binding to the Ser230 phosphorylated regulatory domain: this binding event occurs at the promoters where Adr1 binds and prevents the formation of the PIC (Braun et al., 2013; Parua et al., 2014).

REGULATION BY CONTROLLING THE INTRACELLULAR AMOUNT OF TRANSCRIPTION FACTOR

Some pathways can be controlled by altering the transcript levels of the TFs instead of their activity.

In *S. cerevisiae*, Gcn4 is a transcription activator of genes expressed during amino acid starvation. It is also involved in the regulation of genes involved in purine biosynthesis, autophagy and multiple stress responses. Levels of Gcn4 in the cell are tightly regulated: under non-starvation

conditions, Gcn4 is rapidly degraded by the ubiquitin pathway, whereas under amino acids starvation, its half-life is increased (Hinnebusch, 1997).

REGULATION BY CONTROLLING THE CELLULAR LOCALIZATION

Finally, the activity of some transcription factors is regulated by their cellular localization. This is the case for TFs such as Swi5 or Yap1.

In *S. cerevisiae*, Swi5 is involved in the recruitment of Mediator and Swi/Snf complexes to DNA and in the activation of expression of certain genes at the M/G1 phase boundary and the G1 phase. Swi5 is also required for the expression of the *HO* gene, controlling mating-type switching. The activity of Swi5 is regulated by its cellular localization, which is itself regulated by phosphorylation. During anaphase and early telophase, Swi5 is phosphorylated, which prevents its entry into the nucleus. Therefore in these phases, Swi5 stays in the cytoplasm. In the late telophase, Swi5 is dephosphorylated by Cdc14, which allows its entry in mother and daughter cells nuclei. Since Swi5 lacks a nuclear export signal, it remains in the nucleus until the early G1 phase, when it gets degraded (Weiss, 2012).

For many transcription factors, the tight regulation of their activity is achieved by using combinations of the previously mentioned mechanisms. This is the case for Swi5, as seen above, for which dephosphorylation of specific sites (regulation by covalent modifications) triggers entry into the nucleus (regulation by cellular localization).

IV. Mechanisms for transcriptional regulation

Transcriptional regulation is a critical process in all living organisms. It is coordinated through a variety of mechanisms involving transcription factors and other proteins to precisely adjust the amount of RNA being synthesized.

IV.1. Transcription factors in transcriptional regulation

TRANSCRIPTION FACTORS RECOGNIZE SPECIFIC DNA SEQUENCES

TFs typically recognize 6-12bp long degenerate DNA sequences. Other rules than simple affinity of an individual TF for DNA are involved in binding events: TF-DNA binding specificity is in fact influenced by various characteristics, including the nucleotide sequence, the three-dimensional structure and flexibility of the TFs and their binding sites, the cooperative DNA-binding of different TFs, and the chromatin accessibility.

The preferential binding of TFs to specific DNA sequences is achieved by two protein-DNA recognition mechanisms: the base readout and the shape readout. The base readout describes the

preference for a given nucleotide at a specific position by physical interactions between the amino-acid side chains of the TF and the accessible functional groups of the bases. The shape readout states that TFs can recognize the sequence-dependent DNA shape of their binding sites, such as the DNA bending and unwinding, which contributes to their sequence specificity (Rohs et al., 2010). Recent structural studies showed that most proteins use the interplay of base and shape readout to recognize their specific binding sites. The contribution of base and shape readout, however, varies across protein families (Slattery et al., 2014).

FUNCTIONAL AND NON-FUNCTIONAL BINDING EVENTS

The study of *in vivo* TF-DNA binding can be achieved by genome-wide chromatin immunoprecipitation combined with sequencing (ChIP-seq) and related approaches. Some ChIP studies showed that TF binding is extremely broad: from hundreds to tens of thousands binding events were observed for several TFs in different model organisms, which considerably exceeds the number of characterized or potential direct target genes (MacQuarrie et al., 2011). Therefore, it seems that binding of a TF does not necessary lead to regulation, and only a small fraction of all binding events might have an impact on gene expression. Indeed, a comparison of TF binding events and expression profiling data (*i.e.* genes being differentially expressed when that TF is non-functional) in yeast revealed that around 50% of genes whose promoter region is bound by a TF are true regulatory targets of that TF (Gao et al., 2004; Ucar et al., 2009). Some studies also indicate that a reasonable fraction of TF binding events may be neutral or non-functional and only reflect chromatin availability (John et al., 2008), while other studies indicate that although some binding events do not lead to a response in gene expression, they may have a functional role in chromatin remodeling or nucleosome positioning (Buck and Lieb, 2006).

MECHANISMS MODULATING BINDING OF TRANSCRIPTION FACTORS

Chromatin accessibility and transcription factors binding

Nuclear DNA is associated in nucleosomes, which are constituted by approximately 147 bp of DNA wrapped around a histone octamer. Nucleosomes facilitate packaging of the DNA in the nucleus and are involved in various regulatory mechanisms: histones can undergo a large number of post-translational modifications which regulates chromatin compaction and affects recruitment and binding of transcriptional regulators (Bai and Morozov, 2010).

TFs were classified into three categories depending on their DNA binding strategies and their impact on DNA accessibility: pioneer, settler and migrant TFs (Magnani et al., 2011; Sherwood et al., 2014; Zaret and Carroll, 2011).

1. **Pioneer TFs** are able to bind inaccessible nucleosome-associated DNA sites and to promote DNA accessibility by creating an open chromatin environment permissive to binding of non-pioneer TFs (**Figure 5A**).
2. **Settler TFs** bind to their DNA target sites only if these sites are in accessible DNA, they cannot bind to inaccessible DNA (**Figure 5B**).
3. **Migrant TFs** only bind to a fraction of their target sites, even if these sites are in accessible DNA. The site selectivity of migrant TFs is probably driven by interactions with extra cofactors (**Figure 5C**).

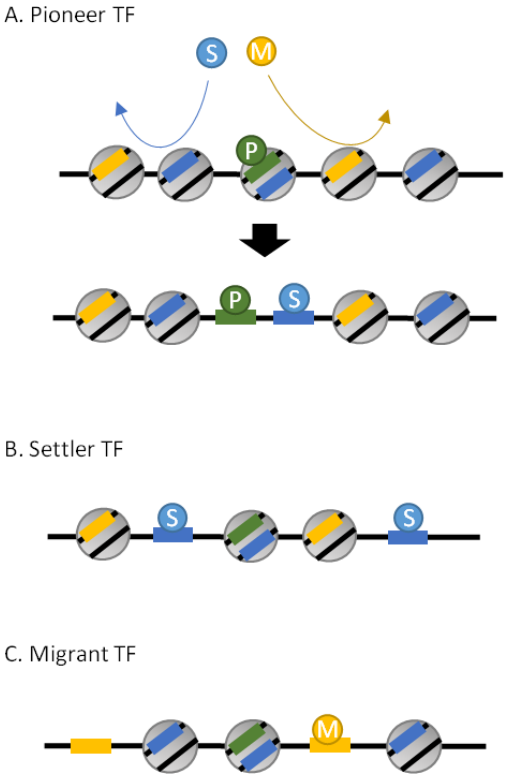


Figure 5: TF-DNA binding strategies.
(A) Pioneer TFs (P, green) bind to inaccessible nucleosome-associated DNA sites and create an environment permissive for the binding of non-pioneer TFs (settler or migrant TFs). (B) Settler TFs (S, blue) bind to all accessible copies of their DNA binding sites. (C) Migrant TFs (M, yellow) bind to only a subset of their accessible DNA binding sites. It seems that DNA accessibility substantially contributes to DNA binding selectivity of most TFs (except pioneer TFs). Adapted from (Slattery et al., 2014)

It was shown that the TF occupancy patterns *in vivo* correlate with nucleosome-depleted regions (Liu et al., 2006). However, it is generally not well-known whether a particular chromatin state permits TF binding, actively directs it or is the result of it. Some studies showed that nucleosome displacement was important for TF occupancy at low affinity motifs, or that chromatin remodeling is required before TF binding in some cases (Buck and Lieb, 2006; John et al., 2008). However, the removal of some TFs results in a decreased size of nucleosome depleted regions (*e.g.* Rap1, (Yarragudi et al., 2004)), and the binding of other factors is sufficient to disrupt a well-

positioned nucleosome (*e.g.* Gal4, (Morse, 1993)). The relationship and interactions between TFs and nucleosomes thus seem to be complex and reciprocal.

Interactions of transcription factors at genomic regulatory regions

Architecture and arrangements of transcription factor binding sites have an effect on the cooperation between different TFs, and the integration of multiple TF inputs is important to direct precise patterns of gene expression. Two properties of the promoter are particularly important for cooperation of TFs:

1. The **motif composition (motif grammar)**, characterized by the presence within promoters of binding sites for specific TFs that are essential for activating or repressing transcription.
2. The **motif positioning**, which corresponds to the orientation, the relative order and the spacing of TF motifs within the promoter. The motif positioning ensures that TFs are arranged appropriately to facilitate interactions and to promote cooperative binding as well as recruitment of cofactors or the transcriptional machinery.

Different models are used to describe assembly and cooperation of TFs at promoters:

1. The **enhanceosome model** proposes that all TFs binding to an enhancer are essential for the occupancy and activation of the enhancer. In this model, the composition of TF-DNA binding sites and their positioning relative to each other (motif grammar) act as a scaffold to recruit all TFs, which form an ordered protein interface to regulate transcription. The enhanceosome assembly does not tolerate alterations in the motif grammar, since it could disrupt protein-protein interactions and cooperativity. True enhanceosome regulations are not common: they may be necessary only under specific regulatory conditions, such as for amplification of signals at promoters regulated by low-abundance TFs (Escalante et al., 2007; Thanos and Maniatis, 1995).
2. The **billboard model** proposes that TFs act cooperatively to direct precise patterns of gene expression, but that their recruitment does not depend on many constraints on the relative positioning of their binding sites. In this model, a subset of TFs bind to the enhancer cooperatively, while other TFs bind in an additive or independent manner (Menoret et al., 2013; Slattery et al., 2014).
3. The **"TF collective" model** proposes that TFs bind to their target regions in an "all-or-nothing" way, the binding being directed by the collective action of many TFs. In this model, the motif grammar is flexible and protein-protein interactions are very important (Junion et al., 2012).

The means by which various different TFs assemble on promoters probably fall on a continuum between the different models presented above. In addition, different TF binding properties lead to diverse types of transcriptional outputs, depending on how TFs interact with each other. Non-cooperative TF binding (or additive TF binding) is better suited for regulating graded gene expression: enhancer activation is proportional to the concentration of the individual TFs, which is often necessary for homeostatic responses. Cooperative binding is more appropriate for switch-like, on/off mode of enhancer activity, which is often seen in developmental contexts. This mechanism allows the same set of TFs to function at different concentration ranges, and can buffer variation in levels of individual TFs.

Indirect cooperation between transcription factors

Many cooperative binding events involve different TFs bound to adjacent DNA sites and having direct protein-protein interactions. However, other indirect modes of TF cooperativity exist:

1. **Transcriptional synergy**: two or more TFs co-bound to adjacent binding sites can recruit a common cofactor or different components of a multiprotein complex, which may increase the affinity of each TF for its binding site, or may increase the retention time of each TF at its binding site (Merika et al., 1998).
2. **Activation of chromatin remodeling**: some TFs can act cooperatively by activating chromatin remodeling. Pioneer TFs, as described before, can bind to inaccessible nucleosome-associated TFBSs and promote DNA accessibility for other TFs by creating an open chromatin environment (Biddie et al., 2011).
3. **Collaborative competition**: co-expression of two TFs competing for the same binding site can lead to an increase occupancy of each TF: the alternating binding of each TF might counteract the nucleosome repositioning (Miller and Widom, 2003).
4. **DNA bending**: binding of a TF can trigger local DNA bending, which can increase affinity of other TFs for adjacent binding sites (Falvo et al., 1995).

ROLES OF TRANSCRIPTION FACTORS IN TRANSCRIPTIONAL REGULATION

Transcription factors bind to promoter DNA and regulate gene expression by stimulating or suppressing transcription. Specifically, TFs can bind to UAS or URS of promoters of the genes they regulate and use different mechanisms to alter gene expression:

1. Some TFs can either strengthen or prevent the binding of RNA polymerase to the promoter, which respectively enhances or represses the transcription.
2. Other TFs are able to recruit protein complexes that will activate or repress the transcription.

3. Finally, some TFs may catalyze or recruit some catalytic enzymes to modify histones. This leads to changes in the chromatin structure, which influences gene expression.

More details on the role of transcription factors in transcription regulation is discussed below.

IV.2. Transcription activation mechanisms

Transcription activation can be modulated by different mechanisms, including by stimulating the formation of the PIC, modifying the chromatin structure in the promoters and post-initiation mechanisms (Hahn, 1998; Struhl, 1999).

STIMULATION OF THE FORMATION OF THE PIC

The recruitment of coactivators and general TFs to promoters increases the assembly of the PIC and is probably a major means of transcription activation (Green, 2005). Different studies using chromatin immunoprecipitation (ChIP) to investigate the level of factors at promoters before and after gene activation showed that the number of coactivators and general TFs increases at promoter regions upon transcriptional stimulation. It seems that transcription activation by recruitment is a mechanism involved in the transcription stimulation of nearly all RNA polymerase II transcribed genes.

CHROMATIN MODIFICATION AND REMODELING

Two types of chromatin are usually distinguished, corresponding to different levels of chromatin compaction: euchromatin and heterochromatin. Euchromatin is sometimes referred to as decompacted chromatin, as it contains most genes that are expressed in the cell. In contrast, heterochromatin contains regions of highly compacted chromatin, with very few active genes. The alteration of the chromatin structure into a repressive state can be achieved by modifying histones: transcriptional activation is generally associated with histone acetylation, whereas transcriptional repression is usually associated with histone deacetylation (Struhl, 1999).

Three classes of proteins associated with the RNA polymerase II are involved in chromatin remodeling: histone-modifying enzymes, chromatin-binding proteins and ATP-dependent nucleosome remodeling proteins. Activators can recruit one or more of these proteins to the promoter, which results in chromatin remodeling. However, the factors modifying the chromatin structure are not by themselves sufficient for transcription activation: artificial recruitment of TFs with only chromatin-remodeling function did not stimulate transcription (Green, 2005). It is thought that the changes in the chromatin structure regulate transcription by altering the accessibility of TFs, RNA polymerase II and some components of the PIC to the DNA.

ENHANCEMENT OF THE STEPS OCCURRING AFTER THE FORMATION OF THE PIC

Post-initiation mechanisms have been well described in higher Eukaryotes. For example, RNA polymerase II pausing shortly after transcription initiation (stalling) was shown to be an important mechanism in gene regulation in mammalian and insect cells (Buratowski, 2009; Core et al., 2008; Zeitlinger et al., 2007).

IV.3. Transcription repression mechanisms

As previously mentioned, transcription factors either activate or repress promoter activity and thereby control transcription initiation.

There are two types of transcriptional repression: general / global repression and gene-specific repression. In general repression, a repressor protein or complex sequesters or modifies a central component of the PIC or RNA polymerase II, so that it is unavailable for transcription. Therefore, general repression down-regulates expression of all genes transcribed by this RNA polymerase. In the case of gene-specific repression, the transcription of a particular gene or set of genes is controlled by a gene-specific repressor or co-repressor.

The repressor proteins involved in the different repression mechanisms constitute a large group of diverse proteins that negatively regulate transcription, and they are not easily defined. Some repressors are TFs that negatively regulate the transcription of specific genes. Some repressors are proteins that do not directly bind DNA, but are rather recruited to promoters by other proteins. Finally, some repressors do not bind to DNA, directly or indirectly, but can for example bind to target activators or components of the PIC (Gaston and Jayaraman, 2003).

Three categories of repression mechanisms have been identified (Herschbach and Johnson, 1993): inhibiting the basal transcriptional machinery, suppressing the activator function and remodeling chromatin.

REPRESSION VIA INHIBITION OF THE BASAL TRANSCRIPTIONAL MACHINERY

Targeting of the transcriptional machinery can result in a global shut-down of transcription. Different mechanisms have been described to inhibit the basal transcription machinery: modifying the RNA polymerase II large subunit, or inhibiting the binding of TBP (a subunit of TFIID) to DNA.

1. **Modification of the RNA polymerase II large subunit**: the C-terminal domain (CTD) of a large subunit of the RNA polymerase II can be a target for repression. In yeast, the kinase Srb10 (also termed Ssn3) represses the transcription of sets of genes involved in meiosis and sugar utilization by phosphorylating the CTD before association of the RNA

polymerase with the promoter. This inhibits the formation of the PIC and hence inhibits the transcription initiation (Hengartner et al., 1998).

2. **Inhibition of the binding of TBP to the TATA box**: many eukaryotic promoters do not contain a TBP binding site (TATA box), but TBP is found at both TATA containing and TATA-less promoters.

   The Mot1 protein (Dasgupta et al., 2002) was described to repress the expression of specific sets of genes. It can interact with TBP, which blocks the binding of TBP/TFIID to DNA and prevents the assembly of the PIC and thus the transcription initiation (Moyle-Heyrman et al., 2012).

REPRESSION VIA ABLATION OF ACTIVATOR FUNCTION

Some transcriptional repressor proteins can regulate the activity or location of transcription activators/co-activators, which influences transcription initiation. This can be achieved by various mechanisms, including regulating the turnover, and thus the levels of an activator, regulating its intracellular localization, inhibiting its DNA-binding activity, or inhibiting protein-protein interactions that the activator makes with the transcriptional machinery (Gaston and Jayaraman, 2003).

REPRESSION BY REMODELLING CHROMATIN

Some histone-modifying enzymes such as the Sir proteins were shown to be required for telomeric silencing and mating type silencing in *S. cerevisiae*. Among these proteins, Sir3 and Sir4 bind to the N-terminal tails of histones, and Sir4 recruits Sir2, a NAD-dependent deacetylase (Hecht et al., 1995; Imai et al., 2000). In addition, ATP-dependent remodeling complexes can recognize histone modifications, and through ATP hydrolysis unwrap, exchange or eject the nucleosomes, which affects chromatin compaction and transcription activity.

Transcription factors are key proteins in the regulation of transcription, affecting transcription activation as well as transcription repression. Given the fundamental role of transcription regulation in the global response of cells to various stimulus and in regulation of protein expression, transcription factors are interesting targets for cell engineering in biotechnology.

V. *Pichia pastoris* in biotechnology

V.1. *Pichia pastoris*, a host for recombinant protein production

The yeast *Pichia pastoris* was first described in 1920 by Alexandre Guillermond and named *Zygosaccharomyces pastoris*. It was renamed *P. pastoris* by Herman Phaff in 1956 (Phaff, 1956), and reclassified in a distinct phylogenetic genus called *Komagataella* in 1995. This genus was later split into several species based on the 26S rRNA sequencing data (Yamada et al., 1995). *P. pastoris* is now a synonym for the two production hosts from the *Komagataella* phylum used in biotechnology: *K. pastoris* and *K. phaffii*. These two species have different chromosome structure and different 26S rRNA sequences, but they are hard to distinguish on the phenotypic level by standard growth tests and fermentation (Kurtzman, 2009). None of the two species have been proved to be superior to the other in recombinant protein production, and the genetic tools function in both. Here the established name *Pichia pastoris* is used as a synonym for all *Komagataella* species used in biotechnology.

The use of *P. pastoris* in biotechnology started in the 1970s: the company Phillips Petroleum used it for the production of single cell protein (SCP) in a process exploiting its ability to grow to high cell density on methanol. However, the price of methanol greatly increased because of the oil crisis in 1973, making SCP production economically unattractive. In the 1980s, *P. pastoris* started to be used as a heterologous protein production host. Strong and tightly regulated promoters regulating genes from the methanol utilization pathway were used for expression of heterologous genes (Cregg et al., 1989). In combination with the fermentation process already established for the SCP production, it led to the production of high levels of heterologous proteins. This expression system was originally patented by Phillips Petroleum, but it was rapidly made available to the research community, which contributed to its popularity as a platform for heterologous protein production.

P. pastoris exhibits advantages for the production of recombinant proteins, which explains its popularity as an expression platform. This yeast has indeed the ability to grow rapidly to high cell densities in a defined minimal medium and to perform post-translational modifications such as proteolytic modifications, disulfide bond formation and glycosylation. *P. pastoris* is also able to reach high product yields and to secrete recombinant proteins in the extracellular medium relatively pure of contaminations due to a low-level secretion of endogenous proteins (Puxbaum et al., 2015).

In recent years, *P. pastoris* was also used for the production of chemicals, and particularly complex secondary metabolites, but the major efforts in metabolic and cell engineering were mostly

focused on boosting heterologous protein production, which remains the main application of *P. pastoris*.

V.2. Cell engineering for improved heterologous protein production

Good results were obtained early on in protein production using *P. pastoris*, but cell engineering was still developed in various areas to increase the yield or the quality of the products.

GENERAL APPROACHES FOR IMPROVED PROTEIN PRODUCTION

New promoters and promoter engineering

Transcription is the first step in protein synthesis, therefore it is the first level that can be addressed in cell engineering for efficient protein production. The use of new promoters was developed in *P. pastoris* on the one hand to obtain methanol-free production systems, and on the other hand to generate promoter libraries spanning a wide range of strength to be used in metabolic or cell engineering.

The first strategy for protein production in *P. pastoris* was based on the *Alcohol oxidase 1* (*AOX1*) promoter, because it is tightly regulated and strongly induced on methanol (Cregg et al., 2000). However, the use of a methanol based production system exhibits some technical limitations: the methanol metabolism requires more oxygen and releases higher metabolic heat than the carbohydrates metabolism, and large amounts of methanol are needed for induction in biotechnological processes, requiring special safety measures since it is a highly flammable chemical. Therefore, alternatives to the methanol based system were investigated (Ahn et al., 2009; Menendez et al., 2003; Shen et al., 1998). Utilization of strong constitutive promoters such as the glyceraldehyde 3-phosphate dehydrogenase (GAP) promoter was developed (Waterham et al., 1997). More recently, systems biology was used to identify new promoters for the establishment of methanol free production systems (Prielhofer et al., 2015; Türkanoğlu Özçelik et al., 2019). Among alternative inducible promoters, P_{G1}, induced by limiting glucose, shows higher expression levels than P_{GAP} (Prielhofer et al., 2013). Another methanol free alternative, P_{THI11} is derived from a gene involved in thiamine biosynthesis and reaches 70% of P_{GAP} activity in the absence of thiamine (Stadlmayr et al., 2010).

Promoter engineering was also developed in order to obtain libraries of promoters having different strength. The goal of promoter engineering is to modulate the transcriptional capacity of a promoter by mutating its DNA sequence using diverse strategies. Regulation of the *AOX1* promoter was modified by random mutagenesis, which resulted in increased promoter activity and abolished glucose repression (Berg et al., 2013). The random mutagenesis approach was also

applied to the GAP promoter: a P_{GAP} library was generated with an activity spanning from 8% to 218% of the wild type P_{GAP} promoter activity (Qin et al., 2011).

Another approach in promoter engineering is the development of synthetic promoters. In *P. pastoris*, synthetic promoters were designed by using a consensus sequence of some selected natural core promoters and incorporation of common transcription factor binding sites. The designed core promoters were fused to the upstream activation sequence (UAS) of P_{AOX1}, which led to the creation of a synthetic promoter library with an expression strength of 10 to 117% of the wild-type P_{AOX1} upon induction with methanol (Vogl et al., 2014).

More recently, promoter engineering approaches manipulating the nucleosome occupancy in promoter regions were used. Many studies showed that the poly (dA:dT) tracts present in the yeast genome are important for transcriptional regulation, and that nucleosomes are strongly depleted from these stretches (Segal and Widom, 2009; Struhl, 1985). Therefore, by modifying the presence and the length of native poly (dA:dT) tracts, it is possible to increase the accessibility to nearby TFBSs covered by nucleosomes, which affects promoter activity (Raveh-Sadka et al., 2012). Deletion or lengthening of native poly (dA:dT) tracts in P_{AOX1} could alter the variants activities ranging from ~0.25 to ~3.5 fold of wild-type *AOX1* promoter activity (Yang et al., 2018).

Other strategies for promoter engineering include modification of transcription factor binding sites (TFBS) and transcriptional engineering (see next section).

Protein folding and secretion

P. pastoris is generally employed for the production of secreted recombinant proteins, because the product obtained after cell removal is quite pure. The secreted recombinant proteins go through the secretory pathway of the production host, which enables a better folding and processing of the products. In yeast, the secretory proteins are translocated to the endoplasmic reticulum (ER) where they are folded with the help of chaperones, then travel to the Golgi apparatus for processing, and are finally secreted via secretory vesicles. All the fore-mentioned steps can be limiting factors for the production of recombinant proteins in high titers.

One strategy to enhance the folding and secretion capacity of the production host is to overexpress some chaperones as well as the disulfide isomerase Pdi1 or peptidyl-prolyl isomerase Cpr5. This approach showed beneficial effects in different yeasts (de Ruijter et al., 2016; Delic et al., 2014a). The overexpression of folding-aiding chaperones in the cytosol also improved secretion in *P. pastoris* (Gasser et al., 2007). Production of recombinant proteins could also be improved by targeting components of the secretory machinery: overexpression of *SEC4* or *SSO2*, two genes encoding proteins involved in the fusion of secretory vesicles to the plasma

membrane, increased secretion of Fab and glucoamylase (Gasser et al., 2007; Liu et al., 2005). Cell engineering to obtain loosened cell wall by knocking out cross-linking or cell wall modifying enzymes was also proven to be beneficial for recombinant protein secretion *in P. pastoris* (Larsen et al., 2013; Marx et al., 2006). However these approaches were usually product dependent, and so far no strategy was generally applicable to increasing the secretion of several different recombinant proteins.

Protein glycosylation

Glycosylation is a common protein post-translational modification and it has various functions. They are for example involved in stabilizing protein folds, and glycans can modulate the biological properties of the protein to which they are attached.

The N-glycans in *P. pastoris* are typically referred to as high-mannose type of glycans because they usually contain 9-16 (most frequently 9, 10 or 11) mannoses with terminal α-1,2-linkages. Unlike *S. cerevisiae*, *P. pastoris* does not hyper-glycosylate recombinant proteins and does not contain potentially immunogenic terminal α-1,3-linked mannoses (Maccani et al., 2014; Vervecken et al., 2004), but the high-mannose-type N-glycosylation might be immunogenic and make the downstream processing more difficult. The conversion of these high-mannose N-glycans to human-like N-glycans was achieved by glycoengineering of *P. pastoris*. A mannosyl transferase was deleted and specific glycosyl transferases were added to obtain fully human-like, sialylated N-glycans (Hamilton et al., 2003; Hamilton et al., 2013). *P. pastoris* O-glycans are characterized by 1-4 mannose residues bound to Serine or Threonine. Reduction of O-mannosylation in *P. pastoris* was obtained by partial deletions of protein mannosyl transferases (Pmt) genes (Nett et al., 2013). Deletion of a specific mannosyl transferase and overexpression of an α-1,2 mannosidase led to single mannose O-glycans which could be further engineered to resemble human-like O-glycans (Hamilton et al., 2013).

TRANSCRIPTION FACTORS APPLIED FOR IMPROVING RECOMBINANT PROTEIN PRODUCTION

The knowledge on specific promoters and on the TFs regulating them was used to develop genetic circuits and induction systems in order to allow a precise control of recombinant protein expression in the bacterium *Escherichia coli* (Correa and Oppezzo, 2011). Some approaches using TFs to increase protein production were also applied in the yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis* (Huang et al., 2017; Liu et al., 2015; Yu et al., 2010) and in Chinese hamster ovary (CHO) cells (Gutiérrez-González et al., 2019). Few approaches using transcription factors were developed in order to improve recombinant protein production in *P. pastoris* as well.

Promoter engineering

As seen previously, promoter engineering is an important area in cell engineering for recombinant protein production as transcription is the first step in protein synthesis.

Promoter strength and regulation are the cumulative effect of TFBSs: these short, distinct nucleotide sequences facilitate binding of the transcriptional machinery, which affects promoter activity. Promoter engineering strategies, such as random mutagenesis or the creation of synthetic promoters, are ultimately based on the addition, deletion or modification of TFBSs to modulate promoter capacity. Indeed, random mutagenesis introduces haphazard mutations within or around TFBSs, and synthetic promoter engineering uses TFBSs as a transposable genetic element to construct novel promoters. Therefore, rational construction of promoter libraries with distinct regulatory characteristics is possible by using the knowledge on TFs and their corresponding TFBSs.

Promoter engineering via TFBSs was implemented in *P. pastoris* P_{AOX1} via addition or deletion of putative TFBSs. The transcriptional capacity of the constructed P_{AOX1} promoter library spanned between 6% and 160% activity of the native promoter and with different regulatory properties (Hartner et al., 2008). The *P. pastoris* P_{GAP} was recently modified by duplication of putative activator binding sites and deletion of putative repressor binding sites. The strength of the obtained P_{GAP} library was ranging from 82% to 190% of the wild-type P_{GAP} promoter activity (Ata et al., 2017).

Alternatively, since TFs have a crucial role in activation and in determination of the specificity of transcription, promoter engineering can be achieved via overexpression or knock-out of selected transcription factors. For instance, overexpression of certain transcriptional regulators (Mit1, Mxr1 and Prm1) led to the activation of P_{AOX1} under derepression conditions (when the repressing carbon source is depleted, but without methanol). In controlled pilot-scale bioreactor cultivations, the Mit1-overexpression strain exhibited a higher space/time yield and a higher absolute yield without methanol induction than the methanol induced parental strain (Vogl et al., 2018b). The P_{GAP} library mentioned in the previous paragraph was also further improved by overexpression or deletion of interesting TF genes. The recombinant protein expression levels varied between 35 to 310% of the wild-type P_{GAP} driven expression in the *P. pastoris* mutant strains (Ata et al., 2017).

Global transcriptional engineering

The production of recombinant proteins can require an extensive reprogramming of the metabolism of *P. pastoris*. The engineering of a desired phenotype often demands to

simultaneously modify the expression levels of many genes, which is difficult to accomplish by sequential multigene modifications. In addition, it is challenging to identify the genes needing an alteration with conventional pathway analysis. By using transcription factors, it is possible to regulate the global transcriptional response of *P. pastoris*, which has been proven to be a promising approach to increase protein productivity. Overexpression of Yap1 enhanced recombinant protein production by modifying intra-cellular redox conditions (Delic et al., 2014b). Hac1, a TF involved in the unfolded protein response (UPR), was also successfully used to increase protein production in *P. pastoris* (Guerfal et al., 2010). The overexpression of Aft1 was shown to improve recombinant protein secretion in fed-batch cultivations (Ruth et al., 2014). Since the general metabolic response to the production of high quantities of recombinant proteins involves a decrease of the intracellular amino acids concentration (Carnicer et al., 2012), the amino acids supply was engineered in *P. pastoris* by overexpressing *GCN4*, which encodes a general transcriptional activator of the amino acid biosynthesis. This approach was used to double the production of a glucose oxidase (Gu et al., 2015). Finally, it was shown more recently that overexpression of Msn4 alone or in combination with chaperones or with Hac1 improved the capacity for recombinant protein secretion. Up to 4.5-fold higher productivity of antibody fragments could be obtained in fed-batch cultivations with certain combinations (Zahrl, 2018).

Current knowledge on TFs in *P. pastoris*

Using approaches with identified TFs and doing TFBS modifications is beneficial for promoter engineering and global transcriptional engineering, however, it requires prior knowledge on the involved TFs and/or their binding sites.

In the past years, transcriptional regulation has been investigated in *P. pastoris* by characterizing TFs and their modes of controlling bioprocess relevant features. TFs involved in various processes such as oxidative stress response (Yap1 (Delic et al., 2014b)), iron uptake (Fep1 (Miele et al., 2007)), protein secretion (Hac1 (Guerfal et al., 2010)) or biotin auxotrophy (Rop1 (Kumar and Rangarajan, 2011; Kumar and Rangarajan, 2012)) have for example been studied. In addition, several TFs related to methanol utilization were characterized. The TF Prm1 (also called Trm1) functions as a positive regulator of genes involved in the methanol utilization (MUT) pathway in *P. pastoris* (Sahu et al., 2014). Mxr1 was shown to play an important role in inducing the transcription of *AOX1* and other genes involved in the MUT pathway and that of *PEX* genes (required for the biogenesis of peroxisomes) in *P. pastoris* (Kranthi et al., 2010; Kranthi et al., 2009; Lin-Cereghino et al., 2006). The TF Mit1 is involved in the strict repression of P_{AOX1} on glycerol and its strong induction on methanol ((Wang et al., 2016b). Finally, Nrg1 was shown to repress the expression of numerous genes involved in methanol utilization and peroxisome biogenesis on glucose and glycerol (Wang et al., 2016a).

A lot of TFs from *P. pastoris* have homologs that are found in other yeast species, however, the function of these TFs cannot always be inferred from related sequences from other organisms. Indeed, TF functions are not always conserved during evolution, which means that a specific TF can regulate different target genes in different yeast species (Hogues et al., 2008; Lavoie et al., 2010; Tuch et al., 2008). For example, the previously mentioned *MXR1*, encoding a TF involved in the activation of numerous genes important for methanol utilization in *P. pastoris* (Lin-Cereghino et al., 2006), is a homologue of *S. cerevisiae ADR1*, which encodes a TF necessary for the growth of this yeast on glycerol, ethanol and oleate (Denis and Young, 1983). The two genes seem to have gained new functions and lost others through evolution and now regulate different sets of genes in the two yeasts. Another such example is found with the TF-encoding gene *CRA1*. Cra1 has sequence similarity to *S. cerevisiae* Gal4, which is a transcriptional regulator of the galactose-induced genes. Since *P. pastoris* does not possess a galactose metabolism (Kurtzman, 2005), this TF is obviously not involved in galactose utilization in this yeast. In fact, overexpression of *CRA1* is sufficient to convert the Crabtree-negative *P. pastoris* into a Crabtree positive yeast, and Cra1 was shown to be involved in the regulation of glycolytic and fermentative genes (Ata et al., 2018). It was hypothesized that the function of Gal4 changed: this TF would originally be generalist TF responsible for the regulation of the central carbon metabolism and, during evolution, would have become a specialist TF responsible for the metabolism of galactose (Choudhury and Whiteway, 2018). Since knowledge on TFs cannot always be transferred from the model yeast *S. cerevisiae* or from other yeasts, specific studies are required in order to characterize TFs of interest in *P. pastoris*.

VI. Aim of the Study

The use of transcription factors in cell engineering for improved protein production in *P. pastoris* has been proven to be beneficial in some cases in the past. In addition, previous studies investigating the responses of *P. pastoris* to protein production and to different carbon sources on a genome-wide level showed that the majority of gene regulation occurs on a transcriptional level. A large number of putative transcriptional regulators were also found among the differentially expressed genes and proteins. However, their target genes and functions are for the majority still unidentified. The goal of this work was therefore to broaden the knowledge on TFs of *P. pastoris*, by characterizing the necessity of some transcription factors in different stress conditions and carbon sources as well as their impact on cellular organization. The final objective of this study was to gain information in order to design cell engineering strategies that could be used for *P. pastoris*, whether it is increased robustness during production processes, enhanced productivity or novel targets for promoter engineering.

I. Strains, primers and plasmids

All the *P. pastoris* strains used in this study (**Appendix 3**) were derived from the wild-type strain CBS7435 (*Komagataella phaffii*). *Escherichia coli* DH10B (Invitrogen) and *P. pastoris* CBS7435 were used for cloning experiments.

The list of primers used in this study is given in **Appendix 1**, and the list of plasmids constructed in this study is given in **Appendix 2**.

II. Sequence analyses, alignment and phylogenetic tree

A protein-BLAST search was conducted on the 'Non-redundant protein sequences (nr)' database of NCBI using 8 functionally characterized Cat8 and Sip4 protein sequences.

CBF88979.1_CAT8_Aspergillus_nidulans_FGSC_A4
XP_453133.1_CAT8_Kluyveromyces_lactis
CAE00852.1_Sip4_Kluyveromyces_lactis
XP_018209149.1_CAT8_Ogataea_polymorpha
CAA55139.1_CAT8_Saccharomyces_cerevisiae
CAA89382.1_Sip4_Saccharomyces_cerevisiae
XP_002491690.1_CAT8_Komagataella_phaffii_GS115
XP_002493979.1_CAT8-2_Komagataella_phaffii_GS115

The search was restricted to *Saccharomycetes* (Yeast; taxid:4891) and a maximum E-Value of 9e-30.

All individual BLAST search results were combined (982) and protein ID duplicates were removed (361). Then, all sequences containing invalid protein characters (B J O U X Z) were removed from the selection (352). All sequences were filtered for a minimum length of 600 amino acids (275). Sequences in the selection of 99% sequence identity or higher were represented in the selection by only one sequence of that cluster to reduce sequence redundancy (no additional phylogenetic information, but computational effort) (157). All characterized Cat8 and Sip4 sequences not present in the selection were added again (161).

Selection was aligned with MAFFT G-INS-I and renamed with SeqScrub according to taxonomy. Alignment was trimmed for positions with >90% gaps and a ML tree was calculated with PhyML (LG, Best of NNI and SPRs, SH-like support). The tree was rooted on midpoint. All the proteins

used in the alignment and their identifier numbers (NCBI accession numbers) are listed in **Appendix 6**.

III. Media

YPD medium included of 10 g/L yeast extract, 20 g/L soy peptone containing 2% glucose as carbon source. YPD agar plates consisted of 10 g/L yeast extract, 20 g/L soy peptone, 2% glucose as carbon source and 20 g/L agar-agar. The YPD liquid medium and the YPD agar plates were supplemented with the appropriate antibiotics (zeocin 50 µg/mL, geneticin 500 µg/mL, nourseothricin 100 µg/mL) when needed.

LB media consisted of 10 g/L soy peptone, 5 g/L yeast extract, 5 g/L NaCl. 20 g/L agar-agar was added to prepare LB agar plates.

ASMv6 medium included 6.3 g/L $(NH_4)_2HPO_4$, 0.8 g/L $(NH_4)_2SO_4$, 0.49 g/L $MgSO_4 \cdot 7H_2O$, 2.64 g/L KCl, 0.0535 g/L $CaCl_2 \cdot 2H_2O$, 22 g/L citric acid monohydrate, 1.47 mL PTM, 2 mL biotin (0.2 g/L), 20 mL NH_4OH (25%) with additional carbon source according to the purpose. For limited glucose condition 25% m2p kit Polysaccharide and 0.078% enzyme was used (m2p-labs GmbH, Germany).

PTM stock solution included 0.08 g/L NaI, 6.0 g/L $CuSO_4 \cdot 5H_2O$, 3.36 g/L $MnSO_4 \cdot H_2O$, 0.2 g/L $Na_2MoO_4 \cdot 2H_2O$, 0.82 g/L $CoCl_2$, 0.02 g/L H_3BO_3, 20.0 g/L $ZnCl_2$, 65.0 g/L $FeSO_4 \cdot 7H_2O$ and 5.0 g/L mL H_2SO_4 (95–98%).

YNB without thiamine consisted of 10 g/L $(NH_4)_2SO_4$, 0.2 g/L biotin, 0.8 mg/L Ca-pantothenate, 0.004 mg/L folic acid, 4 mg/L inositol, 0.8 mg/L niacin, 0.4 mg/L p-aminobenzoic acid, 0.8 mg/L pyridoxine HCl, 0.4 mg/L riboflavin, 1 mg/L H_3BO_3, 0.08 mg/L $CuSO_4$, 0.2 mg/L KI, 0.4 mg/L $FeCl_3$, 0.8 mg/L $MnSO_4 \cdot H_2O$, 0.4 mg/L $Na_2MoO_4 \cdot 2H_2O$, 0.8 mg/L $ZnSO_4$, 2 g/L KH_2PO_4, 1 g/L $MgSO_4$, 200 mg/L NaCl, 200 mg/L $CaCl_2$ with additional carbon source according to the purpose. For limited glucose condition 25% m2p kit Polysaccharide and 0.078% enzyme was used (m2p-labs GmbH, Germany).

Phosphate buffered saline (PBS) consisted of 1.8 g/L Na_2HPO_4*$2H_2O$, 0.24 g/L KH_2PO_4, 8 g/L NaCl, 0.2 g/L KCl.

IV. Genomic DNA extraction and PCR

Genomic DNA was extracted from overnight cultures using the Wizard genomic DNA purification kit (Promega Corp., USA) according to the protocol of the manufacturer. All PCRs were performed

using the Q5 polymerase (New England Biolabs, Inc., USA) following the recommendations of the manufacturer or using the OneTaq 2x master mix with GC buffer (New England Biolabs, Inc., USA). The PCRs were used to amplify DNA for cloning and for verifying transformants.

V. Construction of overexpression and knock-out strains

In this study, 12 different TFs were selected as targets for overexpression or knock-out. The sequences of these genes were retrieved from http://pichiagenome-ext.boku.ac.at. Chromosomal regions of the selected transcription factors are given in **Table 1**.

Table 1: Chromosomal locations of the coding regions of the selected transcription factors.

Transcription factor	Chromosomal location of the gene
AFU1	PP7435_Chr1-0281
CAT8-1	PP7435_Chr2-0516
CAT8-2	PP7435_Chr4-0434
MIT1	PP7435_Chr3-0349
PP7435_Chr1-0006 (PAS_chr1-3_0010)	PP7435_Chr1-0006
PP7435_Chr1-0170 (PAS_chr1-3_0166)	PP7435_Chr1-0170
PP7435_Chr4-0940 (PAS_chr4_0077)	PP7435_Chr4-0940
SGF29	PP7435_Chr1-0772
SKO1	PP7435_Chr1-0475
SWI5	PP7435_Chr1-0101
YGR067C	PP7435_Chr3-0964
YPR022C-1	PP7435_Chr1-0680

Golden Gate Assembly (GGA; (Engler et al., 2008)) was used for the construction of the overexpression and knock-out cassettes using the Golden*Pi*CS vector series (Prielhofer et al., 2017). Internal *BsaI* or *BpiI* sites within the CDS or homologous regions for integration of knock-out cassettes were eliminated by designing primers which enable to overlap the modified regions by PCR or ordering in-vitro synthesized gBlocks where these nucleotides were mutated without altering the originally encoded amino acids.

V.1. Overexpression

For overexpression cassettes, the CDS of the corresponding genes were amplified from the *P. pastoris* CBS7435 genome by PCR and cloned into a plasmid carrying the KanMX marker cassette and a region for homologous integration into the *AOX1* terminator. The *THI11* promoter, the

promoter of a gene which encodes a protein involved in synthesis of the thiamine precursor hydroxymethylpyrimidine (HMP), was used for the overexpression of selected TFs (Delic et al., 2013). The expression capacity of this promoter can be controlled by the presence or absence of thiamine (Landes et al., 2016). For all overexpression cassettes, the *RPS3tt* transcription terminator was used. Primers used for the generation of overexpression cassettes are given in **Appendix 1**. The transformants were verified by colony PCR and gene copy number determination (positive clones had two copies of the gene of interest: the native one and the one of the overexpression cassette).

V.2. Knock-out

P. pastoris knock-out strains were constructed by using CRISPR/Cas9-based homology-directed genome editing (Gassler et al., 2018). The homologous regions were amplified by PCR from *P. pastoris* CBS7435 genome. These homologous regions were selected from upstream (5') and downstream (3') of the target gene with an approximate 1000 bp length. The primers used for amplification of these homologous fragments and elimination of internal *Bsa*I sites are given in **Appendix 1**. The flanking upstream and downstream homologous regions of the target gene were assembled with each other in a plasmid by GGA.

A single guide RNA was designed and amplified based on a protospacer adjacent motif (PAM) sequence identified in 50-200 bp upstream of the CDS of the selected TFs. The guide RNA was cloned under the control of the *GAP* promoter and the *RPS25A*tt terminator into a plasmid containing the humanized Cas9 CDS under the control of P_{LAT1} or P_{PFK300} and the *ScCYC*tt terminator by GGA. For the integration of the knock-out cassettes, the plasmids carrying the fused homologous regions were used as the templates to amplify the fragments by PCR. 3-5 µg of amplified homologous DNA and 0.5-1 µg of circular CRISPR/Cas9 plasmid DNA were simultaneously transformed into *P. pastoris* CBS7435 by electroporation. The knock-out strains were controlled by two PCRs using (1) primers binding in the genome outside of the targeted deletion sites and (2) binding in the CDS of the targeted TFs. After confirmation of the TF deletions, true transformants were passaged at least three times on YPD to lose the CRISPR/Cas9 plasmid.

VI. Preparation of electro-competent *Pichia pastoris* cells

A strain of *P. pastoris* was inoculated in 100 mL YPD media and incubated for 16-20 hours (25°C; 180 rpm). Cells were then harvested when the optical density (OD_{600}) reached 1.2 – 2.5 by splitting the culture into two 50 mL falcon tubes and centrifuging (5 min, 1500 *g*, 4°C). Cell pellets

were resuspended in 10 mL of pre-treating solution (0.6 M sorbitol, 10 mM Tris-HCl pH 7.5, 10 mM DTT, 100 mM lithium acetate) and incubated for 30 min (25°C, 180 rpm). The cell suspension was diluted in 40 mL ice-cold 1M sorbitol and again centrifuged (5min, 1500 *g*, 4°C). The pellets were combined and resuspended in 45 mL ice-cold 1M sorbitol and centrifuged (5min, 1500 *g*, 4°C). The pellet was resuspended in 45 mL ice-cold 1M sorbitol and centrifuged again (5min, 1500 *g*, 4°C). Finally, the pellet was resuspended in 1 mL ice cold 1M sorbitol. 80 µL of the competent cell suspension were then aliquoted in pre-chilled 1.5 mL Eppendorf tubes. The aliquoted electro-competent cells were stored at -80°C until further use.

VII. Transformation of *Pichia pastoris*

Prior to *P. pastoris* transformation, the overexpression plasmids were linearized within the genome integration locus and purified (innuPREP DOUBLEpure Kit, Analytik Jena, Germany). PCR-amplified homologous regions of knock-out cassettes were directly purified without linearizing. *P. pastoris* transformation was performed by electroporation (BioRad Gene Pulser, 2000 V, 25 µF and 200 Ω) by using 0.5–1 µg of each linearized overexpression plasmid or 3-5 µg of purified knock-out fragments and 0.5-1 µg of circular CRISPR/Cas9 plasmid. Transformed cells were then regenerated by incubation at 30°C for 1.5–3 h in YPD medium (280 rpm) and then plated on YPD plates including the appropriate antibiotics concentration (zeocin 50 µg/mL, geneticin 500 µg/mL, nourseothricin 100 µg/mL). After 48-72 h at 30°C, randomly selected transformants were streaked on selective YPD plates and incubated 48 h at 30°C.

VIII. Gene copy number determination

Gene copy number (GCN) was determined by quantitative real-time PCR (qPCR). Genomic DNA was extracted from overnight cultures using the Wizard genomic DNA purification kit (Promega Corp., USA). The GCN was determined by the relative quantification of the TF of interest sequence compared to wild type *P. pastoris* CBS7435 (carrying a single native copy of the TF gene of interest). The amplifications were carried out using 4.5 µL of genomic DNA solution at a concentration of 1.777 ng/µL with 0.25 µL of both forward and reverse primers (final concentration: 10 µM) and 5 µL of 2x qPCR S'Green BlueMix (Biozym Scientific GmbH, Germany). Amplifications were done in a Rotor-Gene Q instrument (QIAGEN GmbH, Germany) with the program given in **Table 2**. The GCN in the *P. pastoris* mutant strains were calculated relative to the corresponding wild-type control using the threshold cycle (ΔΔ*CT*) method. All signals were normalized to *ACT1* (PP7435_Chr3-0993). The primers used for qPCR analysis are provided in **Appendix 1**.

Table 2: Program used for the qPCR in the gene copy number analysis

Cycles	Temperature	Time	Notes
1	95°C	3 min	Initial denaturation and enzyme activation
40	95°C 60°C	5 seconds 30 seconds	Denaturation Annealing/extension
Melt analysis	65 to 99°C rising by 0.5°C each step	2 seconds	

IX. Growth assay in liquid medium

The wild-type, overexpression and deletion strains were inoculated at OD_{600} 0.01 in 100 µL of YNB without thiamine containing either 2% glucose, 2% glycerol, 2% ethanol or 1% methanol in a 96-well sterile microtiter plate. The plate was incubated in a TECAN Sunrise plate reader at 30°C for 24-48 hours with constant shaking. The absorbance at 600 nm in each well was measured every 15 minutes. The parameters used for measuring the OD in the TECAN Sunrise are given in **Table 3**.

For each strain, three to four biological replicates were cultivated each in three wells. The blank value (OD_{600} of the media without cells) was subtracted from the raw OD values to obtain the corrected ODs. The corrected OD_{600} of the replicates for a given strain were then averaged. The average corrected OD_{600} were plotted against the time to obtain the growth curves. For the calculation of the growth rates, average corrected OD_{600} were divided by the average initial OD and natural logarithm was applied. The growth rate is given by the slope of the log transformed ODs, the maximal growth rate being identified as the maximum value of the slope.

Table 3: Parameters used for measuring the OD_{600} during the growth assay in the TECAN Sunrise.

General	Absorbance	
Wavelengths: fixed	measured	600
	Reference	0
Measured parameters	normal read mode	
Kinetics	number of cycles	99
	Interval	14:39
	estimated run time without stacker	09 s
	run time	~24 h
Temperature valid range	29.5 – 30.5°C	
Shaking	before measurement	5 s, inside, normal
	between cycles	870 s, inside, normal

X. Spotting assay

Phenotypic characterization was performed on YNB without thiamine agar plate (2% agar-agar) containing 2% glucose (w/v) supplemented with the different stressors in serial 10-fold dilutions (initial OD_{600} 0.3). The stressors used were cell wall damaging agents (1 and 5 µg/mL Calcofluor white, 0.00005% and 0.0001% Congo Red, 0.01% SDS), osmotic stressors (1M KCl, 1M NaCl) and oxidative stressor (0.5 mM and 1 mM H_2O_2). Temperature sensitivity was also assessed for some genes by incubating plates containing either 2% glucose or 1% methanol at 25, 30, 35 and 37°C. Plates were incubated 3 to 7 days and were scanned using an EPSON perfection V750 PRO scanner.

XI. RNA extraction and transcript levels analysis

The wild-type, overexpression and deletion strains were grown on ASMv6 medium with limiting glucose to OD_{600} 7.0-8.0, washed twice in PBS, inoculated at OD_{600} 3.5-4.5 in ASMv6 medium containing either 2% glucose, 2% glycerol, 2% ethanol or 1% methanol and grown for 5 hours. Samples were collected by centrifugation at full speed at 4°C, and cell pellets were resuspended in 1 mL TRI reagent solution (Invitrogen) and stored at −70°C until further use. Cells were mechanically disrupted using 500 µL of glass beads in a ribolyzer (5.5 m/s for 40 seconds), and the total RNA extraction was performed according to the TRI reagent protocol. RNA concentrations and purity were analyzed with a Nanodrop spectrophotometer. DNAse treatment of isolated RNA samples was performed with DNA-free kit (Invitrogen) and cDNA was synthesized using oligo$(dT)_{23}$ primers (New England Biolabs, Inc., USA) and the Biozym cDNA synthesis kit according to directions of the manufacturer (Biozym Scientific GmbH, Germany). The Real-time PCR reactions were performed on a Rotor-Gene Q instrument (QIAGEN GmbH, Germany) using Blue S'Green qPCR Mix (Biozym Scientific GmbH, Germany) according to manufacturer's instructions. Changes in transcript levels in *the P. pastoris mutant* strains were calculated relative to the corresponding *wild-type* control using the threshold cycle (ΔΔ*CT*) method. All signals were normalized to *ACT1* (PP7435_Chr3-0993) expression. Used primers for qPCR analysis are provided in **Appendix 1**.

XII. Construction of the HA-tagged Cat8-1 and Cat8-2 strains

Initially, the tagging of Cat8-1 and Cat8-2 was tested with a 3xFLAG tag at the native locus using the CRISPR/Cas9-based homology-directed genome editing (Gassler et al., 2018), but no clones were obtained after transformation. Another approach was tested by cloning the coding sequence of *CAT8-1* and *CAT8-2* with a 3xFLAG tag before their STOP codons under the control of their

respective native promoters on two distinct plasmids carrying the *AOX1* terminator homologous region (for integration into the *P. pastoris* genome). The linearized plasmids were transformed into *P. pastoris cat8-1Δ* (for Cat8-1-3xFLAG) and *cat8-2Δ* (for Cat8-2-3xFLAG) deletion strains. Again no clones were obtained. Since both single deletion strains of *CAT8-1* and *CAT8-2* were viable, it seemed that introducing a FLAG-tagged copy of Cat8-1 and Cat8-2 was detrimental to the cells since no clones were obtained in both attempts to tag Cat8-1 and Cat8-1 with a 3xFLAG tag. Therefore tagging of Cat8-1 and Cat8-2 with the HA tag was tested, and was proven to be a successful strategy.

To construct the HA-tagged Cat8-1 and Cat8-2 strains, promoter regions (upstream 1000 bp regions), coding sequences and terminator regions of *CAT8-1* and *CAT8-2* were amplified from *P. pastoris* CBS7435 genomic DNA by PCR. Reverse primers for the amplification on the coding sequences were designed so there is insertion of the HA tag (encoded by TACCCATACGATGTTCCAGATTACGC) before the STOP codons of the two genes. The PCR fragments were assembled by GGA in a vector carrying the *AOX1* terminator homologous regions (for integration into the *P. pastoris* genome) and the KanMX marker cassette. 0.5-1µg of linearized plasmids were transformed into *P. pastoris cat8-1Δ* (for Cat8-1-HA) and *cat8-2Δ* (for Cat8-2-HA) deletion strains. Transformants were controlled by colony PCR and qPCR (for gene copy number determination). The selected clones were also grown on YNB liquid medium supplemented with 1% ethanol to make sure that the insertion of the tagged version of the proteins in the deletion strains restored their growth on this carbon source.

XIII. Enhanced GFP reporter assay

XIII.1. Cloning of eGFP under the control of P_{CAT8-1} and P_{CAT8-2}

Promoter regions (upstream 1000 bp regions) of *CAT8-1* and *CAT8-2* were amplified from *P. pastoris* CBS7435 genomic DNA by PCR (Primers in **Appendix 1**). The respective promoters, the eGFP coding sequence and the *ScCYC*tt were assembled by GGA in a vector containing *AOX1* terminator homologous regions (for integration into *P. pastoris* genome) and the KanMX marker cassette. 0.5-1 µg of the linearized plasmids were transformed into *P. pastoris* CBS7435 (WT), *cat8-1Δ*, *cat8-2Δ*, and *cat8-1Δcat8-2Δ* deletion strains. Transformants were controlled by colony PCR and qPCR for gene copy number determination.

XIII.2. Reporter assay

The wild-type and positive transformants of the wild-type, *cat8-1Δ*, *cat8-2Δ*, and *cat8-1Δcat8-2Δ* deletion strains carrying the eGFP coding sequence under the control of the respective promoters

were grown on ASMv6 medium with limiting glucose (polysaccharide solution) at 25°C to OD_{600} 7.0-8.0, washed twice in PBS and inoculated at OD_{600} 3.5-4.5 in ASMv6 medium containing either 2% glucose, 2% glycerol, 2% ethanol or 1% methanol and grown for 5 hours at 25°C, to have induction on glucose, glycerol, ethanol and methanol, respectively. After induction, cells were analyzed by flow cytometry.

XIII.3. Flow cytometry

Cells were diluted in PBS (KH_2PO_4 0.24 g/L, Na_2HPO_4*2 H_2O 1.8 g/L, KCl 0.2 g/L, NaCl 8 g/L) to OD_{600} 0.2. The forward and side scatter of 10 000 cells for each sample as well as their green fluorescence (FL1 channel, 505-545 nm) were then measured on a CytoFLEX flow cytometer (Beckman Coulter). The Kaluza analysis software (Beckman Coulter) was used to analyze the data. GFP-positive cells were gated using the WT_$P_{CAT8\text{-}1}$_eGFP and WT_$P_{CAT8\text{-}2}$_eGFP strains as reference.

XIV. Screening of eGFP production under P_{AOX1} in TF mutant strains

XIV.1. Cloning of eGFP under the control of P_{AOX1}

The *AOX1* promoter, the eGFP coding sequence and the *ScCYC*tt were assembled by GGA in a vector containing the *RGI2* homologous regions (for integration into *P. pastoris* genome) and the natMX marker cassette. 0.5-1 µg of the linearized plasmids were transformed into *P. pastoris* CBS7435 (WT) and the *SWI5_OE* mutant strain.

XIV.2. Screening procedure

11 clones of the WT_P_{AOX1}_eGFP and 11 clones of the SWI5_OE_P_{AOX1}_eGFP were inoculated and incubated for 24 hours in 2 mL of selective YP media in a 24 deep well plate (280 rpm, 25°C). Cell cultures were then harvested by centrifuging the 24 deep wells plate (2000 *g*, 5 min, 4°C) and carefully removing the supernatant. The pellets were then resuspended in 1mL ASMv6 medium and the OD_{600} was measured. For the main culture, cells were inoculated at OD_{600} 2.70 in 2 mL ASMv6 medium with limiting glucose (25% m2p kit Polysaccharide and 0.35% enzyme, m2p-labs GmbH, Germany) in another 24 deep well plate and incubated for 48 hours (280 rpm, 25°C). 3 hours after the beginning of the main cultures, cells were fed with 0.5% methanol (v/v). 19 hours, 27 hours and 43 hours after the beginning of the main culture, cells were fed with 1% methanol (v/v). 48 hours after inoculation of the main culture, samples were processed for flow cytometry.

XIV.3. Flow cytometry

Cells were diluted in PBS (KH_2PO_4 0.24 g/L, Na_2HPO_4*2 H_2O 1.8 g/L, KCl 0.2 g/L, NaCl 8 g/L) to OD_{600} 0.2. The forward and side scatter of 10 000 cells for each sample as well as their green fluorescence (FL1 channel, 505-545 nm) were then measured on a CytoFLEX flow cytometer (Beckman Coulter). The Kaluza analysis software (Beckman Coulter) was used to analyze the data. GFP-positive cells were gated using the WT_P_{AOX1}_eGFP strains as reference.

XV. Microscopy

The microscopy was achieved using a Zeiss Axio Observer.Z1/7 microscope using a LCI Plan-Neofluar 63x/1.3 water immersion objective. The cell morphology was analyzed in bright-field mode.

CHAPTER 1: PHENOTYPIC CHARACTERIZATION OF TRANSCRIPTION FACTORS FROM *Pichia pastoris*

I. Introduction

The overall goal of this PhD work was to study specific TFs of the yeast *P. pastoris*. The first step of the work was therefore to select different interesting TFs for further characterization. For each of these TFs, OE and KO mutants were generated, and these strains were arrayed on different media representing different growth conditions. The results of the growth assays then enabled to select some TFs to be more precisely characterized or that could have potential applications for production of recombinant proteins in *P. pastoris*.

II. Selection of TFs and generation of TF overexpression and knock-out mutants

II.1. Selection of transcription factors in *P. pastoris*

To select the most interesting TFs for further analysis, a list of all *P. pastoris* proteins containing domains of transcriptional regulators (e.g. bZIP domain, leucine zipper, $Zn(II)_2Cys_6$ Zn finger) was overlaid with available transcriptomics data studying gene regulation in *P. pastoris:* 1. cells cultivated at different growth rates, 2. on different carbon sources, 3. producing different recombinant proteins (Burgard et al., 2017; Prielhofer et al., 2015; Rebnegger et al., 2014; Russmayer et al., 2015). Focus was made on proteins whose genes were transcriptionally regulated in the different conditions. In total, 12 candidate genes, including eight Zinc finger (among them, two Cys_2His_2 Zn finger and five $Zn(II)_2Cys_6$ Zn finger finger), two Leucine zipper and one helix-loop-helix transcriptional regulators as well as one protein binding methylated histones (involved in transcriptional regulation through chromatin modifications) were selected because of their interesting regulation patterns (**Table 4**).

Table 4: Selected transcription factors in *P. pastoris* and their regulation patterns under various conditions.
+ indicates up-regulation, - down-regulation and 0 no regulation of a gene under a given condition.

Name	DNA binding domain	Low growth rate	Carbon sources	Protein production	Putative function
AFU1	$Zn(II)_2Cys_6$ Zn finger	0	+ in glucose fed-batch, + in methanol fed-batch	-	Activates fatty acid utilization (*POR1* in *Y. lipolytica*) (Poopanitpan et al., 2010)
CAT8-1	$Zn(II)_2Cys_6$ Zn finger	+	+ in methanol fed-batch	-	Binds carbon sources responsive elements (*S. cerevisiae*, *K. lactis*) (Mehlgarten et al., 2015)
CAT8-2	$Zn(II)_2Cys_6$ Zn finger	+	+ in glucose fed-batch, + in glycerol batch, + in methanol fed-batch	-	Binds carbon sources responsive elements (*S. cerevisiae*, *K. lactis*) (Mehlgarten et al., 2015)
MIT1	$Zn(II)_2Cys_6$ Zn finger	+	+ in glucose fed-batch, + in glycerol fed-batch, + in methanol fed-batch	+/- (depends on protein)	Regulation of methanol utilization genes (*P. pastoris*) (Wang et al., 2016b)
PP7435_Chr1-0006	$Zn(II)_2Cys_6$ Zn finger	-	+ in glucose fed-batch	-	No information
PP7435_Chr1-0170	Basic-leucine zipper	+	+ in glucose fed-batch, + in methanol fed-batch	-	Similar to putative AP1 (*Candida maltosa*)
PP7435_Chr4-0940	Helix-loop-helix	-	+ in methanol fed-batch	+/-	No information
SGF29	Binds methylated histones (H3K4me)	+	+ in glucose fed-batch	-	Component of SAGA, SLIK, ADA complexes (histone modifying complexes) (*S. cerevisiae*) (Bian et al., 2011; Shukla et al., 2012)
SKO1	Leucine zipper	+	+ in glucose fed-batch	-	Osmotic and oxidative stress response (*S. cerevisiae*)(Rep et al., 2001)
SWI5	Cys_2His_2 Zn finger	-	+ in glucose fed-batch, + in glycerol fed-batch, - in methanol fed-batch	-	Involved in mating type switching by recruitment of Mediator and Swi/Snf complexes (*S. cerevisiae*) (Bhoite et al., 2001; Dohrmann et al., 1992)
YGR067C	Zn finger	+	+ in glucose fed-batch, + glycerol fed-batch, + in methanol fed-batch	-	No information
YPR022C-1	Cys_2His_2 Zn finger	+	+ in glucose fed-batch	-	No information

Out of the twelve selected transcription factors of *P. pastoris*, six are homologs of transcription factors in other yeasts, which gives us potential indications of the functions of these TFs in *P. pastoris*. In addition, at the beginning of this project the TF Mit1 was uncharacterized, but its function in the regulation of the methanol utilization genes in *P. pastoris* has been published in the meantime. The other selected TFs seem to be homologs of TFs involved in a variety of cellular processes, including regulation of carbon source utilization, osmotic and oxidative stress response as well as recruitment of various protein complexes to the DNA (**Table 4**).

II.2. Generation of TF overexpression and knock-out TF mutant strains

For analysis of the twelve selected transcription factors, overexpression and knock-out mutants of each TF were generated.

For overexpression of the selected TFs in *P. pastoris*, the genes encoding the TFs were initially placed under the control of the strong promoter P_{SPI1} (*SPI1* encodes a cell wall protein), and six of them were transformed into the wild type strain CBS7435. Eight clones per TF were screened to check the correct integration of the overexpression construct by colony PCR, and their gene copy number determination was assessed by qPCR. The results of the screening showed that the transformed yeasts integrated only the resistance marker, but not the extra copy of the TF to be overexpressed. As the chosen promoter for the overexpression was very strong (more than 1000x of the native transcript levels), we speculated that overexpression of the TFs with this promoter was too high and thus toxic to the cells. Therefore, another promoter was selected for overexpression: the *THI11* promoter, which is weaker than the previously used promoter and whose expression can be controlled by the presence or absence of thiamine (Delic et al., 2013; Landes et al., 2016). Using P_{THI11}, it was possible to generate overexpression mutant strains for the twelve selected TFs.

For creating the gene knock-outs, CRISPR/Cas9-based homology-directed genome editing was used. For each TF, a guide RNA (gRNA) was designed according to (Gao and Zhao, 2014). The first 20-nucleotide sequence at the 5'-end of the gRNA is complementary to the target sequences located 150-200 bp upstream of the genes to knock-out. Each gRNA was cloned on an episomal plasmid containing also the Cas9 expression cassette (Gassler et al., 2018). Additionally, the 1000 bp regions upstream and downstream of each gene encoding the TFs were assembled to form 2000 bp template DNA for homologous recombination (homology regions HR1 and HR2). For each TF, the corresponding template DNA and plasmid containing gRNA and Cas9 were co-transformed into *P. pastoris*, and the clones obtained were checked by colony PCR (with one pair of primers binding outside of the homology regions and a second pair of primers binding inside the gene to check for potential reintegration of the target genes) (**Figure 6**). Positive knock-out

clones were confirmed for eight out of the twelve selected TFs. Despite testing additional guide RNAs and trying different selection conditions, even after several rounds of transformations it was not possible to obtain knock-out mutants of the following TF genes: *AFU1*, PP7435_Chr1-0170, PP7435_Chr1-0006 and *YGR067C*.

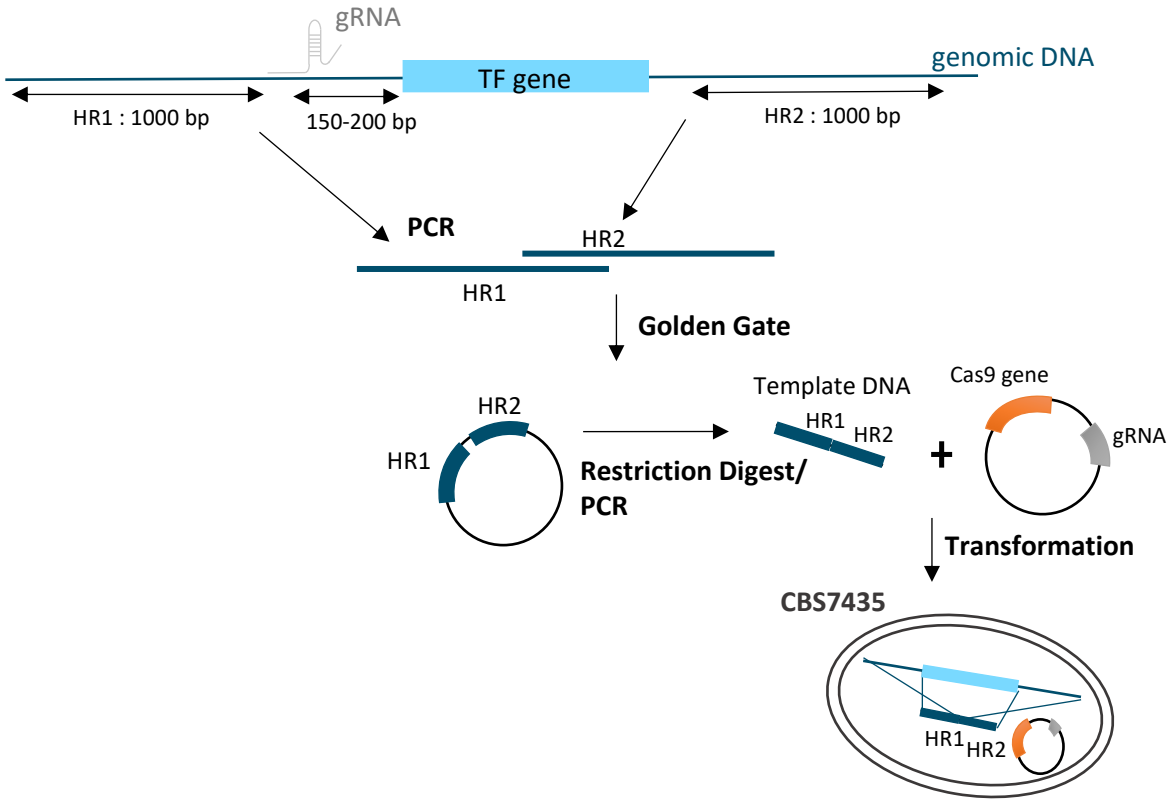


Figure 6: Scheme of knock-out mutant generation.

III. Arraying the TF mutants on different media representing different growth conditions

Once obtained, the mutant strains with overexpression or knock-out of the selected transcriptional regulators were cultivated on media representing different growth and stress conditions, and their ability to grow in these media were determined.

III.1. Arraying of *P. pastoris* TF mutants on different stressors

To array the TF mutants on different stressors, a spotting assay was used. *P. pastoris* wild-type CBS7435 as well as at least three confirmed clones for each TF mutant were spotted on YNB-glucose agar plates (YNB without thiamine in case of the overexpression strains) at decreasing concentrations (in 1:10 dilutions, starting with a similar defined OD_{600}). After two to three days of incubation, the difference in growth for the wild-type and the mutant strains was evaluated.

The spotting assay was performed using the following stressors:

- Cell wall damaging agents (1 and 5 μg/mL calcofluor white, 0.00005% and 0.0001% congo red, 0.01% SDS)
- Agents creating osmotic stress (1M NaCl, 1M KCl)
- Oxidative stress (0.5 mM and 1 mM H_2O_2)

Photos of the individual plates obtained when arraying the mutants on various stressors are in **Appendix 4**. A summary of the results is given in **Table 5**.

Some TF mutants such as overexpression of *SKO1*, *AFU1*, *YGR067C* and *MIT1* as well as the knock-outs of *MIT1* and *PP7435_Chr4-0940* show increased tolerance to osmotic stress (1 M KCl or 1M NaCl). There is no obvious impact by oxidative stress. Several TF mutants have decreased growth in the presence of cell wall damaging agents such as calcofluor white and SDS (KO of *CAT8-1*, and *YPR022C-1*; overexpression of *SKO1, AFU1* and *PP7435_Chr1-0006*). Interestingly, the KO of *SWI5* and the OE of *MIT1* show an adverse behavior on the two tested agents, having higher growth on calcofluor white and lower on SDS.

Table 5: Summary of the spotting assay to array the TF mutants on different stressors. Yellow indicates that no difference was observed between the TF mutant and the wild-type in the given condition. Orange indicates that the TF mutant showed a growth defect compare to the wild-type in the given condition. Green indicates that the TF mutant was growing better than the wild-type on the given condition.

Name of TFs		Osmotic stress		Oxidative stress		Cell wall damaging agents				
		KCl 1M	NaCl 1M	H_2O_2		Calcofluor white		SDS	Congo Red	
				0.5 mM	1 mM	1 µg/mL	5 µg/mL	0,1%	0.00005%	0.0001%
AFU1	OE									
CAT8-1	OE									
	KO									
CAT8-2	OE									
	KO									
MIT1	OE									
	KO									
PP7435_Chr1-0006	OE									
PP7435_Chr1-0170	OE									
PP7435_Chr4-0940	OE									
	KO									
SGF29	OE									
	KO									
SKO1	OE									
	KO									
SWI5	OE									
	KO									
YGR067C	OE									
YPR022C-1	OE									
	KO									

III.2. Arraying of *P. pastoris* TF mutants on different carbon sources

A high-throughput method for cultivation and growth assessment of *P. pastoris* in liquid medium was developed using the wild type strain CBS7435. Cultures in YPD were performed in a 96-well plate and incubated at 30°C in a microplate reader (Tecan Sunrise) for 14.5 to 15 hours, with an OD_{600} measurement every 15 minutes. Initially different culture volumes and inoculation densities were tested and the best parameters were determined (inoculation at OD_{600} 0.01 in 100 µL of medium). Then, cultures in the microplate reader (biological triplicates) were compared to cultures in YPD done in biological duplicates in shake flasks at 30°C, with an OD_{600} measurement every hour for 14 hours, and a final point at 25 hours. OD_{600} of cultures in shake flasks were measured using a desktop photometer as well as the microplate reader used for incubation of the 96 well plates.

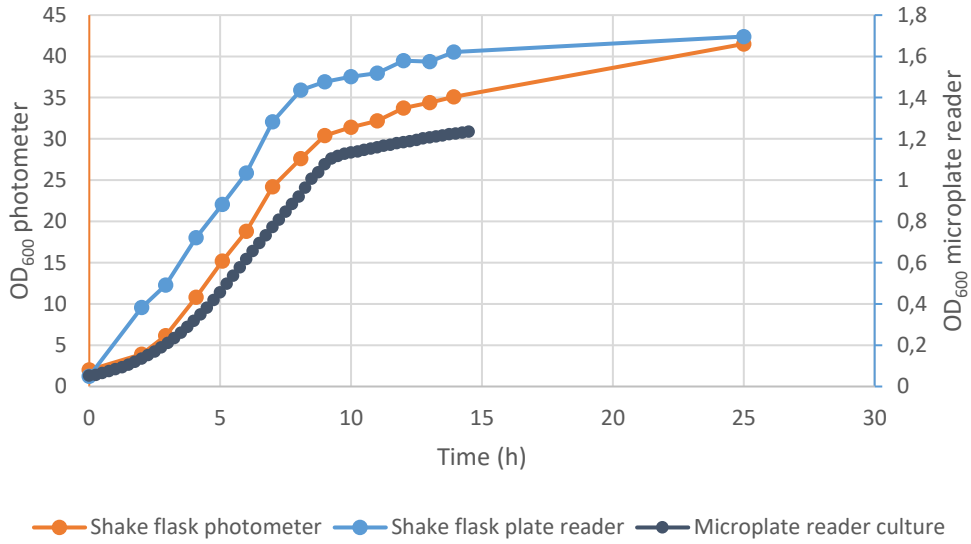


Figure 7: Growth curves of *P. pastoris* CBS7435 cultivated in YPD in shake flasks or in 96-well plate incubated in a microplate reader.

Growth curves of CBS7435 in YPD in shake flasks and in the 96-well plates have a similar shape (**Figure 7**). The OD_{600} obtained after 15 hours of cultures in the 96-well plates is around 1.2, which corresponds to an OD_{600} of around 35 when measured with a small photometer. Only the OD_{600} after 15 hours of growth in shake flasks is a bit higher, around 40 or 1.4 when measured with a photometer or with the microplate reader, respectively. In addition, the maximum growth rates calculated in shake flasks and in the 96-well plates are relatively close (0.36 h^{-1} and 0.38 h^{-1} respectively). The reproducibility of the growth curves and calculated growth rates was also assessed with parallel cultures and was quite good (not shown). Therefore, a culture in a 96-well plate that is incubated in a plate reader seems to be a good method for assaying growth of the future mutant strains in a high throughput manner. This method was therefore used to assess the

growth of the TF mutant strains on minimal media with different carbon sources (2% glucose, 2% glycerol, 1% methanol and 1% ethanol).

Due to technical problems with the lid of the used Tecan reader, we were not able to assess the growth on different carbon sources in liquid cultures with the method presented here for all of the TF mutants. The mutant strains that could not be tested in liquid cultures were instead tested in a spotting assay with YNB agar plates (YNB without thiamine in case of the overexpression strains) supplemented with 2% glucose, 2% glycerol, 1% ethanol or 1% methanol to assess the growth on glucose, glycerol, ethanol or methanol, respectively.

The photos of the plates used to assay the growth of the TF mutants on different carbon sources are in **Appendix 5** and results are summarized in **Table 6**.

The growth of the *SKO1* OE is impaired on all the carbon sources tested, therefore it seems that overexpressing this TF impacts the general growth of *P. pastoris*. The growth of *AFU1* OE is impaired on glucose. Interestingly, *YPR022C-1* OE clones exhibit a different growth on glycerol compare to the WT: the clones reached a similar OD_{600} at the end of the culture, while the growth rate of these clones was lower than for the WT on this carbon source (**Appendix 5**). Both *CAT8-1* and *CAT8-2* KO clones exhibit a lower growth on ethanol, and the growth of *CAT8-2* OE is impaired on glycerol.

The growth of the *AFU1* OE clones was also tested on oleic acid, as a homolog of this gene is involved in fatty acid utilization in *Yarrowia lipolytica*, but no difference was observed compared to the WT *P. pastoris*. This indicates that Afu1 in *P. pastoris* is most probably not playing a major role in fatty acid utilization, but has another yet unknown function.

The growth of the overexpressions of *SKO1*, *SWI5, YPR022C-1* and *MIT1* on methanol is reduced at 30 °C, while the KOs of *SGF29* and *CAT8-2* are only showing a methanol-sensitive phenotype when grown at 37°C. The growth of the *CAT8-1* KO clones is impaired on methanol on both tested temperatures in the spotting assay. However liquid cultures of *CAT8-1* KO on methanol at 30°C did not show any growth defect compare to the WT on this carbon source. The different conditions of the two assays are probably the reasons why there are conflicting results for *CAT8-1* KO on methanol.

Table 6: Summary of the growth assays to array the TF mutants on different carbon sources. Yellow indicates that no difference was observed between the TF mutant and the wild-type in the given condition. Orange indicates that the TF mutant showed a growth defect compare to the wild-type in the given condition. Green indicates that the TF mutant was growing better than the wild-type on the given condition. White indicates that the condition was not tested. "Liquid" indicates that the growth was assessed in a liquid culture, whereas "spotting" indicates that the growth was assessed with a spotting assay.

Name of TFs		Glucose 2%	Glycerol 2%	Ethanol 1%	Methanol 1%		Oleic acid
					30°C	37°C	
AFU1	OE	liquid	liquid	liquid	spotting		spotting
CAT8-1	OE	liquid	liquid	liquid	spotting	spotting	
	KO	liquid	liquid	liquid	liquid / spotting	spotting	
CAT8-2	OE	liquid	liquid	liquid	spotting	spotting	
	KO	liquid	liquid	liquid	liquid/spotting	spotting	
MIT1	OE	liquid	liquid	spotting	spotting		
	KO	spotting	spotting	spotting	spotting		
PP7435_Chr1-0006	OE	liquid	liquid	spotting	spotting		
PP7435_Chr1-0170	OE	spotting	spotting	spotting	spotting		
PP7435_Chr4-0940	OE	liquid	liquid	spotting	spotting		
	KO	Liquid	spotting	spotting	spotting		
SGF29	OE	spotting	spotting	spotting	spotting		
	KO	liquid	liquid	liquid	spotting	spotting	
SKO1	OE	spotting	spotting	spotting	spotting		
	KO	liquid	liquid	liquid	spotting	spotting	
SWI5	OE	spotting	spotting	spotting	spotting		
	KO	liquid (WCW)	liquid	liquid	spotting	spotting	
YGR067C	OE	liquid	liquid	spotting	spotting		
YPR022C-1	OE	liquid	liquid	liquid	spotting	spotting	
	KO	spotting	spotting	spotting	spotting	spotting	

III.3. Phenotype of *SWI5* mutant clones

The overexpression and knock-out clones of the TF *SWI5* exhibit an interesting phenotype on solid as well as on liquid media. On solid medium, the wild-type exhibit a smooth colony morphology, whereas roughened colonies are observed for the *SWI5* knock-out strain (**Figure 8A**).

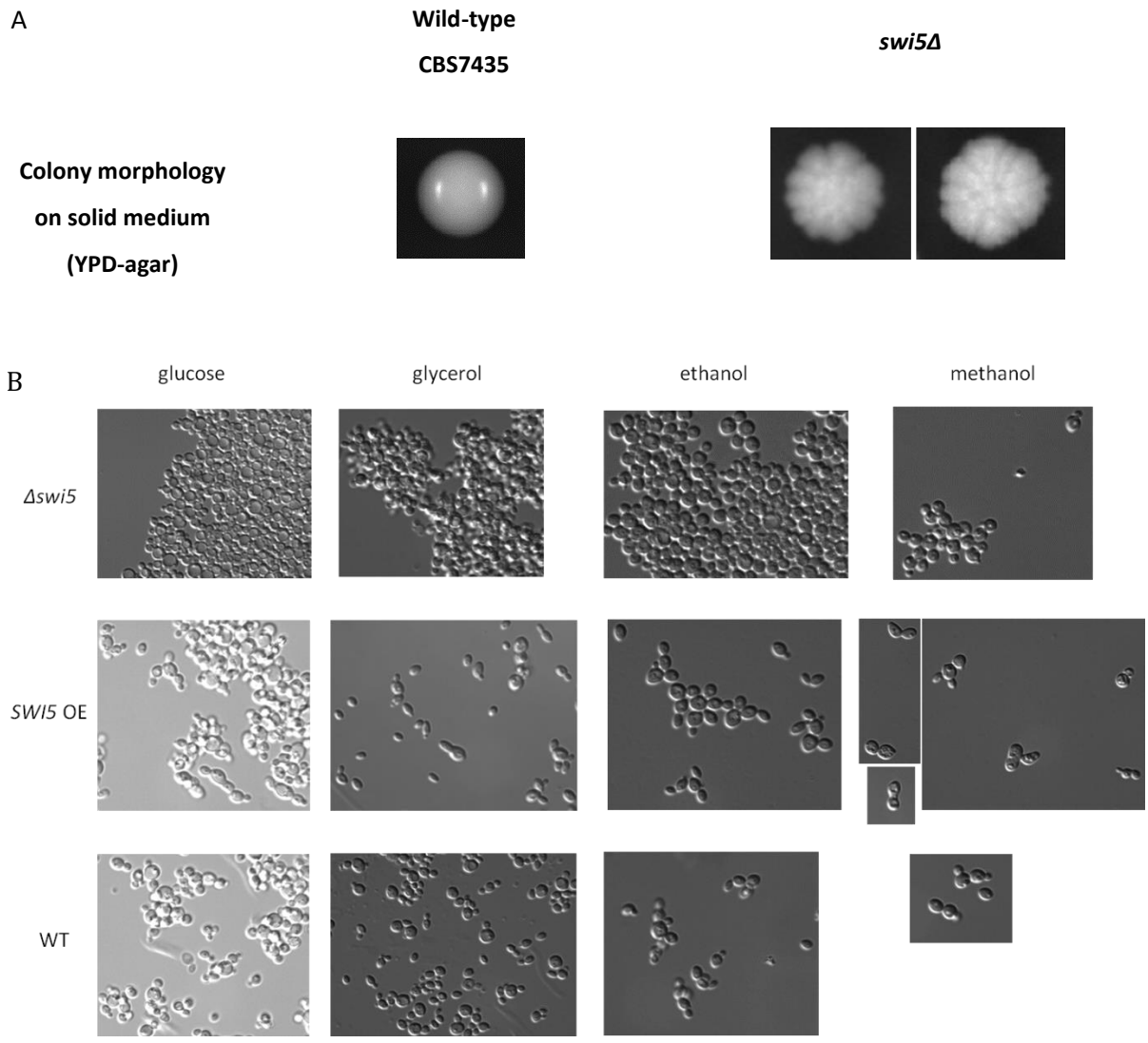


Figure 8: Altered colony morphology and cell shape of the *SWI5* overexpression and knock-out strains on solid and liquid media. (A) Pictures showing colony morphologies of *P. pastoris* wild type CBS7435 and *SWI5* knock-out clones on YPD-agar. (B) Pictures of liquid cultures of *SWI5* knock-out and overexpression strains and CBS7435 (WT) on different carbon sources (2% glucose, 2% glycerol, 1% ethanol and 1% methanol).

The cell morphology of the *SWI5* KO and OE clones was also assessed by growing the clones on different carbon sources in liquid cultures and comparing their morphology to the WT in the same conditions (**Figure 8B**). The knock-out strains show the same phenotype on all carbon sources:

intense flocculation. However, the cell shape of the *SWI5* OE mutants are different depending on the carbon source. On glucose, *SWI5* OE clones show an elongated shape with multiple buds still attached to the mother cell. On glycerol, some of the cells have an elongated shape, others are small and have an ovoid morphology. On ethanol, *SWI5* OE clones flocculate, whereas on methanol, some cells have an elongated shape and did not seem to bud correctly (**Figure 8B**).

IV. Influence of TF mutants on *AOX1* promoter activity

IV.1. *AOX1* transcript levels measurement

The alcohol oxidase I (*AOX1*) promoter is very popular for production of recombinant proteins in *P. pastoris* because it is tightly regulated and strongly transcribed on methanol. Therefore, many studies focused on optimization of the *AOX1* promoter (see introduction for more details on promoter engineering), and finding new targets to alter the *AOX1* promoter capacity can be beneficial for promoter engineering strategies.

Some TF mutants exhibit a growth defect when arrayed on methanol at 30°C (**Table 6**): the OE clones of *SKO1*, *SWI5*, *YPR022C-1* and *MIT1* as well as the KO clone of *MIT1*. For these TFs, the transcript levels of *AOX1* were measured in the OE and KO mutants on different carbon sources: the TF mutants were cultivated in biological duplicates on minimal medium with limiting glucose (m2p media development kit) until OD_{600} reached 4, and then shifted to 2% glycerol and 1% methanol for 5 h to measure the P_{AOX1} induction on glycerol and methanol, respectively. At the end of the induction time, cells were harvested and RNA was extracted for rt-qPCR.

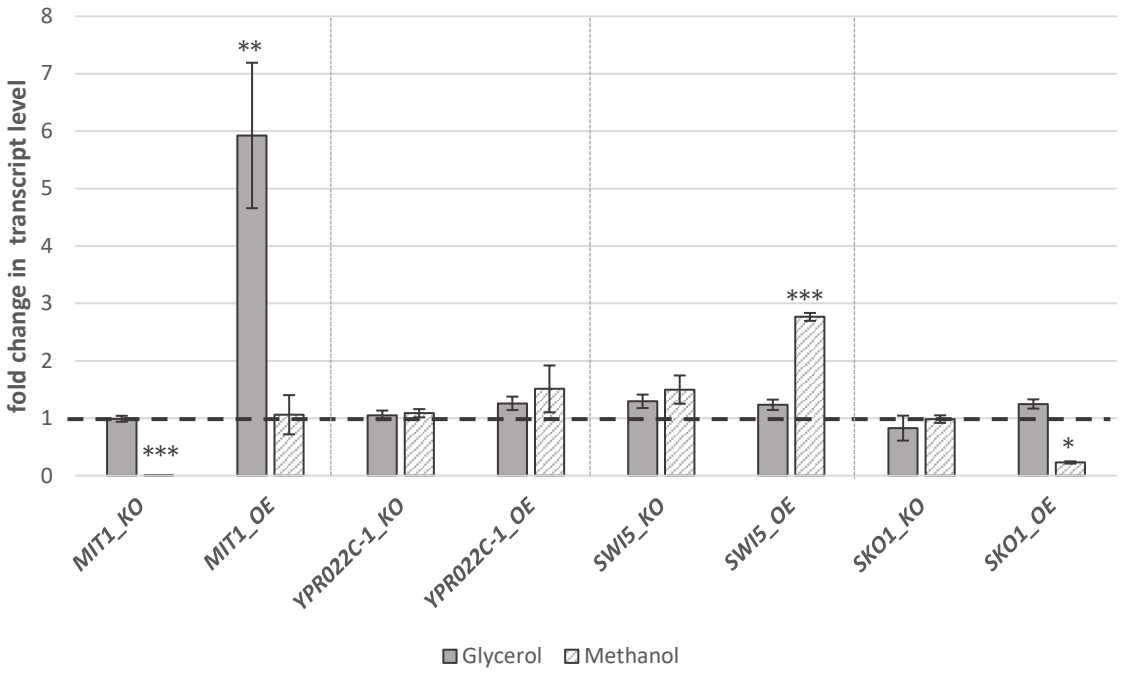


Figure 9: Transcript levels of the *AOX1* gene in the *MIT1*, *YPR022C-1*, *SWI5* and *SKO1* mutant strains on glycerol and methanol were determined by qRT-PCR. Gene expression levels were normalized to the reference gene *ACT1* and quantified relative to wild-type levels (set to 1.0; dashed line). Error bars represent the standard deviations of two independent biological samples each measured in technical triplicates. The statistically significant differences compared to the input-DNA are indicated with asterisks (Student's t-test; * $p<0.05$, ** $p<0.01$, *** $p<0.001$).

MIT1 KO mutant exhibits a decrease in *AOX1* transcript level on methanol, whereas *AOX1* transcript level increases in *MIT1* OE on glycerol compared to the wild-type. These regulation patterns fit to what was recently described in the literature for this TF (Wang et al., 2017b; Wang et al., 2016b). No significant difference was observed in the *AOX1* transcript levels in the *YPR022C-1* mutant strains on glycerol and methanol. The *SWI5* OE mutant exhibits an increase in *AOX1* transcript levels on methanol, whereas on the same carbon source, the *SKO1* OE mutant exhibits a decrease in *AOX1* transcript levels (**Figure 9**).

IV.2. Effect of *SWI5* overexpression on P_{AOX1}

In order to confirm the regulation of *AOX1* by Swi5 on the protein level, an eGFP reporter construct was generated which expresses eGFP under the control of the *AOX1* promoter. This construct was transformed into *SWI5* OE and the wild-type CBS7435. The obtained transformants were screened for production of eGFP with methanol induction.

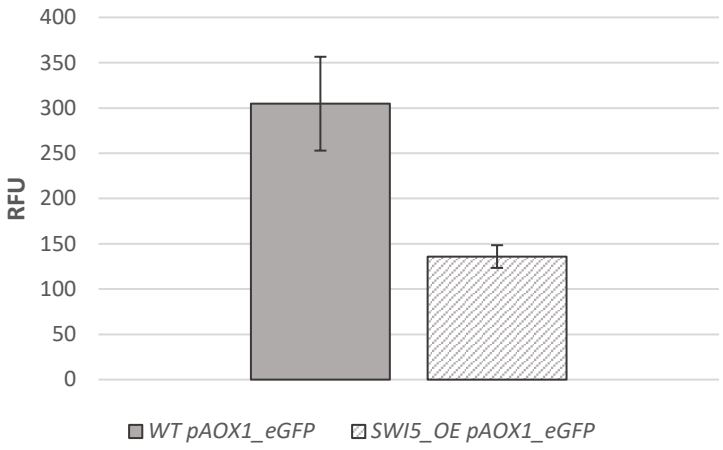


Figure 10: Production of eGFP under the *AOX1* promoter in the *P. pastoris* wild-type and in the *SWI5* OE mutant strains. The error bars represent the standard deviations obtained with seven independent biological samples.

The eGFP levels are lower in the *SWI5* OE clones than in the WT (**Figure 10**), which does not correlate with the results obtained for the transcript levels analysis (**Figure 9**). Since the cell morphology of the *SWI5* OE clones was shown to be strongly altered on all carbon sources and specifically on methanol, this could have an impact on the measurement of eGFP fluorescence, as well as on the capacity of the cell to produce correct recombinant proteins.

V. Discussion

In total, 12 TFs from *P. pastoris* were selected due to their interesting regulation patterns in industrially relevant conditions. OE and KO mutant strains were generated for each of these TFs, and their growth was assessed on different media to gain insights on the potential functions of these TFs in *P. pastoris,* and to select TFs for more precise characterization.

A summary of the observed phenotype for each TF mutant is given in **Table 7**.

Table 7: Summary of the putative function of the selected TFs in other species and the phenotypes observed when these TFs were overexpressed or knocked-out in *P. pastoris* CBS7435.

Name	Putative function	Mutant	Observed phenotypes in *P. pastoris* CBS7435
AFU1	Activates fatty acid utilization (*POR1* in *Y. lipolytica*) (Poopanitpan et al., 2010)	OE	- increased tolerance to osmotic stress - decreased resistance to SDS - decreased growth on glucose - no difference in growth on oleic acid
		KO	- mutant not obtained
CAT8-1	Binds carbon sources responsive elements (*S. cerevisiae*, *K. lactis*) (Mehlgarten et al., 2015)	OE	- no significant phenotype
		KO	- decreased tolerance to osmotic stress - decreased tolerance to calcofluor white - decreased growth on ethanol and on methanol at 37°C
CAT8-2	Binds carbon sources responsive elements (*S. cerevisiae*, *K. lactis*) (Mehlgarten et al., 2015)	OE	- decreased growth on glycerol
		KO	- increased tolerance to osmotic stress - decreased growth on ethanol and on methanol at 37°C
MIT1	Regulation of methanol utilization genes (*P. pastoris*) (Wang et al., 2016b)	OE	- increased resistance to osmotic stress - increased resistance to cell wall damaging agents - decreased growth on glycerol and methanol - increased *AOX1* transcript levels on glycerol
		KO	- increased resistance to osmotic stress - decreased growth on methanol - decreased *AOX1* transcript levels on methanol
PP7435_Chr1-0006	No information	OE	- decreased tolerance to calcofluor white
		KO	- mutant not obtained
PP7435_Chr1-0170	Similar to putative AP1 (*Candida maltose*)	OE	- increased tolerance to oxidative stress
		KO	- mutant not obtained
PP7435_Chr4-0940	No information	OE	- no significant phenotype
		KO	- increased growth on glucose - increased tolerance on osmotic stress
SGF29	Component of SAGA, SLIK, ADA complexes (histone modifying complexes) (*S. cerevisiae*) (Bian et al., 2011; Shukla et al., 2012)	OE	- no significant phenotype
		KO	- decreased growth on glucose
SKO1	Osmotic and oxidative stress response (*S. cerevisiae*)(Rep et al., 2001)	OE	- decreased tolerance to calcofluor white - decreased growth on all carbon sources tested - decreased AOX1 transcript levels on methanol
		KO	- increased resistance to osmotic stress
SWI5	Involved in mating-type switching and pseudohyphal growth (*S. cerevisiae*) (Bhoite et al., 2001; Dohrmann et al., 1992; Pan and Heitman, 2000)	OE	- increased resistance to osmotic stress - decreased growth on methanol - altered cell morphology depending on the carbon source - increased *AOX1* transcript levels on methanol
		KO	- increased resistance to CCW, decreased tolerance to SDS - rough colony morphology on solid medium, intense flocculation in liquid cultures
YGR067C	No information	OE	- increased resistance to osmotic stress - decreased resistance to oxidative stress
		KO	- mutant not obtained
YPR022C-1	No information	OE	- decreased growth on glycerol and methanol - increased tolerance on osmotic stress - decreased tolerance to SDS
		KO	- decreased tolerance to cell wall damaging agents

SKO1 was shown in *S. cerevisiae* to encode a transcriptional regulator involved in the regulation of genes from the osmotic and oxidative stress response (Rep et al., 2001). More specifically, Sko1 binds as a heterodimer to the consensus sequence TGACGTCA and represses transcription of stress defense genes by recruiting the Cyc8(Ssn6)-Tup1 general repressor complex to promoters (Proft and Struhl, 2002). Upon osmotic shock, Sko1 is phosphorylated by the Hog1 kinase (a mitogen-activated protein kinase involved in osmoregulation), which inactivates the repressor complex and turns Sko1 into a transcriptional activator (Proft et al., 2001; Proft and Struhl, 2002). Sko1 then activates a regulatory network by binding promoters of other transcription factors (including *MSN2*, involved in the general stress response; *MOT3*, having a role in cellular adjustment to osmotic stress and *ROX1*, which regulates the expression of hypoxia-induced genes). Sko1 also activates the transcription of genes encoding proteins that directly relieve osmotic stress, such as cell wall proteins (*SED1*, *CWP1*), vacuolar or cytoplasmic transporters (*STL1*, *HXT5*) as well as oxidoreductases implicated in the repair of oxidative damages (*AHP1*, *SFA1*) (Proft et al., 2005; Rep et al., 2001). In addition, Sko1 was shown to have a role in the recruitment of the nucleosome-remodeling complexes SAGA and Swi/Snf to osmotic inducible promoters (Proft and Struhl, 2002). Finally, the nuclear localization of Sko1 is regulated by its phosphorylation by the protein kinase A (PKA): Sko1 is present in the nucleus of unstressed cells, whereas upon severe salt stress it redistributes to the cytosol (Pascual-Ahuir et al., 2001). In *S. cerevisiae*, *sko1Δ* deletion strains exhibit an increase resistance to osmotic stress (on NaCl and LiCl) (Proft and Serrano, 1999), which is also seen in *P. pastoris* (on NaCl and KCl). We also observed a growth defect of the *SKO1* OE mutants on all carbon sources and a decreased tolerance to a cell wall damaging agent. Given the observed phenotypes in the mutants of this TF in *P. pastoris*, Sko1 may have similar activities in this yeast as in *S. cerevisiae*, although a lot more investigations would be necessary to confirm or infirm this hypothesis. For example, transcript levels of selected genes shown to be regulated by Sko1 in *S. cerevisiae* could be measured in the *P. pastoris SKO1* OE and KO strains in order to know if they are also regulated by Sko1 in *P. pastoris*. The potential phosphorylation of the Sko1 protein by the Hog1 kinase (encoded by *HOG1*, PP7435_Chr1-0238) could also be investigated in *P. pastoris*. Finally, tagging Sko1 with a fluorescent protein such as GFP could enable to investigate the subcellular localization of this TF in *P. pastoris* in unstressed or osmotically stressed cells using fluorescence microscopy.

In *S. cerevisiae*, Swi5 was shown to activate the transcription of the *HO* gene, which encodes an endonuclease initiating mating-type switching (Bhoite et al., 2001). In addition, a paralog of *SWI5*, termed *ACE2* is found in *S. cerevisiae*. *SWI5* and *ACE2* have 37% similarity in their full length nucleotide sequences, and they share 95% similarity in their DNA binding domains. The TFs Swi5 and Ace2 regulate a shared set of genes in *S. cerevisiae* and bind identical DNA motifs *in vitro* with

comparable affinities (Dohrmann et al., 1996; McBride et al., 1999), but they also show distinct promoter specificities (Dohrmann et al., 1992; Dohrmann et al., 1996). In particular, Ace2 was shown to regulate the expression of genes encoding enzymes responsible for locally specific cell wall degradation, which enables proper separation of mother and daughter cells during mitosis (Dohrmann et al., 1992; Kuranda and Robbins, 1991). The *ace2Δ* knock-out mutants in *S. cerevisiae* have an altered colony morphology (Voth et al., 2005) and they tend to flocculate a lot in liquid media (Dohrmann et al., 1992). Swi5, in addition to mating-type switching, was shown to be involved in the regulation of pseudohyphal growth in *S. cerevisiae* (Pan and Heitman, 2000). In *P. pastoris*, the paralog *ACE2* is not present, since it originates from the whole genome duplication event in *S. cerevisiae*. Genes often diverge and adopt different more specific functions after being duplicated compared to an ancestral gene in a pre-whole genome duplication species such as *P. pastoris*. In addition, considering the phenotypes observed in *P. pastoris* in the *SWI5* mutants (rough colony morphology on solid medium and intense flocculation in liquid cultures for the *swi5Δ* mutants, altered cell morphology depending on the carbon source for the *SWI5* OE mutants), *PpSwi5* might have similar activities to Swi5 and/or Ace2 from *S. cerevisiae*.

Since starting this project, the TF Mit1 was also studied by Wang and coworkers in *P. pastoris*. This TF was shown to regulate many genes involved in the methanol utilization pathway, including *AOX1*. Especially, it was shown that *MIT1* is important for activation of *AOX1* on methanol (Wang et al., 2016b). Furthermore as we observed in our transcript levels analysis of *AOX1*, knocking-out *MIT1* suppresses the activation of *AOX1* transcription on methanol, while overexpressing this TF leads to increased transcription of *AOX1* on glycerol (also seen in (Wang et al., 2017b)). This TF was also used for promoter engineering of P_{AOX1} to develop methanol-free induction systems in *P. pastoris* (Vogl et al., 2018a; Wang et al., 2017b). We also saw in our assays that the *MIT1* mutants have an increased resistance to osmotic stress and cell wall damaging agents, suggesting that this TF might be involved in other processes than just the regulation of the methanol utilization pathway.

AFU1 was described in the yeast *Yarrowia lipolytica* to activate fatty acids utilization. In *P. pastoris*, no difference was observed between the *AFU1* OE mutant and the WT when grown on oleic acid. Either overexpressing *AFU1* in *P. pastoris* does not give a phenotype on oleic acid even if this TF is involved in fatty acids utilization, or the function of this TF is different in *P. pastoris* and in *Y. lipolytica.* Indeed, it is known that TF functions are not always conserved during evolution, which means that a specific TF (characterized through its DNA binding domain) can regulate different target genes in different yeast species (Hogues et al., 2008; Lavoie et al., 2010; Tuch et al., 2008). Here we cannot rule out any of the two possibilities, as further experiments were not done.

Regarding the two TFs Cat8-1 and Cat8-2, growth defects on ethanol were observed in the KO mutants for both TFs. In addition, the KO mutants of *CAT8-1* showed decreased tolerance to cell wall damaging agents and osmotic stress, whereas the KO mutants of *CAT8-2* exhibit an increased resistance on osmotic stress reagents. The two TFs Cat8-1 and Cat8-2 are homologs of Cat8, which was shown to be involved in the regulation of utilization of C2 carbon sources such as ethanol and acetate in the yeasts *S. cerevisiae* and *K. lactis*. We further investigated the activities of Cat8-1 and Cat8-2 in *P. pastoris*, especially regarding the regulation of ethanol utilization, as presented in Chapter 2 of this thesis.

For the 6 other TFs, either no information was available on their putative functions, or only sparse information was available. The growth and spotting assays were not enough to obtain much information on their activities in *P. pastoris*.

Finally, we investigated the influence of 4 of the selected TFs on the *AOX1* promoter activity, a promoter widely used for recombinant protein production in *P. pastoris*, but were not able to identify new interesting targets for potential use in P_{AOX1} promoter engineering.

CHAPTER 2: THE Cat8 TFs REGULATE CARBON SOURCE UTILIZATION IN *P. pastoris*

I. Introduction

As seen in the previous chapter, the overexpression and knock-out mutant strains for the transcriptional regulators Cat8-1 and Cat8-2 showed interesting phenotypes, especially regarding growth on ethanol. The function of these two TFs regarding the regulation of ethanol utilization was therefore studied in more depth.

The yeast *P. pastoris* adapts to different growth conditions through various mechanisms, including reprogramming of gene expression and protein synthesis (Prielhofer et al., 2015). The release from glucose and catabolite repression alters the transcription of genes involved in numerous cellular processes, such as glycolysis, gluconeogenesis, tricarboxylic acid (TCA) cycle and metabolism of alternative carbon sources (Prielhofer et al., 2015). Growth on non-fermentable carbon sources such as glycerol and ethanol require enzymes from the gluconeogenesis and the glyoxylate cycle, among others. In the yeasts *S. cerevisiae* and *K. lactis*, the genes encoding these enzymes are activated through upstream activation sites (UAS) found in their promoters, such as the carbon source responsive elements (CSREs) (Mehlgarten et al., 2015; Turcotte et al., 2010). The CSREs are under the control of two transcriptional regulators, which are members of the binuclear zinc cluster family: Cat8 and Sip4 (Roth et al., 2004; Vincent and Carlson, 1998).

Cat8 (CATabolite repression) and Sip4 (Snf1 interacting protein) possess a highly similar N-terminal zinc cluster ($Zn(II)_2Cys_6$) binding domain (Rahner et al., 1996), but they share little similarity in the rest of their protein sequences (Mehlgarten et al., 2015; Turcotte et al., 2010). Although both Cat8 and Sip4 were shown to bind the carbon source responsive element (CSRE) (Roth et al., 2004; Vincent and Carlson, 1998), Sip4 recognizes and binds to a more specific CSRE motif than Cat8, which probably explains why Cat8 and Sip4 contribute unequally to gene activation via binding to this motif (Hiesinger et al., 2001). In *S. cerevisiae* and *K. lactis*, Cat8 and Sip4 were described to be activators of transcription, but their mechanism of action in these two yeast are slightly different (Mehlgarten et al., 2015).

In *S. cerevisiae*, the expression and activities of *CAT8* and *SIP4* were shown to be regulated by glucose, in a process mediated by the Snf1 kinase (Hardie et al., 1998). This kinase seems to have a fundamental role in glucose derepression through the activation of various transcriptional activators and the deactivation of the Cys_2His_2 zinc finger protein Mig1. In the presence of glucose, the TF Mig1 binds to the promoter of genes such as *CAT8*, which represses their expression. On

derepressing conditions, the Snf1 kinase phosphorylates Mig1, which leads to its inactivation and the consequent induction of *CAT8* transcription (Carlson, 1999; Schüller, 2003). The Cat8 protein is then activated via phosphorylation by Snf1, and induces the transcription of various genes involved in the growth on non-fermentable carbon sources such as genes of the C2 anabolism, the glyoxylate cycle and the gluconeogenesis (De Vit et al., 1997; Haurie et al., 2001; Hedges et al., 1995; Lesage et al., 1996; Randez-Gil et al., 1997; Tachibana et al., 2005). *Sccat8* knock-out mutants are unable to grow on gluconeogenic carbon sources such as glycerol, ethanol, lactate and acetate (Hedges et al., 1995; Rahner et al., 1996), whereas *Scsip4* knock-out mutants have no apparent growth phenotype on any of the tested carbon sources. This indicates that in *S. cerevisiae*, Sip4 plays a minor role in CSRE-dependent regulation (Lesage et al., 1996). In addition, the *SIP4* promoter contains CSRE motifs which were shown to be bound by Cat8, indicating that Cat8 is regulating the transcription of *SIP4* in *S. cerevisiae* (Vincent and Carlson, 1998).

In the yeast *K. lactis*, both Cat8 and Sip4 are present but the regulatory networks are different than in *S. cerevisiae*. The *Klcat8* knock-out mutants can grow on glycerol but not on C2 carbon sources such as acetate and ethanol, which shows that unlike in *S. cerevisiae*, the gluconeogenesis encoding genes are not regulated by Cat8 nor Sip4 in *K. lactis* (Georis et al., 2000). On the other hand, contrary to *S. cerevisiae* where Cat8 but not Sip4 is required for growth on ethanol, both *Klcat8* and *Klsip4* knock-out mutants exhibit a growth defect on C2 carbon sources (Mehlgarten et al., 2015). It was also shown that only *Kl*Sip4 binds to the CSRE motifs in the promoters of the glyoxylate pathway genes and the carnitine shuttle encoding genes (Mehlgarten et al., 2015; Rodicio et al., 2008). In *K. lactis*, as in *S. cerevisiae*, Cat8 was shown to activate the transcription of Sip4. *Kl*Cat8 was also shown to be regulated via phosphorylation of a conserved serine residue (Ser-661) by the Snf1 kinase (Charbon et al., 2004).

Homologs of the Cat8 transcription factor were also found in other yeasts, although their function was not studied in as much depth as in *S. cerevisiae* or *K. lactis*. *Candida albicans* Cat8 knock-out mutants have a similar phenotype to the wild-type in terms of gluconeogenesis, glyoxylate shunt and ethanol utilization pathway, and *Ca*Cat8 does not seem to regulate the gluconeogenic gene *PCK1* nor *ICL1* (encoding isocitrate lyase, an enzyme of the glyoxylate cycle) (Ramirez and Lorenz, 2009). Cat8 knock-out mutants in *Ogataea (Hansula) polymorpha* exhibit a growth defect on glycerol, ethanol and xylose, and have a higher ethanol production from xylose fermentation (Ruchala et al., 2017). Finally, in *Pichia guillermondii*, knocking-out the *CAT8* gene triggers the respiro-fermentative metabolism of this Crabtree-negative yeast (Qi et al., 2014).

In *Pichia pastoris*, two putative transcription factors homolog of Cat8 termed *CAT8-1* (PP7435_ Chr2-0516) and *CAT8-2* (PP7435_ Chr4-0434) are found. The *CAT8-2* gene was shown to be induced on limiting glucose (about 39-fold up-regulated compared to excess glucose) and on

methanol (about 7-fold up-regulated compared to excess glucose). No up- or down-regulation of *CAT8-1* was observed in the same conditions (Prielhofer et al., 2015). It is yet unclear which genes are regulated by the TFs Cat8-1 and Cat8-2 and if one of them is the homolog of Sip4. In order to study the role of *CAT8-1* and *CAT8-2* in *P. pastoris*, overexpression and knock-out mutants of each of these two TF genes were generated and the ability of the TF mutant strains to grow on different nutrient sources was tested (see chapter 1). Based on the findings, the transcript levels of some selected genes from the carbon metabolism were quantified to see how their regulation patterns on different carbon sources were affected in the absence of either or both Cat8 homologs. Furthermore, the promoter activity of *CAT8-1* and *CAT8-2* was measured with an eGFP reporter assay to assess if one TF is regulating the expression of the other. Finally, it was also investigated if there is an impact on the regulation of *CAT8-1* and *CAT8-2* by the transcription factors Mig1-1 and Mig1-2 in *P. pastoris*.

II. Phylogenetic analysis of Cat8 and Sip4 homologs in *Saccharomycetes*

The transcription factors Cat8 and Sip4 both belong to the family of binuclear zinc cluster proteins. These proteins are found only in fungi and typically possess a zinc cluster motif with the following consensus sequence: $CysX_2CysX_6CysX_{5\text{-}12}CysX_2CysX_{6\text{-}8}Cys$. The six conserved cysteine residues are involved in the folding of the zinc cluster domain, which is important for DNA recognition. Usually, the DNA binding domain of binuclear zinc cluster proteins is located near the N-terminus, while the activation domain is found at the C-terminus. Within the DNA binding domain, a variant linker region connects the zinc cluster motif to a dimerization domain, and contributes to DNA binding specificity. The dimerization region consists of a structure similar to the leucine zipper heptad repeat, and is involved in protein-protein interactions (Todd and Andrianopoulos, 1997).

As mentioned previously, *CAT8-1* and *CAT8-2* are two homologs of *CAT8,* which was characterized in *S. cerevisiae* and *K. lactis*. Both Cat8-1 and Cat8-2 exhibit the characteristic binuclear zinc cluster $Zn(II)_2Cys_6$ binding domain. To find out if one of these two genes could encode a protein structurally related to Sip4, phylogenic analysis was performed.

161 sequences obtained by BLAST search on the NCBI database using 6 functionally characterized Cat8 and Sip4 protein sequences from various yeasts as well as Cat8-1 and Cat8-2 were aligned, and a phylogenic tree was calculated (**Figure 11**).

Generally, a high sequence variation was observed within the Cat8 and Sip4 variants from the different yeasts. Only one domain was conserved in all the sequences: the DNA binding domain. Specifically, the six cysteine residues involved in forming the domain are well conserved in the 8

functionally characterized Cat8 and Sip4 homologs (**Figure 11A**). The adjacent Leucine zipper dimerization domain in the C-terminal region of the binuclear cluster is conserved as well (**Figure 11A**).

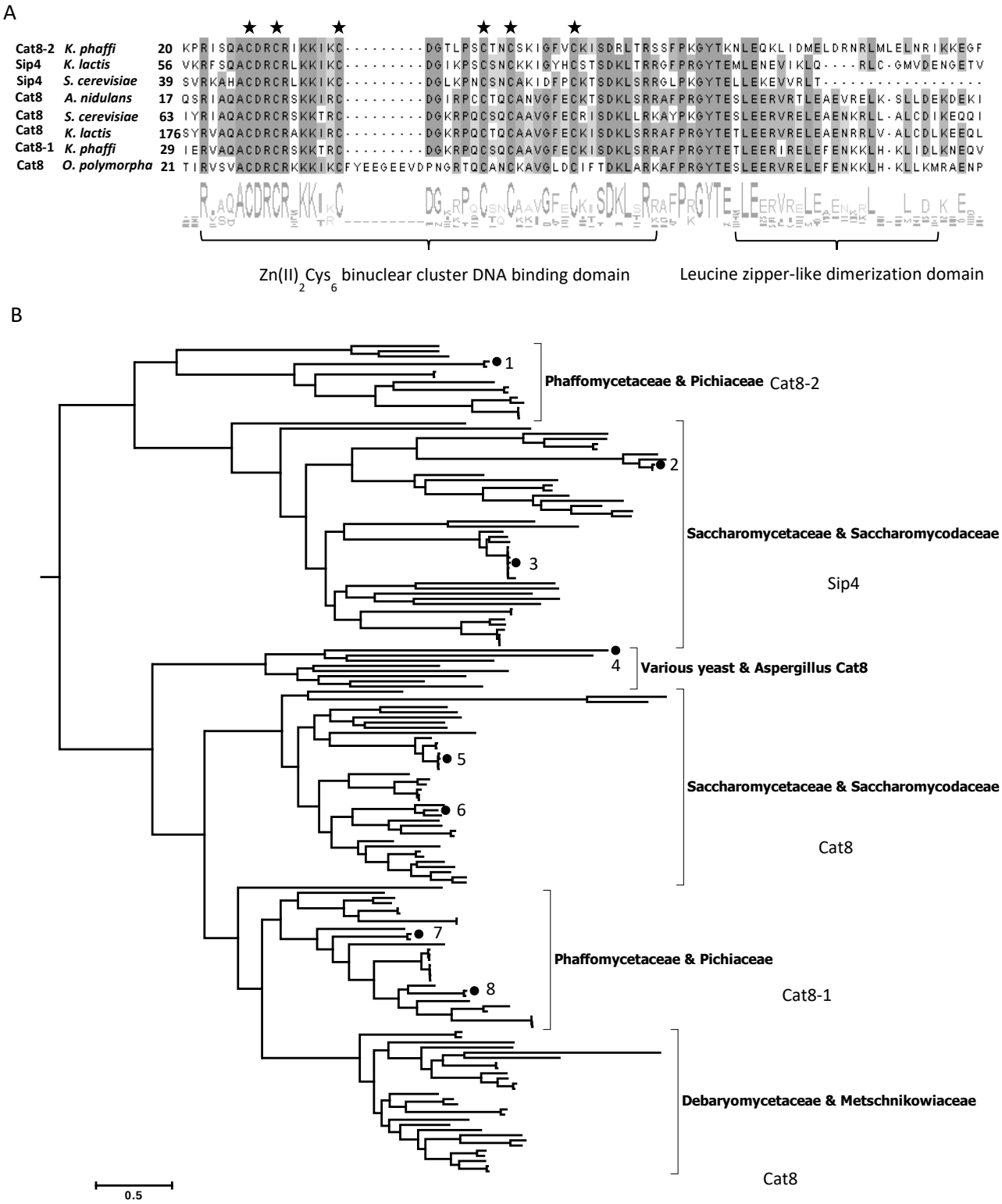


Figure 11: Phylogenetic analyses of 161 sequence homologs of Cat8 and Sip4. (A) Multiple sequence alignments of the fungal $Zn(II)_2Cys_6$ DNA binding domain of the 6 functionally characterized Cat8 and Sip4 as well as Cat8-1 and Cat8-2 homologs in *Saccharomycetes* (Yeast; taxid:4891). Identical residues are shown in grey boxes and the cysteine residues involved in forming the DNA binding domain are marked with a star. (B) Phylogenetic tree based on full length amino acid sequences of Cat8, Sip4, Cat8-1 and Cat8-2 homologs in *Saccharomycetes* (Yeast; taxid:4891). Full circles (•) represent the positions of the 6 functionally characterized Cat8 and Sip4 homolog as well as Cat8-1 and Cat8-2 protein sequences that were used for BLAST search: 1. CAT8-2_Komagataella_phaffii_GS115, 2. Sip4_Kluyveromyces_lactis, 3. Sip4_Saccharomyces_cerevisiae, 4. CAT8_Aspergillus_nidulans_FGSC_A4, 5. CAT8_Saccharomyces_cerevisiae, 6. CAT8_Kluyveromyces_lactis, 7. CAT8_Komagataella_phaffii_GS115 and 8. CAT8_Ogataea_polymorpha.

Since only few sequences were functionally characterized among the ones used for the alignment and the calculation of the tree, the annotations are uncertain. Five big clades cluster were nonetheless identified according to functionality and fungal genus: the Cat8-2 clade of *Phaffomycetaceae* and *Pichiaceae*, the Sip4 clade of *Saccharomycetaceae* and *Saccharomycodaceae*, the Cat8 clade of *Saccharomycetaceae* and *Saccharomycodaceae*, the Cat8 clade of *Phaffomycetaceae* and *Pichiaceae*, and finally the Cat8 clade of *Debaryomycetaceae* and *Metschnikowiaceae*. In addition, a small clade of 8 sequences of 5 different yeast genes and one *Aspergillus* Cat8 is also observed (**Figure 11B**). In *Debaryomycetaceae* and *Metschnikowiaceae*, a Cat8 homolog was found, but not a Sip4 homolog.

In 36 of the 44 Sip4 sequences of the *Saccharomycetaceae* and *Saccharomycodaceae* cluster, a conserved region of 16 amino acids is observed (data not shown). This regions contains a lot of charged amino acids, is rather short and does not seem to be a functional domain. It could be a surface interaction domain, but so far it does not have a clear function. Interestingly, this region seems to be Sip4 specific, as it is not found in any of the sequences forming the Cat8-2 clade of *Phaffomycetaceae* and *Pichiaceae*.

Based on full-length protein sequences, the Cat8-1 protein from *P. pastoris* clusters with Cat8, whereas the Cat8-2 protein clusters with Sip4, although constituting a different group from the characterized *S. cerevisiae* and *K. lactis* Sip4 homologs (**Figure 11B**).

III. *CAT8-1* and *CAT8-2* are essential for ethanol assimilation in *P. pastoris*

To identify in which cellular processes the TFs Cat8-1 and Cat8-2 are involved, overexpression and knock-out strains of *CAT8-1* and *CAT8-2* were generated (for details see Chapter 1). For the overexpression, both genes were cloned under the control of the tunable *THI11* promoter, which is repressed in presence of thiamine in the growth medium and induced in thiamin-depleted conditions (Delic et al., 2013; Landes et al., 2016), and transformed in the wild-type *P. pastoris* strain CBS7435. The knock-out strains of *CAT8-1* and *CAT8-2* were generated using CRISPR/Cas9-based homology directed genome editing (Gassler et al., 2018). Furthermore, a double knock-out *cat8-1Δcat8-2Δ* was generated. The growth of the overexpression and knock-out strains as well as the *P. pastoris* wild-type was assessed in liquid cultures with YNB (without thiamine for the

overexpression stains) containing either 2% glucose, 2% glycerol, 2% ethanol or 1% methanol (**Figure 12**).

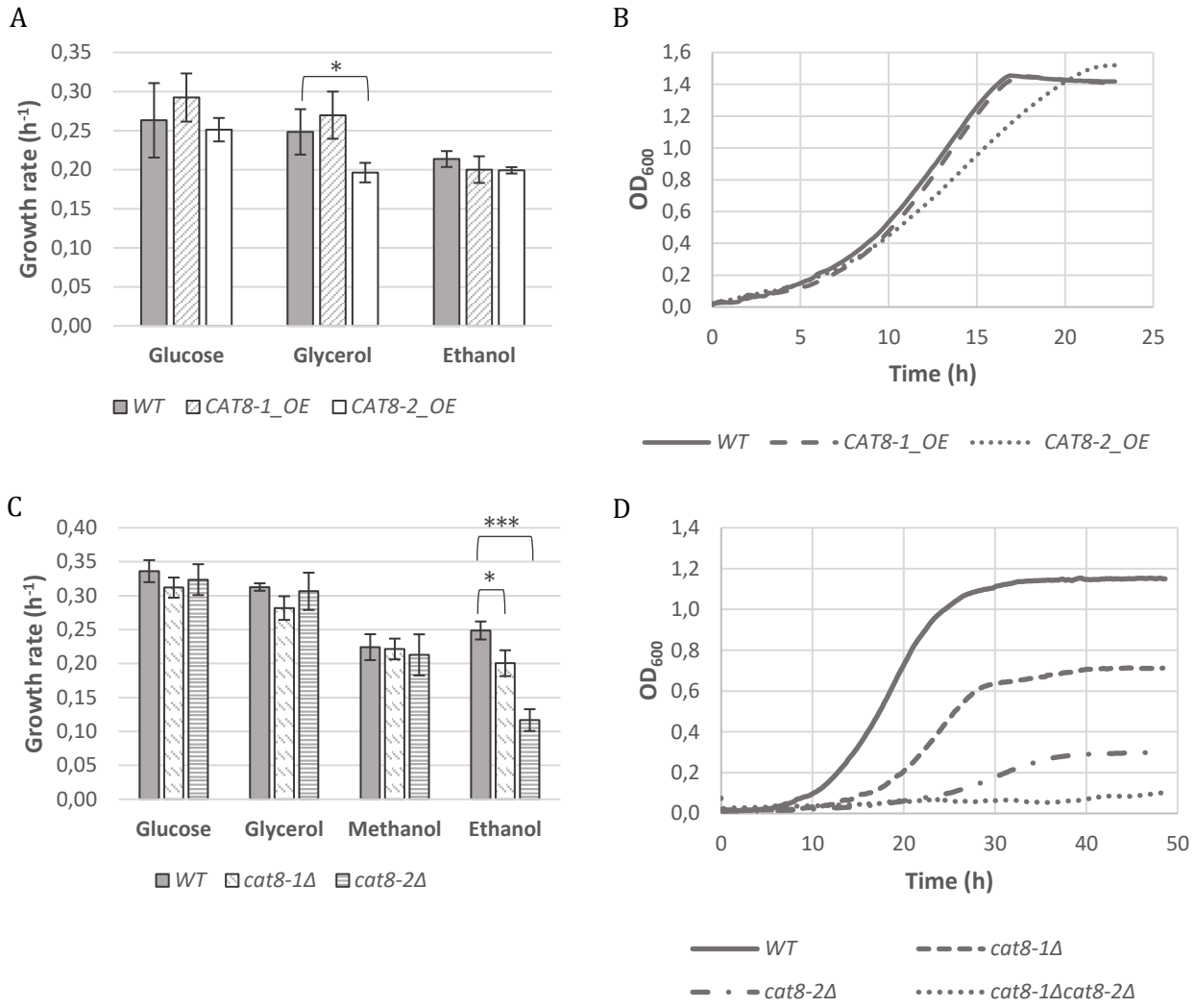


Figure 12: Influence of *CAT8-1* and *CAT8-2* overexpression and deletion on carbon source utilization.
(A) Growth rates of the *CAT8-1* and *CAT8-2* overexpression mutants and the *P. pastoris* wild-type (WT) on YNB without thiamine with 2% glucose, 2% glycerol and 1% ethanol. (B) Growth curves of the *CAT8-1* and *CAT8-2* overexpression mutants and the WT on YNB without thiamine with 2% glycerol. (C) Growth rates of *cat8-1Δ*, *cat8-2Δ* and the WT on 2% glucose, 2% glycerol, 1% methanol and 1% ethanol. (D) Growth curves of *cat8-1Δ*, *cat8-2Δ*, the double knock-out *cat8-1Δ cat8-2Δ* and the WT on 1% ethanol. Error bars represent the standard deviations of three to four independent biological samples each measured in technical triplicates. The statistically significant differences compared to the input-DNA are indicated with asterisks (Student's t-test; * $p<0.05$, ** $p<0.01$, *** $p<0.001$).

No significant difference is observed between the growth rates of the *CAT8-1* overexpression strains and the ones of the WT on the different carbon sources (**Figure 12A**). For the *CAT8-2* overexpression strains, no significant difference is observed with the growth rates of the WT on glucose and methanol. However, the growth rate of the *CAT8-2* overexpression strain is reduced on glycerol (**Figure 12A**), and reaches the stationary phase later than the WT on this carbon source (**Figure 12B**).

No significant difference is observed between the growth rates of *cat8-1Δ* and *cat8-2Δ* compared to the ones of the WT on glucose, glycerol and methanol (**Figure 12C**). However, *cat8-1Δ* and *cat8-2Δ* have lower growth rates than the WT on ethanol (0.20 h^{-1} for *cat8-1Δ* and 0.12 h^{-1} for *cat8-2Δ* against 0.25 h^{-1} for the WT), and they reach a lower OD_{600} at the end of the culture (**Figure 12D**). In addition, the *cat8-1Δcat8-2Δ* double knock-out is unable to grow on ethanol (**Figure 12D**). When one copy of the *CAT8-1* or the *CAT8-2* genes under the control of their native promoters was introduced in the *cat8-1Δ* or *cat8-2Δ* mutant strains, respectively, the growth was similar to that of the WT, showing that the phenotype observed on ethanol is caused only by the lack of these two genes (**Figure 13**). Therefore, Cat8-1 and Cat8-2 are two essential transcription factors for the growth *P. pastoris* on ethanol.

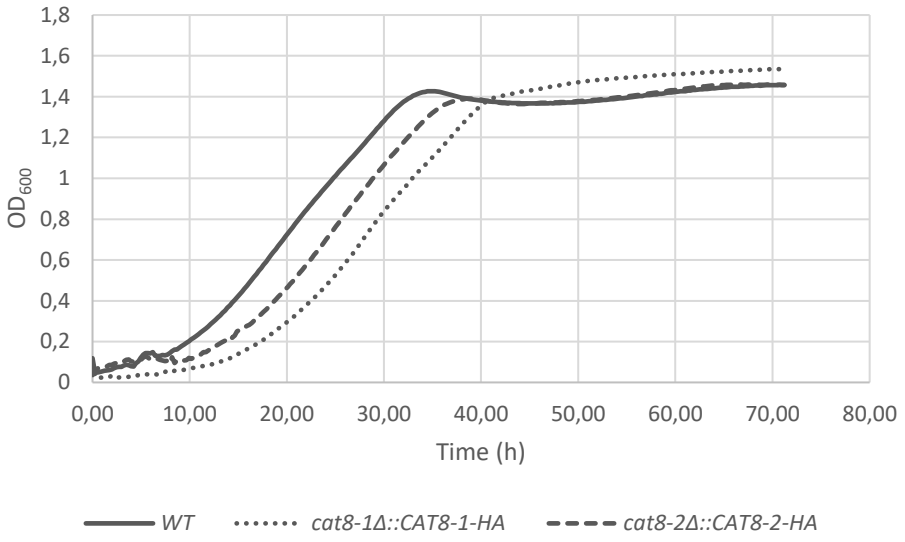


Figure 13: Growth of the *cat8-1Δ::CAT8-1-HA* and *cat8-2Δ::CAT8-2-HA* on ethanol. Plamids carrying the promoter regions (upstream 1000 bp regions), C-terminally tagged coding sequences and terminator regions of *CAT8-1* or *CAT8-2* were transformed into *P. pastoris cat8-1Δ* (for Cat8-1-HA) and *cat8-2Δ* (for Cat8-2-HA) deletion strains. The growth of the obtained strains was assessed on 1% ethanol.

IV. *CAT8-1* and *CAT8-2* expressions are higher on ethanol and methanol than on glucose and glycerol

The transcript levels of *CAT8-1* and *CAT8-2* in the wild-type induced on different carbon sources were determined.

Transcript levels of *CAT8-1* in the *P. pastoris* wild-type were slightly increased on ethanol and methanol compared to glucose and glycerol (**Figure 14A**). Generally, expression of *CAT8-1* remained quite low: around 0.04 and 0.06 of *ACT1* expression on ethanol and methanol, respectively (**Figure 14B**). Transcript levels of *CAT8-2* in the *P. pastoris* wild-type were higher on glycerol, ethanol and methanol than on glucose (**Figure 14A**). Generally, the expression of *CAT8-*

2 is higher than the expression of *CAT8-1*: *CAT8-2* reached around 0.6 of *ACT1* expression on ethanol and methanol (**Figure 14B**), which was more than 10 times higher than the expression levels reached by *CAT8-1*.

In addition, there was around 15-fold up-regulation of *CAT8-1* in the *CAT8-1* overexpression strain on ethanol, whereas up-regulation was 150-fold in same strain on methanol compared to the WT (**Figure 14C**). For *CAT8-2*, we observed 2.34-fold up-regulation of *CAT8-2* in the *CAT8-2* overexpression strain on ethanol and a 40-fold up-regulation on methanol (**Figure 14C**). We can therefore conclude from these observations that the overexpression of *CAT8-1* and *CAT8-2* with P_{THI11} is more efficient on methanol than ethanol.

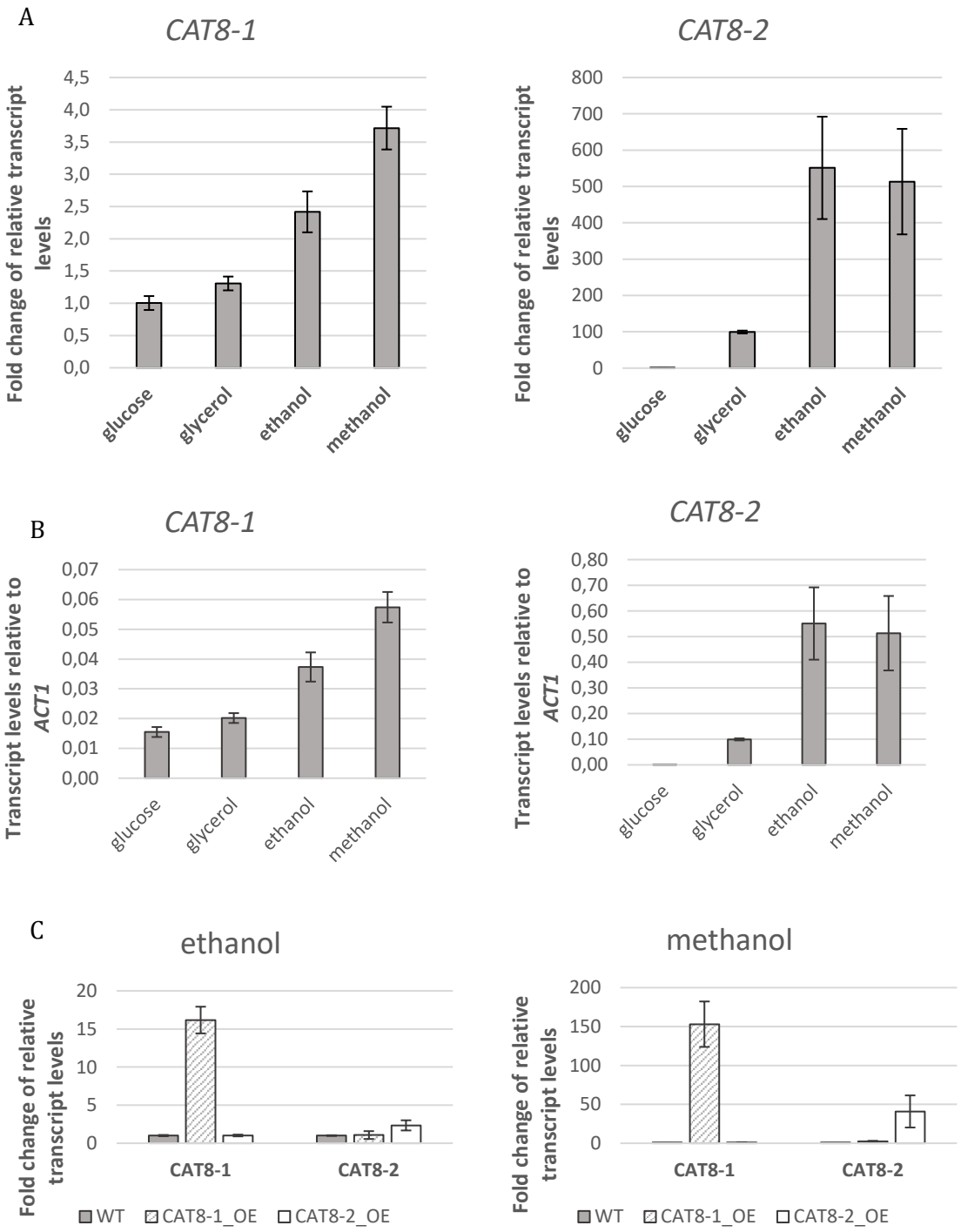


Figure 14: Influence of the carbon source on *CAT8-1* and *CAT8-2* gene expression. (A) Transcript levels of *CAT8-1* and *CAT8-2* in the *P. pastoris* wild-type CBS7435 induced on different carbon sources. (B) Transcript levels of CAT8-1 and CAT8-2 on glucose, glycerol, ethanol and methanol relative to the ACT1 gene. Gene expression levels were normalized to the reference gene *ACT1* and quantified relative to the levels on glucose (set to 1.0, dashed line) for each carbon source. Error bars represent the standard deviations obtained with two biological samples each measured in technical triplicates in two independent experiments. (C) mRNA levels of *CAT8-1* and *CAT8-2* in overexpression mutants and the *P. pastoris* wild-type on ethanol and methanol. Gene expression levels were normalized to the reference gene *ACT1* and quantified relative to wild type levels (set to 1.0, dashed line) for each carbon source. Error bars represent the standard deviations of two biological samples each measured in technical triplicates in two independent experiments.

V. Cat8-1 and Cat8-2 regulate the expression of genes important for ethanol assimilation

To identify the target genes of Cat8-1 and Cat8-2, mRNA levels of genes important for the growth of *P. pastoris* on non-fermentable carbon sources were analyzed in the overexpression and knock-out mutants of *CAT8-1* and *CAT8-2* as well as in the wild-type induced on different carbon sources.

The samples for the transcript level analysis were obtained by growing the mutant strains and the wild-type in liquid cultures on minimal medium with limited glucose and then shifting the cultures to 2% glucose, 2% glycerol, 1% methanol or 2% ethanol for 5 hours to have induction on either glucose, glycerol, methanol and ethanol, respectively. Cells were then harvested and RNA was extracted for rt-qPCR (**Figure 15A**).

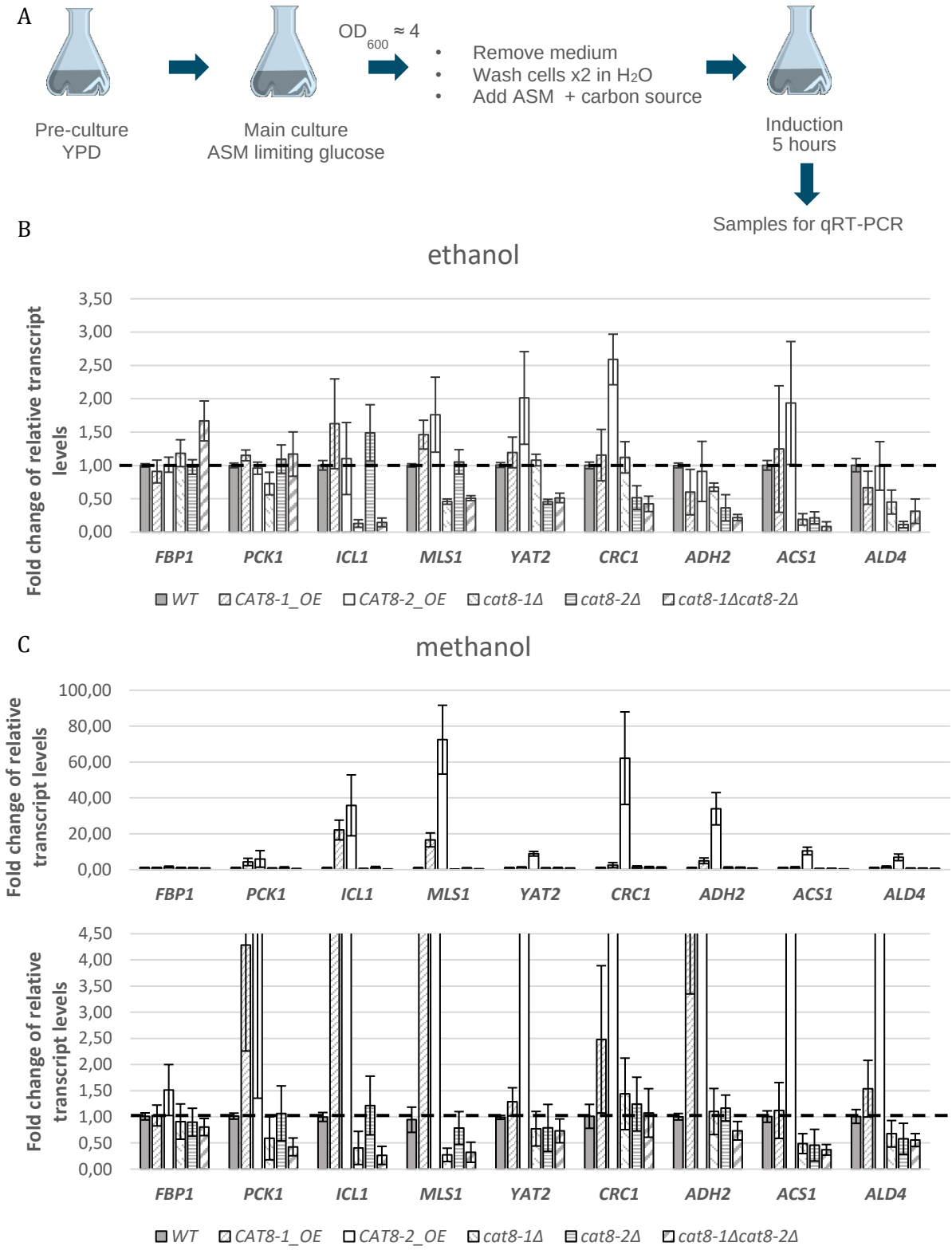


Figure 15: Influence of *CAT8-1* and *CAT8-2* overexpression and knock-out on transcript levels of selected genes on ethanol and methanol. (A) Scheme of the sampling for the transcript level analysis. (B) Transcript levels of selected genes in *CAT8-1*_OE, *CAT8-2*_OE, *cat8-1Δ*, *cat8-2Δ* and the double knock-out *cat8-1Δcat8-2Δ* induced on 2% ethanol, determined by qRT-PCR. (C) Transcript levels of selected genes in *CAT8-1*_OE, *CAT8-2*_OE, *cat8-1Δ*, *cat8-2Δ* and *cat8-1Δcat8-2Δ* induced on 1% methanol, determined by qRT-PCR. Gene expression levels were normalized to the reference gene *ACT1* and quantified relative to WT levels (WT set to 1.0, dash line). Error bars represent the standard deviations of two independent biological samples each measured in technical triplicates in two independent experiments.

On glucose and glycerol, no significant difference in transcript levels was observed in the different strains for the selected genes compared to the wild-type (data not shown).

On ethanol, the expression of the two gluconeogenic genes *FBP1* and *PCK1* (encoding fructose-1,6-bisphosphatase and phosphoenolpyruvate carboxykinase, respectively) were similar to that of the wild-type in the different mutant strains, showing that Cat8-1 and Cat8-2 are not involved in the regulation of these two genes (**Figure 15B**). In *cat8-1Δ*, *cat8-2Δ*, and in the *cat8-1Δcat8-2Δ* double knock-out mutants, the transcript levels of three genes involved in ethanol assimilation (*ADH2* encoding alcohol dehydrogenase II, *ACS1* encoding acetyl-coenzyme A synthetase I and *ALD4* encoding aldehyde dehydrogenase IV) are decreased (**Figure 15B**), showing that both Cat8-1 and Cat8-2 are involved in the regulation of the expression of these three genes. The expression of two genes from the glyoxylate shunt (*ICL1* encoding isocitrate lyase I, and *MLS1* encoding malate synthase I) is decreased in *cat8-1Δ* and in the *cat8-1Δcat8-2Δ* double knock-out mutant, whereas it is similar to that of the wild-type in *cat8-2Δ.* Therefore, Cat8-1 seems to be necessary for the activation of the genes from the glyoxylate shunt in *P. pastoris*. The expression of two genes from the carnitine shuttle (*YAT2* encoding cytosolic carnitine acetyl transferase and *CRC1* encoding carnitine carrier I) is decreased in *cat8-2Δ* and in the *cat8-1Δcat8-2Δ* double knock-out mutant whereas it is similar to that of the wild-type in *cat8-1Δ.* Therefore, Cat8-2 seems to be necessary for the activation of the genes from the carnitine shuttle. Additionally, the expression of genes important for methanol assimilation (*AOX1* encoding alcohol oxidase I, *DAS1* encoding dihydroxyacetone synthase I and *PEX5* encoding a peroxisomal membrane receptor) was assessed in the TF mutant strains on ethanol, but no specific regulation pattern was observed (**Figure 16**): due to the very weak expression of these genes on ethanol, the small differences observed in expression were not significant.

On methanol, the only similar regulation pattern to what is observed compared to ethanol is the necessity of Cat8-1 for the activation of *ICL1* and *MLS1*, the two genes of the glyoxylate shunt (**Figure 15C**). In addition, we observe an increase in transcript levels for both *ICL1* and *MLS1* in the *CAT8-2_OE* mutant, but no significant difference in *cat8-2Δ* (**Figure 15C**).

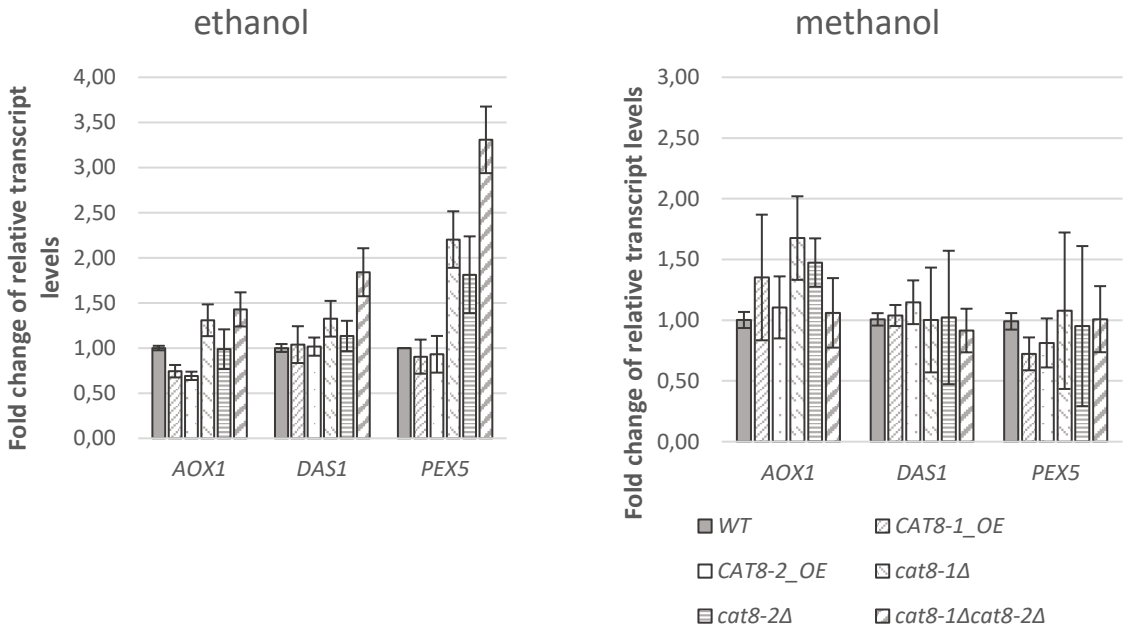


Figure 16: Influence of *CAT8-1* and *CAT8-2* overexpression and knock-out on transcript levels of methanol utilization genes on ethanol and methanol. Transcript levels of *AOX1*, *DAS1* and *PEX5* in *CAT8-1_OE*, *CAT8-2_OE*, *cat8-1Δ*, *cat8-2Δ* and *cat8-1Δcat8-2Δ* induced on 2% ethanol and 1% methanol, determined by qRT-PCR. Gene expression levels were normalized to the reference gene *ACT1* and quantified relative to WT levels (WT set to 1.0, dash line). Error bars represent the standard deviations of two independent biological samples each measured in technical triplicates in two independent experiments.

VI. The *CAT8-2* gene expression is autoregulated by the Cat8-2 protein

In *K. lactis* and *S. cerevisiae*, the Cat8 protein regulates the expression of Sip4, and Sip4 autoregulates the activity of its own promoter (Mehlgarten et al., 2015). In order to analyze a possible regulation of the *CAT8-1* and *CAT8-2* genes by the Cat8-1 and Cat8-2 proteins in *P. pastoris*, eGFP reporter strains were generated. The eGFP gene was expressed under the control of the *CAT8-1* and *CAT8-2* promoters, respectively. These constructs were transformed into the different knock-out mutants (*cat8-1Δ*, *cat8-2Δ* and *cat8-1Δcat8-2Δ*) and the wild-type. The strains obtained were cultivated in biological triplicates on minimal medium with limiting glucose until OD_{600} reached 4, and then shifted to either 2% glucose, 2% glycerol, 1% methanol or 2% ethanol for 5 h to have induction on glucose, glycerol, methanol and ethanol, respectively. At the end of the induction time, samples were diluted to OD_{600} 0.2 in PBS in 96-well fluorescence microtiter-plates. The eGFP fluorescence (505–545 nm) of 10,000 cells per sample was measured by flow cytometry (**Figure 17**).

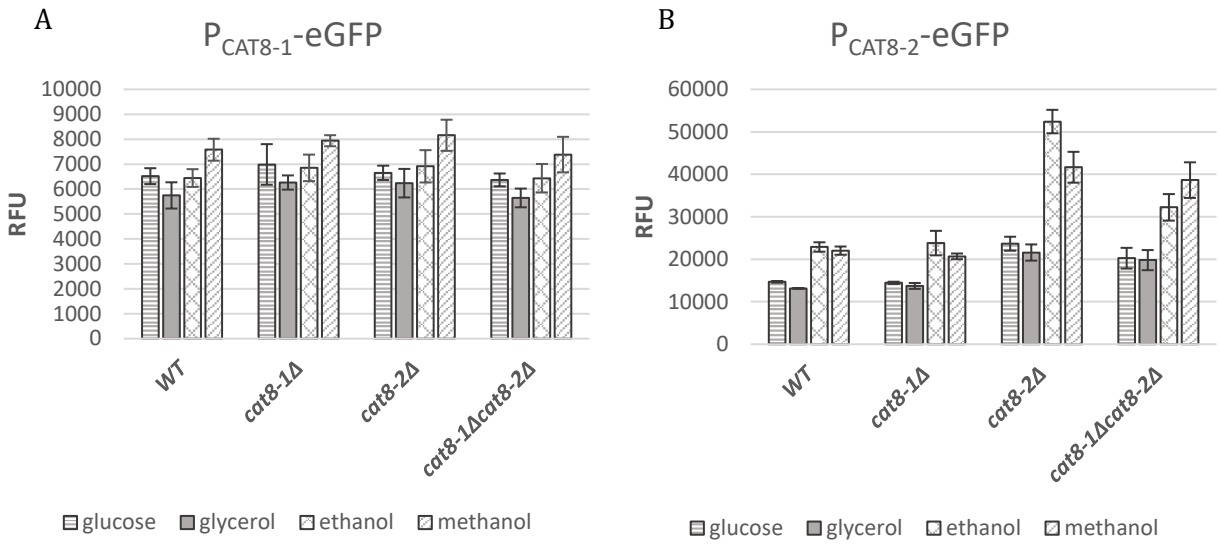


Figure 17: Influence of *CAT8-1* and *CAT8-2* deletions on (A) *CAT8-1* and (B) *CAT8-2* promoters' activity. The *CAT8-1* and *CAT8-2* promoters were fused to the eGFP reporter gene and transformed into wild-type cells (CBS7435) and the knock-out mutants *cat8-1Δ*, *cat8-2Δ* and *cat8-1Δcat8-2Δ*, respectively. All mutants carrying the eGFP reporter gene as well as the non-transformed wild-type control were grown in biological triplicates on minimal media with limiting glucose and shifted to glucose, glycerol, ethanol or methanol for 5 hours. The eGFP fluorescence (505–545 nm) of 10,000 cells per sample was measured by flow cytometry. Mean values and standard deviation for the three biological replicates are presented.

In all strains and on all the carbon sources, the eGFP fluorescence is lower when the eGFP gene is under the control of the *CAT8-1* promoter compared to when the eGFP gene is under the control of the *CAT8-2* promoter (compare axes in **Figure 17A** and **17B**). With the *CAT8-1* promoter, no significant difference is observed in fluorescence levels in the different strains on the different carbon sources (**Figure 17A**). The expression of *CAT8-1* does not seem to be affected by neither Cat8-1 nor Cat8-2. With the *CAT8-2* promoter in the wild type, the eGFP fluorescence is lower on glucose and glycerol than on ethanol and methanol (**Figure 17B**), which fits to the results obtained for the transcript level analysis of the *CAT8-2* gene on these four carbon sources (**Figure 14A**). In the *cat8-1Δ* mutant, the eGFP levels are similar to the ones measured in the wild type on each the four carbon sources (**Figure 17B**). However in the *cat8-2Δ* and *cat8-1Δcat8-2Δ* mutants, the eGFP levels are higher than the ones of the wild type on glucose, glycerol, ethanol and methanol (**Figure 17B**). Therefore, it seems that Cat8-1 is not regulating the expression of *CAT8-2*, but that the Cat8-2 protein is repressing/autoregulating its own promoter on glucose, glycerol, ethanol and methanol.

VII. The *CAT8-2* expression is regulated by the Mig1-2 transcription factor

The transcription of *CAT8* was described to be regulated by the transcription factor Mig1 in *S. cerevisiae* (Carlson, 1999; Schüller, 2003). *P. pastoris* possesses two gene homologs of this transcription factor, termed *MIG1-1* (PP7435_Chr4-0661) and *MIG1-2* (PP7435_Chr1-1325). Knock-out and overexpression strains were obtained for both of these transcription factors (Ata et al., 2017). The Mig1-1 and Mig1-2 mutant strains were cultivated in biological duplicates on minimal medium with limiting glucose until OD_{600} reached 4, and then shifted to 2% glucose, 2% glycerol, 1% methanol or 2% ethanol for 5h to have induction on glucose, glycerol, methanol and ethanol, respectively. At the end of the induction, samples were processed and the transcript levels of *CAT8-1* and *CAT8-2* were measured in the different strains induced on the different carbon sources (**Figure 18**).

The *CAT8-1* transcript levels are similar to that of the wild type in all mutant strains and on all carbon sources (**Figure 18**). Therefore *CAT8-1* does not seem to be regulated by neither Mig1-1 nor Mig1-2 in the conditions studied. The expression of *CAT8-2* is around 30-fold higher in the *mig1-2Δ* mutant than in the wild type on glucose (**Figure 18A**). It is also lower on glycerol in the *MIG1-2_OE* mutant (**Figure 18B**). Finally, the expression of *CAT8-2* is lower than in the wild type on ethanol in the *MIG1-2_OE* mutant. *CAT8-2* therefore seems to be repressed by Mig1-2 on glucose and glycerol, but not on ethanol unless *MIG1-2* is artificially overexpressed. Transcript levels of genes regulated by Cat8-1 and Cat8-2 were also measured in the Mig1-1 and Mig1-2 mutant strains, but no significant difference was observed compared to the wild-type (**Figure 19**).

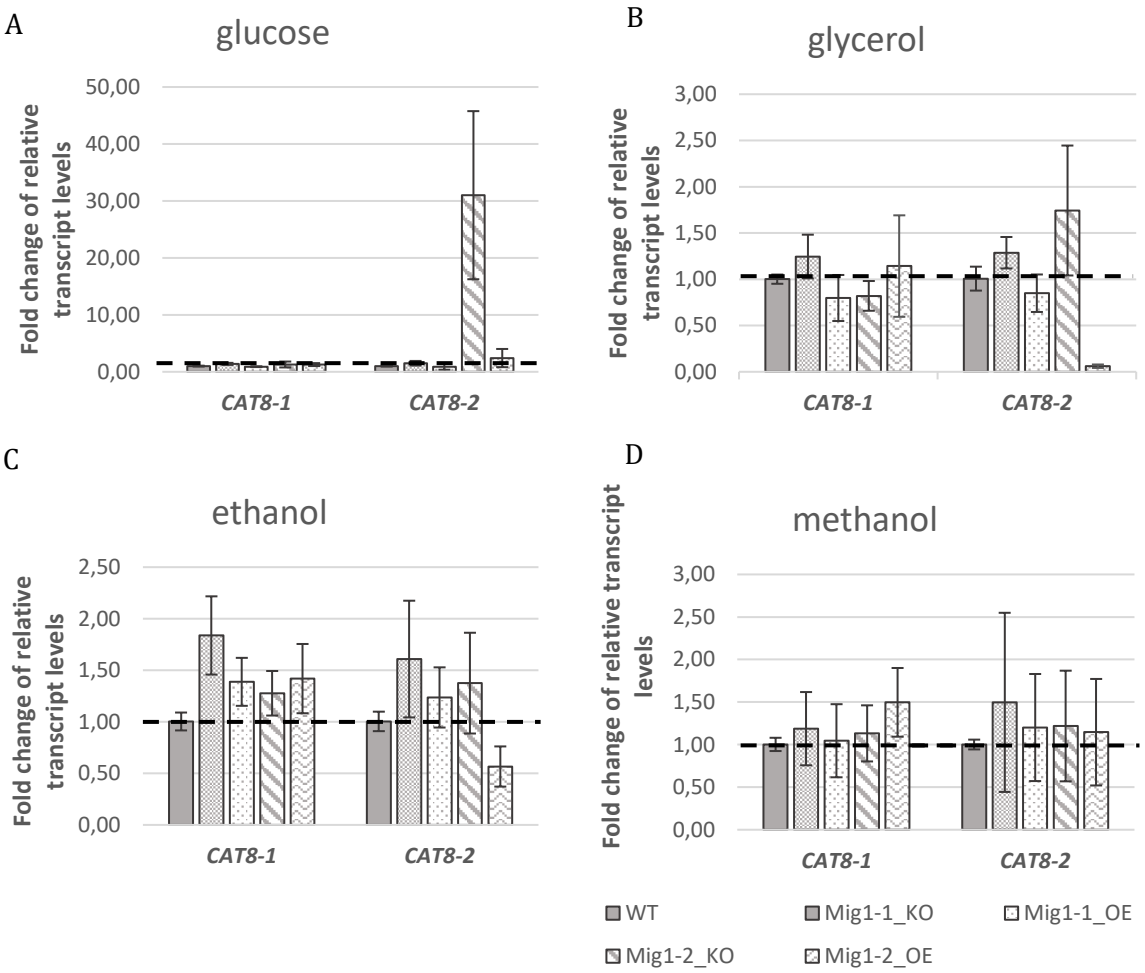


Figure 18: Influence of *MIG1-1* and *MIG1-2* deletion and overexpression on *CAT8-1* and *CAT8-2* transcript levels. Transcript levels of *CAT8-1* and *CAT8-2* in *mig1-1Δ*, *MIG1-1_ OE*, *mig1-2Δ* and *MIG1-2_OE* mutants induced on (A) glucose, (B) glycerol, (C) ethanol or (D) methanol relative to wild type strain were determined by qRT-PCR. Gene expression levels were normalized to the reference gene *ACT1* and quantified relative to wild type levels (set to 1.0, dashed line) for each carbon source. Error bars represent the standard deviations of with two independent biological samples each measured in technical triplicates in three independent experiments.

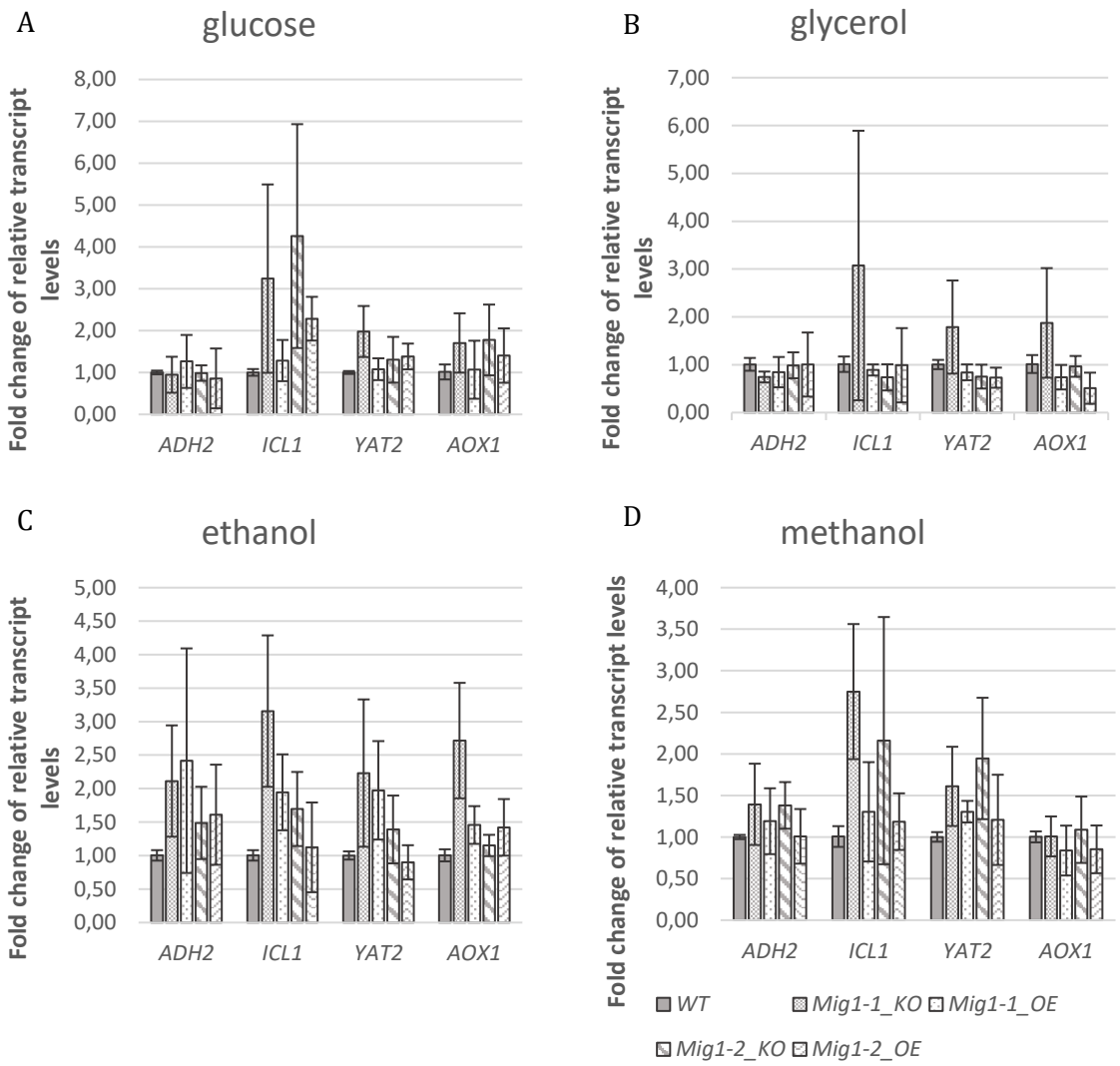


Figure 19: Influence of *MIG1-1* and *MIG1-2* deletion and overexpression on the transcript leves of selected genes. Transcript levels of *ADH2*, *ICL1*, *YAT2* and *AOX1* in *mig1-1Δ*, *MIG1-1_ OE*, *mig1-2Δ* and *MIG1-2_OE* mutants induced on (A) glucose, (B) glycerol, (C) ethanol or (D) methanol relative to wild type strain were determined by qRT-PCR. Gene expression levels were normalized to the reference gene *ACT1* and quantified relative to wild type levels (set to 1.0) for each carbon source. Mean values and standard deviation for two independent biological samples each measured in technical triplicates in three independent experiments are presented.

VIII. Discussion

Here, we studied the functions of two transcription factors of *P. pastoris*: Cat8-1 and Cat8-2. Both were shown to be essential for the growth of *P. pastoris* on ethanol. Cat8-1 seems to be necessary for the activation of the genes encoding enzymes of the glyoxylate cycle (*MLS1* and *ICL1*), whereas Cat8-2 seems to be necessary for the activation of the carnitine shuttle encoding genes (*YAT2* and *CRC1*). Both transcription factors are required for the activation of *ADH2*, *ALD4* and *ACS1*, three genes encoding enzymes important for assimilation of ethanol. In addition, we showed that the *CAT8-2* gene is repressed by the Mig1-2 transcription factor (a homolog of the *S. cerevisiae* Mig1) on glucose, and repressed/autoregulated by the Cat8-2 protein on glucose, glycerol, ethanol and methanol.

VIII.1. Cat8-2 is a Sip4 equivalent in *P. pastoris*

The Cat8 and Sip4 homologous proteins used in the phylogenetic analysis generally have the DNA binding domain and the Leucine-zipper dimerization domain conserved, but have a very high variation in the rest of the sequences in the different organisms. From the phylogeny, the Cat8-1 protein sequence from *P. pastoris* was shown to be closer to Cat8 but still different than the Cat8 proteins in the *Saccharomycetaceae* and *Saccharomycodaceae* clade. Cat8-2 clusters with the Sip4 protein homologs but still form a separate group with related proteins from *Phafomyceteceae* and *Pichiaceae*. Furthermore, Cat8-2 and Sip4 share common features in the regulation of non-fermentable carbon sources in *P. pastoris*, *S. cerevisiae* and *K. lactis*. Therefore, Cat8-2 is probably a Sip4 homolog in *P. pastoris*.

VIII.2. Role of Cat8-1 and Cat8-2 in the regulation of ethanol utilization

When shifted to a nonfermentable carbon source, the yeast cells undergo a massive reprogramming of their gene expression, including genes involved in gluconeogenesis, the glyoxylate cycle and tricarboxylic acid cycle. The ethanol is metabolized by the alcohol dehydrogenase (encoded by *ADH2*) to acetaldehyde, which is subsequently converted into acetate by the aldehyde dehydrogenase (*ALD*). Acetate is then transformed into acetyl-Coenzyme A by acetyl-CoA synthetase (*ACS*) in the cytoplasm. Then, acetyl-CoA must be transferred to the mitochondria for the production of energy (Schmalix and Bandlow, 1993; Stemple et al., 1998). Two pathways exist to transport acetyl-CoA into the mitochondria: (1) acetyl-CoA is converted into glyoxylate cycle intermediates which are transported to the mitochondria (Palmieri et al., 1997) and (2) acetyl-CoA is converted into acetylcarnitine, which is transported into mitochondria via the carnitine shuttle. In addition of the carnitine shuttle and the glyoxylate cycle, the gluconeogenesis is fundamental for the growth on non-fermentable carbon sources: it is

essential for the production of glucose-6-phosphate, which is crucial for cell growth (Barnett and Entian, 2005).

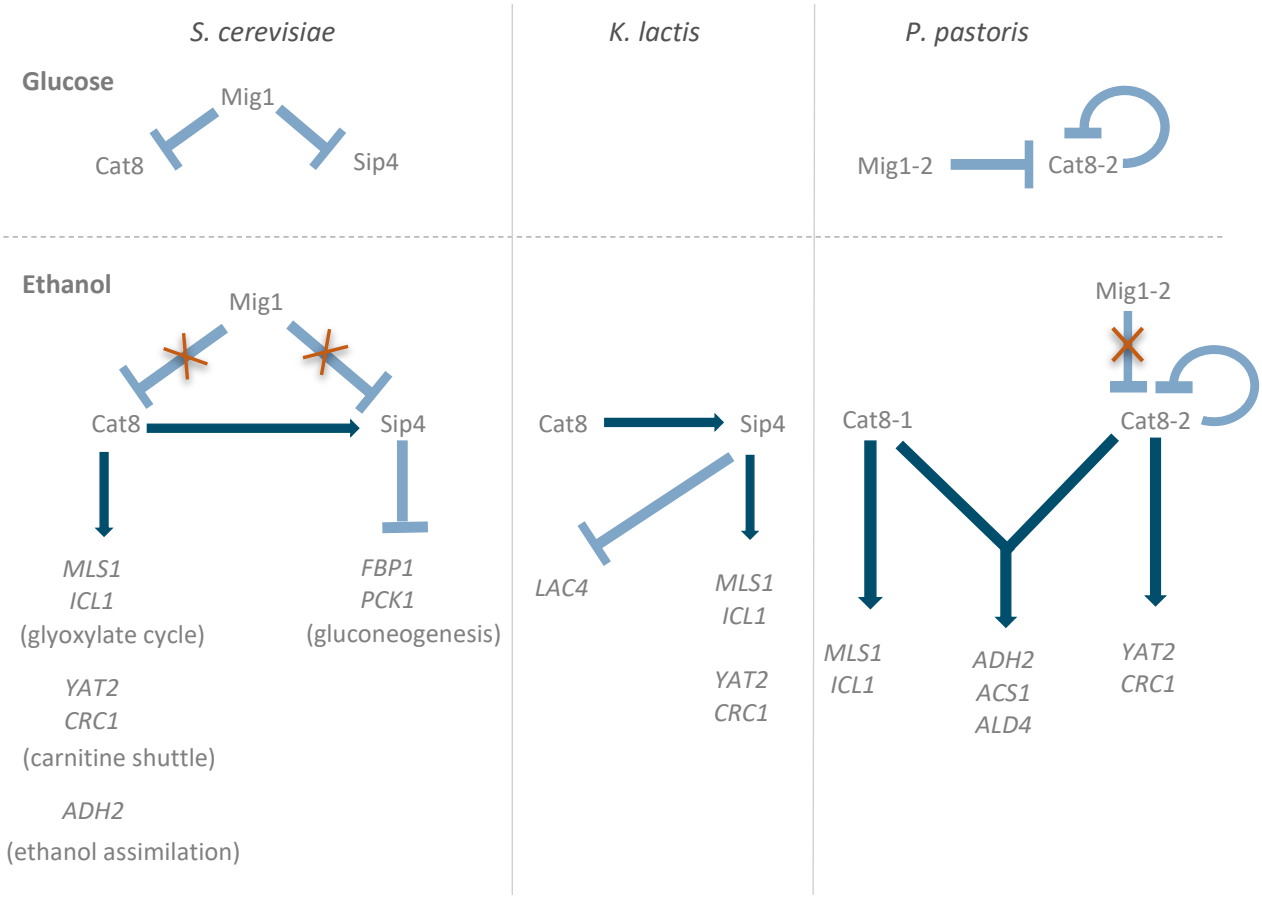


Figure 20: Comparison of the Cat8-Sip4 regulatory networks in *S. cerevisiae*, *K. lactis* and *P. pastoris*. Schematic representation of the regulation by the transcription factors Cat8 and Sip4 for selected genes. Arrows in dark blue indicate transcriptional activation whereas arrows in light blue indicate transcriptional repression.

We can observe three major differences in regulation of genes for ethanol utilization by Cat8 and Sip4 or Cat8-1 and Cat8-2 in the three yeast species *S. cerevisiae*, *K. lactis* and *P. pastoris* (**Figure 20**): (1) Cat8 is involved in the regulation of the gluconeogenesis in *S. cerevisiae* but not in *K. lactis* nor in *P. pastoris*, (2) the expression of Sip4 is activated by Cat8 in *S. cerevisiae* and *K. lactis* but not in *P. pastoris*, and (3) the dissimilar importance of Cat8 and Sip4 in the different yeasts: although the same set of genes are targeted by Cat8/Sip4 and Cat8-1/Cat8-2 (except the gluconeogenesis genes), the TF directly regulating them varies in the different species.

REGULATION OF THE GLUCONEOGENESIS

In *S. cerevisiae*, Cat8 and Sip4 are involved in the regulation of genes of the gluconeogenesis, therefore knocking out the genes encoding these two TFs results in a growth defect on glycerol (Hedges et al., 1995; Rahner et al., 1996). In *P. pastoris* as well as in *K. lactis*, when *CAT8-1* and

CAT8-2 (or *CAT8* and *SIP4*) are knocked-out, there is no impairment of the growth on glycerol. In addition, transcript level analysis in *P. pastoris* showed that *CAT8-1* and *CAT8-2* deletions did not have a major influence on *FBP1* and *PCK1* expression on any of the carbon source tested (**Figure 15** for ethanol and methanol, not shown for glucose and glycerol). Therefore, the regulation of the gluconeogenesis seems to be achieved by a different regulatory network in *P. pastoris* compared to *S. cerevisiae*. Another TF, Rds2, was shown to have partially overlapping functions with those of Cat8 in *S. cerevisiae* and to be a major regulator the gluconeogenesis. Specifically, Rds2 was shown to directly activate the expression of the gluconeogenic genes while repressing the negative regulators of this pathway, possibly through binding of the CSREs (Soontorngun et al., 2007). The homolog of Rds2 in *P. pastoris* (PP7435_Chr2-1080) was not studied so far and could be involved in the regulation of the gluconeogenesis in this yeast.

ACTIVATION OF *SIP4* EXPRESSION BY Cat8

In both *S. cerevisiae* and *K. lactis*, Cat8 has an important role in activating Sip4 expression on ethanol. This phenomenon is not observed in *P. pastoris*, suggesting that a rewiring of the regulation network of the carbon metabolism happened.

DIRECT TARGETS OF Cat8 AND Sip4 VARY IN THE DIFFERENT YEAST SPECIES

In *S. cerevisiae*, Cat8 is the main activator for the growth on non-fermentable carbon sources. In fact, the *SIP4* deletion mutants do not exhibit any growth defect on any carbon source (Lesage et al., 1996), and the role of *Sc*Sip4 is still unclear. *Sc*Cat8 was shown to activate expression of genes from the gluconeogenesis, the glyoxylate cycle, ethanol assimilation and the carnitine shuttle. In *K. lactis*, Cat8 activates the expression of Sip4 but does not seem to directly activate the expression of genes important for the growth on non-fermentable carbon sources. Instead, *Kl*Sip4 is important for the regulation of the glyoxylate cycle and the carnitine shuttle. In *P. pastoris*, some genes seem to be specifically regulated by either Cat8-1 or Cat8-2, while others seem to be overlapping targets of these two TFs. Indeed, Cat8-1 seems to be necessary for the activation of the genes from the glyoxylate cycle, while Cat8-2 seems to be important for the activation of genes from the carnitine shuttle. In addition, Cat8-1 and Cat8-2 are both involved in the regulation of *ADH2*, *ALD4* and *ACS1*, three genes important for the conversion of ethanol to acetyl-CoA (**Figure 15**), although the contribution of each TF for the regulation of these three genes remains to be investigated.

VIII.3. Cat8-1 and Cat8-2 possibly recognize different subsets of CSRE motifs

It was shown in *S. cerevisiae* that mutant CSREs show differential activation by Cat8 and Sip4 (Roth et al., 2004). In addition, the purified DNA binding domain of the TF Rds2 was shown to

bind to the CSREs of the *PCK1* and *FBP1* promoters (Soontorngun et al., 2007). It was hypothesized that these three TFs bind subsets of CSREs with diverging affinities, which would allow for the regulation of both distinct and common target genes. This phenomenon could also be present in *P. pastoris*: transcript level analysis showed that on methanol in the *CAT8-2_OE* strain, there is increased transcript levels of *YAT2* and *CRC1* as well as *ICL1* and *MLS1*, two genes mainly regulated by Cat8-1, whereas in *CAT8-1_OE*, there is an increase only in *ICL1* and *MLS1* transcript levels (**Figure 15**). This suggests that Cat8-1 and Cat8-2 could have each specific binding sites, but that on methanol and when overexpressed, Cat8-2 could bind the motif specific for Cat8-1. However, further experiments such as a chromatin immunoprecipitation (ChiP) assay should be performed to investigate the respective DNA binding sites of Cat8-1 and Cat8-2.

VIII.4. Mechanisms of Cat8 and Sip4 activation

In presence of glucose, the expressions of *CAT8* and *SIP4* are inhibited by the TF Mig1 and the Ssn6/Tup1 corepressor complex in *S. cerevisiae* (Carlson, 1999; Schüller, 2003). *P. pastoris* possess two homologs of *Sc*Mig1: Mig1-1 and Mig1-2. These two TFs were shown to be involved in the regulation of the methanol metabolism in *P. pastoris*: derepression of *AOX1* was observed in *mig1-1Δ* and *mig1-1Δmig1-2Δ* knock-out strains on glycerol (Wang et al., 2017c), and the genes of the methanol metabolism as well as the peroxisomes biogenesis pathway were upregulated on glycerol in *mig1-1Δmig1-2Δ* (Shi et al., 2018). Finally, Mig1-1 and Mig1-2 localized in the nucleus of cells grown on glucose or glycerol, but when *P. pastoris* cells were transferred to methanol, Mig1-1 and Mig1-2 predominantly localized to the cytoplasm (Wang et al., 2017a).

In our experiments, we observed a repression of *CAT8-2* on glucose by Mig1-2. The transcript levels of genes shown to be regulated by Cat8-1 and Cat8-2 were also measured in the Mig1-1 and Mig1-2 mutant strains on glucose, glycerol, ethanol and methanol, but no significant difference was observed compared to their levels in the WT. This could be explained by the involvement of other TFs in the regulation of these genes. In addition, this could partly be caused by the necessity of post-translational activation of Cat8-1 and/or Cat8-2. Indeed, Cat8 and Sip4 are both phosphorylated upon activation in *S. cerevisiae*: Cat8 was shown to be phosphorylated by the yeast homolog of AMPK, the Snf1 kinase complex, whereas Sip4 was shown to be phosphorylated by the Ssn3 kinase. Both Cat8-1 and Cat8-2 were tagged with a HA-tag for investigation of possible post-translational regulation, but the full length proteins could not be seen on Western Blot, although the functionality of the tagged version of Cat8-1 and Cat8-2 was verified by growth complementation on ethanol of *cat8-1Δ* and *cat8-2Δ* upon introduction of *CAT8-1-HA* and *CAT8-2-HA*, respectively. Since the TF Mig1 was also shown to be phosphorylated in *S. cerevisiae* (Carlson, 1999; Schüller, 2003), phosphorylation of Mig1-2 was investigated in *P. pastoris* in a

master's thesis project. Three kinases of *P. pastoris* were selected for deletion: Snf1, Snf1-2 (both homologs of *Sc*Snf1, encoded by PP7435_Chr2-0772 and PP7435_Chr1-0450, respectively), and Ssn3 (homolog of *Sc*Ssn3 encoded by PP7435_Chr1-1091). The generation of the *SNF1-2* and *SSN3* KO mutants was successful, but not that of *SNF1*. Growth assays in liquid media showed no difference for the *snf1-2Δ* deletion strain compared to the WT, while there was a growth defect for *ssn3Δ* on glucose but not on glycerol, ethanol or methanol. Finally, Western Blot and mass spectrometry analyses showed that Mig1-2 is phosphorylated on repressing (4% glucose) and derepressing (0.2% glucose) conditions, and most likely degraded on 2% ethanol.

The present study focused mainly on elucidating the involvement in Cat8-1 and Cat8-2 in the regulation of the ethanol assimilation pathway, the carnitine shuttle and the glyoxylate shunt. However, the CSRE motif, which was shown to be recognized by Cat8 and Sip4 in *S. cerevisiae* and *K. lactis,* was found in around 90% of the total promoters in *P. pastoris* (not shown). Furthermore, as shown in the chapter 1 of this thesis, knocking-out *CAT8-1* and *CAT8-2* influences the ability of *P. pastoris* to tolerate osmotic stress and cell wall damaging agents. This results suggest that these two TFs might have other targets and/or a broader role in *P. pastoris,* both of which remain to be investigated. Regarding the role of Cat8-1 and Cat8-2 in the regulation of ethanol utilization, the effect of each individual TF for the regulation of the ethanol assimilation pathway remains to be investigated, as well a possible cooperation of Cat8-1 and Cat8-2 with other TF(s) and the post-translational modifications that could affect Cat8-1 and Cat8-2.

CONCLUSION

Knowledge on TFs can be used for cell engineering purposes, such as promoter engineering or global transcriptional engineering strategies. A lot of the TFs of *P. pastoris* that were characterized so far are involved in the regulation of the genes from the methanol utilization pathway, and a large number of putative transcriptional regulators in this yeast are still of unknown function. The aim of this project was therefore to study and characterize some transcription factors from *P. pastoris*. To do so, twelve TFs were selected because of their interesting regulation patterns in industrially relevant conditions. Overexpression and knock-out mutants were generated for each of the selected transcription factors. These TF mutants were arrayed on different media representing different growth and stress conditions. In addition, the influence of TF overexpressions and knock-outs on recombinant protein production under the control of the *AOX1* promoter was assessed. From the twelve initially selected TFs, the two proteins Cat8-1 and Cat8-2 were further characterized. These two TFs are essential for the growth of *P. pastoris* on ethanol, as they regulate the expression of genes that are of primary importance for the metabolism of this carbon source. Cat8-1 was indeed shown to be necessary for the activation of genes of the glyoxylate cycle, whereas Cat8-2 was shown to be necessary for the activation of genes of the carnitine shuttle when *P. pastoris* is grown on ethanol. Both Cat8-1 and Cat8-2 are required for activation of genes of the ethanol utilization pathway. Finally, the *CAT8-2* gene was shown to be repressed by the TF Mig1-2 on glucose, and autoregulated by the Cat8-2 protein on all carbon sources. The knowledge obtained in this study therefore provides a basis for further research on TFs and possible applications using TFs for cell engineering of *P. pastoris*.

I. Appendix 1: List of primers used in this study

Table 8: Primers used for the generation of the overexpression mutant strains.

Number	Name	Target	Sequence	Purpose
prBD_0001	AFU1_Primer_OE_Fw1	AFU1	GATTAGGTCTCCCATGGATAAAAGTAAAATCAGGTACAAGAGATCAAGATCGG	GG cloning
prBD_0002	AFU1_Primer_OE_Rv1	AFU1	GTCCTGTTTATTATGGGGGACCATGGATGAACG	GG cloning
prBD_0003	AFU1_Primer_OE_Fw2	AFU1	CGTTCATCCATGGTCCCCCATAATAAACAGGAC	GG cloning
prBD_0004	AFU1_Primer_OE_Rv2	AFU1	GTCGATAGTCAATGGTCTACCAAAAGCTACCGAAATAC	GG cloning
prBD_0005	AFU1_Primer_OE_Fw3	AFU1	GTATTTCGGTAGCTTTTGGTAGACCATTGACTATCGAC	GG cloning
prBD_0006	AFU1_Primer_OE_Rv3	AFU1	CCATCGGAATGAGATTGTCCTCATTAGTGAAACCAG	GG cloning
prBD_0007	AFU1_Primer_OE_Fw4	AFU1	CTGGTTTCACTAATGAGGACAATCTCATTCCGATGG	GG cloning
prBD_0008	AFU1_Primer_OE_Rv4	AFU1	GTTAGGTCCATAGACGGTTTCAGGATTAAGGATATTAC	GG cloning
prBD_0009	AFU1_Primer_OE_Fw5	AFU1	GTAATATCCTTAATCCTGAAACCGTCTATGGACCTAAC	GG cloning
prBD_0010	AFU1_Primer_OE_Rv5	AFU1	GATTAGGTCTCCAAGCCTATTGGAAAAAGTTGAACCAGTCTGATCTGGC	GG cloning
prBD_0011	CAT8-1_Primer_OE_Fw	CAT8-1	GATTAGGTCTCCCATGATGCCGGAGGAAC	GG cloning
prBD_0012	CAT8-1_Primer_OE_Rv	CAT8-1	GATTAGGTCTCCAAGCCTAAAGTCCGAATAAACTCCC	GG cloning
prBD_0013	PAS_chr1-3_0010_Primer_OE_Fw	PAS_chr1-3_0010	GATTAGGTCTCCCATGGAGGCAGGAGTACAC	GG cloning
prBD_0014	PAS_chr1-3_0010_Primer_OE_Rv	PAS_chr1-3_0010	GATTAGGTCTCCAAGCTCAACCCGAATCGCC	GG cloning
prBD_0015	PAS_chr1-3_0166_Primer_OE_Fw1	PAS_chr1-3_0166	GATTAGGTCTCCCATGACGACAATACAGCCTTTGGCC	GG cloning
prBD_0016	PAS_chr1-3_0166_Primer_OE_Rv1	PAS_chr1-3_0166	GGCAAACAACGAAGTTTTCTTGAAGGCATTTGTG	GG cloning

prBD_0017	PAS_chr1-3_0166_Primer_OE_Fw2	PAS_chr1-3_0166	CACAAATGCCTTCAAGAAAACTTCGTTGTTTGCC	GG cloning
prBD_0018	PAS_chr1-3_0166_Primer_OE_Rv2	PAS_chr1-3_0166	GTTGATGATAGGCGGCAGACTTGTCTCGGCTAATAGTTC	GG cloning
prBD_0019	PAS_chr1-3_0166_Primer_OE_Fw3	PAS_chr1-3_0166	GAACTATTAGCCGAGACAAGTCTGCCGCCTATCATCAAC	GG cloning
prBD_0020	PAS_chr1-3_0166_Primer_OE_Rv3	PAS_chr1-3_0166	GCTGCATGGTATGAACACGTTATCCTGAATTTTG	GG cloning
prBD_0021	PAS_chr1-3_0166_Primer_OE_Fw4	PAS_chr1-3_0166	CAAAATTCAGGATAACGTGTTCATACCATGCAGC	GG cloning
prBD_0022	PAS_chr1-3_0166_Primer_OE_Rv4	PAS_chr1-3_0166	GATTAGGTCTCCAAGCTTATGCATTACTAAACTTTCGGTCCAGTTCCCTC	GG cloning
prBD_0023	PAS_chr3_0836_Primer_OE_Fw1	PAS_chr3_0836	GATTAGGTCTCCCATGAGTACCGCAGCCCCAATC	GG cloning
prBD_0024	PAS_chr3_0836_Primer_OE_Rv1	PAS_chr3_0836	AATCGATGAGTTCGAGGACTGGTCCTTAGGAATG	GG cloning
prBD_0025	PAS_chr3_0836_Primer_OE_Fw2	PAS_chr3_0836	ATTCCTAAGGACCAGTCCTCGAACTCATCGATTTTC	GG cloning
prBD_0026	PAS_chr3_0836_Primer_OE_Rv2	PAS_chr3_0836	GTTGTCGTGCTAGTTTTCTTGGTGAAAAATGGC	GG cloning
prBD_0027	PAS_chr3_0836_Primer_OE_Fw3	PAS_chr3_0836	GCCATTTTTCACCAAGAAAACTAGCACGACAAC	GG cloning
prBD_0028	PAS_chr3_0836_Primer_OE_Rv3	PAS_chr3_0836	GATTAGGTCTCCAAGCCTATTCTTCAACATTCCAGTAGTCAATTAACTCCTTGCC	GG cloning
prBD_0029	PAS_chr4_0077_Primer_OE_Fw1	PAS_chr4_0077	GATTAGGTCTCCCATGTCCAACAACAACATACCCCCACC	GG cloning
prBD_0030	PAS_chr4_0077_Primer_OE_Rv1	PAS_chr4_0077	GTTGTTGAATGGTGTCTCAGGAGAGAAAAGG	GG cloning
prBD_0031	PAS_chr4_0077_Primer_OE_Fw2	PAS_chr4_0077	CCTTTTCTCTCCTGAGACACCATTCAACAAC	GG cloning

prBD_0032	PAS_chr4_0077_Primer_OE_Rv2	PAS_chr4_0077	GATTAGGTCTCCAAGCTCAATTTAGCAGACTATGAATGCTGGTTTTCTGCTTCATCAG	GG cloning
prBD_0033	SGF29_Primer_OE_Fw	SGF29	GATTAGGTCTCCCATGGAAAGCAGTTGGGATG	GG cloning
prBD_0034	SGF29_Primer_OE_Rv	SGF29	GATTAGGTCTCCAAGCTCAACGCCCAGGAAAAG	GG cloning
prBD_0035	SKO1_Primer_OE_Fw1	SKO1	GATTAGGTCTCCCATGAACTCCGTTAGTCCTCCTAACGATTTGGGAAAAACAACC	GG cloning
prBD_0036	SKO1_Primer_OE_Rv1	SKO1	CTCTTCGTCAGCAGGCGTTTTCTTTGAGTCAGGCTCAGC	GG cloning
prBD_0037	SKO1_Primer_OE_Fw2	SKO1	GCTGAGCCTGACTCAAAGAAAACGCCTGCTGACGAAGAG	GG cloning
prBD_0038	SKO1_Primer_OE_Rv2	SKO1	GATTAGGTCTCCAAGCCTAGCTTTTCCTCTCTCCCAAACTCGTATTGTTAGAAGCC	GG cloning
prBD_0039	SWI5_Primer_OE_Fw1	SWI5	GATTAGGTCTCCCATGGACTCCAACTCTTACTGGAGTTCACAAG	GG cloning
prBD_0040	SWI5_Primer_OE_Rv1	SWI5	GATATCAAAGTTTGTGAAATCTTCGGTTGGAAGTGGAAC	GG cloning
prBD_0041	SWI5_Primer_OE_Fw2	SWI5	GTTCCACTTCCAACCGAAGATTTCACAAACTTTGATATC	GG cloning
prBD_0042	SWI5_Primer_OE_Rv2	SWI5	GAACAGGGAGGGACCGCATCTGAACAG	GG cloning
prBD_0043	SWI5_Primer_OE_Fw3	SWI5	CTGTTCAGATGCGGTCCCTCCCTGTTC	GG cloning
prBD_0044	SWI5_Primer_OE_Rv3	SWI5	GATACTCCACTATCAGGCCTCCTATCTGGGGTATCTG	GG cloning
prBD_0045	SWI5_Primer_OE_Fw4	SWI5	CAGATACCCCAGATAGGAGGCCTGATAGTGGAGTATC	GG cloning
prBD_0046	SWI5_Primer_OE_Rv4	SWI5	CAGCTTCAGATTTTTGGAGTCCTCACTTTGTTTCGATTG	GG cloning
prBD_0047	SWI5_Primer_OE_Fw5	SWI5	CAATCGAAACAAAGTGAGGACTCCAAAAATCTGAAGCTG	GG cloning
prBD_0048	SWI5_Primer_OE_Rv5	SWI5	GATTAGGTCTCCAAGCCTAATAAGCAAACTGTGTGCTACGGAAACTGAG	GG cloning
prBD_0049	YGR067C_Primer_OE_Fw1	YGR067C	GATTAGGTCTCCCATGGGTCCTCAGAAAAGATACATTTGTTCCTTCTG	GG cloning
prBD_0050	YGR067C_Primer_OE_Rv1	YGR067C	CAGTAAAAGTTTGTTACTAGGCAGACTTAAAAACGACTTGATACTC	GG cloning
prBD_0051	YGR067C_Primer_OE_Fw2	YGR067C	GAGTATCAAGTCGTTTTTAAGTCTGCCTAGTAACAAACTTTTACTG	GG cloning
prBD_0052	YGR067C_Primer_OE_Rv2	YGR067C	AGGATAGGTGTGTCCTCCTCTGTATCTTCCAC	GG cloning
prBD_0053	YGR067C_Primer_OE_Fw3	YGR067C	TGGAAGATACAGAGGAGGACACACCTATCCTG	GG cloning

prBD_0054	YGR067C_Primer_OE_Rv3	YGR067C	GATTAGGTCTCCAAGCCTAATTAAACCCCCATAAAAGGCCATCCTGTTTAC	GG cloning
prBD_0055	YPR022C-1_Primer_OE_Fw	YPR022C-1	GATTAGGTCTCCCATGATTACTTCTGAAGCATCG	GG cloning
prBD_0056	YPR022C-1_Primer_OE_Rv	YPR022C-1	GATTAGGTCTCCAAGCCTAGCCGTACATATTTTGACAAG	GG cloning
prBD_0178	CAT8-2_OE_gBlock	CAT8-2	GATTAGGTCTCCTACAAAAGACCATCTTCAATAGTCAAACAAGAACCAACAATCAACCCGAGGTCAAATGGCACTAACACCGATAGCAATCTATTCGATACCTTTAATGATTCTATCAAAGGCTCTTTGAATAACGGTTTGAAGAAGTTGAAAGATATCAGATGCAATTCTGTCGTGGAAAGATCTCACTCTTCCCAAAGAAATGATTTCTTGATGGATCAAGAGGACAGTATAACCAAGGAGACAATCAACTTTTCTGAGCTTTTCACCTGCGGAACTCCAACTGCGTCTCAGAGTATTGACAGATCTCCCAAGTCACTGCTGTTAAATGACTTAGCTATAGCTCCCGATACTTTGGTCATCAAACCAGACGCTGAAGATCTGGACAGATTGAAAAACAAAATCAGATCTGTCAAATCAACTGTTCACTAGGCTTCGAGACCGATTA	GG cloning
prBD_0179	CAT8-2_OE_Fw1	CAT8-2	GATTAGGTCTCGCATGAAAGAGAACCAAGCCTCC	GG cloning
prBD_0180	CAT8-2_OE_Rv1	CAT8-2	GAGTCTAGACTATTGTTATCTTCTGCAAATTTTTCGTTAAG	GG cloning
prBD_0181	CAT8-2_OE_Fw2	CAT8-2	CTTAACGAAAAATTTGCAGAAGATAACAATAGTCTAGACTC	GG cloning
prBD_0182	CAT8-2_OE_Rv2	CAT8-2	CTTGGGTCAATGTGTAGACGGAGTGAAG	GG cloning
prBD_0183	CAT8-2_OE_Fw3	CAT8-2	CTTCACTCCGTCTACACATTGACCCAAG	GG cloning
prBD_0184	CAT8-2_OE_Rv3	CAT8-2	GAACATATTCATCAAATCTTCATGAGCAGCAGAG	GG cloning
prBD_0185	CAT8-2_OE_Fw4	CAT8-2	CTCTGCTGCTCATGAAGATTTGATGAATATGTTC	GG cloning
prBD_0186	CAT8-2_OE_Rv4	CAT8-2	GATTAGGTCTCTTGTAACTTCCAAAAAGCGAG	GG cloning
prBD_0227	OE_AOX_amp_fwd	AOXtt	ATGGGGTGGTGTTTTGGA	Colony PCR OE
prBD_0228	OE_Kan_amp_Rv	Kan resistance gene	AAGCTTTTGCCATTCTCAC	Colony PCR OE

Table 9: Primers used for the generation of the knock-out mutant strains.

Number	Name	Target	Sequence	Purpose
prBD_0060	CAT8-2_KO_HA1_Fw	CAT8-2_HA1	GATTAGGTCTCCCATGCGTCTCGGGCAGAAAGAAAGTACTAATATG	GG cloning
prBD_0061	CAT8-2_KO_HA1_Rv	CAT8-2_HA1	GATTAGGTCTCCTAAACAAAATAAGAGATACTCTAAAAAAACAAG	GG cloning
prBD_0062	CAT8-2_KO_HA2_Fw1	CAT8-2_HA2	GATTAGGTCTCCTTTAGAGTAAAAGAGTTCCCAAATGAATG	GG cloning

prBD_0063	CAT8-2_KO_HA2_Rv1	CAT8-2_HA2	TAATCTACGGTCTAGTGATGGTGATC	GG cloning
prBD_0064	CAT8-2_KO_HA2_Fw2	CAT8-2_HA2	ATCACCATCACTAGACCGTAGATTAC	GG cloning
prBD_0065	CAT8-2_KO_HA2_Rv2	CAT8-2_HA2	GATTAGGTCTCGAAGCCGTCTCAGGATGGTCCAAGCGTC	GG cloning
prBD_0066	SWI5_KO_HA1_Fw	SWI5_HA1	GATTAGGTCTCCCATGCGTCTCTGGAACTTCCAACAAGTCAC	GG cloning
prBD_0067	SWI5_KO_HA1_Rv	SWI5_HA1	GATTAGGTCTCCTGTGCTATGAACCTCATTTAAC	GG cloning
prBD_0068	SWI5_KO_HA2_Fw	SWI5_HA2	GATTAGGTCTCCCACATAACATATTCATTTCGATTAGTTTCATTTTAG	GG cloning
prBD_0069	SWI5_KO_HA2_Rv	SWI5_HA2	GATTAGGTCTCCAAGCCGTCTCTGGCTAAAATTGAAGTTCTTCG	GG cloning
prBD_0070	SKO1_KO_HA1_Fw	SKO1_HA1	GATTAGGTCTCCCATGCGTCTCTGGCATTCAAGAGGTGG	GG cloning
prBD_0071	SKO1_KO_HA1_Rv	SKO1_HA1	GATTAGGTCTCCAACGCCAATCAATCCCG	GG cloning
prBD_0072	SKO1_KO_HA2_Fw1	SKO1_HA2	GATTAGGTCTCCCGTTATTGAAACAAAGATTCGCGATCACAGAAAAACG	GG cloning
prBD_0073	SKO1_KO_HA2_Rv1	SKO1_HA2	CATACCTCCCAACGCGACCCTAACGTG	GG cloning
prBD_0074	SKO1_KO_HA2_Fw2	SKO1_HA2	CACGTTAGGGTCGCGTTGGGAGGTATG	GG cloning
prBD_0075	SKO1_KO_HA2_Rv2	SKO1_HA2	GATTAGGTCTCCAAGCCGTCTCTGGTTCAGTTGAAATAACTCGTCCTGC	GG cloning
prBD_0076	YPR022C-1_KO_HA1_Fw1	YPR022C-1_HA1	GATTAGGTCTCCCATGCGTCTCTGGCAATCCACCGATGAAGAGG	GG cloning
prBD_0077	YPR022C-1_KO_HA1_Rv1	YPR022C-1_HA1	CCCAACAATTTCAGGGACCTATCCAAAC	GG cloning
prBD_0078	YPR022C-1_KO_HA1_Fw2	YPR022C-1_HA1	GTTTGGATAGGTCCCTGAAATTGTTGGG	GG cloning
prBD_0079	YPR022C-1_KO_HA1_Rv2	YPR022C-1_HA1	GATTAGGTCTCCGGTACTGCAAGTTGACCCCACTCTATTG	GG cloning
prBD_0080	YPR022C-1_KO_HA2_Fw	YPR022C-1_HA2	GATTAGGTCTCCTACCTAATAAGTAGATGAACGACAACAC	GG cloning
prBD_0081	YPR022C-1_KO_HA2_Rv	YPR022C-1_HA2	GATTAGGTCTCCAAGCCGTCTCTGGATGTGAACGAGATTCATAAAC	GG cloning
prBD_0132	CAT8-1_KO_HA1_Fw	CAT8-1_HA1	GATTAGGTCTCCCATGGAAGACCTGCTCGTCGTGTCCTAAC	GG cloning
prBD_0133	CAT8-1_KO_HA1_Rv	CAT8-1_HA1	GATTAGGTCTCCGGGAGAGCTAGTGAGAAAC	GG cloning
prBD_0134	CAT8-1_KO_HA2_Fw1	CAT8-1_HA2	GATTAGGTCTCCTCCCGAGAAGGTAGAGCGGATAG	GG cloning
prBD_0135	CAT8-1_KO_HA2_Rv1	CAT8-1_HA2	CAATGAATGGACTCTCTTAACGGTTAATC	GG cloning

prBD_0136	CAT8-1_KO_HA2_Fw2	CAT8-1_HA2	GATTAACCGTTAAGAGAGTCCATTCATTG	GG cloning
prBD_0137	CAT8-1_KO_HA2_Rv2	CAT8-1_HA2	GATTAGGTCTCCAAGCGAAGACCAGAAAGATGCTGAGTGAC	GG cloning
prBD_0138	SGF29_KO_HA1_Fw	SGF29_HA1	GATTAGGTCTCGAAACTTCCCTGTCATAATTATACCAAAAG	GG cloning
prBD_0139	SGF29_KO_HA1_Rv	SGF29_HA1	GATTAGGTCTCGAAGCGAAGACGGATCAAACTGAGTTTCGG	GG cloning
prBD_0140	SGF29_KO_HA2_Fw1	SGF29_HA2	GATTAGGTCTCGCATGGAAGACTCGTTCGGTCCCTTCTTC	GG cloning
prBD_0141	SGF29_KO_HA2_Rv1	SGF29_HA2	CAAATTCAGAAGGCAGCCAGAAAG	GG cloning
prBD_0142	SGF29_KO_HA2_Fw2	SGF29_HA2	CTTTCTGGCTGCCTTCTGAATTTG	GG cloning
prBD_0143	SGF29_KO_HA2_Rv2	SGF29_HA2	GATTAGGTCTCGGTTTTGTCGTTCCTCACTC	GG cloning
prBD_0148	PAS_chr1-3_0010_KO_HA1_Fw	PAS_chr1-3_0010_HA1	GATTAGGTCTCCCATGCGTCTCAGAACGGACCACAAAAAG	GG cloning
prBD_0149	PAS_chr1-3_0010_KO_HA1_Rv	PAS_chr1-3_0010_HA1	GATTAGGTCTCCACGCAAAACGAAAGAGAG	GG cloning
prBD_0150	PAS_chr1-3_0010_KO_HA2_Fw	PAS_chr1-3_0010_HA2	GATTAGGTCTCCGCGTTACAGGGAAGGGGAGCTA	GG cloning
prBD_0151	PAS_chr1-3_0010_KO_HA2_Rv	PAS_chr1-3_0010_HA2	GATTAGGTCTCGAAGCCGTCTCGGGATGAAGTTCAGGAGCAA	GG cloning
prBD_0152	PAS_chr1-3_0166_KO_HA1_Fw1	PAS_chr1-3_0166	GATTAGGTCTCCCATGCGTCTCCCGTTGCTCGACAAAAACAAATACG	GG cloning
prBD_0153	PAS_chr1-3_0166_KO_HA1_Rv1	PAS_chr1-3_0166	TTCTGACAGCGTCTGCCCATGACGTTTC	GG cloning
prBD_0154	PAS_chr1-3_0166_KO_HA1_Fw2	PAS_chr1-3_0166	AACGTCATGGGCAGACGCTGTCAGAATAG	GG cloning
prBD_0155	PAS_chr1-3_0166_KO_HA1_Rv2	PAS_chr1-3_0166	GATTAGGTCTCCAGTACCCATTAAGTGGGCTAAGCCATTG	GG cloning
prBD_0156	PAS_chr1-3_0166_KO_HA2_Fw1	PAS_chr1-3_0166	GATTAGGTCTCCTACTCACTGATCAATCTGATAAAATTTCCCTTCTTAG	GG cloning
prBD_0157	PAS_chr1-3_0166_KO_HA2_Rv1	PAS_chr1-3_0166	CACATTACTGAGGTCTACACAACGAATG	GG cloning
prBD_0158	PAS_chr1-3_0166_KO_HA2_Fw2	PAS_chr1-3_0166	CATTCGTTGTGTAGACCTCAGTAATGTG	GG cloning
prBD_0159	PAS_chr1-3_0166_KO_HA2_Rv2	PAS_chr1-3_0166	GATTAGGTCTCCAAGCCGTCTCAAATGTCCCATTAAACTCAGGACTATTTC	GG cloning
prBD_0160	AFU1_KO_HA1_Fw	AFU1_HA1	GATTAGAAGACCGCATGCGTCTCTGTGGGTCAGCTAAAAAAAC	GG cloning

prBD_0161	AFU1_KO_HA1_Rv	AFU1_HA1	GATTAGAAGACGCATAACTAGCAGACTTTGGC	GG cloning
prBD_0162	AFU1_KO_HA2_Fw	AFU1_HA2	GATTAGAAGACGCTTATGTTTTTATCAAGGACATGGGG	GG cloning
prBD_0163	AFU1_KO_HA2_Rv	AFU1_HA2	GATTAGAAGACGCAAGCCGTCTCTTTCTCGTTTACTTTTCTCTAACTTC	GG cloning
prBD_0187	YGR067C_KO_KO_HA1_Fw	YGR067C_HA1	GATTAGGTCTCCCATGCGTCTCAGACAAGTCAAAACAACCG	GG cloning
prBD_0188	YGR067C_KO_KO_HA1_Rv	YGR067C_HA1	GATTAGGTCTCCGTGTTTAATTATTGTTCTTCTATGCTC	GG cloning
prBD_0189	YGR067C_KO_KO_HA2_Fw	YGR067C_HA2	GATTAGGTCTCCACACAAACCACCCGCATTATAC	GG cloning
prBD_0190	YGR067C_KO_KO_HA2_Rv	YGR067C_HA2	GATTAGGTCTCCAAGCCGTCTCTAGATTTCTGTTCTCCACTC	GG cloning
prBD_0191	PAS_chr3_0836_KO_HA1_Fw	PAS_chr3_0836_HA1	GATTAGGTCTCCCATGCGTCTCAAAACCTTTCCTACCAACTC	GG cloning
prBD_0192	PAS_chr3_0836_KO_HA1_Rv	PAS_chr3_0836_HA1	GATTAGGTCTCCAGTACGAAGAATCTGAGAGAG	GG cloning
prBD_0193	PAS_chr3_0836_KO_HA2_Fw	PAS_chr3_0836_HA2	GATTAGGTCTCCTACTGAGGGCTGAAACAAAAAAAAC	GG cloning
prBD_0194	PAS_chr3_0836_KO_HA2_Rv	PAS_chr3_0836_HA2	GATTAGGTCTCCAAGCTGCAGGAGACGTTCACAC	GG cloning
prBD_0195	PAS_chr4_0077_KO_HA1_Fw	PAS_chr4_0077_HA1	GATTAGGTCTCCCATGCGTCTCATGAATAAATGAGCTGGCTG	GG cloning
prBD_0196	PAS_chr4_0077_KO_HA1_Rv	PAS_chr4_0077_HA1	GATTAGGTCTCCAACATCTTGCTAACTTGGG	GG cloning
prBD_0197	PAS_chr4_0077_KO_HA2_Fw	PAS_chr4_0077_HA2	GATTAGGTCTCCTGTTATGCCTCAACTGGTTTAAAG	GG cloning
prBD_0198	PAS_chr4_0077_KO_HA2_Rv	PAS_chr4_0077_HA2	GATTAGGTCTCCAAGCCGTCTCAAGTAAATGGTGGTAGGAAG	GG cloning
prBD_0237	YGR067C_KO_HA2_Fw	YGR067C_KO_HA2	GATTAGGTCTCCACACTAGATTTCTGTTCTCCACTC	GG cloning
prBD_0238	YGR067C_KO_HA2_Rv	YGR067C_KO_HA2	GATTAGGTCTCCAAGCCGTCTCAAACCACCCGCATTATAC	GG cloning
prBD_0082	CAT8-2_gRNA_1	CAT8-2	GATAGGTCTCCCATGCTTTGTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0083	CAT8-2_gRNA_2	CAT8-2	AAACGAGTAAGCTCGTCACAAAGAAGTAACAAATAAAgttttagagctagaaatagcaag	gRNA assembly
prBD_0084	SWI5_gRNA_1	SWI5	GATAGGTCTCCCATGTTCTTGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0085	SWI5_gRNA_2	SWI5	AAACGAGTAAGCTCGTCCAAGAAGATCTAAGCAAAGTgttttagagctagaaatagcaag	gRNA assembly
prBD_0086	SKO1_gRNA_1	SKO1	GATAGGTCTCCCATGACGTTACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0087	SKO1_gRNA_2	SKO1	AAACGAGTAAGCTCGTCTAACGTAATGACGACGTCATgttttagagctagaaatagcaag	gRNA assembly

prBD_0088	YPR022C-1_gRNA_1	YPR022C-1	GATAGGTCTCCCATGCTCACACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0089	YPR022C-1_gRNA_2	YPR022C-1	AAACGAGTAAGCTCGTCTGTGAGGGGCAACATGATTCgttttagagctagaaatagcaag	gRNA assembly
prBD_0144	CAT8-1_gRNA_1	CAT8-1	GATAGGTCTCCCATGTACTAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0145	CAT8-1_gRNA_2	CAT8-1	AAACGAGTAAGCTCGTCTTAGTACGAAGCCAGCTCGCgttttagagctagaaatagcaag	gRNA assembly
prBD_0146	SGF29_gRNA_1	SGF29	GATAGGTCTCCCATGGCTGTACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0147	SGF29_gRNA_2	SGF29	AAACGAGTAAGCTCGTCTACAGCAAGATTGTTGCTGTgttttagagctagaaatagcaag	gRNA assembly
prBD_0164	PAS_chr1-3_0010_gRNA_1	PAS_chr1-3_0010	GATAGGTCTCCCATGTCTAGACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0165	PAS_chr1-3_0010_gRNA_2	PAS_chr1-3_0010	AAACGAGTAAGCTCGTCTCTAGAATTTTTGACATTTAgttttagagctagaaatagcaag	gRNA assembly
prBD_0166	PAS_chr1-3_0166_gRNA_1	PAS_chr1-3_0166	GATAGGTCTCCCATGTATTAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0167	PAS_chr1-3_0166_gRNA_2	PAS_chr1-3_0166	AAACGAGTAAGCTCGTCTTAATATATATTTTATACGAgttttagagctagaaatagcaag	gRNA assembly
prBD_0168	AFU1_gRNA_1	AFU1	GATAGGTCTCCCATGATTAATCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0169	AFU1_gRNA_2	AFU1	AAACGAGTAAGCTCGTCATTAATCTCAGAGTTTTGGCgttttagagctagaaatagcaag	gRNA assembly
prBD_0199	YGR067C_KO_gRNA1	YGR067C	GATAGGTCTCCCATGCTTTTGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0200	YGR067C_KO_gRNA2	YGR067C	AAACGAGTAAGCTCGTCCAAAAGGAGGAGGCGGTTGAgttttagagctagaaatagcaag	gRNA assembly
prBD_0201	PAS_chr3_0836_KO_gRNA1	PAS_chr3_0836	GATAGGTCTCCCATGGCAAGCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0202	PAS_chr3_0836_KO_gRNA2	PAS_chr3_0836	AAACGAGTAAGCTCGTCGCTTGCCCAATTATATATATgttttagagctagaaatagcaag	gRNA assembly
prBD_0203	PAS_chr4_0077_KO_gRNA1	PAS_chr4_0077	GATAGGTCTCCCATGGCAGTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0204	PAS_chr4_0077_KO_gRNA2	PAS_chr4_0077	AAACGAGTAAGCTCGTCAACTGCAACTTTGGCACACGgttttagagctagaaatagcaag	gRNA assembly
prBD_0229	SGF29_KO_gRNA_2_1	SGF29	GATAGGTCTCCCATGTTCGAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0230	SGF29_KO_gRNA_2_2	SGF29	AAACGAGTAAGCTCGTCTTCGAATGTGGGTGATTCGTgttttagagctagaaatagcaag	gRNA assembly
prBD_0235	OE_AOXtt_gRNA1	AOXtt	GATAGGTCTCCCATGGAGAATCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0236	OE_AOXtt_gRNA2	AOXtt	AAACGAGTAAGCTCGTCATTCTCGAGTATCGGCGCGCgttttagagctagaaatagcaag	gRNA assembly
prBD_0239	AFU1_gRNA_2_1	AFU1	GATAGGTCTCCCATGCGGACCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly

prBD_0240	AFU1_gRNA_2_2	AFU1	AAACGAGTAAGCTCGTCGGTCCGATCAAATTGTCGATgttttagagctagaaatagcaag	gRNA assembly
prBD_0241	0010_gRNA_2_1	PAS_chr1-3_0010	GATAGGTCTCCCATGCTAGAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0242	0010_gRNA_2_2	PAS_chr1-3_0010	AAACGAGTAAGCTCGTCTTCTAGAGATTCAAAAGAATgttttagagctagaaatagcaag	gRNA assembly
prBD_0243	0077_gRNA_2_1	PAS_chr4_0077	GATAGGTCTCCCATGGATACCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0244	0077_gRNA_2_2	PAS_chr4_0077	AAACGAGTAAGCTCGTCGGTATCAAACAGCTCAAATTgttttagagctagaaatagcaag	gRNA assembly
prBD_0247	PAS_chr1-3_0166_gRNA_2_1	PAS_chr1-3_0166	GATAGGTCTCCCATGATGCATCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0248	PAS_chr1-3_0166_gRNA_2_2	PAS_chr1-3_0166	AAACGAGTAAGCTCGTCATGCATAAGAGGATATATATGttttagagctagaaatagcaag	gRNA assembly
prBD_0249	PAS_chr3_0836_gRNA_2_1	PAS_chr3_0836	GATAGGTCTCCCATGACTCTGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0250	PAS_chr3_0836_gRNA_2_2	PAS_chr3_0836	AAACGAGTAAGCTCGTCCAGAGTAGAAATAGTTACAGgttttagagctagaaatagcaag	gRNA assembly
prBD_0251	YGR067C_gRNA_2_1	YGR067C	GATAGGTCTCCCATGATTGTCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly
prBD_0252	YGR067C_gRNA_2_2	YGR067C	AAACGAGTAAGCTCGTCGACAATGAACTTGTAGAATTgttttagagctagaaatagcaag	gRNA assembly
prBD_0108	SWI5_KO_outHR_Fw	SWI5_OutHR	CTTTTTTCGTCTTGCTTGGT	Colony PCR KO
prBD_0109	SWI5_KO_outHR_Rv	SWI5_OutHR	TGGTATTTCGGGGAAGCTG	Colony PCR KO
prBD_0110	SWI5_KO_Fw	SWI5	ATGACCCAAACTCTAACGG	Colony PCR KO
prBD_0111	SWI5_KO_Rv	SWI5	GTGCTACGGAAACTGAGG	Colony PCR KO
prBD_0112	YPR022C-1_KO_outHR_Fw	YPR022_OutHR	TGTCCATTCCAGCAGTC	Colony PCR KO
prBD_0113	YPR022C-1_KO_outHR_Rv	YPR022_OutHR	CTCAGATAAGGAAGCGATT	Colony PCR KO
prBD_0114	YPR022C-1_KO_In_Fw	YPR022C-1	GAAGGAGGATGGAAGGT	Colony PCR KO
prBD_0115	YPR022C-1_KO_In_Rv	YPR022C-1	CTCTTTCCTAGACTTCGTT	Colony PCR KO
prBD_0116	SKO1_KO_outHR_Fw	SKO1_OutHR	TTGGAAGCAGGAAGAGA	Colony PCR KO
prBD_0117	SKO1_KO_outHR_Rv	SKO1_OutHR	AACTTAGGTGCCCTTG	Colony PCR KO
prBD_0118	SKO1_KO_In_Fw	SKO1	ACCCTAACTCCATTTGTTC	Colony PCR KO
prBD_0119	SKO1_KO_In_Rv	SKO1	CAAGCCTTCTAGCGATAC	Colony PCR KO

prBD_0170	CAT8-2_KO_outHR_Fw	CAT8-2_OutHR	CGGGAGAACACTTTTGATG	Colony PCR KO
prBD_0171	CAT8-2_KO_outHR_Rv	CAT8-2_OutHR	CTGTTGGGTGTCTATCTAGG	Colony PCR KO
prBD_0172	SGF29_KO_outHR_Fw	SGF29_OutHR	ATTCAACCTCAAAGTCTTCC	Colony PCR KO
prBD_0173	SGF29_KO_outHR_Rv	SGF29_OutHR	GTTACCACATCATCCACTTTC	Colony PCR KO
prBD_0213	CAT8-1_KO_OutHR_Fw	CAT8-1_OutHR	AATACCAATAACCAGCACC	Colony PCR KO
prBD_0214	CAT8-1_KO_OutHR_Rv	CAT8-1_OutHR	ATTATCGCCTTTTTCCTGAG	Colony PCR KO
prBD_0215	AFU1_KO_OutHR_Fw	AFU1_OutHR	GCCAATCCCCTCTATCAC	Colony PCR KO
prBD_0216	AFU1_KO_OutHR_Rv	AFU1_OutHR	TTCGGCGCAATCTCATAC	Colony PCR KO
prBD_0217	PAS_chr1-3_0166_OutHR_Fw	PAS_chr1-3_0166_OutHR	GTTTTTGGGGGTATCCTTG	Colony PCR KO
prBD_0218	PAS_chr1-3_0166_OutHR_Rv	PAS_chr1-3_0166_OutHR	GCTCTGAACTGGTTGTAAATC	Colony PCR KO
prBD_0219	PAS_chr1-3_0010_OutHR_Fw	PAS_chr1-3_0010_OutHR	TTTGTAGGAACCAGATTGAG	Colony PCR KO
prBD_0220	PAS_chr1-3_0010_OutHR_Rv	PAS_chr1-3_0010_OutHR	TTGCAAAAGGTGACAGAG	Colony PCR KO
prBD_0221	YGR067C_KO_OutHR_Fw	YGR067C_KO_OutHR	CCAAGAACAGACAACAAC	Colony PCR KO
prBD_0222	YGR067C_KO_OutHR_Rv	YGR067C_KO_OutHR	AGACGGATCTACCAGAAG	Colony PCR KO
prBD_0223	PAS_chr3_0836_KO_OutHR_Fw	PAS_chr3_0836_KO_OutHR	ATCATTAGTCTGTTCCTGC	Colony PCR KO
prBD_0224	PAS_chr3_0836_KO_OutHR_Rv	PAS_chr3_0836_KO_OutHR	AAACTGGGGTAAAGAACTC	Colony PCR KO
prBD_0225	PAS_chr4_0077_KO_OutHR_Fw	PAS_chr4_0077_KO_OutHR	CTGTTTGAGACTCTTGTTACC	Colony PCR KO
prBD_0226	PAS_chr4_0077_KO_OutHR_Rv	PAS_chr4_0077_KO_OutHR	ATGTATTGGATACAGAAGTTTCTC	Colony PCR KO

Table 10: Primers used for rtPCR.

Number	Name	Target	Sequence	Purpose
prBD_0098	AFU1_qPCR_Fw	AFU1	TGCCGATGTTTGATGTTCCT	qPCR
prBD_0099	AFU1_qPCR_Rv	AFU1	TCTTCAACACTGCTTCCACTT	qPCR
prBD_0100	PAS_chr1-3_0166_qPCR_Fw	PAS_chr1-3_0166	CCCTCCTTGATGCTCTTTTCT	qPCR
prBD_0101	PAS_chr1-3_0166_qPCR_Rv	PAS_chr1-3_0166	AGACACTCCCACGACCAA	qPCR

	ACT1_up	ACT1	CCTGAGGCTTTGTTCCACCCATCT	qPCR
	ACT1_low	ACT1	GGAACATAGTAGTACCACCGGACATAACGA	qPCR
prBD_0120	YGR067C_qPCR_Fw	YGR067C	CCTGGGATTGGTGACTGTA	qPCR
prBD_0121	YGR067C_qPCR_Rv	YGR067C	ATCACAACAACAACAACGCT	qPCR
prBD_0122	PAS_chr1-3_0010_qPCR_Fw	PAS_chr1-3_0010	GTGTGCTCAACCTTCCTTT	qPCR
prBD_0123	PAS_chr1-3_0010_qPCR_Rv	PAS_chr1-3_0010	ATCTACAACAGCAAGACCTC	qPCR
prBD_0124	PAS_chr3_0836_qPCR_Rv	PAS_chr3_0836	CTAGCAGCCAAAGTGAAC	qPCR
prBD_0125	PAS_chr3_0836_qPCR_Fw	PAS_chr3_0836	ATTAACTCCTTGCCCTTCCA	qPCR
prBD_0126	SWI5_qPCR_Fw	SWI5	TGCATAGGAGGTTTTGATAGTG	qPCR
prBD_0127	SWI5_qPCR_Rv	SWI5	CATCCATTTCGTTCACCAT	qPCR
prBD_0128	YPR022C-1_qPCR_Fw	YPR022C-1	ATCATCACCTATATCGACCACA	qPCR
prBD_0129	YPR022C-1_qPCR_Rv	YPR022C-1	ATAACCTTCCTCCCTTGCT	qPCR
prBD_0130	SKO1_qPCR_Fw	SKO1	GGCGTTAGTACAGTGGGAGA	qPCR
prBD_0131	SKO1_qPCR_Rv	SKO1	ATTGGAGCCTAACCCGTTT	qPCR
prBD_0174	SGF29_qPCR_Fw	SGF29	TCGCCTTTCAACCTCAGT	qPCR
prBD_0175	SGF29_qPCR_Rv	SGF29	CCCCAACGGATTTAAAGTTCT	qPCR
prBD_0176	CAT8-2_qPCR_Fw	CAT8-2	GGTCTTCAGCGTCTGGTT	qPCR
prBD_0177	CAT8-2_qPCR_Rv	CAT8-2	ACTTTTCTGAGCTTTTCACCTG	qPCR
prBD_0205	CAT8-1_qPCR_Fw	CAT8-1	TAATGTATCTCCTCCCAATAGTGAAAG	qPCR
prBD_0206	CAT8-1_qPCR_Rv	CAT8-1	GTCCGAATAAACTCCCAGCAG	qPCR
prBD_0207	AFU1_qPCR_Fw	AFU1	AGTGGAAGCAGTGTTGAAGA	qPCR
prBD_0208	AFU1_qPCR_Rv	AFU1	AAGTCGAATCCCTCCATGTTAG	qPCR
prBD_0209	PAS_chr1-3_0166_qPCR_Fw	PAS_chr1-3_0166	AGCCTCAACCAGCATTCC	qPCR
prBD_0210	PAS_chr1-3_0166_qPCR_Rv	PAS_chr1-3_0166	GCCCCACTACCCAAAATTCAG	qPCR
prBD_0211	PAS_chr4_0077_qPCR_Fw	PAS_chr4_0077	TGCTGGTCTTCTGCTTCATC	qPCR
prBD_0212	PAS_chr4_0077_qPCR_RV	PAS_chr4_0078	TTGTGAGTTGCCCCCGTTA	qPCR
prBD_0253	AOX1_rtPCR_Fw	AOX1	TTGAAGGTTGGTGACTTGTCC	qPCR
prBD_0254	AOX1_rtPCR_Rv	AOX1	AGGAACAGTCATGTCTAAGGC	qPCR
prBD_0255	AOX2_rtPCR_Fw	AOX2	CTTCTCGTAAGTGCCCAAC	qPCR
prBD_0256	AOX2_rtPCR_Rv	AOX2	TAACACCTACACCACCGC	qPCR
prBD_0257	DAS1_rtPCR_Fw	DAS1	CGGTAAGTCTCTTCCTGTTG	qPCR
prBD_0258	DAS1_rtPCR_Rv	DAS1	TTGGTTTTCCCTTCAAGTCG	qPCR
prBD_0259	DAS2_rtPCR_Fw	DAS2	GGTTTTCCCTTCAAATCGGTG	qPCR
prBD_0260	DAS2_rtPCR_Rv	DAS2	TTTGGTAAGTCTCTTCCTGTTGAG	qPCR
prBD_0261	FDH1_rtPCR_Fw	FDH1	CTTGAGGTCTGTAGTCAAACTTC	qPCR
prBD_0262	FDH1_rtPCR_Rv	FDH1	GGAGAGATATGAGAAACAAGTACGG	qPCR
prBD_0263	PEX5_rtPCR_Fw	PEX5	CTCTCCTATTCATACCCAAAAATGC	qPCR
prBD_0264	PEX5_rtPCR_Rv	PEX5	AGGTTGAAGGTGTTGATGC	qPCR
prBD_0265	PEX14_rtPCR_Fw	PEX14	CTCTTTTGGAGCTGGTGC	qPCR
prBD_0266	PEX14_rtPCR_Rv	PEX14	CGAGGAGGACGAGAAAAAGG	qPCR
prBD_0267	FLD_rtPCR_Fw	FLD	TTGGAGGTATCAAGGGACG	qPCR

prBD_0268	FLD_rtPCR_Rv	FLD	CTTTGTTGATGTTGTCCAGGTC	qPCR
prBD_0269	ICL1_rtPCR_Fw	ICL1	CAGAAATGGTCAGGAGCCG	qPCR
prBD_0270	ICL1_rtPCR_Rv	ICL1	AATGGTCCTTGAATTGATCTTCAG	qPCR
prBD_0271	ADH2_rtPCR_Fw	ADH2	CCAGCCTCCATCTGTTCGTA	qPCR
prBD_0272	ADH2_rtPCR_Rv	ADH2	GTTCTGAAGTCCATCGAGATCAAG	qPCR
prBD_0320	MLS1_rtPCR_Fw	MLS1	ACTGGCGAGAAGATCACA	qPCR
prBD_0321	MLS1_rtPCR_Rv	MLS1	CTCCACTAATCTCCTTCTTCAAAC	qPCR
prBD_0322	YAT2_rtPCR_Fw	YAT2	GGCAGTGCAACCAACTTA	qPCR
prBD_0323	YAT2_rtPCR_Rv	YAT2	TGGAAGAAAGAGGTGGAGAGTGAA	qPCR
prBD_0324	CRC1_rtPCR_Fw	CRC1	ATGCCTTAACCCCACCTTT	qPCR
prBD_0325	CRC1_rtPCR_Rv	CRC1	ACAATGGCAGGAGGTTTT	qPCR
prBD_0326	FBP1_rtPCR_Fw	FBP1	TGAAGAAACCCCAAGCAAAC	qPCR
prBD_0327	FBP1_rtPCR_Rv	FBP1	TGGAGTCTGCTGGATAGC	qPCR
prBD_0328	PCK1_rtPCR_Fw	PCK1	CCAACCTTCGGTCTACAAAT	qPCR
prBD_0329	PCK1_rtPCR_Rv	PCK1	CTCGGTGTTGAAGTTGTCT	qPCR
prBD_0365	ACS1_qPCR_Fw	ACS1	GGTTTTGGCTGGAGAGGAAGA	qPCR
prBD_0366	ACS1_qPCR_Rv	ACS1	TTGCGGGCATCCCTTTT	qPCR
prBD_0367	ALD4_qPCR_Fw	ALD4	CGGTCTTGCTGCTGGTAT	qPCR
prBD_0368	ALD4_qPCR_Rv	ALD4	TTTGGTGGAAATCGTTGTAGGT	qPCR

Table 11: Primers used for cloning the eGFP coding gene under the control of P_{CAT8-1} and P_{CAT8-2}.

Number	Name	Target	Sequence	Purpose
prBD_0309	CAT8-1_GG_prom_Fw	P_{CAT8-1}	GATTAGAAGACCCGGAGACTCCAACAGAGAGTGAATTG	Prom CAT8-1_eGFP
prBD_0344	CAT8-1_GG_prom_Rv	P_{CAT8-1}	GATTAGAAGACCCCATGTTTAAGGAGATGAATCGAGACAC	Prom CAT8-1_eGFP
prBD_0307	CAT8-2_GG_prom_Fw	P_{CAT8-2}	GATTAGAAGACCCGGAGGGTGATCATGAGTTGCATCC	Prom CAT8-2_eGFP
prBD_0319	CAT8-2_GG_prom_Rv	P_{CAT8-2}	GATTAGAAGACCCCATGGAACTGGGTGTAGTGGGTAAAAC	Prom CAT8-2_eGFP

II. Appendix 2: List of plasmids generated and used in this study

Table 12: List of plasmids generated and used in this study.

Number	Name	Backbone	Promoter_Gene_Terminator	Resistance	Purpose
pIBD_0001	AFU1_BB1	BB1_L_23_syn_BsaI_1	_AFU1_	Kan	OE
pIBD_0002	CAT8-1_BB1	BB1_L_23_syn_BsaI_1	_CAT8-1_	Kan	OE
pIBD_0003	PAS_chr1-3_0010_BB1	BB1_L_23_syn_BsaI_1	_PAS_chr1-3_0010_	Kan	OE
pIBD_0004	PAS_chr1-3_0166_BB1	BB1_L_23_syn_BsaI_1	_PAS_chr1-3_0166_	Kan	OE
pIBD_0005	PAS_chr3_0836_BB1	BB1_L_23_syn_BsaI_1	_PAS_chr3_0836_	Kan	OE
pIBD_0006	PAS_chr4_0077_BB1	BB1_L_23_syn_BsaI_1	_PAS_chr4_0077_	Kan	OE
pIBD_0007	SGF29_BB1	BB1_L_23_syn_BsaI_1	_SGF29_	Kan	OE
pIBD_0008	SKO1_BB1	BB1_L_23_syn_BsaI_1	_SKO1_	Kan	OE
pIBD_0009	SWI5_BB1	BB1_L_23_syn_BsaI_1	_SWI5_	Kan	OE
pIBD_0010	YGR067C_BB1	BB1_L_23_syn_BsaI_1	_YGR067C_	Kan	OE
pIBD_0011	YPR0022C-1_BB1	BB1_L_23_syn_BsaI_1	_YPR0022C-1_	Kan	OE
pIBD_0012	AFU1_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_AFU1_RPS3TT	Kan/G418	OE
pIBD_0013	CAT8-1_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_CAT8-1_RPS3TT	Kan/G418	OE
pIBD_0014	PAS_chr1-3_0010_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_PAS_chr1-3_0010_RPS3TT	Kan/G418	OE
pIBD_0015	PAS_chr1-3_0166_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_PAS_chr1-3_0166_RPS3TT	Kan/G418	OE
pIBD_0016	PAS_chr3_0836_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_PAS_chr3_0836_RPS3TT	Kan/G418	OE
pIBD_0017	PAS_chr4_0077_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_PAS_chr4_0077_RPS3TT	Kan/G418	OE
pIBD_0018	SGF29_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_SGF29_RPS3TT	Kan/G418	OE
pIBD_0019	SKO1_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_SKO1_RPS3TT	Kan/G418	OE
pIBD_0020	SWI5_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_SWI5_RPS3TT	Kan/G418	OE
pIBD_0021	YGR067C_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_YGR067C_RPS3TT	Kan/G418	OE
pIBD_0022	YPR0022C-1_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	PCS1_YPR0022C-1_RPS3TT	Kan/G418	OE
pIBD_0023	gRNA_CAT8-2_Cas9	BB3_pGAP_PCR_Linker_FS23_BbsI(intern)__RPS25Att_pTEF_Cas9_CYCt_BC	pGAP_gRNA CAT8-2_RPS25Att_pTEF_Cas9_CYCt	Kan/G418	KO

pIBD_0024	gRNA_SWI5_Cas9	BB3_pGAP_PCR_Linker_FS23_BbsI(intern)__RPS25Att_pTEF_Cas9_CYCt_BC	pGAP_gRNA SWI5_RPS25Att_pTEF_Cas9_CYCt	Kan/G418	KO
pIBD_0025	gRNA_SKO1_Cas9	BB3_pGAP_PCR_Linker_FS23_BbsI(intern)__RPS25Att_pTEF_Cas9_CYCt_BC	pGAP_gRNA SKO1_RPS25Att_pTEF_Cas9_CYCt	Kan/G418	KO
pIBD_0026	gRNA_YPR022C-1_Cas9	BB3_pGAP_PCR_Linker_FS23_BbsI(intern)__RPS25Att_pTEF_Cas9_CYCt_BC	pGAP_gRNA YPR022C-1_RPS25Att_pTEF_Cas9_CYCt	Kan/G418	KO
pIBD_0027	gRNA_CAT8-2_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_gRNA CAT8-2_	Kan	KO
pIBD_0028	gRNA_SWI5_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_gRNA SWI5_	Kan	KO
pIBD_0029	gRNA_SKO1_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_gRNA SKO1_	Kan	KO
pIBD_0030	gRNA_YPR022C-1_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_gRNA YPR022C-1_	Kan	KO
pIBD_0031	HR_SKO1_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_SKO1 HA1_SKO1 HA2_	Kan	KO
pIBD_0032	HR_YPR022C-1_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_YPR022C-1 HA1_YPR022C-1 HA2_	Kan	KO
pIBD_0033	HR_SWI5_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_SWI5 HA1_SWI5 HA2_	Kan	KO
pIBD_0038	AFU1_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_AFU1_RPS3TT	Kan/G418	OE
pIBD_0039	SKO1_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_SKO1_RPS3TT	Kan/G418	OE
pIBD_0040	SWI5_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_SWI5_RPS3TT	Kan/G418	OE
pIBD_0041	YPR0022C-1_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_YPR0022C-1_RPS3TT	Kan/G418	OE
pIBD_0042	AFU1_pLAT_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pLAT_AFU1_RPS3TT	Kan/G418	OE
pIBD_0043	SKO1_pLAT_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pLAT_SKO1_RPS3TT	Kan/G418	OE
pIBD_0044	SWI5_pLAT_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pLAT_SWI5_RPS3TT	Kan/G418	OE
pIBD_0045	YPR0022C-1_pLAT_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pLAT_YPR0022C-1_RPS3TT	Kan/G418	OE
pIBD_0046	Cas9_gRNA_CAT8-1_BB3	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA CAT8-1_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0047	Cas9_gRNA_SGF29_BB3	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA SGF29_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0048	gRNA_CAT8-1_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_gRNA CAT8-1_	Kan/G418	KO
pIBD_0049	gRNA_SGF29_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_gRNA SGF29_	Kan/G418	KO
pIBD_0050	HR_CAT8-2_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_CAT8-2 HA1_CAT8-2 HA2_	Kan	KO
pIBD_0051	HR_SGF29_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_SGF29 HA1_SGF29 HA2_	Kan	KO
pIBD_0052	HR_CAT8-1_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_CAT8-1 HA1_CAT8-1 HA2_	Kan	KO

pIBD_0053	HR_PAS_chr1-3_0010_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_PAS_chr1-3_0010 HA1_PAS_chr1-3_0010 HA2_	Kan	KO
pIBD_0054	HR_PAS_chr4_0077_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_PAS_chr4_0077 HA1_PAS_chr4_0077 HA2_	Kan	KO
pIBD_0055	HR_YGR067C_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_YGR067C HA1_YGR067C HA2_	Kan	KO
pIBD_0056	HR_PAS_chr1-3_0166_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_PAS_chr1-3_0166 HA1_PAS_chr1-3_0166 HA2_	Kan	KO
pIBD_0057	HR_AFU1_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_AFU1 HA1_AFU1 HA2_	Kan	KO
pIBD_0058	CAT8-2_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_CAT8-2_	Kan	OE
pIBD_0059	Cas9_gRNA_PAS_chr3_0836_BB3	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA PAS_chr3_0836_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0060	Cas9_gRNA_PAS_chr1-3_0010_BB3	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA PAS_chr1-3_0010_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0061	Cas9_gRNA_PAS_chr1-3_0166_BB3	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA PAS_chr1-3_0166_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0062	Cas9_gRNA_YGR067C_BB3	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA YGR067C_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0063	Cas9_gRNA_SGF29_BB3_2	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA SGF29_2_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0064	Cas9_gRNA_PAS_chr4_0077_BB3	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA PAS_chr4_0077_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0065	Cas9_gRNA_AFU1_BB3	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA AFU1_RPS25Att_ppfk1_Cas9_CYCt	Kan/G418	KO
pIBD_0066	Cas9_gRNA_AOXtt_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_gRNA AOXtt_	Kan	OE
pIBD_0067	Cas9_gRNA_AOXtt_BB3	BB3shorcut_pMDH3_Linker_BbsI(intern)_RPS25Att_pLAT1_Cas9_RPS3TT_PseudoNTC	pMDH3_gRNA AOXtt_RPS25Att_pLAT1_Cas9_CycTT	NTC	OE
pIBD_0068	HR_PAS_chr3_0836_BB1	BB1_L_23_syn_BsaI_1_v2 (070)	_PAS_chr3_0836 HA1_PAS_chr3_0836 HA2_	Kan/G418	KO
pIBD_0069	HR_YGR067C_BB1_2	BB1_L_23_syn_BsaI_1_v2 (070)	_YGR067C HA1_YGR067C HA2_	Kan	KO
pIBD_0070	Cas9_gRNA_PAS_chr3_0836_BB3_2	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA PAS_chr3_0836_RPS25Att_pLAT1_Cas9_CYCt	Kan/G418	KO
pIBD_0071	Cas9_gRNA_PAS_chr1-3_0010_BB3_2	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA PAS_chr1-3_0010_RPS25Att_pLAT1_Cas9_CYCt	Kan/G418	KO

pIBD_0072	Cas9_gRNA_PAS_chr1-3_0166_BB3_2	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA PAS_chr1-3_0166_RPS25Att_pLAT1_Cas9_CYCt	Kan/G418	KO
pIBD_0073	Cas9_gRNA_YGR067C_BB3_2	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA YGR067C_RPS25Att_pLAT1_Cas9_CYCt	Kan/G418	KO
pIBD_0074	Cas9_gRNA_PAS_chr4_0077_BB3_2	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA PAS_chr4_0077_RPS25Att_pLAT1_Cas9_CYCt	Kan/G418	KO
pIBD_0075	Cas9_gRNA_AFU1_BB3_2	GaT_B3_025_pGAP_Linker_BbsI_RPS25tt_pPppfk1_Cas9_CYCtt_KanR	pGAP_gRNA AFU1_RPS25Att_pLAT1_Cas9_CYCt	Kan/G418	KO
pIBD_0076	PAS_chr1-3_0010_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_PAS_chr1-3_0010_RPS3TT	Kan/G418	OE
pIBD_0077	PAS_chr3_0836_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_PAS_chr3_0836_RPS3TT	Kan/G418	OE
pIBD_0078	SGF29_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_SGF29_RPS3TT	Kan/G418	OE
pIBD_0079	CAT8-2_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_CAT8-2_RPS3TT	Kan/G418	OE
pIBD_0080	PAS_chr1-3_0166_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_PAS_chr1-3_0166_RPS3TT	Kan/G418	OE
pIBD_0081	PAS_chr4_0077_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_PAS_chr4_0077_RPS3TT	Kan/G418	OE
pIBD_0082	YGR067C_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_YGR067C_RPS3TT	Kan/G418	OE
pIBD_0083	CAT8-1_pTHI11_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pTHI11_CAT8-1_RPS3TT	Kan/G418	OE
pIBD_0091	CAT8-1_1HATag_AOXtt_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pCAT8-1(native)_CAT8-1_1HATag_CAT8-1TT(native)	Kan/G418	HA tag
pIBD_0092	CAT8-2_1HATag_AOXtt_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pCAT8-2(native)_CAT8-2_1HATag_CAT8-2TT(native)	Kan/G418	HA tag
pIBD_0093	CAT8-2_prom_eGFP_CycTT_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pCAT8-2(native)_eGFP_CycTT	Kan/G418	eGFP assay
pIBD_0094	CAT8-1_prom_eGFP_CycTT_BB3	BB3_KAN_FS1_FS4_AOXfragment (001)	pCAT8-1(native)_eGFP_CycTT	Kan/G418	eGFP assay

III. Appendix 3: List of strains used in this study

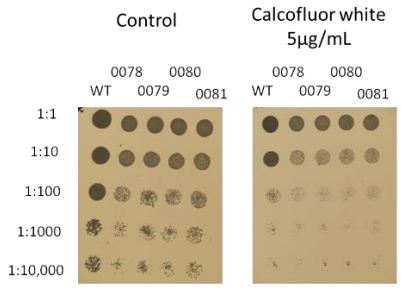
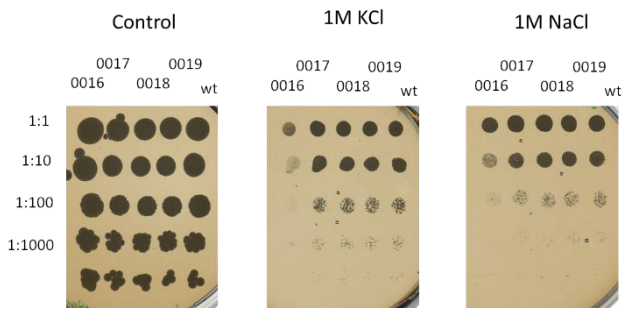
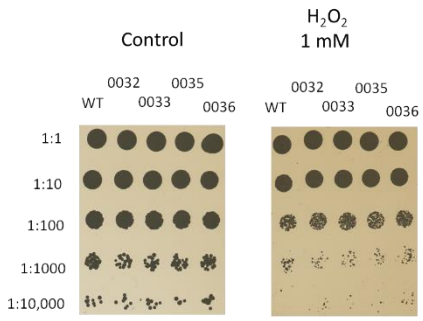
Table 13: List of strains used in this study.

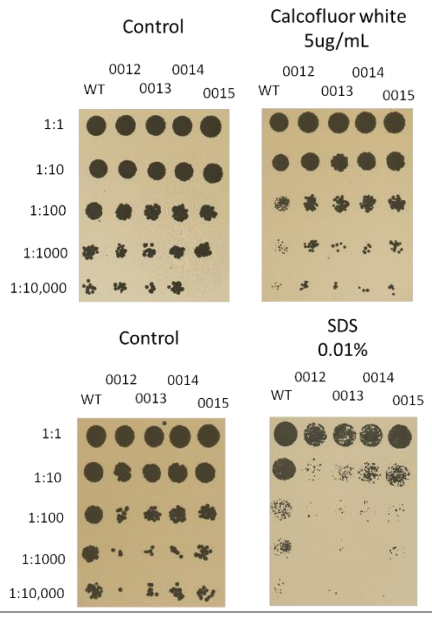
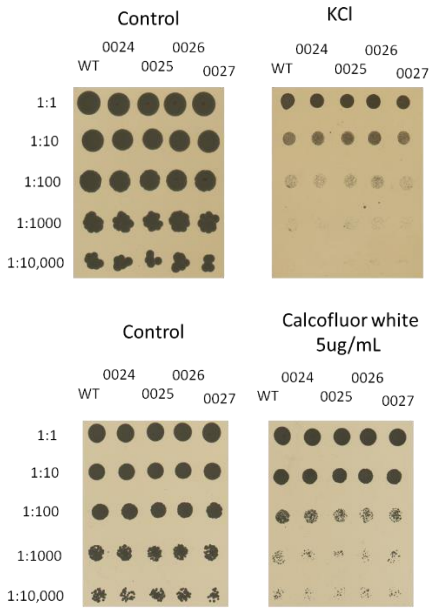
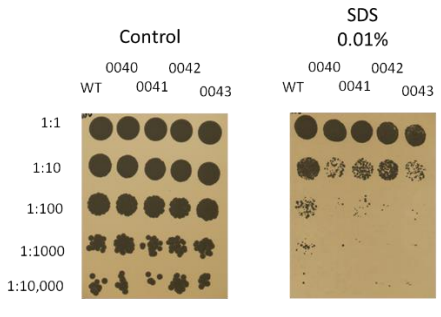
Number	Name	Target (ORF name)	Target (Short Name)	Target type (OE/ KO/Prom)	Plasmid(s) name	Source
stBD_0001-0011	YPR022C-1_KO	PP7435_Chr1-0680	YPR022C-1	KO	plBD_0026 - fgt plBD_0032	this study
stBD_0012-0015	SWI5_KO	PP7435_Chr1-0101	SWI5	KO	plBD_0024 - fgt plBD_0033	this study
stBD_0016-0019	SKO1_KO	PP7435_Chr1-0475	SKO1	KO	plBD_0025 - fgt plBD_0032	this study
stBD_0020-0023	CAT8-2_KO	PP7435_Chr4-0434	CAT8-2	KO	plBD_0023-fgt plBD_0050	this study
stBD_0024-0027	CAT8-1_KO	PP7435_Chr2-0516	CAT8-1	KO	plBD_0046-fgt plBD_0052	this study
stBD_0028-0030	SGF29_KO	PP7435_Chr1-0772	SGF29	KO	plBD_0063-fgt plBD_0051	this study
stBD_0031	AFU1_KO	PP7435_Chr1-0281	AFU1	KO	plBD_0065-fgt plBD_0057	this study
stBD_0032-0036	SWI5_OE	PP7435_Chr1-0101	SWI5	OE	plBD_0040	this study
stBD_0039-0042	AFU1_OE	PP7435_Chr1-0281	AFU1	OE	plBD_0038	this study
stBD_0043-0046	0166_OE	PP7435_Chr1-0170	PAS_chr1-3_0166	OE	plBD_0080	this study
stBD_0047-0050	SGF29_OE	PP7435_Chr1-0772	SGF29	OE	plBD_0078	this study
stBD_0051-0054	CAT8-2_OE	PP7435_Chr4-0434	CAT8-2	OE	plBD_0079	this study
stBD_0055-0058	CAT8-1_OE	PP7435_Chr2-0516	CAT8-1	OE	plBD_0083	this study
stBD_0059-0061	YPR022C-1_OE	PP7435_Chr1-0680	YPR022C-1	OE	plBD_0041	this study
stBD_0062-0065	PAS_chr1-3_0010_OE_2	PP7435_Chr1-0006	PAS_chr1-3_0010	OE	plBD_0076	this study
stBD_0066-0069	PAS_chr3_0836_OE	PP7435_Chr3-0349	PAS_chr3_0836	OE	plBD_0077	this study
stBD_0070-0073	YGR067C_OE	PP7435_Chr3-0964	YGR067C	OE	plBD_0082	this study
stBD_0074-0076	PAS_chr4_0077_OE	PP7435_Chr4-0940	PAS_chr4_0077	OE	plBD_0081	this study
stBD_0078-0081	SKO1_OE	PP7435_Chr1-0475	SKO1	OE	plBD_0039	this study
stBD_0082-0085	WT_BB3	-	-	OE	001_BB3aK_FS1_FS4	this study

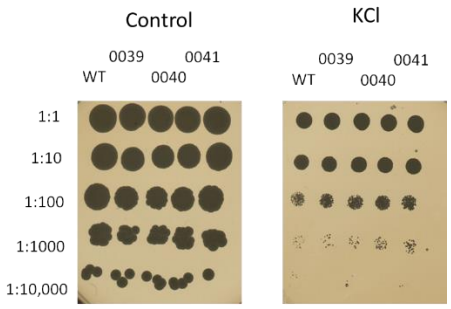
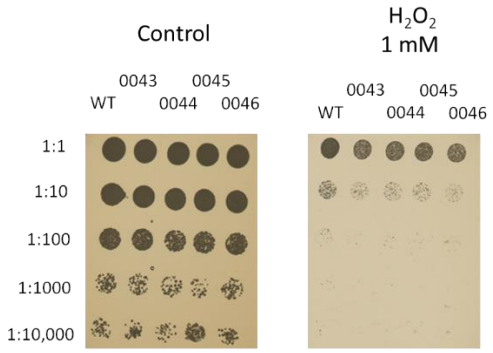
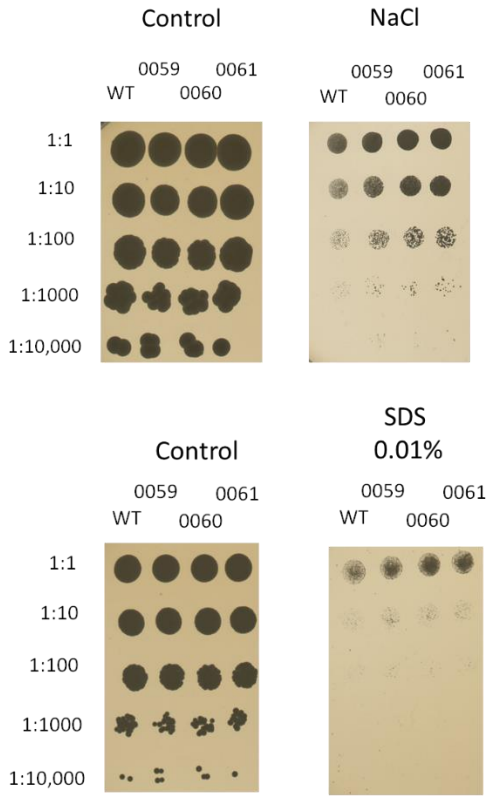
stBD_0086-0089	PAS_chr4_0077_KO	PP7435_Chr4-0940	PAS_chr4_0077	KO	plBD_0070-fgt plBD_0068	this study
stBD_0090-0092	PAS_chr3_0836_KO	PP7435_Chr3-0349	PAS_chr3_0836	KO	plBD_0074-fgt plBD_0054	this study
stBD_0093-0096	CAT8dKO	PP7435_Chr2-0516 / PP7435_Chr4-0434	CAT8-1 / CAT8-2	KO	plBD_0023-fgt plBD_0050	this study
stBD_0107-0110	CAT8-1_Hatag	PP7435_Chr2-0516	CAT8-1	TAG	plBD_0091	this study
stBD_0111-0112	CAT8-2_Hatag	PP7435_Chr4-0434	CAT8-2	TAG	plBD_0092	this study
stBD_0113-0115	pCAT8-2_eGFP_CAT8-1KO	PP7435_Chr4-0434	CAT8-2	Prom	plBD_0093	this study
stBD_0116-0118	pCAT8-2_eGFP_CAT8-2KO	PP7435_Chr4-0434	CAT8-2	Prom	plBD_0093	this study
stBD_0119-0121	pCAT8-2_eGFP_WT	PP7435_Chr4-0434	CAT8-2	Prom	plBD_0093	this study
stBD_0122-0123	pCAT8-2_eGFP_CAT8dKO	PP7435_Chr4-0434	CAT8-2	Prom	plBD_0093	this study
stBD_0125-0127	pCAT8-1_eGFP_WT	PP7435_Chr2-0516	CAT8-1	Prom	plBD_0094	this study
stBD_0128-0131	pCAT8-1_eGFP_CAT8-1KO	PP7435_Chr2-0516	CAT8-1	Prom	plBD_0094	this study
stBD_0132-0133	pCAT8-1_eGFP_CAT8-2KO	PP7435_Chr2-0516	CAT8-1	Prom	plBD_0094	this study
stBD_0134-0136	pCAT8-1_eGFP_CAT8dKO	PP7435_Chr2-0516	CAT8-1	Prom	plBD_0094	this study
stBD_0137-0147	pAOX1mutS_eGFP_wt	PP7435_Chr3-0805	AOX1	Prom	plBD_0104	this study
stBD_0148-0158	pAOX1mutS_eGFP_swi5OE	PP7435_Chr3-0805	AOX1	Prom	plBD_0104	this study
-	MIG1-1_KO	PP7435_Chr4-0661	MIG1-1	KO	-	Ata 2017
-	MIG1-1_OE	PP7435_Chr4-0661	MIG1-1	OE	-	Ata 2017
-	MIG1-2_KO	PP7435_Chr1-1325	MIG1-2	KO	-	Ata 2017
-	MIG1-2_OE	PP7435_Chr1-1325	MIG1-2	OE	-	Ata 2017

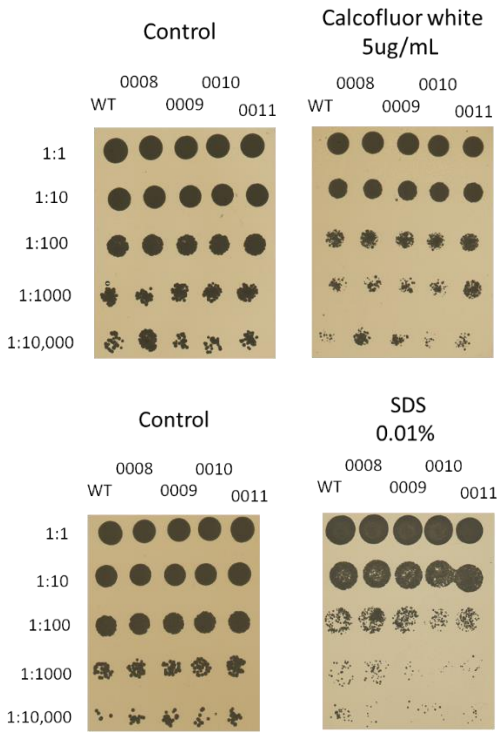
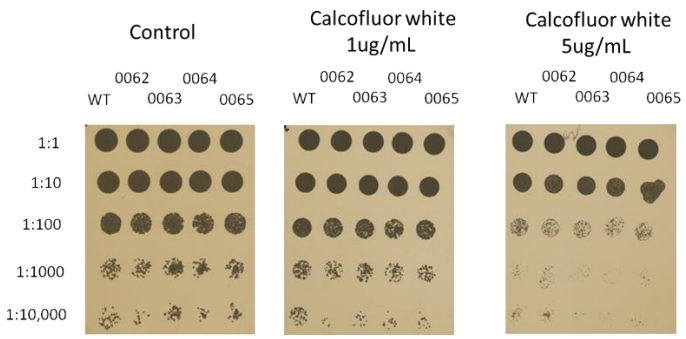
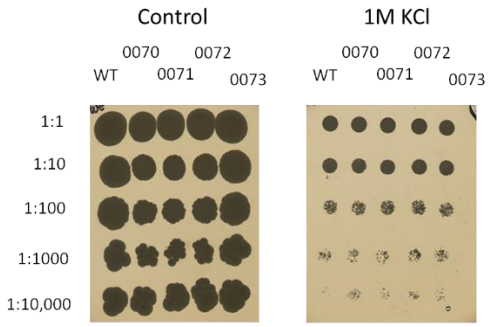
IV. Appendix 4: Results of the spotting assay

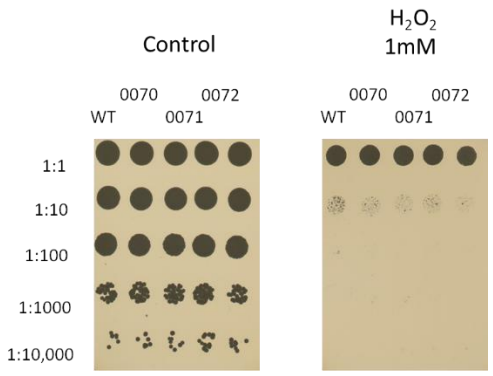
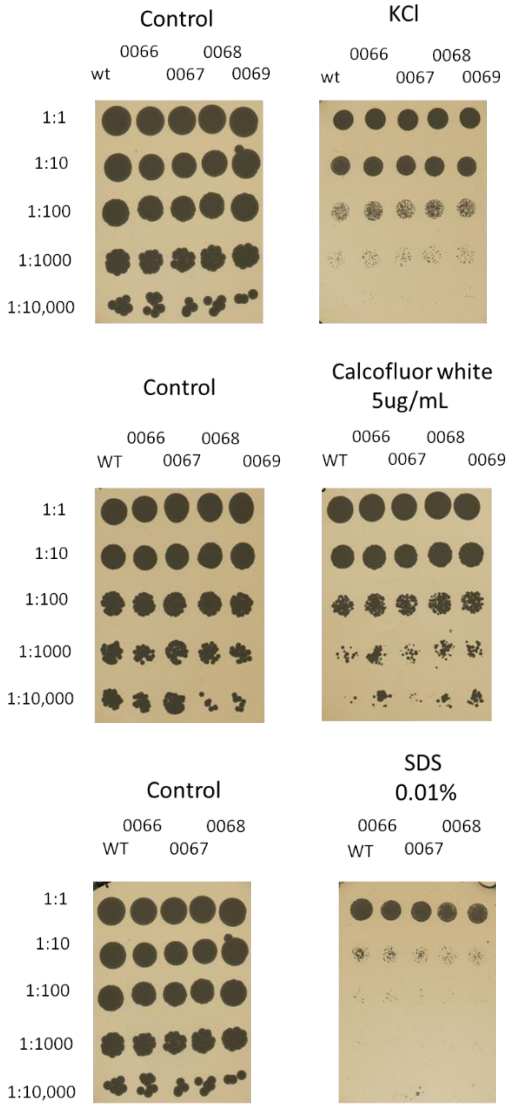
Table 14: Results of the spotting assay. The numbers 00XX indicate the strain numbers of the mutant used for a specific TF. The control plate is YNB-agar supplemented with 2% glucose. It was incubated in the same conditions than the plate testing the growth with a stressor. Only the plates where a growth difference with the WT was observed are presented.

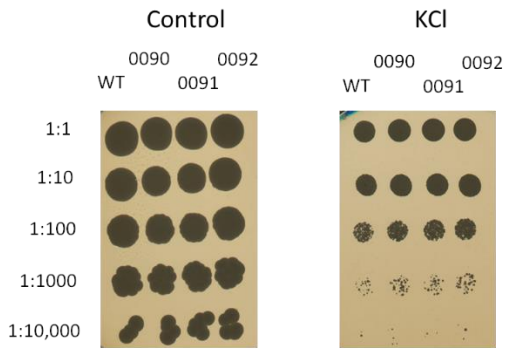
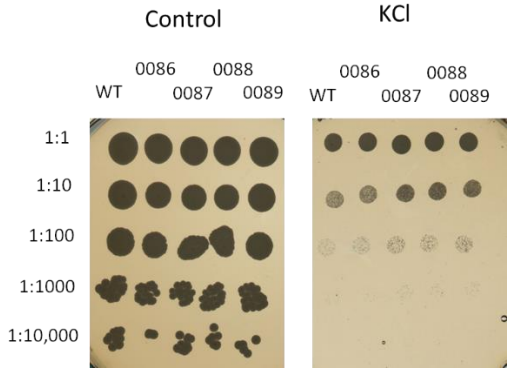
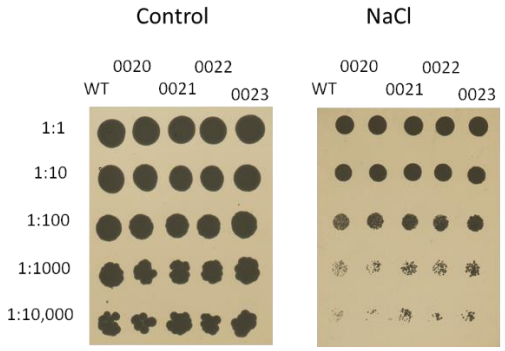
TF		Pictures
SKO1	OE	
	KO	
SWI5	OE	

	KO	
CAT8-1	KO	
AFU1	OE	

		
PP7435_Chr1-0170	OE	
YPR022C-1	OE	

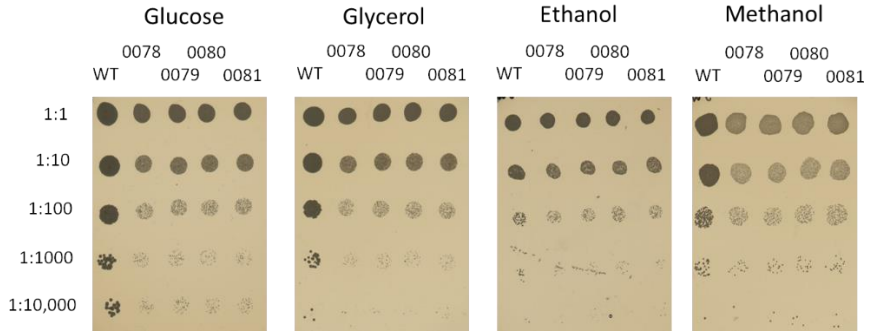
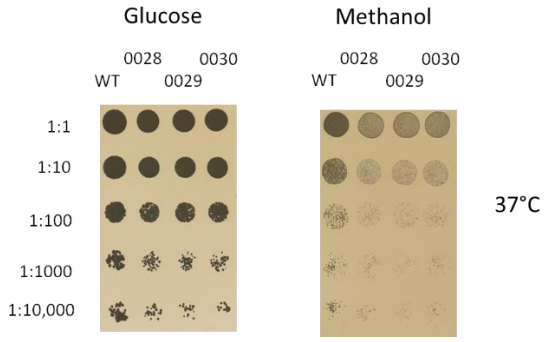
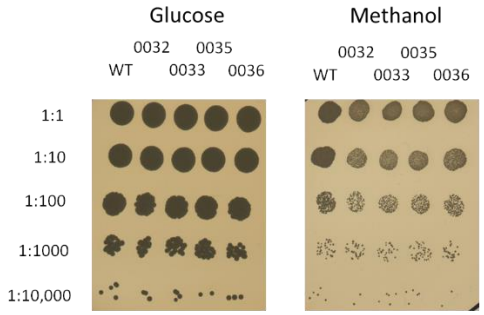
	KO	
PP7435_Chr1-0006	OE	
YGR067C	OE	

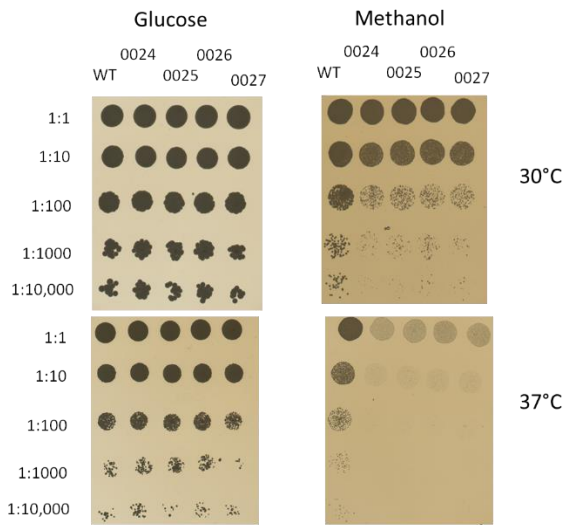
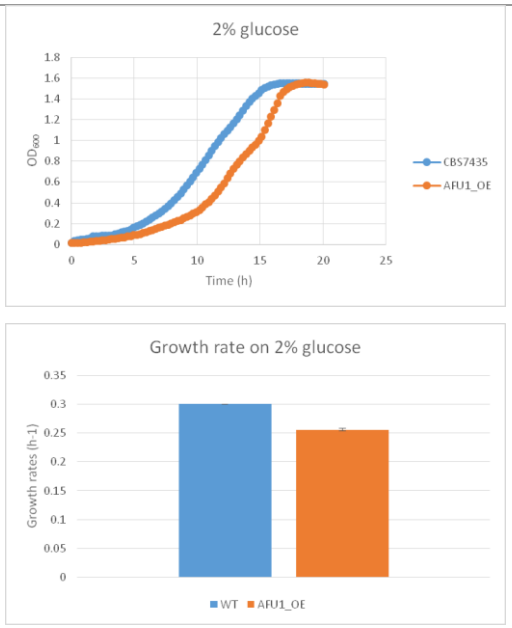
		 Control H_2O_2 1mM 0070 0072 WT 0071 0070 0072 WT 0071 1:1 1:10 1:100 1:1000 1:10,000
MIT1	OE	 Control KCl 0066 0068 wt 0067 0069 0066 0068 wt 0067 0069 1:1 1:10 1:100 1:1000 1:10,000 Control Calcofluor white 5ug/mL 0066 0068 WT 0067 0069 0066 0068 WT 0067 0069 1:1 1:10 1:100 1:1000 1:10,000 Control SDS 0.01% 0066 0068 WT 0067 0066 0068 WT 0067 1:1 1:10 1:100 1:1000 1:10,000

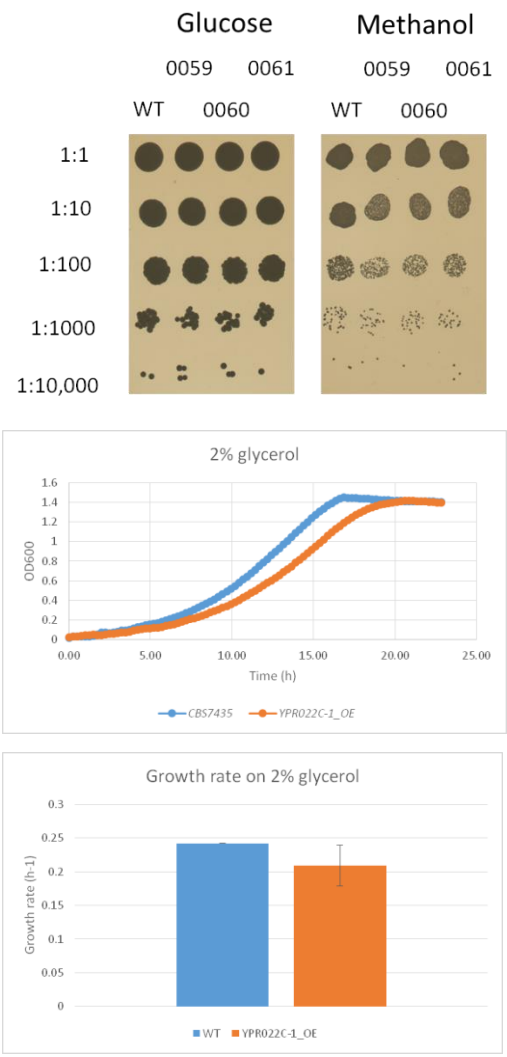
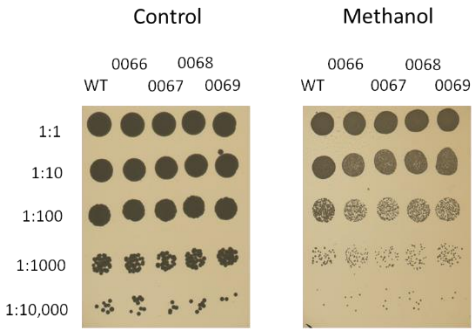
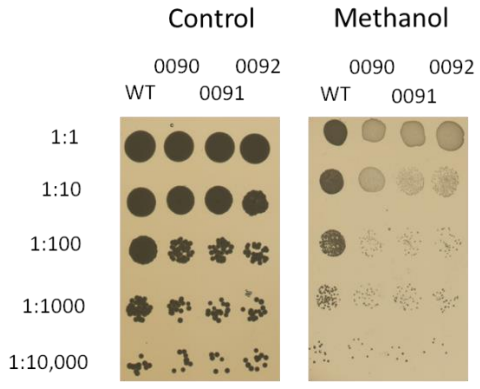
	KO	
PP7435_Chr4-0940	KO	
CAT8-2	KO	

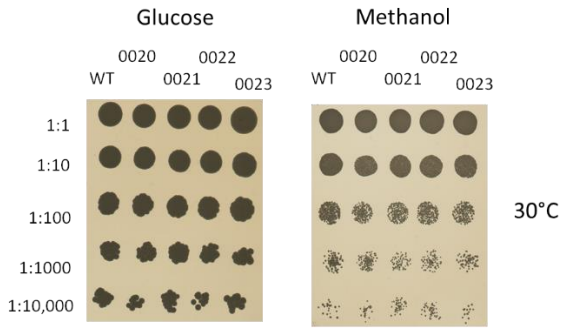
V. Appendix 5: Results of the growth of the TF mutants on different carbon sources

Table 15: Results of the growth of the TFs mutants on different carbon sources. The numbers (00XX) indicate the strain number of the TF mutants. Only the conditions where a growth difference compare to the WT is observed are shown.

TF		
SKO1	OE	
SGF29	KO	
SWI5	OE	

CAT8-1	KO	
AFU1	OE	

YPR022C-1	OE	
MIT1	OE	
	KO	

CAT8-2	KO	 Glucose Methanol WT 0020 0021 0022 0023 1:1 1:10 1:100 1:1000 1:10,000 30°C

VI. Appendix 6: List of proteins their identifiers (NCBI accession numbers) used for the phylogenetic analyses of Cat8-1 and Cat8-2

Cat8-2_Phaffomycetaceae_and_Pichiaceae_clade

XP_011276078.1_Saccharomycetales_Phaffomycetaceae_Wickerhamomyces_Wickerhamomyces_ciferrii
CEP22994.1_Saccharomycetales_Phaffomycetaceae_Cyberlindnera_Cyberlindnera_jadinii
CDR44725.1_Saccharomycetales_Phaffomycetaceae_Cyberlindnera_Cyberlindnera_fabianii
Cat8_2_XP_002493979.1_Saccharomycetales_Phaffomycetaceae_Komagataella_Komagataella_phaffii_GS115
ANZ78000.1_Saccharomycetales_Phaffomycetaceae_Komagataella_Komagataella_pastoris
OWB49142.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
OWB81660.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
ODV87249.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_arabinofermentans_NRRL_YB-2248
XP_018208536.1_Saccharomycetales_Pichiaceae_Ogataea_Ogataea_polymorpha
XP_013937395.1_Saccharomycetales_Pichiaceae_Ogataea_Ogataea_parapolymorpha_DL-1
GAV28280.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_membranifaciens
XP_019017762.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_membranifaciens_NRRL_Y-2026
KGK38021.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_kudriavzevii
OUT21345.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_kudriavzevii
AWU75108.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_kudriavzevii

Sip4_Saccharomycetaceae_and_Saccharomycodaceae_clade

XP_004180554.1_Saccharomycetales_Saccharomycetaceae_Tetrapisispora_Tetrapisispora_blattae_CBS_6284
SSD61869.1_Saccharomycetales_Saccharomycodaceae_Saccharomycodes_Saccharomycodes_ludwigii
XP_017986333.1_Saccharomycetales_Saccharomycetaceae_Eremothecium_Eremothecium_sinecaudum
XP_003647533.1_Saccharomycetales_Saccharomycetaceae_Eremothecium_Eremothecium_cymbalariae_DBVPG7215
AGO13243.1_Saccharomycetales_Saccharomycetaceae_Saccharomycetaceae_sp._Ashbya_aceri
NP_985643.2_Saccharomycetales_Saccharomycetaceae_Eremothecium_Eremothecium_gossypii_ATCC_10895
XP_022675308.1_Saccharomycetales_Saccharomycetaceae_Kluyveromyces_Kluyveromyces_marxianus_DMKU3-1042
CDO96066.1_Saccharomycetales_Saccharomycetaceae_Kluyveromyces_Kluyveromyces_dobzhanskii_CBS_2104
Sip4_CAE00852.1_Saccharomycetales_Saccharomycetaceae_Kluyveromyces_Kluyveromyces_lactis
XP_455723.1_Saccharomycetales_Saccharomycetaceae_Kluyveromyces_Kluyveromyces_lactis
SCW00936.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_fermentati
SCV00608.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_mirantina
XP_002552880.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_thermotolerans_CBS_6340
CUS20723.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_quebecensis
SCU90333.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_nothofagi_CBS_11611
SCU83125.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_dasiensis_CBS_10888
SCU78707.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_meyersii_CBS_8951
XP_022629831.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_lanzarotensis
SCU77698.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_sp._CBS_6924
XP_001643936.1_Saccharomycetales_Saccharomycetaceae_Vanderwaltozyma_Vanderwaltozyma_polyspora_DSM_70294
XP_003687217.1_Saccharomycetales_Saccharomycetaceae_Tetrapisispora_Tetrapisispora_phaffii_CBS_4417
XP_018221239.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_eubayanus
EJT44735.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_kudriavzevii_IFO_1802
EJS43144.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_arboricola_H-6
AJR71411.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae_YJM1447
KQC43050.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_sp._boulardii
AJR60093.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae_YJM195
Sip4_CAA89382.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae
AJP39617.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae_YJM1078
PTN14399.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae
EHN06472.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae_x_Saccharomyces_kudriavzevii_VIN7
XP_003980212.1_Saccharomycetales_Saccharomycetaceae_Naumovozyma_Naumovozyma_dairenensis_CBS_421
XP_003673880.1_Saccharomycetales_Saccharomycetaceae_Naumovozyma_Naumovozyma_castellii_CBS_4309
XP_022462873.1_Saccharomycetales_Saccharomycetaceae_Kazachstania_Kazachstania_naganishii_CBS_8797
SMN21725.1_Saccharomycetales_Saccharomycetaceae_Kazachstania_Kazachstania_saulgeensis
XP_003954721.1_Saccharomycetales_Saccharomycetaceae_Kazachstania_Kazachstania_africana_CBS_2517
XP_448919.1_Saccharomycetales_Saccharomycetaceae_Nakaseomyces_Candida_glabrata
SLM12920.1_Saccharomycetales_Saccharomycetaceae_Nakaseomyces_Candida_glabrata
GAV56003.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_rouxii

GAV51207.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_rouxii
CDH15222.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_bailii_ISA1307
SJM83910.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_bailii
AQZ12288.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_parabailii
CDF90410.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_bailii_CLIB_213

Cat8_Various-yeasts_and_Aspergillus_clade

Cat8_CBF88979.1_Eurotiales_Aspergillaceae_Aspergillus_Aspergillus_nidulans_FGSC_A4
ODV92727.1_Saccharomycetales_Trigonopsidaceae_Tortispora_Tortispora_caseinolytica_NRRL_Y-17796
ODQ74993.1_Saccharomycetales_Lipomycetaceae_Lipomyces_Lipomyces_starkeyi_NRRL_Y-11557
ODQ66190.1_Saccharomycetales_Nadsonia_Nadsonia_fulvescens_var._elongata_DSM_6958
RDW25834.1_Saccharomycetales_Dipodascaceae_Yarrowia_Yarrowia_lipolytica
XP_018737583.1_Saccharomycetales_Trichomonascaceae_Sugiyamaella_Sugiyamaella_lignohabitans
CDO52471.1_Saccharomycetales_Dipodascaceae_Geotrichum_Geotrichum_candidum
CDO54999.1_Saccharomycetales_Dipodascaceae_Geotrichum_Geotrichum_candidum

Cat8_Saccharomycetaceae_and_Saccharomycodaceae_clade

SSD60226.1_Saccharomycetales_Saccharomycodaceae_Saccharomycodes_Saccharomycodes_ludwigii
SGZ41833.1_Saccharomycetales_Saccharomycodaceae_Hanseniaspora_Hanseniaspora_guilliermondii
OBA25299.1_Saccharomycetales_Saccharomycodaceae_Hanseniaspora_Hanseniaspora_valbyensis_NRRL_Y-1626
XP_003675398.1_Saccharomycetales_Saccharomycetaceae_Naumovozyma_Naumovozyma_castellii_CBS_4309
XP_003672473.1_Saccharomycetales_Saccharomycetaceae_Naumovozyma_Naumovozyma_dairenensis_CBS_421
XP_022466353.1_Saccharomycetales_Saccharomycetaceae_Kazachstania_Kazachstania_naganishii_CBS_8797
XP_003955826.1_Saccharomycetales_Saccharomycetaceae_Kazachstania_Kazachstania_africana_CBS_2517
SMN22030.1_Saccharomycetales_Saccharomycetaceae_Kazachstania_Kazachstania_saulgeensis
XP_449478.1_Saccharomycetales_Saccharomycetaceae_Nakaseomyces_Candida_glabrata
XP_018220327.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_eubayanus
EHN00712.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae_x_Saccharomyces_kudriavzevii_VIN7
EJT42916.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_kudriavzevii_IFO_1802
EWH16862.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae_P283
Cat8_CAA55139.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae
AJS92705.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae_YJM1418
AJS81799.1_Saccharomycetales_Saccharomycetaceae_Saccharomyces_Saccharomyces_cerevisiae_YJM1273
XP_003679393.1_Saccharomycetales_Saccharomycetaceae_Torulaspora_Torulaspora_delbrueckii
GAV55967.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_rouxii
XP_002498603.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_rouxii
AQZ14297.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_parabailii
CDH08542.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_bailii_ISA1307
AQZ10426.1_Saccharomycetales_Saccharomycetaceae_Zygosaccharomyces_Zygosaccharomyces_parabailii
BAP73751.1_Saccharomycetales_Saccharomycetaceae_Kluyveromyces_Kluyveromyces_marxianus
Cat8_XP_453133.1_Saccharomycetales_Saccharomycetaceae_Kluyveromyces_Kluyveromyces_lactis
CDO93457.1_Saccharomycetales_Saccharomycetaceae_Kluyveromyces_Kluyveromyces_dobzhanskii_CBS_2104
XP_017987283.1_Saccharomycetales_Saccharomycetaceae_Eremothecium_Eremothecium_sinecaudum
XP_003646462.1_Saccharomycetales_Saccharomycetaceae_Eremothecium_Eremothecium_cymbalariae_DBVPG7215
NP_982826.2_Saccharomycetales_Saccharomycetaceae_Eremothecium_Eremothecium_gossypii_ATCC_10895
AGO10335.1_Saccharomycetales_Saccharomycetaceae_Saccharomycetaceae_sp._Ashbya_aceri
SCW03288.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_fermentati
SCU99863.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_mirantina
XP_002552389.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_thermotolerans_CBS_6340
CUS21419.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_quebecensis
SCV01935.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_nothofagi_CBS_11611
SCU97224.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_dasiensis_CBS_10888
SCV00984.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_meyersii_CBS_8951
XP_022627556.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_lanzarotensis
SCU89845.1_Saccharomycetales_Saccharomycetaceae_Lachancea_Lachancea_sp._CBS_6924

Cat8_Single-sequence_Debaryomycetaceae_Babjeviella_inositovora (no clade)

XP_018983956.1_Saccharomycetales_Debaryomycetaceae_Babjeviella_Babjeviella_inositovora_NRRL_Y-12698

Cat8_Phaffomycetaceae_and_Pichiaceae_clade

XP_020044763.1_Saccharomycetales_Ascoideaceae_Ascoidea_Ascoidea_rubescens_DSM_1968
XP_019041802.1_Saccharomycetales_Phaffomycetaceae_Wickerhamomyces_Wickerhamomyces_anomalus_NRRL_Y-366-8
XP_011271777.1_Saccharomycetales_Phaffomycetaceae_Wickerhamomyces_Wickerhamomyces_ciferrii
XP_020071922.1_Saccharomycetales_Phaffomycetaceae_Cyberlindnera_Cyberlindnera_jadinii_NRRL_Y-1542
CEP21463.1_Saccharomycetales_Phaffomycetaceae_Cyberlindnera_Cyberlindnera_jadinii
CDR37533.1_Saccharomycetales_Phaffomycetaceae_Cyberlindnera_Cyberlindnera_fabianii
ONH69489.1_Saccharomycetales_Phaffomycetaceae_Cyberlindnera_Cyberlindnera_fabianii
ODV94686.1_Saccharomycetales_Saccharomycetaceae_Pachysolen_Pachysolen_tannophilus_NRRL_Y-2460
Cat8_1_XP_002491690.1_Saccharomycetales_Phaffomycetaceae_Komagataella_Komagataella_phaffii_GS115
ANZ75826.1_Saccharomycetales_Phaffomycetaceae_Komagataella_Komagataella_pastoris
XP_022460283.1_Saccharomycetales_Kuraishia_Kuraishia_capsulata_CBS_1993
OWB67751.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
OWB86927.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
OUM55387.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
OWB81137.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
OWB63700.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
OWB74921.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
OWB59026.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_boidinii
ODV82578.1_Saccharomycetales_Pichiaceae_Ogataea_Candida_arabinofermentans_NRRL_YB-2248
Cat8_XP_018209149.1_Saccharomycetales_Pichiaceae_Ogataea_Ogataea_polymorpha
XP_013936787.1_Saccharomycetales_Pichiaceae_Ogataea_Ogataea_parapolymorpha_DL-1
VEU21323.1_Saccharomycetales_Pichiaceae_Brettanomyces_Brettanomyces_naardenensis
GAV29036.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_membranifaciens
XP_019016004.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_membranifaciens_NRRL_Y-2026
XP_020542872.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_kudriavzevii
AWU74705.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_kudriavzevii
OUT20191.1_Saccharomycetales_Pichiaceae_Pichia_Pichia_kudriavzevii

Cat8_Debaryomycetaceae_and_Metschnikowiaceae_clade

RLV87758.1_Saccharomycetales_Debaryomycetaceae_Meyerozyma_Meyerozyma_sp._JA9
XP_001485621.1_Saccharomycetales_Debaryomycetaceae_Meyerozyma_Meyerozyma_guilliermondii_ATCC_6260
XP_006683732.1_Saccharomycetales_Debaryomycetaceae_Yamadazyma_Yamadazyma_tenuis_ATCC_10573
XP_002617846.1_Saccharomycetales_Metschnikowiaceae_Clavispora_Clavispora_lusitaniae_ATCC_42720
RKP29934.1_Saccharomycetales_Metschnikowiaceae_Metschnikowia_Metschnikowia_bicuspidata
XP_018713241.1_Saccharomycetales_Metschnikowiaceae_Metschnikowia_Metschnikowia_bicuspidata_var._bicuspidata_NRRL_YB-4993
SGZ53145.1_Saccharomycetales_Metschnikowiaceae_Clavispora_Candida_intermedia
SGZ58541.1_Saccharomycetales_Metschnikowiaceae_Clavispora_Candida_intermedia
PSK79111.1_Saccharomycetales_Metschnikowiaceae_Clavispora_Candida_auris
XP_025341395.1_Saccharomycetales_Metschnikowiaceae_Clavispora_Candida_haemulonis
XP_024715689.1_Saccharomycetales_Metschnikowiaceae_Clavispora_Candida_pseudohaemulonis
XP_025338733.1_Saccharomycetales_Metschnikowiaceae_Clavispora_Candida_duobushaemulonis
XP_020075091.1_Saccharomycetales_Debaryomycetaceae_Hyphopichia_Hyphopichia_burtonii_NRRL_Y-1933
XP_460549.2_Saccharomycetales_Debaryomycetaceae_Debaryomyces_Debaryomyces_hansenii_CBS767
XP_015467971.1_Saccharomycetales_Debaryomycetaceae_Debaryomyces_Debaryomyces_fabryi
CCE83004.1_Saccharomycetales_Debaryomycetaceae_Millerozyma_Millerozyma_farinosa_CBS_7064
CCE82081.1_Saccharomycetales_Debaryomycetaceae_Millerozyma_Millerozyma_farinosa_CBS_7064
XP_020066064.1_Saccharomycetales_Debaryomycetaceae_Suhomyces_Suhomyces_tanzawaensis_NRRL_Y-17324
XP_001383208.2_Saccharomycetales_Debaryomycetaceae_Scheffersomyces_Scheffersomyces_stipitis_CBS_6054
RLV93997.1_Saccharomycetales_Debaryomycetaceae_Spathaspora_Spathaspora_sp._JA1
XP_001528147.1_Saccharomycetales_Debaryomycetaceae_Lodderomyces_Lodderomyces_elongisporus_NRRL_YB-4239
CCE43097.1_Saccharomycetales_Debaryomycetaceae_Candida_Candida_parapsilosis
XP_003866499.1_Saccharomycetales_Debaryomycetaceae_Candida_Candida_orthopsilosis_Co_90-125
XP_002417419.1_Saccharomycetales_Debaryomycetaceae_Candida_Candida_dubliniensis_CD36
KGR14049.1_Saccharomycetales_Debaryomycetaceae_Candida_Candida_albicans_P57072
EMG49803.1_Saccharomycetales_Debaryomycetaceae_Candida_Candida_maltosa_Xu316
XP_026596208.1_Saccharomycetales_Debaryomycetaceae_Candida_Candida_viswanathii
XP_026593335.1_Saccharomycetales_Debaryomycetaceae_Candida_Candida_viswanathii

REFERENCES

Ahn, J., Hong, J., Park, M., Lee, H., Lee, E., Kim, C., Lee, J., Choi, E. S., Jung, J. K., 2009. Phosphate-responsive promoter of a Pichia pastoris sodium phosphate symporter. Appl Environ Microbiol. 75**,** 3528-34.

Akhtar, M. S., Heidemann, M., Tietjen, J. R., Zhang, D. W., Chapman, R. D., Eick, D., Ansari, A. Z., 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol Cell. 34**,** 387-93.

Albert, I., Mavrich, T. N., Tomsho, L. P., Qi, J., Zanton, S. J., Schuster, S. C., Pugh, B. F., 2007. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. Nature. 446**,** 572-6.

Ata, O., Prielhofer, R., Gasser, B., Mattanovich, D., Calik, P., 2017. Transcriptional engineering of the glyceraldehyde-3-phosphate dehydrogenase promoter for improved heterologous protein production in *Pichia pastoris*. Biotechnol Bioeng.

Ata, Ö., Rebnegger, C., Tatto, N. E., Valli, M., Mairinger, T., Hann, S., Steiger, M. G., Çalık, P., Mattanovich, D., 2018. A single Gal4-like transcription factor activates the Crabtree effect in *Komagataella phaffii*. Nat Commun. 9**,** 4911.

Bai, L., Morozov, A. V., 2010. Gene regulation by nucleosome positioning. Trends Genet. 26**,** 476-83.

Barnett, J. A., Entian, K. D., 2005. A history of research on yeasts 9: regulation of sugar metabolism. Yeast. 22**,** 835-94.

Basehoar, A. D., Zanton, S. J., Pugh, B. F., 2004. Identification and distinct regulation of yeast TATA box-containing genes. Cell. 116**,** 699-709.

Berg, L., Strand, T. A., Valla, S., Brautaset, T., 2013. Combinatorial mutagenesis and selection to understand and improve yeast promoters. Biomed Res Int. 2013**,** 926985.

Bernstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O., Schreiber, S. L., 2004. Global nucleosome occupancy in yeast. Genome Biol. 5**,** R62.

Bhoite, L. T., Yu, Y., Stillman, D. J., 2001. The Swi5 activator recruits the Mediator complex to the *HO* promoter without RNA polymerase II. Genes Dev. 15**,** 2457-69.

Bian, C., Xu, C., Ruan, J., Lee, K. K., Burke, T. L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B. O., Fleharty, B. E., Paulson, A., Allali-Hassani, A., Zhou, J.-Q., Mer, G., Grant, P. A., Workman, J. L., Zang, J., Min, J., 2011. Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. EMBO J. 30**,** 2829-2842.

Biddie, S. C., John, S., Sabo, P. J., Thurman, R. E., Johnson, T. A., Schiltz, R. L., Miranda, T. B., Sung, M. H., Trump, S., Lightman, S. L., Vinson, C., Stamatoyannopoulos, J. A., Hager, G. L., 2011. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Mol Cell. 43**,** 145-55.

Bird, A. J., Zhao, H., Luo, H., Jensen, L. T., Srinivasan, C., Evans-Galea, M., Winge, D. R., Eide, D. J., 2000. A dual role for zinc fingers in both DNA binding and zinc sensing by the Zap1 transcriptional activator. EMBO J. 19**,** 3704-13.

Braun, K. A., Parua, P. K., Dombek, K. M., Miner, G. E., Young, E. T., 2013. 14-3-3 (Bmh) proteins regulate combinatorial transcription following RNA polymerase II recruitment by binding at Adr1-dependent promoters in *Saccharomyces cerevisiae*. Mol Cell Biol. 33**,** 712-24.

Bricmont, P. A., Daugherty, J. R., Cooper, T. G., 1991. The DAL81 gene product is required for induced expression of two differently regulated nitrogen catabolic genes in *Saccharomyces cerevisiae*. Mol Cell Biol. 11**,** 1161-6.

Buck, M. J., Lieb, J. D., 2006. A chromatin-mediated mechanism for specification of conditional transcription factor targets. Nat Genet. 38**,** 1446-51.

Buratowski, S., 2009. Progression through the RNA polymerase II CTD cycle. Mol Cell. 36**,** 541-6.

Burgard, J., Valli, M., Graf, A. B., Gasser, B., Mattanovich, D., 2017. Biomarkers allow detection of nutrient limitations and respective supplementation for elimination in *Pichia pastoris* fed-batch cultures. Microb Cell Fact. 16**,** 117.

Böhm, S., Frishman, D., Mewes, H. W., 1997. Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. Nucleic Acids Res. 25**,** 2464-9.

Carey, M., Smale, S. T., 2001. Transcriptional Regulation in Eucaryotes. CSHL Press.

Carlson, M., 1999. Glucose repression in yeast. Curr Opin Microbiol. 2**,** 202-7.

Carnicer, M., Ten Pierick, A., van Dam, J., Heijnen, J. J., Albiol, J., van Gulik, W., Ferrer, P., 2012. Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions. Microb Cell Fact. 11**,** 83.

Charbon, G., Breunig, K. D., Wattiez, R., Vandenhaute, J., Noël-Georis, I., 2004. Key role of Ser562/661 in Snf1-dependent regulation of Cat8p in *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. Mol Cell Biol. 24**,** 4083-91.

Choudhury, B. I., Whiteway, M., 2018. Evolutionary Transition of GAL Regulatory Circuit from Generalist to Specialist Function in Ascomycetes. Trends Microbiol. 26**,** 692-702.

Core, L. J., Waterfall, J. J., Lis, J. T., 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 322**,** 1845-8.

Correa, A., Oppezzo, P., 2011. Tuning different expression parameters to achieve soluble recombinant proteins in *E. coli*: advantages of high-throughput screening. Biotechnol J. 6**,** 715-30.

Cramer, P., Armache, K. J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C. D., Lehmann, E., Leike, K., Sydow, J. F., Vannini, A., 2008. Structure of eukaryotic RNA polymerases. Annu Rev Biophys. 37**,** 337-52.

Cregg, J. M., Cereghino, J. L., Shi, J., Higgins, D. R., 2000. Recombinant protein expression in *Pichia pastoris*. Mol Biotechnol. 16**,** 23-52.

Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G. P., Stillman, C. A., 1989. Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol Cell Biol. 9**,** 1316-23.

Dasgupta, A., Darst, R. P., Martin, K. J., Afshari, C. A., Auble, D. T., 2002. Mot1 activates and represses transcription by direct, ATPase-dependent mechanisms. Proc Natl Acad Sci U S A. 99**,** 2666-71.

de Ruijter, J. C., Koskela, E. V., Frey, A. D., 2016. Enhancing antibody folding and secretion by tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum. Microb Cell Fact. 15**,** 87.

De Vit, M. J., Waddle, J. A., Johnston, M., 1997. Regulated nuclear translocation of the Mig1 glucose repressor. Mol Biol Cell. 8**,** 1603-18.

Delic, M., Gongrich, R., Mattanovich, D., Gasser, B., 2014a. Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins. Antioxid Redox Signal. 21**,** 414-37.

Delic, M., Graf, A. B., Koellensperger, G., Haberhauer-Troyer, C., Hann, S., Mattanovich, D., Gasser, B., 2014b. Overexpression of the transcription factor Yap1 modifies intracellular redox conditions and enhances recombinant protein secretion. Microb Cell. 1**,** 376-386.

Delic, M., Mattanovich, D., Gasser, B., 2013. Repressible promoters - A novel tool to generate conditional mutants in *Pichia pastoris*. Microb Cell Fact. vol. 12, England, pp. 6.

Denis, C. L., Young, E. T., 1983. Isolation and characterization of the positive regulatory gene ADR1 from *Saccharomyces cerevisiae*. Mol Cell Biol. 3**,** 360-70.

Dohrmann, P. R., Butler, G., Tamai, K., Dorland, S., Greene, J. R., Thiele, D. J., Stillman, D. J., 1992. Parallel pathways of gene regulation: homologous regulators SWI5 and ACE2 differentially control transcription of *HO* and chitinase. Genes Dev. 6**,** 93-104.

Dohrmann, P. R., Voth, W. P., Stillman, D. J., 1996. Role of negative regulation in promoter specificity of the homologous transcriptional activators Ace2p and Swi5p. Mol Cell Biol. 16**,** 1746-58.

Engler, C., Kandzia, R., Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS One. 3**,** e3647.

Escalante, C. R., Nistal-Villán, E., Shen, L., García-Sastre, A., Aggarwal, A. K., 2007. Structure of IRF-3 bound to the PRDIII-I regulatory element of the human interferon-beta enhancer. Mol Cell. 26**,** 703-16.

Falvo, J. V., Thanos, D., Maniatis, T., 1995. Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y). Cell. 83**,** 1101-11.

Feaver, W. J., Svejstrup, J. Q., Henry, N. L., Kornberg, R. D., 1994. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell. 79**,** 1103-9.

Gao, F., Foat, B. C., Bussemaker, H. J., 2004. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. BMC Bioinformatics. 5**,** 31.

Gao, Y., Zhao, Y., 2014. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J Integr Plant Biol. 56**,** 343-9.

Gasser, B., Sauer, M., Maurer, M., Stadlmayr, G., Mattanovich, D., 2007. Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. Appl Environ Microbiol. 73**,** 6499-507.

Gassler, T., Heistinger, L., Mattanovich, D., Gasser, B., Prielhofer, R., 2018. CRISPR/Cas9-mediated homology directed genome editing in P*ichia pastoris*. Meth Mol Biol.

Gaston, K., Jayaraman, P. S., 2003. Transcriptional repression in eukaryotes: repressors and repression mechanisms. Cellular and Molecular Life Sciences (CMLS). 60**,** 721-741.

Georis, I., Krijger, J. J., Breunig, K. D., Vandenhaute, J., 2000. Differences in regulation of yeast gluconeogenesis revealed by Cat8p-independent activation of *PCK1* and *FBP1* genes in *Kluyveromyces lactis*. Mol Gen Genet. 264**,** 193-203.

Green, M. R., 2005. Eukaryotic transcription activation: right on target. Mol Cell. 18**,** 399-402.

Gu, L., Zhang, J., Liu, B., Du, G., Chen, J., 2015. High-level extracellular production of glucose oxidase by recombinant *Pichia pastoris* using a combined strategy. Appl Biochem Biotechnol. 175**,** 1429-47.

Guarente, L., Hoar, E., 1984. Upstream activation sites of the *CYC1* gene of *Saccharomyces cerevisiae* are active when inverted but not when placed downstream of the "TATA box". Proc Natl Acad Sci U S A. 81**,** 7860-4.

Guerfal, M., Ryckaert, S., Jacobs, P. P., Ameloot, P., Van Craenenbroeck, K., Derycke, R., Callewaert, N., 2010. The HAC1 gene from Pichia pastoris: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. Microb Cell Fact. 9**,** 49.

Gutiérrez-González, M., Latorre, Y., Zúñiga, R., Aguillón, J. C., Molina, M. C., Altamirano, C., 2019. Transcription factor engineering in CHO cells for recombinant protein production. Crit Rev Biotechnol. 39**,** 665-679.

Hahn, S., 1998. Activation and the role of reinitiation in the control of transcription by RNA polymerase II. Cold Spring Harb Symp Quant Biol. 63**,** 181-8.

Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol. 11**,** 394-403.

Hahn, S., Maurer, P., Caesar, S., Schlenstedt, G., 2008. Classical NLS proteins from Saccharomyces cerevisiae. J Mol Biol. 379**,** 678-94.

Hamilton, S. R., Bobrowicz, P., Bobrowicz, B., Davidson, R. C., Li, H., Mitchell, T., Nett, J. H., Rausch, S., Stadheim, T. A., Wischnewski, H., Wildt, S., Gerngross, T. U., 2003. Production of complex human glycoproteins in yeast. Science. 301**,** 1244-6.

Hamilton, S. R., Cook, W. J., Gomathinayagam, S., Burnina, I., Bukowski, J., Hopkins, D., Schwartz, S., Du, M., Sharkey, N. J., Bobrowicz, P., Wildt, S., Li, H., Stadheim, T. A., Nett, J. H., 2013. Production of sialylated O-linked glycans in *Pichia pastoris*. Glycobiology. 23**,** 1192-203.

Hampsey, M., 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev. 62**,** 465-503.

Hardie, D. G., Carling, D., Carlson, M., 1998. The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell? Annu Rev Biochem. 67**,** 821-55.

Hartner, F. S., Ruth, C., Langenegger, D., Johnson, S. N., Hyka, P., Lin-Cereghino, G. P., Lin-Cereghino, J., Kovar, K., Cregg, J. M., Glieder, A., 2008. Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. Nucleic Acids Res. 36**,** e76.

Haurie, V., Perrot, M., Mini, T., Jenö, P., Sagliocco, F., Boucherie, H., 2001. The transcriptional activator Cat8p provides a major contribution to the reprogramming of carbon metabolism during the diauxic shift in *Saccharomyces cerevisiae*. J Biol Chem. 276**,** 76-85.

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M., Grunstein, M., 1995. Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell. 80**,** 583-92.

Hedges, D., Proft, M., Entian, K. D., 1995. CAT8, a new zinc cluster-encoding gene necessary for derepression of gluconeogenic enzymes in the yeast *Saccharomyces cerevisiae*. Mol Cell Biol. 15**,** 1915-22.

Hengartner, C. J., Myer, V. E., Liao, S. M., Wilson, C. J., Koh, S. S., Young, R. A., 1998. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Mol Cell. 2**,** 43-53.

Herschbach, B. M., Johnson, A. D., 1993. Transcriptional repression in eukaryotes. Annu Rev Cell Biol. 9**,** 479-509.

Hiesinger, M., Roth, S., Meissner, E., Schuller, H. J., 2001. Contribution of Cat8 and Sip4 to the transcriptional activation of yeast gluconeogenic genes by carbon source-responsive elements. Curr Genet. 39**,** 68-76.

Hinnebusch, A. G., 1997. Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. J Biol Chem. 272**,** 21661-4.

Hogues, H., Lavoie, H., Sellam, A., Mangos, M., Roemer, T., Purisima, E., Nantel, A., Whiteway, M., 2008. Transcription Factor Substitution during the Evolution of Fungal Ribosome Regulation. Mol Cell. 29**,** 552-562.

Huang, M., Bao, J., Hallström, B. M., Petranovic, D., Nielsen, J., 2017. Efficient protein production by yeast requires global tuning of metabolism. Nat Commun. 8**,** 1131.

Imai, S., Armstrong, C. M., Kaeberlein, M., Guarente, L., 2000. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. Nature. 403**,** 795-800.

Jindrich, K., Degnan, B. M., 2016. The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity. BMC Evol Biol. 16**,** 28.

John, S., Sabo, P. J., Johnson, T. A., Sung, M. H., Biddie, S. C., Lightman, S. L., Voss, T. C., Davis, S. R., Meltzer, P. S., Stamatoyannopoulos, J. A., Hager, G. L., 2008. Interaction of the glucocorticoid receptor with the chromatin landscape. Mol Cell. 29**,** 611-24.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E. H., Birney, E., Furlong, E. E., 2012. A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell. 148**,** 473-86.

Kranthi, B. V., Kumar, H. R., Rangarajan, P. N., 2010. Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast *Pichia pastoris*. Yeast. 27**,** 705-11.

Kranthi, B. V., Kumar, R., Kumar, N. V., Rao, D. N., Rangarajan, P. N., 2009. Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in *Pichia pastoris*. Biochim Biophys Acta. 1789**,** 460-8.

Krishna, S. S., Majumdar, I., Grishin, N. V., 2003. Structural classification of zinc fingers: survey and summary. Nucleic Acids Res. 31**,** 532-50.

Kumar, N. V., Rangarajan, P. N., 2011. Catabolite repression of phosphoenolpyruvate carboxykinase by a zinc finger protein under biotin- and pyruvate carboxylase-deficient conditions in *Pichia pastoris*. Microbiology. 157**,** 3361-9.

Kumar, N. V., Rangarajan, P. N., 2012. The zinc finger proteins Mxr1p and repressor of phosphoenolpyruvate carboxykinase (ROP) have the same DNA binding specificity but regulate methanol metabolism antagonistically in *Pichia pastoris*. J Biol Chem. 287**,** 34465-73.

Kuranda, M. J., Robbins, P. W., 1991. Chitinase is required for cell separation during growth of *Saccharomyces cerevisiae*. J Biol Chem. 266**,** 19758-67.

Kurtzman, C. P., 2005. Description of *Komagataella phaffii* sp. nov. and the transfer of *Pichia pseudopastoris* to the methylotrophic yeast genus *Komagataella*. Int J Syst Evol Microbiol. 55**,** 973-6.

Kurtzman, C. P., 2009. Biotechnological strains of *Komagataella (Pichia) pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. J Ind Microbiol Biotechnol. 36**,** 1435-8.

Landes, N., Gasser, B., Vorauer-Uhl, K., Lhota, G., Mattanovich, D., Maurer, M., 2016. The vitamin-sensitive promoter PTHI11 enables pre-defined autonomous induction of recombinant protein production in Pichia pastoris. Biotechnol Bioeng.

Larsen, S., Weaver, J., de Sa Campos, K., Bulahan, R., Nguyen, J., Grove, H., Huang, A., Low, L., Tran, N., Gomez, S., Yau, J., Ilustrisimo, T., Kawilarang, J., Lau, J., Tranphung, M., Chen, I., Tran, C., Fox, M., Lin-Cereghino, J., Lin-Cereghino, G. P., 2013. Mutant strains of *Pichia pastoris* with enhanced secretion of recombinant proteins. Biotechnol Lett. 35**,** 1925-35.

Latchman, D. S., 1997. Transcription factors: an overview. Int J Biochem Cell Biol. 29**,** 1305-12.

Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., Whiteway, M., 2010. Evolutionary tinkering with conserved components of a transcriptional regulatory network. PLoS Biol. 8**,** e1000329.

Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D., Lieb, J. D., 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat Genet. 36**,** 900-5.

Lee, H. C., Zhang, L., 2009. A unique mechanism of chaperone action: heme regulation of Hap1 activity involves separate control of repression and activation. Protein Pept Lett. 16**,** 642-9.

Lee, T. A., Jorgensen, P., Bognar, A. L., Peyraud, C., Thomas, D., Tyers, M., 2010. Dissection of combinatorial control by the Met4 transcriptional complex. Mol Biol Cell. 21**,** 456-69.

Lesage, P., Yang, X., Carlson, M., 1996. Yeast SNF1 protein kinase interacts with SIP4, a C6 zinc cluster transcriptional activator: a new role for SNF1 in the glucose response. Mol Cell Biol. 16**,** 1921-8.

Lin-Cereghino, G. P., Godfrey, L., de la Cruz, B. J., Johnson, S., Khuongsathiene, S., Tolstorukov, I., Yan, M., Lin-Cereghino, J., Veenhuis, M., Subramani, S., Cregg, J. M., 2006. Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in *Pichia pastoris*. Mol Cell Biol. 26**,** 883-97.

Liu, L., Zhang, Y., Liu, Z., Petranovic, D., Nielsen, J., 2015. Improving heterologous protein secretion at aerobic conditions by activating hypoxia-induced genes in *Saccharomyces cerevisiae*. FEMS Yeast Res. 15.

Liu, S. H., Chou, W. I., Sheu, C. C., Chang, M. D., 2005. Improved secretory production of glucoamylase in *Pichia pastoris* by combination of genetic manipulations. Biochem Biophys Res Commun. 326**,** 817-24.

Liu, X., Bushnell, D. A., Kornberg, R. D., 2013. RNA polymerase II transcription: structure and mechanism. Biochim Biophys Acta. 1829**,** 2-8.

Liu, X., Lee, C. K., Granek, J. A., Clarke, N. D., Lieb, J. D., 2006. Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. Genome Res. 16**,** 1517-28.

Lubliner, S., Keren, L., Segal, E., 2013. Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. Nucleic Acids Res. 41**,** 5569-81.

Maccani, A., Landes, N., Stadlmayr, G., Maresch, D., Leitner, C., Maurer, M., Gasser, B., Ernst, W., Kunert, R., Mattanovich, D., 2014. *Pichia pastoris* secretes recombinant proteins less efficiently than Chinese hamster ovary cells but allows higher space-time yields for less complex proteins. Biotechnol J. 9**,** 526-37.

MacPherson, S., Larochelle, M., Turcotte, B., 2006. A fungal family of transcriptional regulators: the zinc cluster proteins. Microbiol Mol Biol Rev. 70**,** 583-604.

MacQuarrie, K. L., Fong, A. P., Morse, R. H., Tapscott, S. J., 2011. Genome-wide transcription factor binding: beyond direct target regulation. Trends Genet. 27**,** 141-8.

Magnani, L., Eeckhoute, J., Lupien, M., 2011. Pioneer factors: directing transcriptional regulators within the chromatin environment. Trends Genet. 27**,** 465-74.

Marx, H., Sauer, M., Resina, D., Vai, M., Porro, D., Valero, F., Ferrer, P., Mattanovich, D., 2006. Cloning, disruption and protein secretory phenotype of the GAS1 homologue of *Pichia pastoris*. FEMS Microbiol Lett. 264**,** 40-7.

McBride, H. J., Yu, Y., Stillman, D. J., 1999. Distinct regions of the Swi5 and Ace2 transcription factors are required for specific gene activation. J Biol Chem. 274**,** 21029-36.

Mehlgarten, C., Krijger, J. J., Lemnian, I., Gohr, A., Kasper, L., Diesing, A. K., Grosse, I., Breunig, K. D., 2015. Divergent Evolution of the Transcriptional Network Controlled by Snf1-Interacting Protein Sip4 in Budding Yeasts. PLoS One. 10**,** e0139464.

Mellor, J., Dobson, M. J., Kingsman, A. J., Kingsman, S. M., 1987. A transcriptional activator is located in the coding region of the yeast PGK gene. Nucleic Acids Res. 15**,** 6243-59.

Menendez, J., Valdes, I., Cabrera, N., 2003. The ICL1 gene of *Pichia pastoris*, transcriptional regulation and use of its promoter. Yeast. 20**,** 1097-108.

Menoret, D., Santolini, M., Fernandes, I., Spokony, R., Zanet, J., Gonzalez, I., Latapie, Y., Ferrer, P., Rouault, H., White, K. P., Besse, P., Hakim, V., Aerts, S., Payre, F., Plaza, S., 2013. Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. Genome Biol. 14**,** R86.

Merika, M., Williams, A. J., Chen, G., Collins, T., Thanos, D., 1998. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Mol Cell. 1**,** 277-87.

Miele, R., Barra, D., Bonaccorsi di Patti, M. C., 2007. A GATA-type transcription factor regulates expression of the high-affinity iron uptake system in the methylotrophic yeast *Pichia pastoris*. Arch Biochem Biophys. 465**,** 172-9.

Miller, J. A., Widom, J., 2003. Collaborative competition mechanism for gene activation in vivo. Mol Cell Biol. 23**,** 1623-32.

Morse, R. H., 1993. Nucleosome disruption by transcription factor binding in yeast. Science. 262**,** 1563-6.

Moyle-Heyrman, G., Viswanathan, R., Widom, J., Auble, D. T., 2012. Two-step mechanism for modifier of transcription 1 (Mot1) enzyme-catalyzed displacement of TATA-binding protein (TBP) from DNA. J Biol Chem. 287**,** 9002-12.

Nett, J. H., Cook, W. J., Chen, M. T., Davidson, R. C., Bobrowicz, P., Kett, W., Brevnova, E., Potgieter, T. I., Mellon, M. T., Prinz, B., Choi, B. K., Zha, D., Burnina, I., Bukowski, J. T., Du, M., Wildt, S., Hamilton, S. R., 2013. Characterization of the *Pichia pastoris* protein-O-mannosyltransferase gene family. PLoS One. 8**,** e68325.

Ogbourne, S., Antalis, T. M., 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem J. 331 (Pt 1)**,** 1-14.

Palmieri, L., Lasorsa, F. M., De Palma, A., Palmieri, F., Runswick, M. J., Walker, J. E., 1997. Identification of the yeast ACR1 gene product as a succinate-fumarate transporter essential for growth on ethanol or acetate. FEBS Lett. 417**,** 114-8.

Pan, X., Heitman, J., 2000. Sok2 regulates yeast pseudohyphal differentiation via a transcription factor cascade that regulates cell-cell adhesion. Mol Cell Biol. 20**,** 8364-72.

Parua, P. K., Dombek, K. M., Young, E. T., 2014. Yeast 14-3-3 protein functions as a comodulator of transcription by inhibiting coactivator functions. J Biol Chem. 289**,** 35542-60.

Pascual-Ahuir, A., Posas, F., Serrano, R., Proft, M., 2001. Multiple levels of control regulate the yeast cAMP-response element-binding protein repressor Sko1p in response to stress. J Biol Chem. 276**,** 37373-8.

Phaff, H. J., 1956. A proposal for amendment of the diagnosis of the genus *Pichia hansen*. Antonie Van Leeuwenhoek. 22**,** 113-6.

Piña, B., Fernández-Larrea, J., García-Reyero, N., Idrissi, F. Z., 2003. The different (sur)faces of Rap1p. Mol Genet Genomics. 268**,** 791-8.

Polish, J. A., Kim, J. H., Johnston, M., 2005. How the Rgt1 transcription factor of Saccharomyces cerevisiae is regulated by glucose. Genetics. 169**,** 583-94.

Poopanitpan, N., Kobayashi, S., Fukuda, R., Horiuchi, H., Ohta, A., 2010. An ortholog of farA of *Aspergillus nidulans* is implicated in the transcriptional activation of genes involved in fatty acid utilization in the yeast *Yarrowia lipolytica*. Biochem Biophys Res Commun. 402**,** 731-735.

Prielhofer, R., Barrero, J. J., Steuer, S., Gassler, T., Zahrl, R., Baumann, K., Sauer, M., Mattanovich, D., Gasser, B., Marx, H., 2017. GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast *Pichia pastoris*. BMC Syst Biol. 11**,** 123.

Prielhofer, R., Cartwright, S. P., Graf, A. B., Valli, M., Bill, R. M., Mattanovich, D., Gasser, B., 2015. *Pichia pastoris* regulates its gene-specific response to different carbon sources at the transcriptional, rather than the translational, level. BMC Genomics. 16**,** 167.

Prielhofer, R., Maurer, M., Klein, J., Wenger, J., Kiziak, C., Gasser, B., Mattanovich, D., 2013. Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. Microb Cell Fact. 12**,** 5.

Proft, M., Gibbons, F. D., Copeland, M., Roth, F. P., Struhl, K., 2005. Genomewide identification of Sko1 target promoters reveals a regulatory network that operates in response to osmotic stress in *Saccharomyces cerevisiae.* Eukaryot Cell. 4**,** 1343-52.

Proft, M., Pascual-Ahuir, A., de Nadal, E., Ariño, J., Serrano, R., Posas, F., 2001. Regulation of the Sko1 transcriptional repressor by the Hog1 MAP kinase in response to osmotic stress. EMBO J. 20**,** 1123-33.

Proft, M., Serrano, R., 1999. Repressors and upstream repressing sequences of the stress-regulated ENA1 gene in *Saccharomyces cerevisiae:* bZIP protein Sko1p confers HOG-dependent osmotic regulation. Mol Cell Biol. 19**,** 537-46.

Proft, M., Struhl, K., 2002. Hog1 kinase converts the Sko1-Cyc8-Tup1 repressor complex into an activator that recruits SAGA and SWI/SNF in response to osmotic stress. Mol Cell. 9**,** 1307-17.

Puxbaum, V., Mattanovich, D., Gasser, B., 2015. Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. Appl Microbiol Biotechnol. 99**,** 2925-38.

Qi, K., Zhong, J. J., Xia, X. X., 2014. Triggering respirofermentative metabolism in the crabtree-negative yeast *Pichia guilliermondii* by disrupting the CAT8 gene. Appl Environ Microbiol. 80**,** 3879-87.

Qin, X., Qian, J., Yao, G., Zhuang, Y., Zhang, S., Chu, J., 2011. GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl Environ Microbiol. 77**,** 3600-8.

Rahner, A., Schöler, A., Martens, E., Gollwitzer, B., Schüller, H. J., 1996. Dual influence of the yeast Cat1p (Snf1p) protein kinase on carbon source-dependent transcriptional activation of gluconeogenic genes by the regulatory gene CAT8. Nucleic Acids Res. 24**,** 2331-7.

Ramirez, M. A., Lorenz, M. C., 2009. The transcription factor homolog CTF1 regulates {beta}-oxidation in *Candida albicans*. Eukaryot Cell. 8**,** 1604-14.

Randez-Gil, F., Bojunga, N., Proft, M., Entian, K. D., 1997. Glucose derepression of gluconeogenic enzymes in *Saccharomyces cerevisiae* correlates with phosphorylation of the gene activator Cat8p. Mol Cell Biol. 17**,** 2502-10.

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., Segal, E., 2012. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat Genet. 44**,** 743-50.

Rebnegger, C., Graf, A. B., Valli, M., Steiger, M. G., Gasser, B., Maurer, M., Mattanovich, D., 2014. In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. Biotechnol J. 9**,** 511-25.

Reinke, A. W., Baek, J., Ashenberg, O., Keating, A. E., 2013. Networks of bZIP protein-protein interactions diversified over a billion years of evolution. Science. 340**,** 730-4.

Rep, M., Proft, M., Remize, F., Tamás, M., Serrano, R., Thevelein, J. M., Hohmann, S., 2001. The *Saccharomyces cerevisiae* Sko1p transcription factor mediates HOG pathway-dependent osmotic regulation of a set of genes encoding enzymes implicated in protection from oxidative damage. Mol Microbiol. 40**,** 1067-83.

Robinson, K. A., Lopes, J. M., 2000. SURVEY AND SUMMARY: *Saccharomyces cerevisiae* basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Res. 28**,** 1499-505.

Rodicio, R., López, M. L., Cuadrado, S., Cid, A. F., Redruello, B., Moreno, F., Heinisch, J. J., Hegewald, A. K., Breunig, K. D., 2008. Differential control of isocitrate lyase gene transcription by non-fermentable carbon sources in the milk yeast *Kluyveromyces lactis*. FEBS Lett. 582**,** 549-57.

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., Mann, R. S., 2010. Origins of specificity in protein-DNA recognition. Annu Rev Biochem. 79**,** 233-69.

Roth, S., Kumme, J., Schüller, H. J., 2004. Transcriptional activators Cat8 and Sip4 discriminate between sequence variants of the carbon source-responsive promoter element in the yeast *Saccharomyces cerevisiae*. Curr Genet. 45**,** 121-8.

Ruchala, J., Kurylenko, O. O., Soontorngun, N., Dmytruk, K. V., Sibirny, A. A., 2017. Transcriptional activator Cat8 is involved in regulation of xylose alcoholic fermentation in the thermotolerant yeast *Ogataea (Hansenula) polymorpha*. Microb Cell Fact. 16**,** 36.

Russmayer, H., Buchetics, M., Gruber, C., Valli, M., Grillitsch, K., Modarres, G., Guerrasio, R., Klavins, K., Neubauer, S., Drexler, H., Steiger, M., Troyer, C., Al Chalabi, A., Krebiehl, G., Sonntag, D., Zellnig, G., Daum, G., Graf, A. B., Altmann, F., Koellensperger, G., Hann, S., Sauer, M., Mattanovich, D., Gasser, B., 2015. Systems-level organization of yeast methylotrophic lifestyle. BMC Biol. 13**,** 80.

Ruth, C., Buchetics, M., Vidimce, V., Kotz, D., Naschberger, S., Mattanovich, D., Pichler, H., Gasser, B., 2014. *Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion. Microb Cell Fact. 13**,** 120.

Sahu, U., Krishna Rao, K., Rangarajan, P. N., 2014. Trm1p, a Zn(II)(2)Cys(6)-type transcription factor, is essential for the transcriptional activation of genes of methanol utilization pathway, in *Pichia pastoris*. Biochem Biophys Res Commun. 451**,** 158-64.

Scazzocchio, C., 2000. The fungal GATA factors. Curr Opin Microbiol. 3**,** 126-31.

Schmalix, W., Bandlow, W., 1993. The ethanol-inducible YAT1 gene from yeast encodes a presumptive mitochondrial outer carnitine acetyltransferase. J Biol Chem. 268**,** 27428-39.

Schüller, H. J., 2003. Transcriptional control of nonfermentative metabolism in the yeast *Saccharomyces cerevisiae.* Curr Genet. 43**,** 139-60.

Scott, S., Abul-Hamd, A. T., Cooper, T. G., 2000. Roles of the Dal82p domains in allophanate/oxalurate-dependent gene expression in *Saccharomyces cerevisiae.* J Biol Chem. 275**,** 30886-93.

Segal, E., Widom, J., 2009. Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr Opin Struct Biol. 19**,** 65-71.

Shen, S., Sulter, G., Jeffries, T. W., Cregg, J. M., 1998. A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. Gene. 216**,** 93-102.

Sherwood, R. I., Hashimoto, T., O'Donnell, C. W., Lewis, S., Barkal, A. A., van Hoff, J. P., Karun, V., Jaakkola, T., Gifford, D. K., 2014. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat Biotechnol. 32**,** 171-178.

Shi, L., Wang, X., Wang, J., Zhang, P., Qi, F., Cai, M., Zhang, Y., Zhou, X., 2018. Transcriptome analysis of *Δmig1Δmig2* mutant reveals their roles in methanol catabolism, peroxisome biogenesis and autophagy in methylotrophic yeast *Pichia pastoris*. Genes Genomics. 40**,** 399-412.

Shlyueva, D., Stampfel, G., Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 15**,** 272-86.

Shukla, A., Lahudkar, S., Durairaj, G., Bhaumik, S. R., 2012. Sgf29p Facilitates the Recruitment of TATA Box Binding Protein but Does Not Alter SAGA's Global Structural Integrity in Vivo. Biochemistry. 51**,** 706-714.

Sidorova, J., Breeden, L., 1993. Analysis of the SWI4/SWI6 protein complex, which directs G1/S-specific transcription in *Saccharomyces cerevisiae*. Mol Cell Biol. 13**,** 1069-77.

Sims, R. J., Mandal, S. S., Reinberg, D., 2004. Recent highlights of RNA-polymerase-II-mediated transcription. Curr Opin Cell Biol. 16**,** 263-71.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R., Rohs, R., 2014. Absence of a simple code: how transcription factors read the genome. Trends Biochem Sci. 39**,** 381-99.

Smale, S. T., Kadonaga, J. T., 2003. The RNA polymerase II core promoter. Annu Rev Biochem. 72**,** 449-79.

Soontorngun, N., Larochelle, M., Drouin, S., Robert, F., Turcotte, B., 2007. Regulation of gluconeogenesis in *Saccharomyces cerevisiae* is mediated by activator and repressor functions of Rds2. Mol Cell Biol. 27**,** 7895-905.

Stadlmayr, G., Mecklenbrauker, A., Rothmuller, M., Maurer, M., Sauer, M., Mattanovich, D., Gasser, B., 2010. Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. J Biotechnol. 150**,** 519-29.

Stemple, C. J., Davis, M. A., Hynes, M. J., 1998. The facC gene of Aspergillus nidulans encodes an acetate-inducible carnitine acetyltransferase. J Bacteriol. 180**,** 6242-51.

Struhl, K., 1985. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. Proc Natl Acad Sci U S A. 82**,** 8419-23.

Struhl, K., 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell. 98**,** 1-4.

Tachibana, C., Yoo, J. Y., Tagne, J. B., Kacherovsky, N., Lee, T. I., Young, E. T., 2005. Combined global localization analysis and transcriptome data identify genes that are directly coregulated by Adr1 and Cat8. Mol Cell Biol. 25**,** 2138-46.

Thanos, D., Maniatis, T., 1995. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell. 83**,** 1091-100.

Todd, R. B., Andrianopoulos, A., 1997. Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genet Biol. 21**,** 388-405.

Tornow, J., Zeng, X., Gao, W., Santangelo, G. M., 1993. GCR1, a transcriptional activator in *Saccharomyces cerevisiae*, complexes with RAP1 and can function without its DNA binding domain. EMBO J. 12**,** 2431-7.

Traven, A., Jelicic, B., Sopta, M., 2006. Yeast Gal4: a transcriptional paradigm revisited. EMBO Rep. 7**,** 496-9.

Tuch, B. B., Galgoczy, D. J., Hernday, A. D., Li, H., Johnson, A. D., 2008. The evolution of combinatorial gene regulation in fungi. PLoS Biol. 6**,** e38.

Turcotte, B., Liang, X. B., Robert, F., Soontorngun, N., 2010. Transcriptional regulation of nonfermentable carbon utilization in budding yeast. FEMS Yeast Res. 10**,** 2-13.

Turpin, P., Ossareh-Nazari, B., Dargemont, C., 1999. Nuclear transport and transcriptional regulation. FEBS Lett. 452**,** 82-6.

Türkanoğlu Özçelik, A., Yılmaz, S., Inan, M., 2019. Pichia pastoris Promoters. Methods Mol Biol. 1923**,** 97-112.

Ucar, D., Beyer, A., Parthasarathy, S., Workman, C. T., 2009. Predicting functionality of protein-DNA interactions by integrating diverse evidence. Bioinformatics. 25**,** i137-44.

van Nimwegen, E., 2003. Scaling laws in the functional content of genomes. Trends Genet. 19**,** 479-84.

Vervecken, W., Kaigorodov, V., Callewaert, N., Geysens, S., De Vusser, K., Contreras, R., 2004. In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. Appl Environ Microbiol. 70**,** 2639-46.

Vincent, O., Carlson, M., 1998. Sip4, a Snf1 kinase-dependent transcriptional activator, binds to the carbon source-responsive element of gluconeogenic genes. EMBO J. 17**,** 7002-8.

Vogl, T., Ruth, C., Pitzer, J., Kickenweiz, T., Glieder, A., 2014. Synthetic core promoters for *Pichia pastoris*. ACS Synth Biol. 3**,** 188-91.

Vogl, T., Sturmberger, L., Fauland, P. C., Hyden, P., Fischer, J. E., Schmid, C., Thallinger, G. G., Geier, M., Glieder, A., 2018a. Methanol independent induction in *Pichia pastoris* by simple derepressed overexpression of single transcription factors. Biotechnol Bioeng. 115**,** 1037-1050.

Vogl, T., Sturmberger, L., Fauland, P. C., Hyden, P., Fischer, J. E., Schmid, C., Thallinger, G. G., Geier, M., Glieder, A., 2018b. Methanol independent induction in *Pichia pastoris* by simple derepressed overexpression of single transcription factors. Biotechnol Bioeng.

Voth, W. P., Olsen, A. E., Sbia, M., Freedman, K. H., Stillman, D. J., 2005. *ACE2,CBK1*, and*BUD4*in Budding and Cell Separation. Eukaryot Cell. 4**,** 1018-1028.

Wang, G., Huang, M., Nielsen, S., 2017a. Exploring the potential of *Saccharomyces cerevisiae* for biopharmaceutical protein production. Current Opinion in Biotechnology. 48**,** 77–84.

Wang, J., Wang, X., Shi, L., Qi, F., Zhang, P., Zhang, Y., Zhou, X., Song, Z., Cai, M., 2017b. Methanol-Independent Protein Expression by *AOX1* Promoter with trans-Acting Elements Engineering and Glucose-Glycerol-Shift Induction in *Pichia pastoris*. Sci Rep. vol. 7.

Wang, J., Wang, X., Shi, L., Qi, F., Zhang, P., Zhang, Y., Zhou, X., Song, Z., Cai, M., 2017c. Methanol-Independent Protein Expression by *AOX1* Promoter with trans-Acting

Elements Engineering and Glucose-Glycerol-Shift Induction in *Pichia pastoris*. Sci Rep. 7**,** 41850.

Wang, X., Cai, M., Shi, L., Wang, Q., Zhu, J., Wang, J., Zhou, M., Zhou, X., Zhang, Y., 2016a. PpNrg1 is a transcriptional repressor for glucose and glycerol repression of *AOX1* promoter in methylotrophic yeast *Pichia pastoris*. Biotechnol Lett. 38**,** 291-8.

Wang, X., Wang, Q., Wang, J., Bai, P., Shi, L., Shen, W., Zhou, M., Zhou, X., Zhang, Y., Cai, M., 2016b. Mit1 Transcription Factor Mediates Methanol Signaling and Regulates the Alcohol Oxidase 1 (*AOX1*) Promoter in*Pichia pastoris*. Journal of Biological Chemistry. 291**,** 6245-6261.

Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V., Cregg, J. M., 1997. Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene. 186**,** 37-44.

Weiss, E. L., 2012. Mitotic exit and separation of mother and daughter cells. Genetics. 192**,** 1165-202.

Werner, F., Grohmann, D., 2011. Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Microbiol. 9**,** 85-98.

Wu, C. Y., Bird, A. J., Chung, L. M., Newton, M. A., Winge, D. R., Eide, D. J., 2008. Differential control of Zap1-regulated genes in response to zinc deficiency in *Saccharomyces cerevisiae*. BMC Genomics. 9**,** 370.

Yamada, Y., Matsuda, M., Maeda, K., Mikata, K., 1995. The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of Komagataella gen. nov. (Saccharomycetaceae). Biosci Biotechnol Biochem. 59**,** 439-44.

Yang, J., Cai, H., Liu, J., Zeng, M., Chen, J., Cheng, Q., Zhang, L., 2018. Controlling *AOX1* promoter strength in *Pichia pastoris* by manipulating poly (dA:dT) tracts. Sci Rep. 8**,** 1401.

Yarragudi, A., Miyake, T., Li, R., Morse, R. H., 2004. Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast *Saccharomyces cerevisiae*. Mol Cell Biol. 24**,** 9152-64.

Yu, J., Jiang, J., Fang, Z., Li, Y., Lv, H., Liu, J., 2010. Enhanced expression of heterologous inulinase in Kluyveromyces lactis by disruption of hap1 gene. Biotechnol Lett. 32**,** 507-12.

Zahrl, R., Reprogramming the yeast *Pichia pastoris* towards a designer cell factory for protein production. Department of Biotechnology, Vol. PhD. BOKU University of Natural Resources and Life Sciences, Vienna, Austria, 2018.

Zaret, K. S., Carroll, J. S., 2011. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 25**,** 2227-41.

Zeitlinger, J., Stark, A., Kellis, M., Hong, J. W., Nechaev, S., Adelman, K., Levine, M., Young, R. A., 2007. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. Nat Genet. 39**,** 1512-6.