Functional Rewiring of Zinc Cluster Transcription Factors Between *Candida albicans* and *Saccharomyces cerevisiae*

Walters Aji Tebung

A Thesis
In The Department
of
Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy (Biochemistry) at
Concordia University
Montreal, Quebec, Canada

July 21st 2016

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Walters Aji Tebung

Entitled: **Functional Rewiring of Zinc Cluster Transcription Factors Between *Candida albicans* and *Saccharomyces cerevisiae***

and submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY (Biochemistry)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Chair

Dr. André Nantel

Dr. Steve Shih — External Examiner

Dr. Peter Pawelek — External to Program

Dr. Paul Joyce — Examiner

Dr. Malcolm Whiteway — Thesis Supervisor

Approved by

Dr. Peter Pawelek, Graduate Program Director

July 21st 2016

Dean André Roy, Faculty of Arts and Science

Abstract

Functional Rewiring of Zinc Cluster Transcription Factors Between *Candida albicans* and *Saccharomyces cerevisiae*

Walters Aji Tebung, PhD

Concordia University, 2016

The prediction of protein function based on the role of orthologous proteins in other species is a common practice in life sciences. While such predictions prove accurate in some cases, a growing number of structurally orthologous proteins with different roles have been reported. Functional differences among structurally orthologous transcription factors create what has been termed transcriptional rewiring. We have characterized some zinc cluster transcription factors (ZCFs) in *Candida albicans* (*C. albicans*) including Ppr1, Stb5 and Put3. We identified cases of total rewiring, partial rewiring and functional conservation (no rewiring) but with some regulatory fine-tuning between *C. albicans* and *S. cerevisiae* for the ZCFs Ppr1, Stb5 and Put3 respectively.

We show that Ppr1 regulates purine catabolism in *C. albicans,* which is different from its role in *S. cerevisiae* where it regulates pyrimidine biosynthesis. We identified that Ppr1 rewired at *Naumovozyma castellii* in the phylogeny, and propose that Ppr1 rewiring is a result of cellular adaptations to permit growth in hypoxic conditions. In another study, we report that Stb5 has gone from regulating drug efflux pumps and glyoxylate cycle enzymes in *C. albicans*, to only regulating drug efflux pumps in *S. cerevisiae*. The glyoxylate cycle is required for *C. albicans* virulence, and our *stb5* null mutants showed sensitivity to the echinocandin class of antifungal drugs. We therefore propose Stb5 as a potential drug target against candidiasis. Finally, we discovered that Put3 has conserved its role in regulating proline catabolism between *C. albicans* and *S. cerevisiae*. However, unlike the case of *S. cerevisiae* where Put3 can only activate transcription of proline catabolism genes in the absence of rich nitrogen sources, there is no equivalent restriction in *C. albicans* where Put3 preserves its ability to activate proline degradation even in the presence of the rich nitrogen source ammonium sulfate, thus allowing for

the use of proline as a carbon source. The robust ability of *C. albicans* to obtain nutrients from a variety of sources adds to its efficiency as a pathogen.

These studies correlate transcriptional rewiring to species requirements, where different species adapt available transcription factors to meet their specific needs.

Acknowledgements

I would like to thank my supervisor Dr. Malcolm Whiteway for allowing me into a MSc. program in biochemistry in his lab and for allowing me to fast track into PhD while moulding me from scratch into a well rounded researcher in the process. The research skills, autonomy, critical thinking, collaboration skills, scientific writing and work ethics that you have instilled in me over these memorable years spent under your supervision will drive me to the very end of my career. You are a great supervisor and a great person and I couldn't ask for more from a supervisor.

I would like to thank Dr. Paul Joyce and Dr. Pawelek who have not only helped me with great input as my committee members, but also trained me during my undergraduate qualifying program and helped point me in the right direction to find a supervisor for my graduate studies. Thank you very much for your very helpful guidance and for walking me all the way from undergraduate to the completion of my PhD in biochemistry.

Special thanks to Dr. Joachim Morschhäuser for his helpful guidance during my PhD studies and for providing the hyperactive zinc cluster transcription factors library that has helped speed up my accomplishments in research by serving me as strains in both ChIP-Chip and transcription profiling experiments. Thank you.

Many thanks to Dr. Baharul I. Choudhury for his bioinformatics contributions in the Ppr1 paper, and to Raha Parvizi Omran for her contributions in the Put3 phenotype studies.

I would like to thank Dr. Faiza Tebbji for the training she gave me in ChIP-Chip, transcription profiling and other experiments, and for her ChIP-Chip and transcription profiling contributions to my research. Your guidance on the bench helped me build a good foundation which has helped me achieve much in a short period of time.

I would also like to thank Dr. Jaideep Mallick for his helpful guidance on immunoprecipitation and western blotting experiments during my early days in the lab. Your input has helped me accomplish much. Thank you.

Special thanks to all my lab members for all their help and support throughout my graduate studies.

I would like to thank the love of my life, Wang Xue. She has been a great and supportive wife throughout my qualifying program at the undergraduate level and throughout my graduate studies, and I couldn't ask for more. She has been a great mom for our two kids Ethan Heng Tebung and Wyatt Kai Tebung, covering for me and herself most of the time. I'm blessed to have you in my

life. Thank you, Ethan and Wyatt for the sacrifices you've made to ensure that I complete my PhD degree in biochemistry.

I would like to thank the people who contributed not just financially but helped build my character as a person, my late dad Jacob Toh Tebung, my mom Lekunze Madia Nanze Epse Djiroda, my grandmother Lekunze Margaret, my entire family and in-laws. Thank you all for being a part of my life journey and for helping me through my PhD degree. Thank you dad for stopping me from dropping out of school at the secondary level, thank you mom for always sacrificing everything to see me succeed, and thank you grandma for keeping me under your roof and seeing me through my early days of school.

Finally, I thank God for all the opportunities and guidance that he has given me throughout my studies and life in general.

Contributions of authors

Chapter 2

Walters Aji Tebung performed experiments, bioinformatic analyses, wrote paper; Baharul I. Choudhury carried out bioinformatic analyses; Faiza Tebbji contributed to the ChIP-Chip experiments; Joachim Morschhäuser provided activated transcription factor reagents; Malcolm Whiteway designed experiments, wrote paper.

Chapter 3

Walters Aji Tebung performed experiments, bioinformatic analyses, wrote paper; Joachim Morschhäuser provided activated transcription factor reagents; Malcolm Whiteway designed experiments, wrote paper

Chapter 4

Walters Aji Tebung performed experiments, bioinformatic analyses, wrote paper; Raha Parvizi Omran contributed to phenotypic experiments; Joachim Morschhäuser provided activated transcription factor reagents; Malcolm Whiteway designed experiments, wrote paper.

Table of Contents

List of Tables

List of Figures

List of abbreviations:

µg	microgram
µl	microlitre
µm	micrometre
bHLH	Basic helix loop helix
bp	base pairs
BSA	Bovine serum albumin
C. albicans	*Candida albicans*
CGD	Candida Genome Database
ChIP	Chromatin immunoprecipitation
dATP	Deoxyadenosine triphosphate
DBD	DNA binding domain
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTPs	Deoxyribonucleotide triphosphates
dTTP	Deoxythymidine triphosphate
GO	Gene ontology
GOF	Gain of function
MeV	Multiexperiment Viewer
ml	millilitre
mM	Millimolar
nm	nanometre
NS	Non-significant
OD	Optical density
Put-GOF	Put3-Gain of function mutant
S. cerevisiae	*Saccharomyces cerevisiae*
SGD	Saccharomyces Genome Database
STB	Sin Three Binding
Stb5-GOF	Stb5-Gain of function mutant
TF	Transcription factor
TFs	Transcription factors

UAP	Urate permease
UOX	Urate oxidase
WT	Wild type
YCB	Yeast carbon base media
ZCF	Zinc cluster transcription factor
ZCFs	Zinc cluster transcription factors

Chapter 1: General introduction

1.1 Zinc cluster transcription factors

Transcription factors (TFs) bind to specific DNA sequences at gene promoters to either activate or repress transcription. Transcription factors have been grouped into families such as the basic helix loop helix (bHLH) class [1], the zinc finger TFs [2], the homeo box family [3], and the leucine zipper TFs family [4]. There are 55 zinc cluster transcription factors (ZCFs) in *Saccharomyces cerevisiae* (*S. cerevisiae*) and 82 ZCFs in *Candida albicans* (*C. albicans*), which makes this class of transcription factors one of the largest in both fungi. ZCFs make up approximately one third of transcription factors in *C. albicans* and are known to play a role in the regulation of diverse processes such as metabolism, meiosis, virulence, and antifungal drug resistance [5-7]. All ZCFs possess the conserved sequence $CX_2CX_6CX_{5-12}CX_2CX_{6-8}C$ in their DNA-binding domain and usually possess a lysine residue (sometimes replaced by arginine, histidine, or glutamine) between the second and third cysteines that directly interacts with the conserved cytosine and guanine bases through hydrogen bonding [8] (Figure 1.1) during DNA binding (Figure 1.1). ZCFs predominantly bind DNA as homodimers using their N-terminal zinc finger motifs to hydrogen bond with CGG nucleotide triplets that are oriented in everted, inverted, or direct repeats [7,8] (Figure 1.2). The binding specificity is based on the spacing of the CGG sequences, which is controlled by the length and folding of a non-conserved linker sequence [7,9-11] (Figure 1.3). The conserved cysteines complex two zinc ions that are important in facilitating the DNA binding of the protein [12,13].

The ZCF Gal4 is a classic model for zinc cluster protein function. Gal4 binds to a typical zinc cluster transcription factor binding site, 5'-CGG(N_{11})CCG-3', at the promoters of *C. albicans* glucose catabolism and subtelomeric TLO genes [14-17] and *S. cerevisiae* galactose catabolism genes or GAL genes (*GAL1*, *GAL2*, *GAL3*, *GAL7*, *GAL10*, and *GAL80*) [18,19] to transcriptionally regulate these processes. *S. cerevisiae* Gal80 on the other hand represses GAL genes by binding the transcriptional activation domain of Gal4 to inhibit the transcription factor from transcriptionally activating GAL genes [15]. To relieve Gal4 inhibition in *S. cerevisiae*, Gal3 sequesters Gal80 in the cytoplasm, in the presence of galactose and ATP, and forms a complex with Gal80 that prevents it from inhibiting Gal4 and thus allowing the activation of

GAL gene expression [15]. Another interesting but unexplained observation is the clustering of the Gal4 target genes, *GAL1*, *GAL7*, and *GAL10*, in *S. cerevisiae* (near the centromere of *S. cerevisiae* chromosome 2) and its close relatives [15].

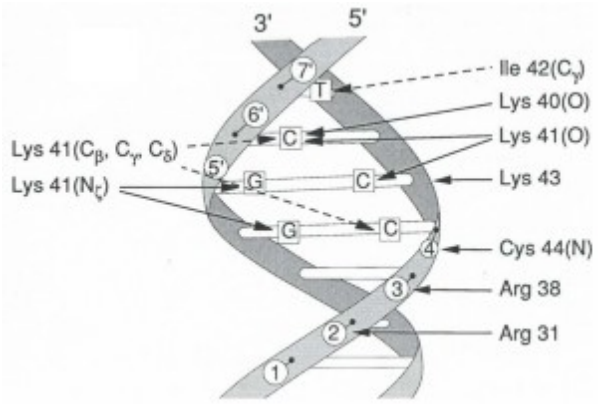


Figure 1.1: Residues of the *S. cerevisiae* zinc cluster transcription factor Ppr1 interacting with DNA during binding. Hydrogen bonding is indicated by arrows and Van Der Waals interactions are indicated by broken lines. Lysine 41 forms the main interaction with the DNA CGG sequence and is almost universally conserved among ZCFs [8].

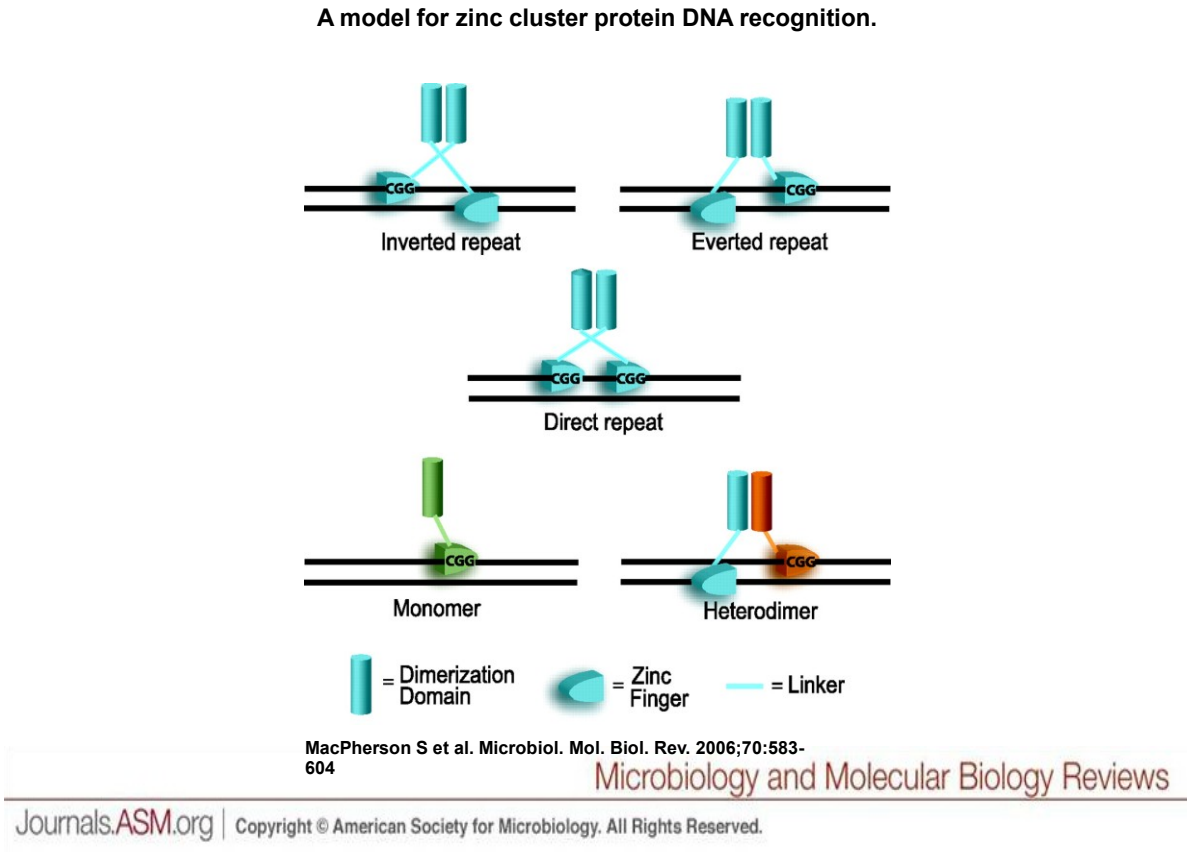


Figure 1.2: The different ways zinc cluster transcription factors bind CGG sequences [7]

Understanding ZCFs and any differences in their roles between *C. albicans* and *S. cerevisiae* could reveal the reason why *C. albicans* is a human pathogen but *S. cerevisiae* is not, and could help improve the treatment of candidiasis and fight drug resistance.

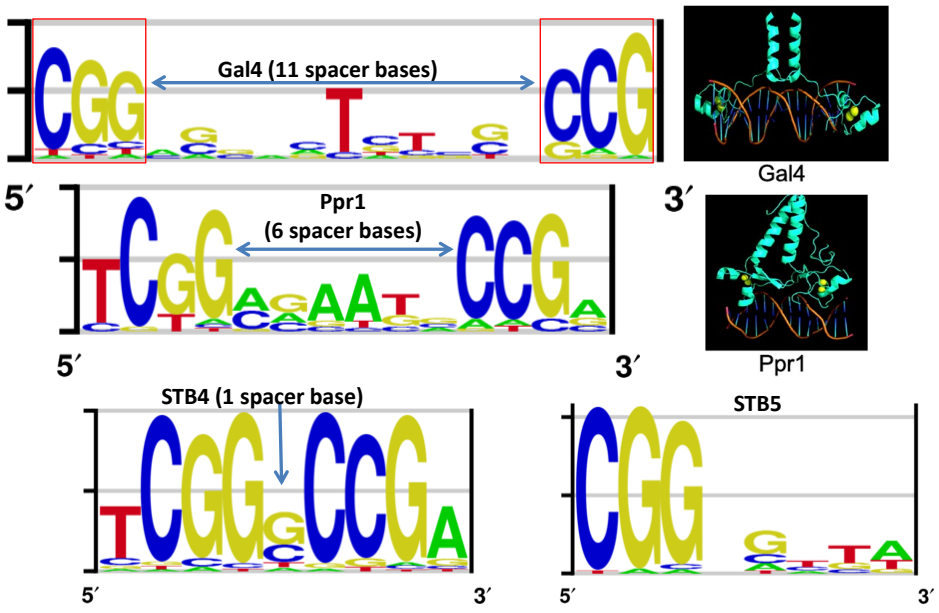


Figure 1.3: DNA binding sequences of some zinc cluster transcription factors. The number of spacer bases helps to determine specificity [7,20]

1.2 Transcriptional rewiring

Different ascomycetes are known to use orthologous transcription factors to regulate varying metabolic circuits [17,21], a process that has been termed transcriptional rewiring [22]. As evident by the growing reports on transcriptional rewiring events, defining the match between transcription factors and metabolic pathways based solely on structural similarity of the transcription factors is not reliable. While the function of orthologous transcription factors remain a good starting point when predicting the role of newly identified orthologs in related species, further investigation is often required to establish cellular processes regulated by newly identified othologous transcription factors.

In addition to being a classically studied TF, Gal4 provides a clear example of this phenomenon of transcriptional rewiring. The *C. albicans* Gal4 protein regulates glucose catabolism [16], while the *S. cerevisiae* Gal4 protein regulates galactose catabolism [14,15] . Although Gal4 controls different circuits in the two species, the DNA binding motif used by the TF to regulate the respective pathways in the species remains the same. Other rewired transcription factors that use similar motifs to carry out different functions in different species

include Cbf1 [23] and Ppr1(Chapter 2 of this thesis). In the absence of large scale comparative location binding studies, we are essentially limited to identifying rewiring events by bioinformatical motif searches, which would miss examples where the actual binding motif of the TF has rewired. Table 1.1 shows clear examples of othologous transcription factors regulating related sets of genes in *C. albicans* and *S. cerevisiae*, but identifies several examples where orthologous transcription factors play different roles in the two species. Table 1.1 was created by taking some *S. cerevisiae* TFs that regulate an obvious pathway or process and have a clear binding motif, and looking in *C. albicans* for this same motif at the promoters of the orthologs of the select targets identified in yeast. Top computationally identified targets were included, as well as experimentally determined targets. Note however that transcription factor functions presented in Table 1.1 are based specifically on the presence or absence of the *S. cerevisiae* transcription factor motif at promoters of *C. albicans* genes and usually require further studies to establish the functions predicted. Table 1.1 suggests that Adr1, for example, regulates the same set of genes in *S. cerevisiae* and *C. albicans*. The *C. albicans ADR1* mutant phenotype, however, suggests a different set of target genes [24-26] compared to targets of the *S. cerevisiae* ortholog; more experimental evidence is therefore required to establish the actual *C. albicans* Adr1 binding sites and regulatory role. On the other hand, the table rightfully predicts that Ppr1 is rewired between *S. cerevisiae* and *C. albicans*; this Ppr1 rewiring prediction has been experimentally confirmed and presented in Chapter 2 of this thesis. Leu3 and Rpn4 are predicted in Table 1.1 to play the same role in *C. albicans* and *S. cerevisiae*, but there is currently no experimental evidence of this functional conservation; it will be important to acquire experimental evidence in support of the *in silico* predictions. A special case is that of Put3 where the DNA binding motif, as identified in *S. cerevisiae*, is not present upstream of the proline degradation genes in *C. albicans* as shown in Table 1.1 and our ChIP-Chip data. One would assume that the transcription factor has been rewired, but our data suggest that Put3 does bind at least one of the two key enzymes of the proline degradation pathway and regulates proline catabolism in *C. albicans* through transcriptional regulation of Put1 and Put2 as reported in Chapter 4 of this thesis. It is not common for a transcription factor to use different DNA binding motifs to regulate the same set of genes in different species, yet not impossible since indirect conservation by creation of one binding site and destruction of another has been shown to result in transcription factors regulating the same genes in different species but using different binding sites [27]. Studies have shown that

the transcription factors CEBPA (CCAAT/enhancer-binding protein alpha) and HNF4A (hepatocyte nuclear factor 4 alpha) have conserved their functions between multiple species including humans, mice, and dogs, but use different species-specific binding sites to carry out the same regulatory function in the different species [28]. It would therefore be interesting to identify the actual Put3 binding site in *C. albicans* for comparison with the *S. cerevisiae* motif. These experimental findings suggest that Put3 plays essentially the same role in *C. albicans* and *S. cerevisiae*.

S. cerevisiae and *C. albicans* are therefore divergent enough to have frequent examples of rewiring, but not so divergent that no conserved circuits can be identified. In some rewiring events, orthologous transcription factors regulate different metabolic pathways, and in other instances, orthologous metabolic pathways are regulated by distinct transcription factors. Gal4 is a classical example of a zinc cluster transcription factor that regulates different processes in the two related ascomycetes *S. cerevisiae* and *C. albicans*. Gal4 and Tye7 have been identified as the regulators of glucose metabolism in *C. albicans* [16], but Gal4 has been rewired to control galactose catabolism in *S. cerevisiae* [14,15], meanwhile Tye7 remains linked to glucose metabolism in both species [16]. In the other hand, the Gcr1 and Gcr2 proteins transcriptionally control glycolytic enzyme expression to regulate glucose metabolism in *S. cerevisiae* [29], but are not present in *C. albicans*. Differences in ribosomal protein gene expression observed between *C. albicans* and *S. cerevisiae* highlight just how flexible metabolic pathways can get through evolution. Tbf1 is the chief regulator of ribosomal protein gene expression in *C. albicans*, a process that is centrally regulated by Rap1 in *S. cerevisiae* [30-32]. This rewiring event is not some strategy to cover for a missing Tbf1 since Tbf1, Rap1, as well as the adapter proteins, Ifh1 and Fhl1, are all present in both species. The connectivity to the ribosomal protein is however rearranged such that different DNA binding proteins are used by *C. albicans* and *S. cerevisiae* to respond similarly to nutrient conditions [32].

Table 1.1 Transcriptional regulatory network rewiring or conservation among orthologous transcription factors

We identified some *S. cerevisiae* TFs that regulate an obvious pathway or process and have a clear binding motif, and then looked for the motif at the promoters of the *C. albicans* orthologs of some of the target genes. Some top computationally identified target genes were used for the analysis, as well as experimentally determined targets. If a motif is over-represented in target genes compared to the control, the *P* value will be highly significant. Motif over-representation was analyzed using 'compare motif analysis' option based on hypergeometric test (significance threshold 0.05) followed by Bonferroni correction as implemented in the MotifLab software package [62]. Motifs showing significant and non-significant (NS) *P* values are based on over-representation analysis in 400 bp promoter regions of the select genes that are directly regulated by each transcription factor in *S. cerevisiae*, or *C. albicans* othologs of each transcription factor's select target genes [21].

TFs	*P* value	
	S. cerevisiae	*C. albicans*
ADR1	8.30E-03	4.70E-02
ARO80	1.18E-07	NS
CUP2	6.60E-04	NS
GAL4	5.82E-09	NS
LEU3	4.39E-09	3.00E-05
PPR1	9.10E-04	NS
PUT3	5.92E-07	NS
RPN4	6.19E-18	1.70E-10

Through bioinformatic analyses of ascomycete fungi, it was shown that the Tbf1 target motif dominates the promoters of the RP genes in *C. albicans* and its relatives, and the Rap1 motif dominates in *S. cerevisiae* and its relatives (Figure 1.4A) [21]. The same bioinformatics analyses for the motifs using relaxed stringency however showed that many species across the entire phylogeny have the potential to adapt their ribosomal protein gene promoters for either Tbf1 or Rap1 binding (Figure 1.4B) [21].

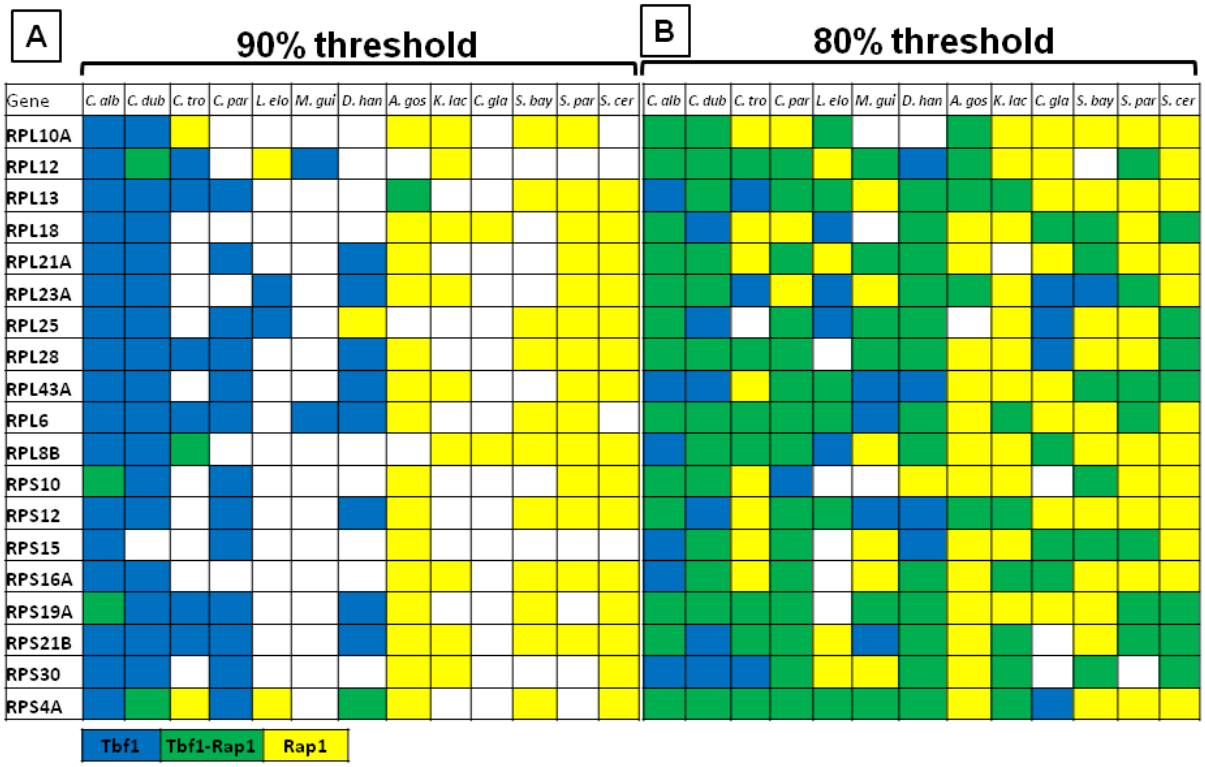


Figure 1.4. Transition of Tbf1-Rap1 DNA binding motif across ascomycetes fungi.

(A) Strict motif scanning (90% threshold) showed dominance of the Rap1 motif in *S. cerevisiae* and its close relatives, whereas Tbf1 dominates in *C. albicans* and its close relatives. (B) Relaxed stringency (80% threshold) for motif scanning showed most species had potential motifs for either Tbf1 or Rap1. Motif scanning was performed using software package MotifLab [62]. Species names: *A. gos*, *A. gossypii*; *C. alb*, *C. albicans*; *C. dub*, *C. dubliniensis*; *C. gla*, *C. glabrata*; *C. tro*, *C. tropicalis*; *C. par*, *C. parapsilosis*; *D. han*, *D. hansenii*; *L. elo*, *L. elongisporus*; *M. gui*, *M. guilliermondii*; *K. lac*, *K. lactis*; *S. bay*, *S. bayanus*; *S. cer*, *S. cerevisiae*; *S. par*, *S. paradoxus* [21].

These findings suggest that switching DNA binding transcription factors can be accomplished with limited changes at any promoter. What is much more difficult to envision is how these changes could be co-ordinated over the multiple promoters of the regulon. Since the overall rewiring in *S. cerevisiae* led to the regulation of highly expressed genes such as the genes for glycolytic enzymes as well as those for ribosomal proteins by Rap1, it could be an indicator that the coordination of highly expressed genes by Rap1 may set the stage for selective a advantage and drive the rewiring from Tbf1 to Rap1 [32]. Other transcriptional rewiring events initially identified among ascomycetes involve a sugar metabolic pathway [15], Mcm1 circuitry [33], coordination of mitochondrial and cytoplasmic ribosomal protein genes [34], cell type

establishment [35], and the cell cycle regulation pathway [36,37]. The rewiring of other cellular pathways has also been reported, including iron assimilation [38], sterol biosynthesis [39], and mating [35,40].

Current knowledge of transcriptional rewiring can only partially address questions such as how and why transcription factors rewire, and whether or not rewiring is random; for example, the fact that the DNA binding sites of Gal4 and Ppr1 moved on from related gene sets to other related gene sets could suggest that the rewiring process is not random. Several transcriptional rewiring driving forces have so far been reported. A transcription factor originating from horizontal gene transfer could hijack a transcriptional regulatory network and force the ancestral transcriptional regulator to change functions or end up extinct from the cell [41]. An example of such transcription factor rewiring is observed between fungi and animals where the transcription factor SBF apparently inserted into the fungi genome from an invading virus, took over the transcriptional regulation of the cell cycle from the original transcriptional regulator E2F. E2F remains the cell cycle regulator in animals and plants. Evidence shows that both transcription factors have the same DNA binding site specificity, which further strengthens this theory [41]. The overall structure of the cell cycle regulon has been conserved across species, but many of the protein components of the regulon have no orthology with other protein components playing a similar role in the cell cycle in other species. In addition to the E2F to SBF transition, Whi5, which originates from an unknown source and inhibits SBF in fungi has no homology to the animal E2F inhibitor, pRb [41]. Gene duplication is another reported driver of transcriptional rewiring because duplicated genes often diverge and evolve to take on new functions [42]. The generation of new genes with new functions might require reorganization and/or extension of the regulatory network to ensure that the new paralogues are properly expressed and that the paralogues do not cross-interfere with each other's specific function [42]. Gene duplication can result from whole genome duplication or independent duplication. The rewiring of MalS and MalR proteins are a result of gene duplication involving both the transcription factor coding gene (*MALR*) and the target gene, *MALS* [42]. *MALS* first duplicated into two paralogous enzymes, one that responds to palatinose and is responsible for palatinose utilization, and another that responds to maltose and is responsible for maltose utilization. The promiscuous ancestral *MALS* regulator, MalR, can bind both CGC and CGG that are randomly distributed in the un-diverged promoters of the homologues of the different genes involved in maltose utilization as well as in

palatinose utilization. This promiscuous binding nature allowed MalR to continue, for a while, the regulation of both *MALS* paralogues before eventually duplicating into two specific *MALR* paralogues [42]. One *MalR* paralogue following a Arg12Cys mutation could no longer bind CGC which are selected for in the promoter of maltose specific genes and therefore specialized in regulating palatinose utilization. The remaining MalR paralogue evolved to lose the ability to activate expression of palatinose-specific genes which are selected to have only one CGG-containing binding site, while conserving the ability to activate maltose-specific promoters that have evolved to contain three CGC binding sites [42]. These findings show how duplication of a promiscuous transcription factor and its target genes led to the development of two separate regulatory networks [42]. The findings could also suggest that transcription factors with an ability to bind multiple regulatory sequences could more easily rewire by specializing in using one of their many binding sites to specifically regulate a new process. We also report for Ppr1 in Chapter 2 of this thesis how changing the inducer of a transcription factor could drive rewiring.

We discuss in this thesis a case of total transcriptional rewiring as well as the likely pressures that led to the rewiring event, a case of partial rewiring of a transcriptional regulator, and a case where a transcriptional regulator has been preserved but a new regulatory pathway has evolved to fine tune its transcriptional role.

1.3 Relationship between transcriptional rewiring and species lifestyle

Transcriptional rewiring is not confined to DNA binding proteins, but has been identified among orthologous general transcription factors such as Mediator, SWI/SNF, SAGA, and NuA4. This is an indication that transcriptional networks can be rewired both through modifications in specific transcription factor regulatory networks and through changes in the subunits and connectivities of general transcription regulators [21]. Differences in post-translational modifications and protein stability have also been reported among ascomycetes. The pressures that drive rewiring and the effects of rewiring on the overall phenotype of species remain less well understood. In this thesis, we have attempted to link rewiring to the different lifestyles of the pathogen *C. albicans* and the bread and wine yeast *S. cerevisiae*. Previous studies had identified differences in carbon metabolism between *C. albicans* and *S. cerevisiae* that appear to correlate with their lifestyles. Such rewiring is observed in the case of glucose metabolism where post-transcriptional control of protein stability differs significantly between the species. Fungi generally shut down expression of the alternative carbon pathway elements and utilize preferred

carbon sources such as glucose when presented with options, a process that has been named carbon catabolite control or CCC. CCC in *S. cerevisiae* is through ubiquitin-mediated proteolysis where enzymes involved in the use of the alternative carbon source are rapidly degraded in the presence of a preferred carbon source such as glucose [43]. In contrast, this proteolysis does not occur in *C. albicans* although the transcriptional regulatory pathway appears to be similar in the two fungi, with the repressor Mig1 being the key player in shutting off the transcriptional pathways required to breakdown the alternative carbon in both species. For example Icl1 and Pkc1 required for oleate utilization in *C. albicans* persist in the pathogen when the carbon source switches from oleate to glucose, even though their gene expression is significantly reduced [44]. *C. albicans'* enzymes lack the ubiquitination target sites found in the orthologous *S. cerevisiae* proteins, so even though the pathogen has the machinery to degrade the proteins, these specific proteins evade this machinery. *S. cerevisiae* proteins expressed in *C. albicans* were destabilized by the same treatment confirming that *C. albicans* has the ability to degrade the proteins. These differences in protein stability between the two fungi dramatically change the cellular response to alternative sugars [45]. Also, *C. albicans* is quite promiscuous in its sugar use and can grow efficiently on a variety of 5-carbon sugars such as xylose [46] and arabinose [47]; sugars that normal *S. cerevisiae* strains are unable to catabolise [48]. Even in the absence of a defined sugar source, *C. albicans* is able to grow using gluconeogenic amino acids to supply carbon for central metabolism [49]. *S. cerevisiae* predominantly uses hexose sugars for energy and growth, while *C. albicans* uses its more advanced carbon metabolism pathways to utilize alternative carbon sources, which could improve its chances of survival in diverse host environments such as upon phagocytosis by macrophages. *C. albicans* can switch from glycolysis to the glyoxylate cycle, gluconeogenesis, and β-oxidation pathways using key enzymes such as Fbp1, Pck1, and Icl1 that allows it to catabolise nonfermentable carbon sources [43,50]. The pathogen also possesses additional genes, such as acyl-CoA oxidase, fatty acid-CoA synthase, and oxidoreductases used in respiratory catabolism, which might enhance its ability to thrive inside the host [51]. We report in this thesis a more liberal use of proline, both as a carbon source and as a nitrogen source by *C. albicans*. Proline is only used as a nitrogen source by *S. cerevisiae* in the absence of a rich nitrogen source such as ammonium sulfate; in contrast our findings establish that *C. albicans* can catabolise proline whether or not ammonium sulfate is present. This could be an indicator that *C.*

albicans is not only flexible in its carbon use but also in its nitrogen use which could further favor its pathogenic lifestyle.

1.4 Thesis aims and organization

This study was designed to characterize zinc cluster transcription factors (ZCFs) in *C. albicans* and to identify transcriptional rewiring between orthologous ZCFs in *C. albicans* and *S. cerevisiae.* There are 55 zinc cluster transcription factors in *S. cerevisiae* and 82 zinc cluster transcription factors in *C. albicans* making up one third of the overall transcription factors in the pathogen. The role that this large class of transcription factors plays in varying cellular processes including metabolism, meiosis, virulence, antifungal drug resistance makes ZCFs suitable for transcriptional rewiring studies that could in turn reveal the reason why *C. albicans* is a pathogen and *S. cerevisiae* is not.

In Chapter 2, we discuss the total rewiring of the ZCF Ppr1 from purine catabolism in *C. albicans* to pyrimidine biosynthesis in *S. cerevisiae*. We identified through this study that Ppr1 has changed functions between the two species, but that the motif bound by Ppr1 in both species to carry out its transcriptional regulation role has been conserved. Dal82 is the transcriptional regulator of purine catabolism in *S. cerevisiae* and no transcriptional regulator has so far been identified for the pyrimidine biosynthesis pathway in *C. albicans*. We also point out that Ppr1 rewired at the lineage leading to *Naumovozyma castellii* concomitant with the DAL cluster formation and the elimination of some oxygen consuming enzymes; events believed to have been driven by pressures to grow in low oxygen milieus. We therefore propose that the adaptation of species in the lineage leading to *Naumovozyma castellii* to grow in hypoxic conditions pressured Ppr1 to rewire at this point in the phylogeny.

In Chapter 3, we discuss the partial rewiring between *C albicans* and *S. cerevisiae* of the ZCF Stb5 and its suitability as a drug target. Stb5 regulates glyoxylate enzymes in *C. albicans* as well as drug efflux pumps, but regulates only drug efflux pumps in *S. cerevisiae* and not glyoxylate enzymes which appear to be under the control of Sip4 in *S. cerevisiae*. Our data show that *stb5* null mutant *C. albicans* strains cannot efficiently utilize acetate as a sole carbon source (acetate catablism requires glyoxylate enzymes) and show sensitivity to the echinocandin drug class. In line with our discovery that Stb5 is required for drug resistance in *C. albicans* and also regulates the glyoxylate cycle that is very critical for virulence, we predict that drugs designed to inhibit Stb5 would prove efficient in treatment of candidiasis.

In Chapter 4, we discuss a ZCF, Put3, that has preserved its primary function and has not been rewired between *C. albicans* and *S. cerevisiae*. We, however, discovered that *C. albicans* Put3 has fewer restrictions on carrying out its functions as compared to its *S. cerevisiae* ortholog. In both species, Put3 activates the transcription of Put1 and Put2, the two key enzymes of the proline degradation pathway, which leads to proline catabolism. Previous studies showed that *S. cerevisiae* can only degrade proline in the absence of a rich nitrogen source such as ammonium sulfate, and as soon as a rich nitrogen source becomes available, the proline degradation pathway gets shut down and degradation of the rich nitrogen source takes over. We report in this study that *C. albicans* Put3 degrades proline in the presence or absence of a rich nitrogen source such as ammonium sulfate in *C. albicans*. These findings highlight the ability of *C. albicans* to more easily obtain nutrients from varying sources, a trait that adds to its ability to survive in the host during virulence.

Chapter 2: Rewiring of the Ppr1 Zinc Cluster Transcription Factor from Purine Catabolism to Pyrimidine Biogenesis in the Saccharomycetaceae

2.1 Preface

The work presented in Chapter 2 is from the following manuscript: Tebung, W. A., Choudhury, B. I., Tebbji, F., Morschhäuser, J. & Whiteway, M., (2016). **Rewiring of the Ppr1 Zinc Cluster Transcription Factor from Purine Catabolism to Pyrimidine Biogenesis in the Saccharomycetaceae.** Curr Biol. 2016 Jun 15. pii: S0960-9822(16)30418-3. doi: 10.1016/j.cub.2016.04.064.

Author contributions:

Walters Aji Tebung performed experiments, bioinformatic analyses, wrote paper; Baharul I. Choudhury carried out bioinformatic analyses; Faiza Tebbji contributed to the ChIP-Chip experiments; Joachim Morschhäuser provided activated transcription factor reagents; Malcolm Whiteway designed experiments, wrote paper.

2.2 Abstract of manuscript

Metabolic pathways are largely conserved in eukaryotes, but the transcriptional regulation of these pathways can sometimes vary between species; this has been termed rewiring. Recently it has been established that in the Saccharomyces lineage starting from *Naumovozyma castellii*, genes involved in allantoin breakdown have been genomically relocated to form the DAL cluster. The formation of the DAL cluster occurred along with the loss of urate permease (UAP) and urate oxidase (UOX), reducing the requirement for oxygen and bypassing the candidate Ppr1 inducer, uric acid. In *Saccharomyces cerevisiae* this allantoin catabolism cluster is regulated by the transcription factor Dal82, which is not present in many of the pre-rearrangement fungal species. We have used ChIP-Chip analysis, transcriptional profiling of an activated Ppr1 protein, bioinformatics, and nitrogen utilization studies, to establish that in *Candida albicans* the zinc cluster transcription factor Ppr1 controls this allantoin catabolism regulon. Intriguingly, in *S. cerevisiae* the Ppr1 ortholog binds the same DNA motif (CGG(N6)CCG) as in *C. albicans*, but serves as a regulator of pyrimidine biosynthesis. This transcription factor rewiring appears to have taken place at the same phylogenetic step as the formation of the rearranged DAL cluster.

This transfer of the control of allantoin degradation from Ppr1 to Dal82, together with the repositioning of Ppr1 to the regulation of pyrimidine biosynthesis, may have resulted from a switch to a metabolism that could exploit hypoxic conditions in the lineage leading to *N. castellii* and *S. cerevisiae*.

2.3 Introduction

The regulation of gene expression is an important factor in the development of organisms, in the evolution of species, and in cellular adaptation to environmental changes. Although cells typically have conserved metabolic machinery, the regulation of genes encoding this machinery can vary, giving rise in part to the differences in phenotype that are observed among species [23]. Key components of eukaryotic gene expression control are the transcription factors (TFs) that act as transcriptional activators or repressors through binding to specific DNA sequences at the promoter regions of genes. These TFs fall into families, such as the basic helix loop helix (bHLH) class [1], the zinc finger TFs [2], the homeo box family [3], and the leucine zipper TFs [4].

The zinc cluster TFs (ZCFs) are a subclass of the zinc-finger proteins found exclusively in fungi. These ZCFs contain the conserved structural motif $CX_2CX_6CX_{5-12}CX_2CX_{6-8}C$ (where C represents cysteine, and X is any amino acid) [5,7,12,52,53]. This cysteine-rich sequence is generally located at the N terminus of the TF, and consists of 6 cysteines with typically a lysine residue between the second and third cysteines [8,9,13]. The entire DNA binding domain (DBD) of the transcription factor consists of the zinc finger, a linker region and a dimerization domain. Typically, ZCFs bind as homodimers using their zinc finger motifs to generate hydrogen bonding major groove interactions to CGG nucleotide triplets that are oriented in everted, inverted, or direct repeats [7,8]. The spacing of the CGG sequences is important for binding specificity, which is controlled by the number of generally non-conserved amino acids specifying both the ZCF's linker length and its folding [7,9-11]. The conserved cysteines serve to complex two zinc ions that are important in facilitating the DNA binding of the protein [12,13].

These zinc cluster transcription factors are major components of fungal-specific regulatory circuits, and are known to control a multitude of processes as varied as metabolism, meiosis, virulence, and antifungal drug resistance [5-7]. A classic zinc cluster TF is the well-studied Gal4 protein of *S. cerevisiae*. This protein served as the basis for the yeast 2-hybrid system [54], and studies have established its mechanism for dimerization [13,55], for binding to the specific DNA

sequence 5'-CGG(N_{11})CCG-3' [13,55], and for interacting with other regulatory proteins to establish the inducibility of the yeast Leloir pathway genes to galactose [18,19]. This level of understanding has made *S. cerevisiae* Gal4 a popular model for systems biology investigations into eukaryotic transcriptional regulation [9,14,19,54]. Intriguingly, outside of the Saccharomyces lineage Gal4 orthologs can perform other functions, although they continue to bind the motif 5'-CGG(N_{11})CCG-3'. For example, in *C. albicans*, the Gal4 ortholog binds glycolytic genes and subtelomeric TLO genes [14-17], and is not involved in galactose regulation. This represents an interesting example of transcription factor rewiring, the process whereby orthologous proteins, interacting with conserved binding sites, control different functions in related species [21]. Because the DNA binding sequences for ZCFs are short, genes can easily move in or out of regulatory circuits through changes in one or two nucleotides in their promoters. However, the driving forces behind such rewiring events are currently poorly understood.

Here we investigate the function of the zinc-cluster transcription factor Ppr1 in *C. albicans*. Ppr1 was initially identified in *S. cerevisiae* where it bound the motif 5'-CGG(N_6)CCG-3' to regulate genes involved in the biosynthesis of uracil [56,57]. We show here that the Ppr1 ortholog in the opportunistic fungal pathogen *C. albicans* uses the same binding motif to regulate genes involved in allantoin degradation. Bioinformatic analysis of the promoter motifs within the fungal phylogeny suggest that this rewiring event resulted in a switch, within the Saccharomyces lineage, of the regulation of allantoin degradation from Ppr1 to Dal82.

2.4 Materials and methods

2.4.1 Strains, media, plasmids and transformation

We followed standard procedures for *C. albicans* cell growth and transformation [58]. YPD was used to culture *C. albicans* strains for transformation, immunoprecipitation, ChIP-Chip analyses and transcriptional profiling experiments. Yeast carbon base media (YCB) at 11.7 mg/ml lacking a nitrogen supplement, or supplemented with ammonium sulfate (37.82 mM), urea (10 mM) or allantoin (10 mM), was used for phenotype studies testing the ability of the *ppr1* null mutant to utilize allantoin as a nitrogen source.

The PPR1-TAP strain was constructed from strain SN148 [59,60] by transformation using a TAP-URA3 PCR product containing 99 base pairs (bp) of homologous sequence immediately

upstream and downstream of the *PPR1* stop codon. The TAP-URA3 portion of the oligomer used for transformation was amplified from the pFA-TAP-URA3 plasmid [60] using the forward primer PPR1_TAP_URA3_F and the reverse primer PPR1_TAP_URA3_R (See table S1 for oligos). Correct integration of the TAP tag was confirmed by PCR using two primer pairs; forward primer PPR1_TAP_URA3_check_F that binds inside the *PPR1* gene and reverse primer FT-U3 that binds inside the *URA3* gene, and forward primer FT-U1 that binds inside the *URA3* gene and reverse primer PPR1_TAP_URA3_check_R that binds downstream of *PPR1*. The *ppr1* null mutant was constructed in strain SN95 by replacing one *PPR1* allele with the *HIS1* marker and the other allele with *ARG4* marker. Oligonucleotides used for *ppr1* null mutant construction were amplified from plasmids pFA-HIS1 for the first allele knockout and pFA-ARG4 for the second allele knockout [60] using the respective forward and reverse primers PPR1_Marker_KO_F that has a 97 bp region with homology just before the *PPR1* start codon and PPR1_Marker_KO_R that has a 95 bp region with homology just after the *PPR1* stop codon. Deletion of the first allele was confirmed by PCR using the forward primer PPR1_KO_Check_F that binds upstream of the *PPR1* gene and the reverse primer PPR1_KO_Check_R that binds downstream of the gene. Deletion of the second allele was confirmed by PCR using two primer pairs; forward primer PPR1_KO_Check_F that binds upstream of the *PPR1* gene and reverse primer FT-H2 that binds upstream of the *HIS1* gene (*HIS1* promoter region) and forward primer FT-U3 that binds inside the *HIS1* gene and reverse primer PPR1_KO_Check_R that binds downstream *PPR1*. The *ppr1* null mutants were confirmed using the forward primer PPR1_KO_Check_Internal_F and the reverse primer PPR1_KO_Check_Internal_R that both bind inside the *PPR1* gene; no band is expected for *ppr1* null mutants for this primer pair. The SCPPR1GAD1A strain was as described previously [5].

2.4.2 Immunoprecipitation

The immunoprecipitation protocol was used as previously described [32] with the following modifications: a 700 µl lysate obtained from the cell culture was incubated with IgG Sepharose beads at 4°C overnight on a lab rotator. After transferring the proteins through Western blotting from a 12% SDS-PAGE gel to a nitrocellulose membrane, the membrane was treated with the primary rabbit polyclonal antibody (1:1000) directed against the TAP-tag, then with the secondary anti rabbit antibody coupled to dye IR 680 (1:5000). Protein bands were detected by an Odyssey scanner at a wavelength of 700 nm.

2.4.3 ChIP-Chip

The strain containing the chromosomally inserted PPR1-TAP fusion as well as the background strain SN148 (untagged) were grown to an optical density of 0.6 at 600 nm (OD 600nm) in 50 ml of YPD. Crosslinking was performed by incubating each 50 ml culture with 1.5 ml of 37% formaldehyde for 30 minutes; modified from 1.4 ml of 37% formaldehyde as described earlier [61]. Cultures were centrifuged for 5 mins in a 4°C centrifuge at 3600 RPM and YPD media removed by decantation, then cells were washed twice in 4°C cold TBS using 40 ml TBS each time. TBS was removed after every wash by decantation following a 5 min centrifugation at 3600 RPM in the centrifuge at 4°C. A final centrifugation was carried out at 13000 RPM for a minute at 4°C before removing residual TBS, leaving cell pellets for the following steps. Cell pellets were resuspended in 700 µl lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) with protease inhibitors, and 0.5 ml acid washed glass beads (425 - 600 µm) were added. Samples were homogenized five times for a minute each at 4,200 RPM using a Bead beater with five minute breaks on ice between homogenization steps; this step was modified from the four rounds of homogenization described previously [61]. The lysate collected for each sample was sonicated using a Fisher Scientific Model FB120 sonicator to obtain DNA with an average length of 300-500 bp (samples were sonicated four times for 20 seconds at Amp 35% and pulse 20, tubes were put on ice for at least a minute between every round of sonication) and all 700 µl of each lysate was incubated with IgG-Sepharose beads overnight, a modification of the 500 µl of lysate described previously [61]. The next day, beads from each sample were washed twice with 1 ml Lysis buffer, twice with 1 ml lysis buffer supplemented with 360 mM NaCl, twice with 1 ml Wash buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP40, 0.5% Nadeoxycholate, 1 mM EDTA) and once with 1 ml TE (10 mM Tris pH 8.0, 1 mM EDTA), then incubated in 50 µl TE/SDS (10 mM Tris pH 8.0, 1 mM EDTA, 1% SDS) overnight at 65°C. 345 µl TE, 3 µl DNase-free RNase A (10 mg/ml in water) and 2 µl glycogen (20 mg/ml) were then added to each supernatant and incubated for 2 hours at 37°C followed by the addition of 15 µl of 10% SDS and 7.5 µl proteinase K (20 mg/ml) and then by a 2 hour incubation at 37°C. DNA isolation involved two 400 µl phenol/chloroform/isoamyl alcohol (25:24:1) extractions as described previously [61]. For DNA amplification and labelling, 40 µl of immunoprecipitated DNA for each sample was incubated for 20 minutes at 12°C with 70 µl of blunting mix (11 µl 10x NEB #2 buffer, 0.5 µl

BSA (10 mg/ml), 0.5 µl dNTPs (20 mM each dATP, dCTP, dGTP and dTTP), 0.2 µl T4 DNA polymerase and 57.8 µl ice-cold water) followed by the previously described DNA extraction protocol [61]. Extracted DNA for each sample was resuspended in 25 µl ice-cold water and incubated overnight at 16°C with 25 µl of ligase mix (8 µl water, 10 µl 5X ligase buffer, 6.7 µl annealed linkers, 0.5 µl T4 DNA ligase) [61]. Amplification PCR and indirect labeling with the appropriate dye, Cy5 or Cy3, was carried out as described [61] with the exception that OneTaq was used instead of Taq polymerase, and Vent polymerase used instead of Pfu polymerase. ChIP DNA from tagged strains was labeled with Cy5 dye and ChIP DNA from untagged strains SN148 was labeled with Cy3 dye, and were then co-hybridized to Agilent 8X15K whole genome arrays containing 14490 60-mer intergenic and intragenic oligonucleotide probes. Microarray hybridization, washing, scanning, and normalization were performed as described [62] with the following modifications: scanning was done using the Axon GenePix 4000B microarray scanner, and data analyses and normalizations were done using GenePix data analysis software and Multiexperiment Viewer (MeV) software, with a 0.05 P-value cut-off for MeV analyses. Scanning settings used were 635 nm for Cy5 and 532 nm for Cy3. The log of ratios Cy5/Cy3 (635 nm/532 nm) with a cut off of at least 1.5 for each spot was considered to be an indicator of significant Ppr1 binding. Data can be accessed online under the accession number GEO: GSE80343.

2.4.4 Transcriptional Profiling Experiments

Transcriptional profiling experiments were carried out with strain SCPPR1GAD1A (Ppr1 gain of function mutant) and the background strain SC5314 (wild type). Fresh colonies of each strain were each inoculated in 50 ml YPD and incubated overnight at 30°C on a 220 rpm shaker. The overnight cultures were diluted to OD600 of 0.1 in YPD and grown to an OD600 of 0.8. Total RNA was extracted using the QIAGEN RNeasy minikit protocol. Cell pellets were resuspended in 1 ml RNeasy buffer RLT with 10 µl β-mercaptoethanol, then 0.6 ml of acid-washed glass beads (425 - 600 µm) were added before homogenizing each sample five times for a minute each at 4,200 RPM using a Bead beater, with five minute breaks on ice between homogenization steps. RNA was extracted from the collected lysate following the Qiagen RNeasy protocol. RNA quantity was determined using a NanoQuant machine. Twenty µg of total RNA from each sample was reverse transcribed using oligo(dT)23VN and Superscript III reverse transcriptase (Invitrogen) in the presence of Cy3 or Cy5 to achieve direct dye incorporation, and

dye swaps were employed for each sample. After cDNA synthesis, template RNA was degraded by adding RNase A (Sigma) to a final concentration of 0.05 mg/ml and 0.05 units/µl RNase H (New England Biolabs) to each sample and incubating for 30 minutes at 37°C before purifying the labeled cDNAs with a QIAquick PCR Purification Kit (QIAGEN). Hybridization, washing, scanning, and normalization were performed as described [62], with the exception that scanning used an Axon GenePix 4000B microarray scanner, and data analyses and normalizations were done using GenePix data analyses software. Scanning settings were 635 nm for Cy5 and 532 nm for Cy3. The median of ratios mutant-Cy5/non-tagged-Cy3 or Mutant-Cy3/non-tagged-Cy5 values were statistically analyzed in the MultiExperiment Viewer (MeV) software using a *p*-value cut-off at 0.05. Positive significant genes (upregulated genes) were candidates for Ppr1 regulation. Data can be accessed online under the accession number GEO: GSE80343.

2.4.5 Bioinformatics

A *C. albicans* whole genome motif scan for the *S. cerevisiae* Ppr1 motif was done using the Candida Genome Database (CGD) tool "Patmatch" (http://www.candidagenome.org/cgi-bin/PATMATCH/nph-patmatch) [63]. Motif scanning was performed using the software package MotifLab [64] for Ppr1 and Dal82 motifs, analyzing 500 bp promoter regions of the appropriate orthologous genes in different Ascomycetes species. Gene ontology analyses were done using the CGD tool "Go Term Finder" (http://www.candidagenome.org/cgi-bin/GO/goTermFinder) [63]. DNA binding motifs were identified from ChIP-Chip data using the online motif search tool MEME (http://meme-suite.org/tools/meme) [65]. Fifty sequences corresponding to the top Ppr1 target genes identified by ChIP-Chip with log of ratios ranging from 2.70 to 9.08 were used for further analysis. We used 500 bp promoter region of the top 50 genes to screen for potential motifs *de novo* using the online motif search tool MEME (http://meme-suite.org/tools/meme) [65]. The validity of predicted motifs was confirmed by performing motif overrepresentation analysis implemented in MotifLab [64]. In this analysis, the upregulated gene list is considered as 'target' and the rest of the genes in the *C. albicans* genome is considered as 'control' to compare the distribution of motif(s) between the two. The hypergeometric statistical test followed by Bonferroni correction as implemented in MotifLab [64] was applied to establish the validity of the candidate motif. Similarly, a potential motif for non-Ppr1 regulated URA genes (*URA1*, *URA2*, *URA3*, *URA4*, *URA6*, *URA10*) was predicted based on 500 bp promoter regions of *C. albicans*, *C. dubliniensis*, *C. tropicalis* and *C.*

parapsilosis using MEME [65] followed by motif overrepresentation analyses for each species. Fungal blast of *C. albicans* Ppr1 was done using the Saccharomyces Genome Database (SGD) fungal blast tool (http://yeastgenome.org/blast-fungal) [66]. Protein sequences were aligned using the SIM Alignment Tool (http://web.expasy.org/sim/) [67]. The LALNVIEW program was used to generate graphical representations of protein alignments [68].

2.4.6 Allantoin utilization assay

The *ppr1* null mutant strain, SCPPR1GAD1A (Ppr1 gain of function strain) and SC5314 (wild type strain) were each cultured in 11.7 mg/ml Yeast Carbon Base (YCB) (Sigma-Aldrich), YCB + allantoin (10 mM), YCB + ammonium sulfate (37.82 mM) and YCB + urea (10 mM) for 3 days at 30°C in a Sunrise plate reader (TECAN) with a 220 rpm shaker. The plate reader was set to collect OD600 data every 10 minutes throughout the incubation period.

2.5 Results

2.5.1 Bioinformatic analysis of Ppr1 function in *C. albicans* uracil biosynthesis

An apparent ortholog of the *S. cerevisiae PPR1* gene is encoded by the *C. albicans* gene C5_04970C_A (*PPR1*). These two orthologs have about 40% sequence identity throughout a 500 amino acid N-terminal region containing the zinc cluster domain (Figure 2.1). The second best hit identified for CaPpr1 in *S. cerevisiae* was Stb5, which has a stretch of 100 amino acids with just over 30% identity located outside of the region containing the zinc cluster domain. *ScPPR1* and C5_04970C_A thus represent reciprocal best hits, supporting C5_04970C_A as the Ppr1 ortholog in *C. albicans*.

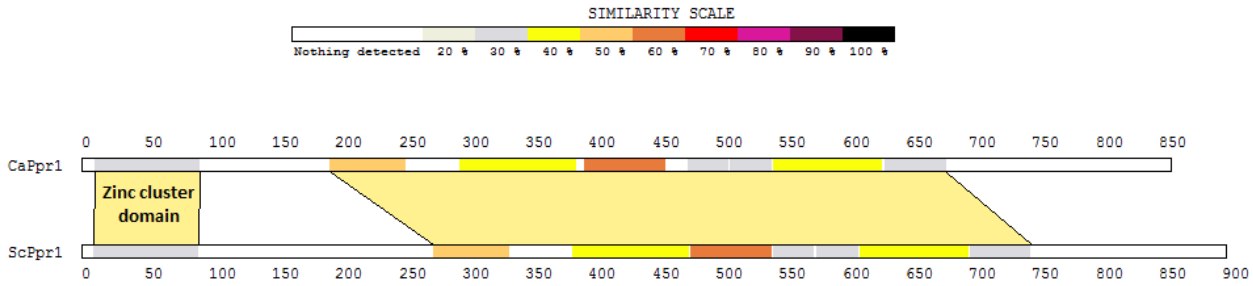


Figure 2.1: *C. albicans* Ppr1 aligned with *S. cerevisiae* Ppr1. The left shaded alignment area has 38.10% identity and includes the zinc cluster domain. The bigger shaded alignment area to the right has 40.73% identity.

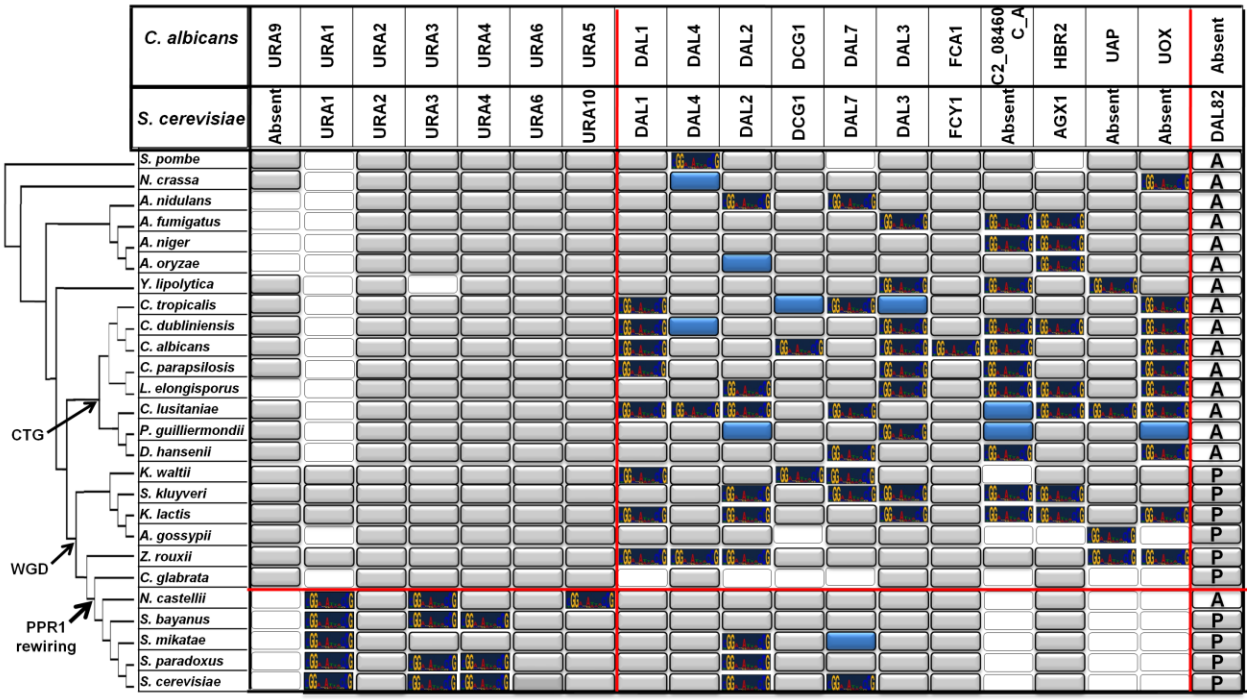
We used a bioinformatic approach to examine whether these proteins perform similar functions. Ppr1 regulates URA genes (*URA1*, *URA2*, *URA3*, *URA4*, *URA6* and *URA10*) in *S. cerevisiae* using the Ppr1 binding motif 5'-CGG(N_6)CCG-3' [56,69]. Motif scanning identified convincing Ppr1-binding sites at the promoter of URA genes in the Saccharomyces lineage from *S. cerevisiae* to *N. castellii*, but these sites were not found at the promoter of URA genes in *C. albicans* and other non-Saccharomyces fungi (Figure 2.2A). Rather, the Ppr1-binding site was found at the promoter of purine catabolism genes in the non-Saccharomyces species (Figure 2.2A and 2.2B). This observation suggested that the regulatory circuit controlled by Ppr1 differs between members of the Saccharomyces lineage and other fungi. Analysis of potential motifs in the promoter regions of the *C. albicans* URA genes has identified a possible enriched motif GryTkGwnTGGT in the promoters of several of the URA genes of Candida species and their close relatives (Figure 2.3).

2.5.2 Direct identification of *C. albicans* Ppr1 target genes using ChIP-Chip

Chromatin immunoprecipitation followed by microarray analysis (ChIP-chip) is a powerful technique for identifying the direct binding targets of transcription factors. For our analysis, a TAP epitope sequence was fused to the endogenous *PPR1* gene in strain SN148; correct DNA fusion was established using PCR and proper TAP-tagged protein expression was confirmed using immunoprecipitation. Chip results from the TAP-tagged strain were consistent with ChIP-Chip and microarray results from an HA-tagged Ppr1 strain; this, together with the phenotype of a *ppr1* null mutant, implied that tagging Ppr1 did not affect its function. After chromatin cross-linking, target binding sequences were identified by amplifying and labeling immunoprecipitated DNA sequences and hybridizing these labeled sequences to Agilent 8X15K *C. albicans* whole genome tiling arrays. Ppr1 target genes were ranked based on their log of

ratios (PPR1-TAP-Cy5/non-tagged-Cy3) values (Table 2.1). Gene ontology assessment of genes corresponding to the top 150 targets with a log of ratios cut-off of at least 1.5 showed enrichment in genes involved in purine nucleobase catabolism (Table 2.1) suggesting that Ppr1 in *C. albicans* may transcriptionally regulate these genes.

A



B

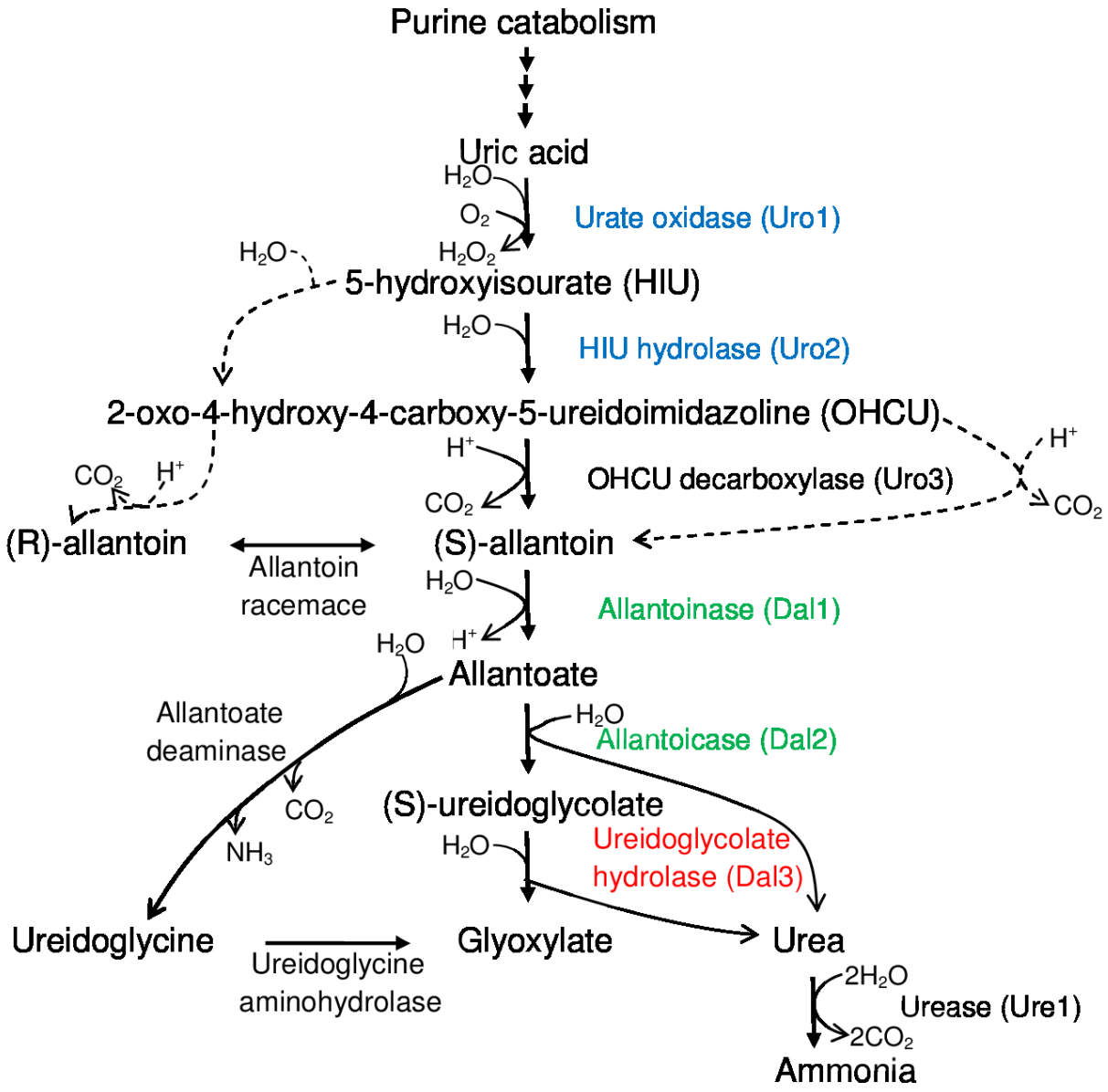


Figure 2.2: Purine catabolism pathway, Ppr1 motif occurrence in the promoters of *de novo* pyrimidine biosynthesis and purine catabolism genes, and the presence or absence of *DAL82* genes across the Ascomycetes.

A) Ppr1 motif occurrence in the promoters of *de novo* pyrimidine biosynthesis (URA) and purine catabolism (includes DAL) genes, and the presence or absence of *DAL82* genes across the Ascomycetes. Dark blue with motif logo represents Ppr1 motif; Grey boxes indicate gene presence; White boxes indicate gene absence; Blue box implies sequence is not available in databases; A= Dal82 Absent; P= Dal82 Present. The phylogenetic tree is based on [70] and Fungal Orthogroups Repository (http://www.broadinstitute.org/regev/orthogroups/).

B) Purine catabolism pathway [71]. Blue enzymes: present in *C. albicans* and regulated by Ppr1 but are missing in *S. cerevisiae*; Green enzymes: regulated by both Ppr1 in *C. albicans* and by Dal82 in *S. cerevisaie*. Red enzymes: regulated by Ppr1 in *C. albicans* and present in *S. cerevisiae* but not regulated by by Dal82 in *S. cerevisiae*. The dashed lines represent non-enzymatic reactions.

2.5.3 Ppr1 DNA binding sequence in *C. albicans*

To identify a potential DNA binding motif for Ppr1 in *C. albicans*, the DNA sequences 220 bp upstream and 220 bp downstream of each positive 60 bp probe, (a total of 500 bp per site), were analyzed [65]. The most probable motif for *C. albicans* Ppr1 detected from the top 50 positive genes (Table S2) was CGG-X_6-CCG (E-value=4.6E-004) (Figure 2.4A), which is identical to the reported motif for *S. cerevisiae* Ppr1 (Figure 2.4B) [56,69]. The motif was further confirmed by motif occurrence analysis that showed significant overrepresentation of the motif in target genes compared to the control (p-value 7.62×10^{-13}). Thus the Ppr1 proteins of the two species bind the same DNA sequence, but this binding controls distinct functions.

2.5.4 Transcriptional profiling experiments using a Ppr1 gain of function mutant

ChIP-Chip and bioinformatics analysis predicted that Ppr1 in *C. albicans* is a transcriptional regulator of genes involved in purine catabolism.

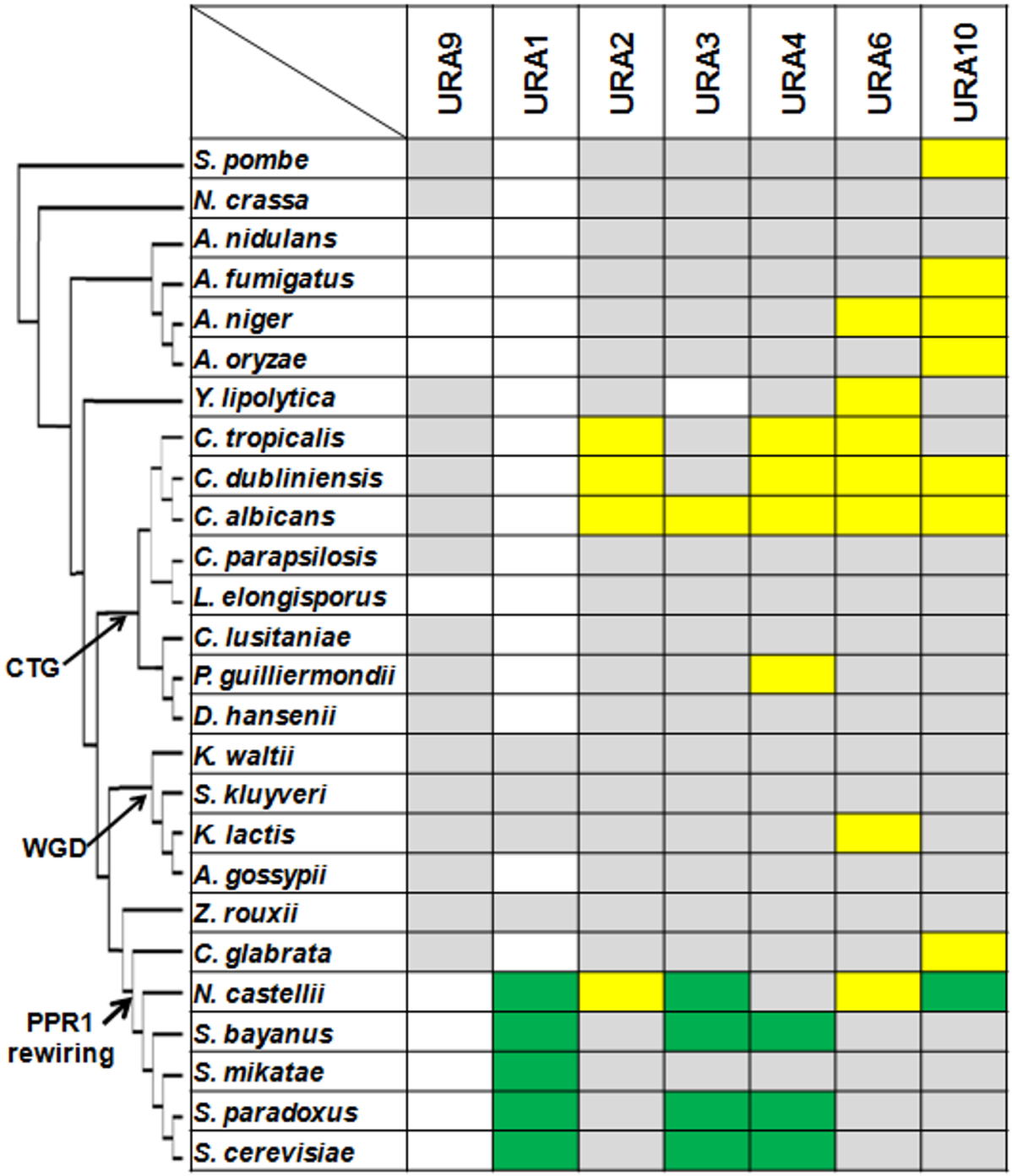


Figure 2.3: Occurrence of Ppr1 and non-Ppr1 motifs across the Ascomycetes fungi.

The Ppr1 motif (green boxes) is present only throughout Saccharomyces species while an alternative GryTkGwnTGGT-like motif (yellow boxes) is present across the non-Saccharomyces species. The presence and absence of a promoter sequence is represented by grey and white boxes respectively. Motif scan was performed at the 90% threshold using the software package Motiflab. The phylogenetic tree is based on [70] and Fungal Orthogroups Repository (http://www.broadinstitute.org/regev/orthogroups/).

Table 2.1 C. albicans Ppr1 binding sites identified by ChIP-Chip

ChIP-Chip data was analyzed using GenePix and MeV, and the top 15 Ppr1 target genes as well as genes in the top 50 ChIP-Chip hits that are involved in purine catabolism or also appear in the transcriptional profiling results are presented in this table. Genes in black font appear in ChIP-Chip only, genes in red font appear as both ChIP-Chip and transcriptional profiling hits, and highlighted genes are involved in purine catabolism. *URA3* frequently appears in ChIP-Chip data of our Tap-tagged strains when transformants are selected using the *URA3* selectable marker; the reason for this is unclear. Primers used for transformation in this study are presented in Table S1.

Gene	orf	Log of ratio Cy5/Cy3	Binding Rank	p-Value	Function/Putative function
	orf19.230	9.08	1	2.76E-06	Ortholog (Pet111) is a mitochondrial translational activator specific for COX2 mRNA
MRV1	orf19.4691	6.87	2	0	Uncharacterized. No predictions.
GCN4	orf19.1358	6.19	3	7.81E-04	bZIP transcription factor; amino acid control response
	orf19.2065	5.42	4	6.90E-07	Ortholog (DAL2) involved in allantoin catabolism
	orf19.2928	5.32	5	0.001922015	Ortholog(s) have COPI-coated vesicle, Golgi apparatus localization
URA3	orf19.1716	4.80	6	6.90E-07	Orotidine-5'-phosphate decarboxylase; pyrimidine biosynthesis
	orf19.77.1	4.73	7	6.21E-06	Protein of unknown function
ETR1	orf19.5450	4.52	8	2.07E-06	Putative 2-enoyl thioester reductase
	orf19.6272	4.36	9	2.76E-06	Putative ubiquitin C-terminal hydrolase
ARG4	orf19.6689	4.18	10	0.010551414	Argininosuccinate lyase, catalyzes the final step in the arginine biosynthesis pathway
	orf19.6480	4.05	11	0.011303658	Ortholog(s) have role in UDP-glucose transport and COPI-coated

					vesicle
SEC1	orf19.6479	4.05	11	0.011303658	Ortholog(s) have SNARE binding activity and role in upregulation of vesicle fusion
	orf19.1043	4.04	12	4.83E-06	Ortholog(s) have lysophosphatidic acid acyltransferase activity
	orf19.155	3.89	13	0.001272602	Functional homolog (Ure2p) regulates nitrogen utilization
PWP1	orf19.4640	3.81	14	3.37E-04	Putative rRNA processing protein
DAL1	orf19.5454	3.81	15	0.006016563	Putative allantoinase
HBR2	orf19.1078	3.63	18	1.52E-05	Putative alanine glyoxylate aminotransferase
	orf19.1681	3.62	19	1.59E-05	Closest protein in A. nidulans involved in purine utilization
AHP2	orf19.6470	3.61	20	1.45E-05	Putative thiol-specific peroxiredoxin
	orf19.2114	3.49	24	1.52E-05	Predicted uricase
	orf19.4889	3.32	29	7.45E-04	Predicted MFS family membrane transporter, member of the drug:proton antiporter
FCA1	orf19.4195.1	3.11	32	4.55E-05	Cytosine deaminase; enzyme of pyrimidine salvage
	orf19.1421	2.91	37	9.73E-05	Ortholog (DAL3) involved in allantoin catabolism
	orf19.3633	2.84	41	1.68E-04	Ortholog(s) have role in purine nucleobase catabolism
CEF3	orf19.4152	2.80	44	2.45E-04	Translation elongation factor 3
	orf19.5449	2.63	56	2.46E-04	Predicted integral membrane protein
	orf19.6316	2.54	62	4.55E-04	Predicted membrane transporter
ALG11	orf19.3468	2.47	68	5.85E-04	Alpha-1,2-mannosyltransferase
	orf19.313	2.244	93	5.54E-04	Putative allantoin permease
	orf19.1105	2.23	96	6.23E-04	Uncharacterized. No predictions.
	orf19.1709	2.22	97	6.83E-04	Sterol carrier domain protein
SAM50	orf19.7358	2.14	110	0.002341615	Predicted part of SAM complex,

					mitochondrial protein import

We transcriptionally profiled a *C. albicans* Ppr1 gain of function (GOF) mutant generated by addition of a Gal4 activation domain that artificially activates the TF [5], to establish the expression network of the transcription factor. Such GOF mutant strains can up-regulate networks in the absence of any stimulatory conditions. As shown in Table 2.2, transcriptional profiling experiments with the Ppr1 (GOF) mutant strain SCPPR1GAD1A identified six genes involved in purine catabolism among the top 11 overexpressed genes. These genes, which include a uricase and a variety of genes involved in allantoin catabolism, also feature among 17 Ppr1 target genes appearing in the ChIP-Chip top 150 hits as well. *PPR1* itself was presumably identified among the up-regulated genes because the hyperactive-allele-construct was driven by the strong *ADH1* promoter.

2.5.5 Allantoin utilization

To test for the role of *PPR1* in purine catabolism predicted by the ChIP-Chip and transcriptional profiling experiments, we cultured *ppr1* null mutants, control SC5314 (wild type *PPR1*) and SCPPR1GAD1A in yeast carbon base (YCB) media with various nitrogen sources.

A

B

Figure 2.4: Ppr1 DNA binding sequence in *C. albicans* and *S. cerevisiae*.

A) *C. albicans* Ppr1 DNA binding motif based on top 50 Ppr1 binding sites from ChIP-Chip. ChIP-Chip identified sequences used to identify *C. albicans* Ppr1 DNA binding motif are presented in Table S2.

B) *S. cerevisiae* Ppr1 DNA binding motif [57].

The *ppr1* null mutant strain showed almost no growth after three days on allantoin as the sole nitrogen source, while SC5314 and SCPPR1GAD1A showed normal growth to saturation (Figure 2.5A). This suggests that *PPR1* is required for the expression of genes involved in allantoin catabolism. The GOF strain has a shorter lag phase than the wild type (Figure 2.5A), perhaps because the hyperactive Ppr1 more efficiently expresses genes required for growth on allantoin. All three strains were also cultured in YCB + ammonium sulfate as the sole nitrogen source (Figure 2.5B), YCB + urea as sole nitrogen source (Figure 2.5C), and YPD as controls. All strains showed normal and similar growth rates up to saturation in three days, suggesting that lack of growth of *ppr1* null mutants in YCB + allantoin is due to the inability of the strain to break down allantoin.

Table 2.2: *C. albicans* Ppr1 upregulated genes identified by expression transcriptional profiling.
Transcriptional profiling data were analyzed using GenePix and MeV, and the top 15 Ppr1 upregulated genes as well as genes appearing both in transcriptional profiling results and in the top 150 ChIP-Chip hits are presented in this table. Genes in black appear in transcriptional profiling hits only, genes in red or green appear in both ChIP-Chip and transcriptional profiling hits, promoters of genes in green are predicted by MEME to have the CGG(N_6)CCG DNA binding motif, and highlighted genes are involved in purine catabolism.

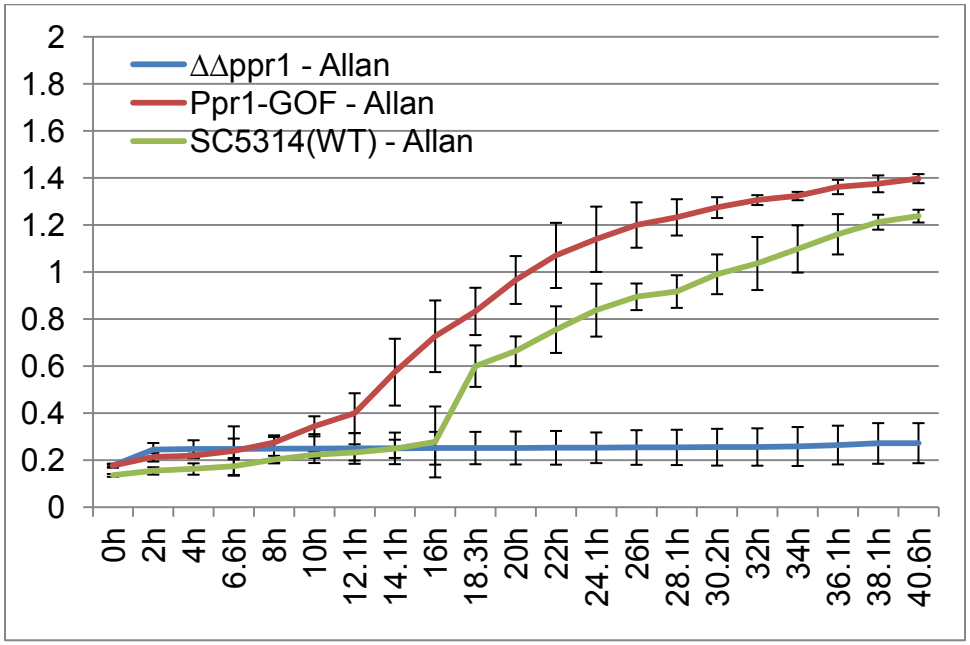
Gene	Orf	Median of ratios	Upregulation rank	p-Values	Function/Putative function
PPR1	orf19.3986	21.879	1	0	ScPPR1 regulates uracil biosynthesis genes
	orf19.2114	16.7755	2	1.55E-06	Predicted uricase
	orf19.1421	16.239	3	1.55E-06	Ortholog (DAL3) involved in allantoin catabolism
	orf19.2065	14.387249	4	1.55E-06	Ortholog (DAL2) involved in allantoin catabolism
	orf19.3458	12.45425	5	1.66E-04	Ortholog(s) involved in late endosome to vacuole transport
DAL1	orf19.5454	11.89975	6	3.10E-06	Putative allantoinase
AHP2	orf19.6470	4.6359997	7	3.56E-05	Putative thiol-specific peroxiredoxin
	orf19.6117	4.5480003	8	0.0082376	predicted auxin family transmembrane transporter ortholog
	orf19.1681	4.19625	9	5.10E-04	Closest protein in A. nidulans involved in purine utilization
	orf19.5713	3.92875	10	6.55E-04	Putative NADH dehydrogenase
	orf19.3633	3.807	11	6.78E-04	Ortholog(s) have role in purine nucleobase catabolism
	orf19.772.1	3.651	12	0.0091529	Uncharacterized. No predictions.
	orf19.2796	3.5855	13	0.010271	Ortholog(s) have DNA polymerase, DNA replication activity.
	orf19.413	3.49225	14	0.0066362	Protein of unknown function

SET2	orf19.1755	3.4105	15	0.0092396	Ortholog(s) have histone methyltransferase activity
ALG11	orf19.3468	3.1995	26	9.43E-04	Alpha-1,2-mannosyltransferase
	orf19.5449	2.13325	92	0.0041753	Predicted integral membrane protein
	orf19.6316	2.085	93	0.0051293	Predicted membrane transporter
CEF3	orf19.4152	2.08375	94	0.0151138	Translation elongation factor 3
	orf19.1709	1.89925	107	0.0228512	Sterol carrier domain protein
	orf19.1105	1.85	111	0.0096097	Uncharacterized. No predictions.
FCA1	orf19.4195.1	1.617	132	0.0210485	Cytosine deaminase; enzyme of pyrimidine salvage
	orf19.4889	1.5947499	135	0.0335326	Predicted MFS family membrane transporter
SAM50	orf19.7358	1.5797501	144	0.0272991	Predicted part of SAM complex, mitochondrial protein import
HBR2	orf19.1078	1.5035	154	0.0406117	Putative alanine glyoxylate aminotransferase

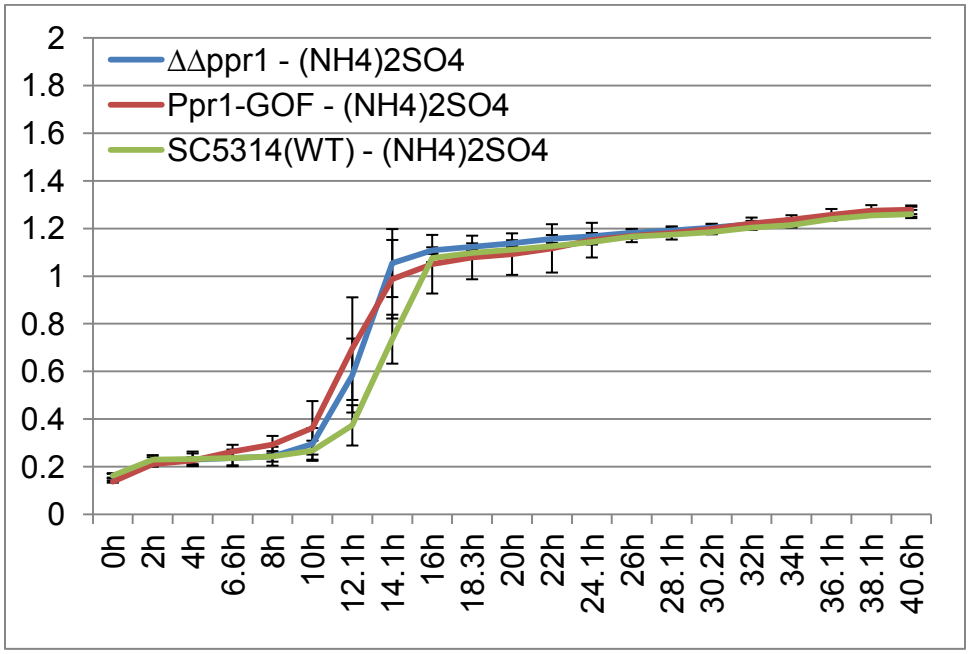
2.5.6 Ppr1 regulation of other functions

In addition to genes implicated in regulation of allantoin catabolism, several other genes appear to be members of the Ppr1 regulon in *C. albicans*. These include genes encoding such diverse functions as membrane transport, amino acid metabolism (aminotransferase function), translation elongation, mitochondrial protein import, sterol transport, glycosylation, antioxidation, and the pyrimidine salvage pathway (Table 2.2). Genes involved in these pathways appear in both transcriptional profiling results and the top 150 ChIP-Chip hits, and the Ppr1 binding motif was identified at the promoter of some of the genes. Phenotype studies will be required to confirm the other predicted Ppr1 functions.

A


ΔΔppr1 - Allan
Ppr1-GOF - Allan
SC5314(WT) - Allan
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0h
2h
4h
6.6h
8h
10h
12.1h
14.1h
16h
18.3h
20h
22h
24.1h
26h
28.1h
30.2h
32h
34h
36.1h
38.1h
40.6h

B


ΔΔppr1 - (NH4)2SO4
Ppr1-GOF - (NH4)2SO4
SC5314(WT) - (NH4)2SO4
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0h
2h
4h
6.6h
8h
10h
12.1h
14.1h
16h
18.3h
20h
22h
24.1h
26h
28.1h
30.2h
32h
34h
36.1h
38.1h
40.6h

C

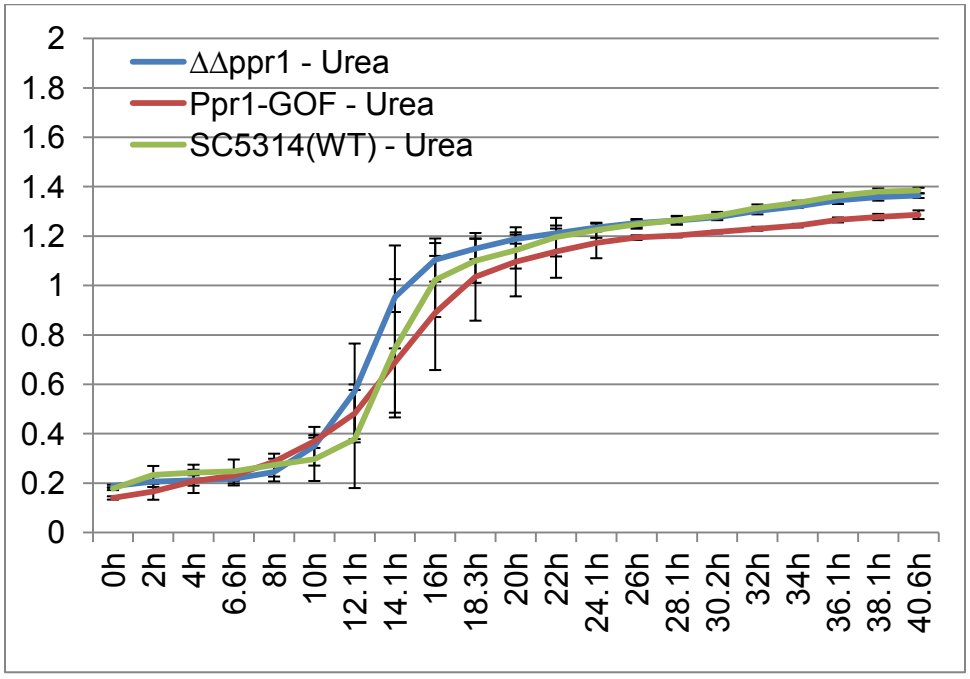


Figure 2.5: Allantoin, ammonium sulfate and urea utilization assays of a *C. albicans ppr1* null mutant strain, a Ppr1 gain of function mutant strain and the *C. albicans* wild type strain SC5314. Error bars are based on the standard deviation of the seven replicates of each data point reading taken every two hours. Primers used for transformation in this study are presented in Table S1.

A) Growth curve of *ppr1* null mutant (ΔΔppr1) strain, Ppr1 gain of function mutant (Ppr1-GOF) strain and SC5314 (WT=wild type) strain in yeast carbon base (YCB) medium with allantoin as the sole nitrogen source.

B) Growth curve of *ppr1* null mutant strain, Ppr1 gain of function mutant strain and SC5314 strain in yeast carbon base medium with ammonium sulfate as the sole nitrogen source.

C) Growth curve of *ppr1* null mutant strain, Ppr1 gain of function mutant strain and SC5314 strain in yeast carbon base medium with urea as the sole nitrogen source.

2.6 Discussion and conclusions

The zinc cluster transcription factor Ppr1 was initially identified in the yeast *S. cerevisiae* as a regulator of pyrimidine (uracil) biosynthesis, through its interaction with the motif 5'-CGG(N_6)CCG-3' found in the promoters of many of the regulated genes [56,69]. Our studies using motif searches, ChIP-Chip analysis and transcriptional profiling suggests that pyrimidine biosynthesis is not under the regulation of Ppr1 in *C. albicans*, and the Ppr1 null mutant strain grows normally on media without uracil. No specific transcription factor has been associated with the URA regulon outside of the Ppr1 regulated species such as *S. cerevisiae*, and it is possible that pyrimidine synthesis is not transcriptionally regulated in these strains. Ppr1 ChIP-Chip and

transcriptional profiling experiments done using SCPPR1GAD1A (Ppr1 in SCPPR1GAD1A strain is HA-tagged) show that Ppr1 does not bind and does not regulate the *URA3* gene. However, motif searches have identified the sequence GryTkGwnTGGT enriched in the promoters of the genes of the URA regulon of *C. albicans* and its close relatives (Figure 3), and thus some yet to be defined transcription factor may be involved in the regulation of uracil biosynthesis in non-Saccharomyces species.

Although Ppr1 does not regulate the URA pathway in *C. albicans*, it has a clear function in this species. The *C. albicans* ortholog of Ppr1, encoded by the gene C5_04970C_A, uses the 5'-CGG(N_6)CCG-3' binding motif to control a regulon involved in purine catabolism, particularly allantoin degradation. ChIP-Chip results suggest that Ppr1 binds to the promoters of allantoin degradation and other purine catabolism genes (Table 2.1), and the upregulation of these target genes as shown in the transcriptional profiling results (Table 2.2) implies that Ppr1 binding leads to transcriptional activation of the target genes. The role of Ppr1 in purine catabolism was further confirmed by the inability of a *ppr1* null mutant *C. albicans* strain to grow in YCB medium with allantoin as the sole nitrogen source (Figure 2.5A); these findings also suggest that Ppr1 could be the sole regulator of purine catabolism in the pathogen. Such dramatic rewiring of transcription factor function has been seen elsewhere within the fungi; examples include Gal4 rewiring [15], Mcm1 regulating cell cycle and mating in most yeast species, but co-regulating ribosomal gene expression with Rap1 in the *C. glabrata*, *K. lactis*, *Y. lipolytica*, and the *A. nidulans* lineages [33], and ribosomal protein gene expression regulated by Tbf1 in *C. albicans* but Rap1 in *S. cerevisiae* [31].

Phylogenetic examination of the candidate motifs upstream of the allantoin degradation regulon and the URA biosynthesis regulon suggests that the shift of the Ppr1-binding motif 5'-CGG(N6)CCG-3' from the promoters of allantoin catabolism genes to pyrimidine biosynthesis genes occurred in the lineage leading to *N. castellii* and the other Saccharomyces species. As shown in Figure 2a, there is essentially no evidence for the motif in the URA gene promoters in species outside the Saccharomyces species, but strong evidence for the motif in many of the genes within these species. By contrast, the motif is commonly associated with the allantoin degradation pathway genes in many fungi outside of the Saccharomyces clade.

The timing of the Ppr1 rewiring event is intriguing, as this phylogenetic divergence corresponds to a massive genomic reorganization of the genes involved in allantoin degradation.

There is an initial scattered distribution in the non-Saccharomyces that shifts to a clustered distribution (the Dal cluster) in the Saccharomyces species [72]. In the clustered regulons transcriptional regulation of allantoin degradation is under the control of Dal82, a transcription factor that arose prior to the whole genome duplication but is not present in species such as *C. albicans*. In *S. cerevisiae* Dal82 forms a heterodimer with Dal81, and the complex binds sequences known as UISALL elements or Dal82 motifs (5'-AAANTTGCG-3') [73-75] to regulate allantoin degradation. Dal82 is induced to activate allantoin degradation genes in *S. cerevisiae* by allophanate, the last intermediate in the allantoin degradation pathway [73]. Although Dal81 is present in both Candida and Saccharomyces species, its role is different in *C. albicans* where it apparently positively regulates cell adhesion [76], and is therefore unlikely to interact with Ppr1 to regulate allantoin degradation. The presence or absence of the Dal82 motif at Dal gene promoters across yeast species suggests that Dal82 may have taken over purine catabolism in the lineage leading to *N. castellii* and relatives at the same time Ppr1 changed its function. In *S. cerevisiae*, nitrogen catabolite pathways are also under the positive regulation of Gln3 and Gat1 and the negative regulation of Dal80 and Deh1 which all bind 5' GATAA 3' sequences to activate or repress genes. Bioinformatic analysis of a candidate nitrogen catabolite repression circuit found that while the members of the circuit can vary among ascomycetes, there is no evidence that this circuit undergoes any transition corresponding to the switch from Ppr1 to Dal82 [77]. Concomitant with the Ppr1 rewiring event was the elimination of some oxygen consuming functions from both the purine catabolism and pyrimidine biosynthesis pathways [72]. The uric acid importer, uric acid permease, together with urate oxidase which uses molecular oxygen to break down uric acid, were both eliminated from the genome of *S. cerevisiae* and close relatives, together with other oxygen-demanding functions [72]. A specific allantoin permease gene, *DAL4*, arose from the duplication of the uracil permease gene, *FUR4*, at the same phylogenic time as the DAL cluster formation and Ppr1 rewiring [72]. Consequently, the Saccharomyces species shifted from importing uric acid, which demanded oxygen for its metabolism, to importing allantoin [72]. Allantoin is an important plant nitrogen storage and transport molecule, and thus would be relatively abundant in the yeast environment. The loss of these oxygen demanding enzymes at this point enhanced the ability of Saccharomyces species to grow effectively in hypoxic conditions. This strategy to minimize oxygen consumption through genomic editing is possibly the reason behind the clustering of the DAL genes [72], and is likely

the driving force that led to the the rewiring of Ppr1 and the introduction of the allophanate-induced Dal82 to the purine catabolism pathway. Uric acid induces the Ppr1 orthologs of *A. nidulans* and *N. crassa*, UaY and PCO-1 respectively, which in both cases lead to the activation of purine catabolism genes [78,79]. Dihydroorotic acid, which is an intermediate in the pyrimidine biosynthesis pathway, has been shown to induce at least one of the activities under control of UaY in *A. nidulans* [78]. The switch from importing uric acid to importing allantoin in the Saccharomyces lineage eliminated the original Ppr1 inducer, uric acid. This could allow Ppr1 to respond to the weaker inducer dihydroorotic acid, resulting ultimately in the switch from the regulation of purine catabolism to the regulation of *de novo* pyrimidine biosynthesis. The Ppr1 function in purine catabolism regulation could be replaced by Dal82, a transcription factor induced by allophanate that is not present in species such as *C. albicans,* in the Saccharomyces lineage. Alternatively, a switch to a low oxygen environment could have blocked the availability of uric acid as a nitrogen source and made the oxidase and permease dispensable in the Saccharomyces species. Overall, these findings suggest that Ppr1 transitioned from regulating purine catabolism to the regulation of pyrimidine biosynthesis at the lineage leading to *N. castellii* and the other Saccharomyces species which coincides with the clustering of Dal genes, the onset of Dal82-regulated purine catabolism and the appearance of fungi with a facultative aerobic metabolism.

Chapter 3: Stb5 is a transcriptional regulator of genes involved in echinocandin resistance and the glyoxylate cycle in *Candida albicans*

3.1 Preface

The work presented in Chapter 3 is from a manuscript under preparation: Tebung, W. A., Morschhäuser, J. & Whiteway, M., (2016). **Stb5 is the transcriptional regulator of genes involved in echinocandin resistance and the glyoxylate cycle in *Candida albicans*.** Target journal: Molecular Microbiology.

Author contributions:

Walters Aji Tebung performed experiments, bioinformatic analyses, wrote paper; Joachim Morschhäuser provided activated transcription factor reagents; Malcolm Whiteway designed experiments, wrote paper.

3.2 Abstract of manuscript

Resistance to drugs is a prevalent problem posed by *Candida albicans* during the treatment of candidiasis. Understanding and countering the mechanisms used by *C. albicans* to fight off drugs is key to reducing the ever-rising annual deaths caused by the pathogen. Stb5 is a zinc cluster transcription factor (ZCF) and one of six STB (Sin Three Binding) proteins initially identified in *Saccharomyces cerevisiae*. Our findings using ChIP-Chip, transcriptional profiling, nutrient utilization assays and drug response assays suggest that the Stb5 ortholog in *C. albicans* regulates drug resistance genes, as well as components of the glyoxylate cycle. The role *C. albicans* Stb5 plays in the regulation of drug pumps correlates with a major role of Stb5 in *S. cerevisiae* where it regulates drug efflux pumps. Although some weak homodimerization has been reported for Stb5 in *S. cerevisiae*, Stb5 mostly binds DNA as a heterodimer with Pdr1 to regulate drug pumps in yeast. The absence of a Pdr1 ortholog in *C. albicans* suggests that Stb5 may have ancestrally regulated drug resistance either as a homodimer or as a heterodimer with a transcription factor other than Pdr1, but evolved in *S. cerevisiae* to recruit the recently arisen Pdr1 for this purpose. Actively pumping out drugs is a common mechanism used by *C. albicans* to evade drugs, making Stb5 a potential target against drug resistance. Stb5 apparently dropped its acetate catabolism regulation role at some point between *C. albicans* and *S. cerevisiae,* but took on other roles such as regulation of the expression of the pentose phosphate pathway genes, the

control of NADPH production for ensuring resistance to oxidative stress and the ergosterol biosynthetic pathway.

3.3 Introduction

Pathogenic organisms are constantly in a battle to survive adverse conditions in the host. Some of the challenges faced by pathogens range from the need to fight off the host immune system, to resist drugs, and to adapt to alternate nutrient sources when preferred sources are limited or absent. *C. albicans* is a successful commensal and opportunistic pathogen that is a major cause of serious fungal infections of humans [80]. Current treatment of such infections (candidiasis) rely on the use of three major drug classes - the azoles, the polyenes and the echinocandins [81], but a growing resistance to these drugs by *C. albicans* calls for a need to identify novel treatment options, and to identify and counter mechanisms used by the pathogen to build resistance to existing drugs. A common azole resistance mechanism in *C. albicans* involves the induction of drug efflux pumps encoded by CDR or MDR genes, which results in the cells actively pumping out azoles to minimize the effect of the drug [82]. The newest class of antifungals, the echinocandins, have a broad spectrum of activity against Candida species [83], but *C. albicans* have been reported to build resistance against this class of antifungals through point mutations in the FKS genes encoding the major subunit of its target enzyme [82], and by actively pumping out echinocandins using Cdr2 [84], and, to a lesser extent, Cdr1 [84]. In addition to its ability to resist antifungals, *C. albicans* also thrives as a pathogen due to its ability to quickly adapt to alternate nutrient sources in the host [85], and its ability to evade the host immune mechanisms [86]. It has been reported that in low glucose environments, *C. albicans* relies on the glyoxylate cycle for survival [87]. The pathogen is able to grow in low nutrient milieus such as inside macrophages and neutrophils by utilizing alternate carbon sources through pathways such as the glyoxylate cycle [85], and has been shown to escape macrophages by switching to hyphal growth and breaking out of the phagocytic cell [88]. The role of the glyoxylate cycle in *C. albicans* pathogenesis has drawn recent attention to enzymes of the pathway such as Icl1 as potential drug targets [85].

We report in this study a zinc cluster transcription factor, Stb5, which regulates the glyoxylate cycle, and confers resistance to echinocandins in *C. albicans* potentially by transcriptionally activating the expression of drug efflux pumps. Stb5 was first characterized in *S. cerevisiae* and is one of the six STB (Sin Three Binding) proteins [89]. STB proteins in *S.*

cerevisiae interact with the transcriptional repressor protein Sin3, hence the name STB. Stb5 in *C. albicans*, however, has so far not been linked to *C. albicans* Sin3 protein and has only adopted its name from the *S. cerevisiae* Stb5 ortholog. In *S. cerevisiae,* Stb5 ZCF is known to regulate multidrug resistance genes [90], genes of the pentose phosphate pathway, genes implicated in NADPH production for ensuring resistance to oxidative stress [91] and genes of the ergosterol biosynthetic pathway. Stb5 has therefore conserved a role in drug resistance between *C. albicans* and *S. cerevisiae,* but has lost its role in the regulation of the glyoxylate cycle which in *S. cerevisiae* now seems to be under the regulation of Sip4. Identifying a drug that can inhibit Stb5 would shut down the glyoxylate cycle making *C. albicans* less virulent, and would shut down drug efflux pumps rendering the pathogen more sensitive to antifungal drugs. Our findings therefore propose Stb5 as a potential drug target against *C. albicans* pathogenesis and drug resistance.

3.4 Materials and methods

3.4.1 Strains, media, plasmids and transformation

We followed standard procedures for *C. albicans* cell growth and transformation [58]. *C. albicans* strains used for transformation, immunoprecipitation, ChIP-Chip analyses and transcriptional profiling experiments were cultured in YPD at 30°C. Yeast nitrogen base medium (YNB) at 6.7 mg/ml supplemented with glucose (20 mg/ml) or acetate (19.6 mg/ml) was used for phenotype studies to test for the ability of the *stb5* null mutants to utilize acetate as a carbon source. Other control media used for *stb5* null mutant phenotype studies were YNB at 6.7 mg/ml without a nitrogen supplement, and YPD.

The STB5-TAP strain was constructed by transforming the strain SN148 [59,60] using a TAP-URA3 PCR product containing 99 base pairs (bp) of homologous sequence immediately upstream and downstream of the *STB5* stop codon. The TAP-URA3 region of the oligomer used to transform SN148 was amplified from the pFA-TAP-URA3 plasmid [60] using the forward primer STB5_TAP_URA3_F and the reverse primer STB5_TAP_URA3_R. Appropriate genomic integration of the oligonucleotide containing the TAP tag was confirmed by PCR using two primer pairs; forward primer STB5_TAP_URA3_check_F that binds inside the *STB5* gene and reverse primer FT-U3 that binds inside the *URA3* gene, and forward primer FT-U1 that binds inside the *URA3* gene and reverse primer STB5_TAP_URA3_check_R that binds downstream of

STB5. The *stb5* null mutant strain was constructed by replacing one *STB5* allele with the *HIS1* marker and the other allele with the *ARG4* marker in the SN95 strain. Oligonucleotides used for *stb5* null mutant construction were amplified from plasmids pFA-HIS1 for the first allele knockout and pFA-ARG4 for the second allele knockout [60] using the respective forward and reverse primers STB5_Marker_KO_F that has a 97 bp region that is complimentary to the sequence just before the *STB5* start codon and STB5_Marker_KO_R that has a 95 bp region that is complimentary to the sequence just after the *STB5* stop codon. Successful deletion of the first allele was confirmed by PCR using the forward primer STB5_KO_Check_F that binds the promoter region of the *STB5* gene and the reverse primer STB5_KO_Check_R that binds downstream of the gene. Deletion of the second allele was confirmed by PCR using two primer pairs; forward primer STB5_KO_Check_F that binds the promoter region of the *STB5* gene and reverse primer FT-H2 that binds upstream of the *HIS1* gene (*HIS1* promoter region) and forward primer FT-H3 that binds inside the *HIS1* gene and reverse primer STB5_KO_Check_R that binds downstream *STB5*. The *STB5* null mutants were confirmed using the forward primer STB5_KO_Check_Internal_F and the reverse primer STB5_KO_Check_Internal_R that both bind inside the *STB5* gene; as expected, no DNA amplification was found for *stb5* null mutants using this primer pair. The SCSTB5GAD1A strain was as described previously [5]. See table S1 for oligos used in these studies.

3.4.2 Immunoprecipitation

We followed the immunoprecipitation protocol previously described [32] with the following modifications: we incubated 700 µl lysate obtained from the cell culture with IgG Sepharose beads at 4°C overnight on a lab rotator. After protein transfer through Western blotting from a 12% SDS-PAGE gel to a nitrocellulose membrane, the membrane was treated with the primary rabbit polyclonal antibody (1:1000) directed against the TAP-tag, followed by treatment with the secondary anti rabbit antibody coupled to dye IR 680 (1:5000). Protein bands were detected using an Odyssey scanner at a wavelength of 700 nm.

3.4.3 ChIP-Chip

We carried out ChIP-Chip experiments as described (Chapter 2 of this thesis) with a few modifications. Briefly, the strain containing the chromosomally inserted STB5-TAP fusion as well as the background strain SN148 (untagged) were grown to an optical density of 0.6 at 600

nm (OD 600nm) in 50 ml of YPD. Crosslinking for each 50 ml culture was done by incubating with 1.5 ml of 37% formaldehyde for 30 minutes, then ChIP was performed as described (Chapter 2 of this thesis). ChIP DNA obtained from tagged strains was labeled with Cy5 dye and ChIP DNA from untagged strains SC5314 was labeled with Cy3 dye, and were then co-hybridized to Agilent 8X15K whole genome arrays containing 14490 60-mer intergenic and intragenic oligonucleotide probes. Microarray hybridization, washing, scanning, and normalization were carried out as described [62] with the following modifications: Axon GenePix 4000B microarray scanner was used for scanning, and data analyses and normalizations were done using GenePix data analyses software and Multiexperiment Viewer (MeV) software; a 0.05 P-value cut-off was used for MeV analyses. Scanning settings used were 635 nm for Cy5 and 532 nm for Cy3. The log of ratios Cy5/Cy3 (635 nm/532 nm) with a cut off of at least 1.5 for each spot was considered to be an indicator of significant Stb5 binding.

3.4.4 Transcriptional Profiling Experiments

Transcriptional profiling experiments were carried out following the procedures described (Chapter 2 of this thesis). Briefly, single colonies of the strains SCSTB5GAD1A (Stb5 gain of function mutant) and SC5314 (wild type) were each inoculated in 50 ml YPD and incubated overnight at 30°C on a 220 rpm shaker. Overnight cultures were then diluted to OD600 of 0.1 in YPD and grown to an OD600 of 0.8. Total RNA was extracted using the QIAGEN RNeasy minikit protocol and quantified using a NanoQuant machine. Reverse transcription of 20 µg of the total RNA from each sample was done using oligo(dT)23VN and Superscript III reverse transcriptase (Invitrogen) in the presence of Cy3 or Cy5 to achieve direct dye incorporation, and dye swaps were done for each sample. Following cDNA synthesis, template RNA was degraded by adding RNase A (Sigma) to a final concentration of 0.05 mg/ml and 0.05 units/µl RNase H (New England Biolabs) to each sample and incubating for 30 minutes at 37°C before purifying the labeled cDNAs with a QIAquick PCR Purification Kit (QIAGEN). Hybridization, washing, scanning, and normalization were performed as described [62], with the exception of scanning chips with an Axon GenePix 4000B microarray scanner, and data analyses and normalizations done using GenePix data analyses software. Scanning settings used were 635 nm for Cy5 and 532 nm for Cy3. The median of ratios mutant-Cy5/non-tagged-Cy3 or Mutant-Cy3/non-tagged-Cy5 values were statistically analyzed in the MultiExperiment Viewer (MeV) software using a *p*-

value cut-off at 0.05. Positive significant genes (upregulated genes) were considered to be regulatory targets of Stb5.

3.4.5 Bioinformatics

Gene ontology analyses were carried out with the CGD tool "Go Term Finder" (http://www.candidagenome.org/cgi-bin/GO/goTermFinder) [63]. Fungal blast of *C. albicans* Stb5 was carried out using the Saccharomyces Genome Database (SGD) fungal blast tool (http://yeastgenome.org/blast-fungal) [66]. Protein sequence alignment was done using the SIM Alignment Tool (http://web.expasy.org/sim/) [67]. Graphical representations of protein alignments were generated using the LALNVIEW program [68].

3.4.6 Acetate Utilization Assay

The *stb5* null mutant strain, SCSTB5GAD1A (STB5 gain of function strain) and SC5314 (wild type strain) were each cultured in 6.7 mg/ml Yeast Nitrogen Base (YNB) (Bishop), YNB + acetate (19.6 mg/ml), YNB + glucose (20 mg/ml) and YPD for three days on a 220 rpm shaker in a 30°C incubator. Cells were cultured in 10 ml of each medium in 50 ml falcon tubes and OD600 data were collected every 24 hours throughout the incubation period.

3.4.7 Drug resistance assay

The *stb5* null mutant strain, SCSTB5GAD1A (STB5 gain of function strain) and SC5314 (wild type strain) were each cultured in YPD, YPD + 2 μg/ml caspofungin, YPD + 0.1 μg/ml anidulafungin and YPD + 1 μg/ml amphotericin B for five days on a 220 rpm shaker in a 30°C incubator. Cells were cultured in 10 ml of each medium in 50 ml falcon tubes and OD600 data were collected every 24 hours throughout the incubation period.

3.5 Results

3.5.1 Confirmation of Stb5 ortholog

Stb5 was initially characterized in *S. cerevisiae* and has an apparent ortholog in *C. albicans* encoded by the gene C1_01200W_A (*STB5*). As shown in Figure 3.1, these two orthologous ZCFs have about 53.5% sequence identity at the 21 amino acid N-terminal region containing the zinc cluster domain, as well as a 444 amino acid stretch with 40.8% identity that spans most of each ortholog at their C-terminal region. The second best hit identified by using blast to compare CaStb5 against all *S. cerevisiae* proteins was *S. cerevisiae* Ppr1, which has a

block of 102 amino acids with 30.4% identity that is located outside of the region containing the zinc cluster domain. These findings establish that C1_01200W_A is the structural ortholog of Stb5 in *C. albicans*.

3.5.2 ChIP-Chip identification of *Candida albicans* Stb5 target genes

To carry out ChIP-Chip analyses, we fused the TAP epitope sequence to the endogenous *STB5* gene in strain SN148; correct DNA fusion was established using PCR and proper TAP-tagged protein expression was confirmed using immunoprecipitation.

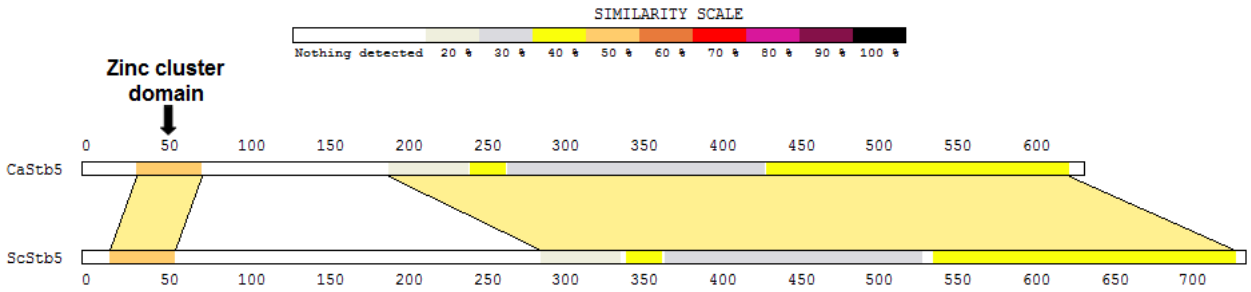


Figure 3.1: *C. albicans* Stb5 aligned with *S. cerevisiae* Stb5. The left shaded alignment area has 53.5% identity and includes the zinc cluster domain. The bigger shaded alignment area to the right has 40.8% identity.

The consistency observed between ChIP-Chip results from the TAP-tagged Stb5 strain, microarray results from an HA-tagged Stb5 strain and the phenotype of a *Stb5* null mutant suggest that tagging Stb5 did not affect its function in any substantive way. Following cross-linking and Stb5 target DNA extraction, target binding sequences were identified by amplifying and labeling the immunoprecipitated DNA sequences and hybridizing these labeled sequences to Agilent 8X15K whole genome tiling arrays representing the *C. albicans* genome. Stb5 target genes were ranked based on their log of ratios (Stb5-TAP-Cy5/non-tagged-Cy3) values (Table 3.1). Gene ontology assessment of genes corresponding to the top 150 targets with a log of ratios cut-off of at least 1.5 showed enrichment in genes involved in acetate catabolism, as well as some drug transport genes (Table 3.1) suggesting that Stb5 in *C. albicans* may transcriptionally regulate genes involved in acetate catabolism and drug transport.

3.5.3 Transcriptional profiling experiments to identify genes regulated by *C. albicans* Stb5

We also performed transcriptional profiling experiments on a *C. albicans* Stb5 gain of function (GOF) mutant generated by addition of a Gal4 activation domain [5], to establish functionally the expression network of the transcription factor.

Table 3.1: *C. albicans* Stb5 regulated genes from ChIP-Chip and transcription profiling. Both ChIP-Chip and transcriptional profiling data were analyzed using GenePix and MeV. Stb5 upregulated genes from transcriptional profiling with median of ratios of at least 2.076 are presented with their binding log of ratios values from ChIP-Chip. ChIP-Chip log of ratios of one or less indicate no binding. Upregulated genes that are not significantly bound by Stb5 are indirectly regulated by Stb5. Gene ontology of genes highlighted in green indicate acetate catabolism, as well as the glyoxylate cycle for *ICL1* and *MLS1*. Gene ontology of genes highlighted in grey indicate drug transport. Genes highlighted in light grey could potentially function in drug transport. N/A means not anayzed by ChIP-Chip. Primers used for transformation in this study are presented in Table S1.

		Transcription profiling			ChIP-Chip	
Gene	orf	Median of ratios	Upregulation rank	P-values	log of ratios	P-values
	orf19.4652	453.042	1	0	N/A	N/A
ICL1	orf19.6844	256.784	2	0	2.790	0.0014
	orf19.6950	119.627	3	0.000662847	-0.394	0.6357
	orf19.7554	118.090	4	6.19483E-06	2.902	0.0011
	orf19.4612	114.078	5	6.19483E-06	2.884	0.0013
	orf19.614	71.619	6	0.002456249	N/A	N/A
	orf19.1431	68.684	7	0.003887254	N/A	N/A
AOX2	orf19.4773	55.171	8	4.64612E-06	0.743	0.1456
YOR1	orf19.1783	37.555	9	4.95586E-05	0.761	0.1376
TES1	orf19.5217	35.540	10	5.73022E-05	0.003	0.3875
MLS1	orf19.4833	33.316	11	0.00013164	2.884	0.0013
QDR2	orf19.6992	30.888	12	0.000322131	1.414	0.0313
	orf19.2195	28.225	13	0.001635435	-0.009	0.5939
ATO7	orf19.1571	23.256	14	0.009409943	0.379	0.3401
	orf19.616	22.776	15	0.008039338	N/A	N/A
	orf19.1867	21.753	16	0.000142481	2.337	0.0046

	orf19.4121	21.727	17	8.67276E-05	0.546	0.2072
CTN1	orf19.4551	19.785	18	9.91172E-05	2.134	0.0537
	orf19.3059.1	19.742	19	0.004570234	0.526	0.2409
	orf19.6723	19.587	20	0.003120644	0.109	0.4962
	orf19.175	18.311	21	0.002951835	0.218	0.4460
	orf19.2049	18.123	22	0.012773735	1.009	0.1006
	orf19.750	17.733	23	0.001973053	0.581	0.2316
	orf19.5218	16.897	24	0.000201332	N/A	N/A
	orf19.1366	16.578	25	0.009025863	0.982	0.0564
TPO3	orf19.341	8.550	79	0.025733313	0.528	0.2505
TPO4	orf19.473	6.301	94	0.00210779	0.222	0.4542
FLU1	orf19.6577	3.964	140	0.006050798	1.110	0.0757
CDR1	orf19.6000	3.638	150	0.007553043	2.740	0.0194
PDR16	orf19.1027	2.840	181	0.012564658	1.619	0.0209
SGE11	orf19.4779	2.407	206	0.021602912	0.395	0.3307
SGE13	orf19.3444	2.155	229	0.027302152	1.595	0.0241
CDR4	orf19.5079	2.076	239	0.030693822	0.959	0.0909

Such transcription factor (GOF) mutant strains can allow for network up-regulation in the absence of stimulatory conditions. In support of the pattern observed by the ChIP-chip analysis, upregulated genes identified in the Stb5 GOF strain were enriched in genes involved in drug resistance, as well as genes of the glyoxylate cycle enzymes, specifically genes involved in acetate catabolism (Table 3.1). Reported acetate catabolic genes such as *ICL1*, *MLS1*, orf19.1867 (*MAE1*) and *CTN1* are highly upregulated, ranking 2nd, 11th, 16th and 18th respectively, among the identified upregulated genes in the Stb5 GOF strain, and also show very strong Stb5 binding signal in the ChIP-Chip data. *ORF19.7554* and *CDR1* appear among at least nine drug export genes identified through transcriptional profiling and also have strong Stb5 binding signals with log of ratios higher than the 1.5 cut-off in our ChIP-Chip data (Table 3.1). Although Stb5 does not bind its own promoter as shown in our ChIP Chip data, it was identified among the up-regulated genes in our transcription profiling data suggesting it indirectly regulates itself. Acetate catabolism genes and some drug pumps, including *CDR1* that has been reported to export echinocandins to a certain extent, were identified through ChIP-Chip and transcriptional profiling

as Stb5 targets, suggesting an acetate catabolism role and a possible drug resistance role of *STB5* in *C. albicans*.

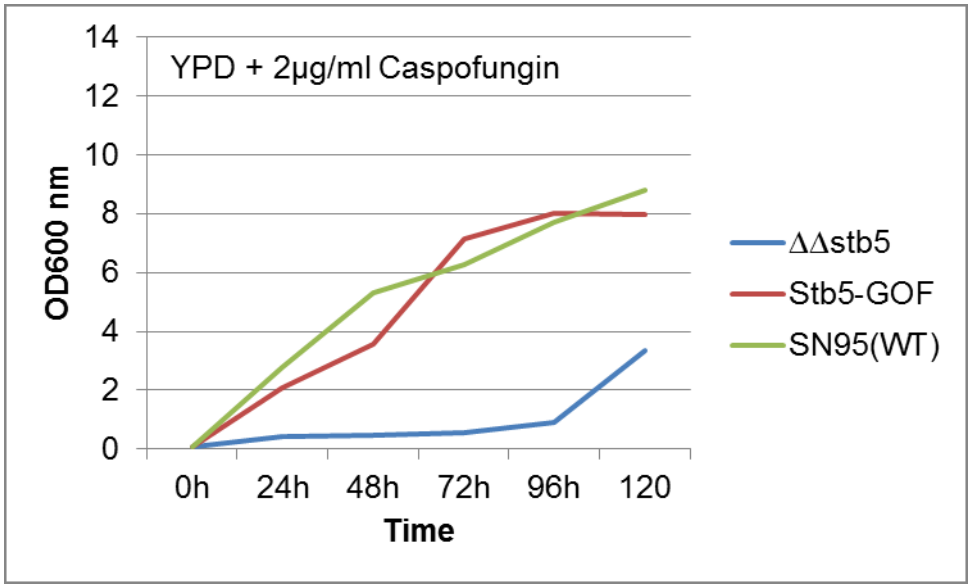
3.5.4 Resistance to echinocandins

ChIP-Chip and transcription profiling results suggested that Stb5 is a transcriptional activator of certain membrane transporters and transmembrane proteins, some of which have been reported to play a role in drug resistance in *C. albicans.* Stb5 is involved in the regulation of the ABC-type plasma membrane transporter Yor1 among other genes involved in multidrug resistance in *S. cerevisiae*; our transcription profiling data show a strong signal that indicates Yor1 (orf19.1783) is regulated by Stb5 in *C. albicans* as well. Other drug transport genes such as *CDR1*, CR_09830W_A (orf19.7554) and *QDR2* appear in both ChIP-Chip and transcription profiling data. We therefore cultured *stb5* null mutants in various drugs at 30°C to identify any drug sensitivity of the strain. *stb5* null mutants showed increased sensitivity to caspofungin (2 µg/ml) (Figure 3.2A) and anidulafungin (0.1µg/ml) (Figure 3.2B) when compared to the wild type SN95, and the Stb5 gain of function mutant strain, SCSTB5GAD1A. The strain (*stb5* null mutant) showed no sensitivity to amphotericin B. Pertubation of carbohydrate catabolism could influence echinocandin sensitivity since echinocandins act on the cell wall, however the upregulation of drug pumps by Stb5 identified in this study for the Stb5-GOF strain show significant correlation with echinocandin sensitivity observed for the *stb5* null mutant strain and could be an indicator that Stb5 upregulates drug pumps in *C. albicans* to confer drug resistance. Stb5 function therefore confers resistance to echinocandins potentially by influencing drug pumps as predicted through ChIP-Chip and transcription profiling. All strains including the *stb5* null mutant strain showed similar growth in YPD at 30°C (Figure 3.2C), a control that confirms that the slow growth observed for *stb5* null mutant in the presence of echinocandins is due to its sensitivity to this class of drugs.

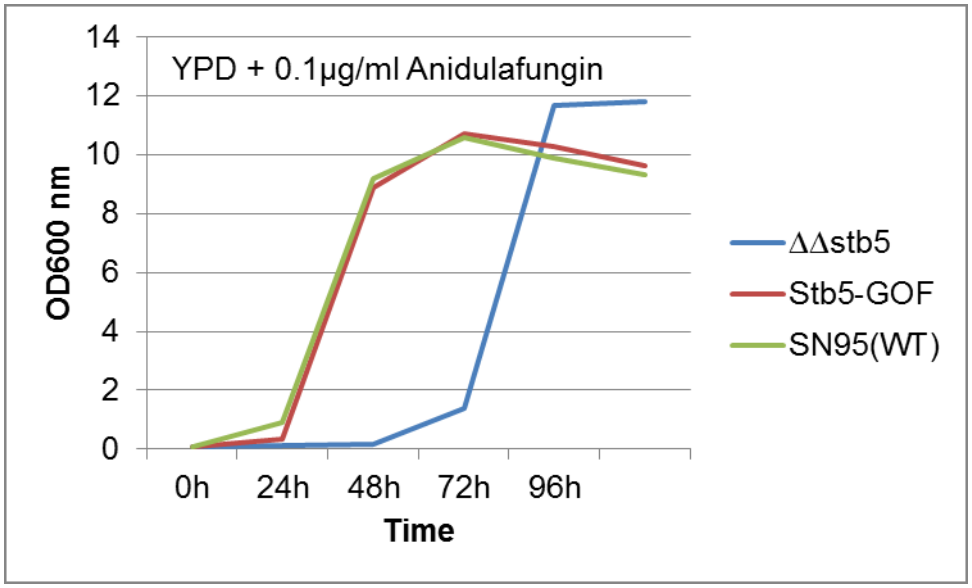
3.5.5 Acetate utilization

We identified through ChIP-Chip and transcriptional profiling that some of the key enzymes of the glyoxylate cycle including Icl1, Mls1 and Ctn1 are regulated by Stb5 in *C. albicans.*

A


YPD + 2µg/ml Caspofungin
OD600 nm
14
12
10
8
6
4
2
0
0h
24h
48h
72h
96h
120
Time
ΔΔstb5
Stb5-GOF
SN95(WT)

B


YPD + 0.1µg/ml Anidulafungin
OD600 nm
14
12
10
8
6
4
2
0
0h
24h
48h
72h
96h
Time
ΔΔstb5
Stb5-GOF
SN95(WT)

C

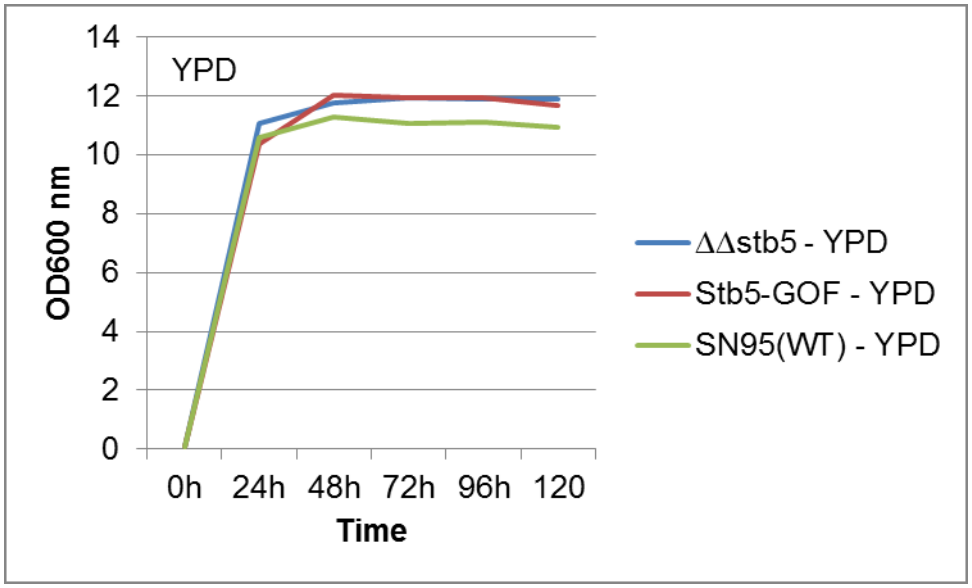


Figure 3.2: Growth curves of ΔΔstb5, SN95 (WT=wild type) and Stb5-GOF with or without echinocandins.

Primers used for transformation in this study are presented in Table S1.

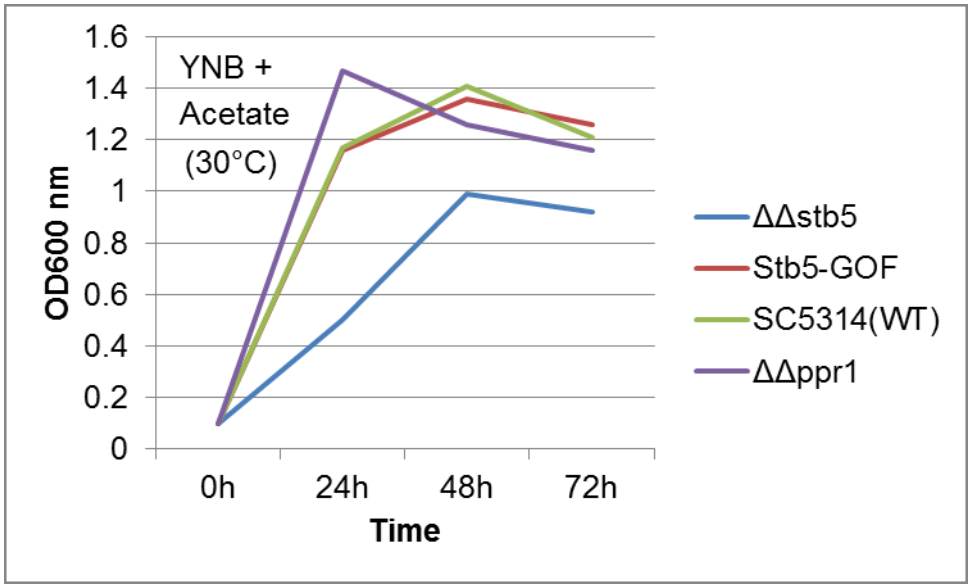
A) Growth curves of ΔΔstb5, SN95 (WT=wild type) and Stb5-GOF in YPD +2μg/ml caspofungin.

B) Growth curves of ΔΔstb5, SN95 (WT=wild type) and Stb5-GOF in YPD +1μg/ml anidulafungin.

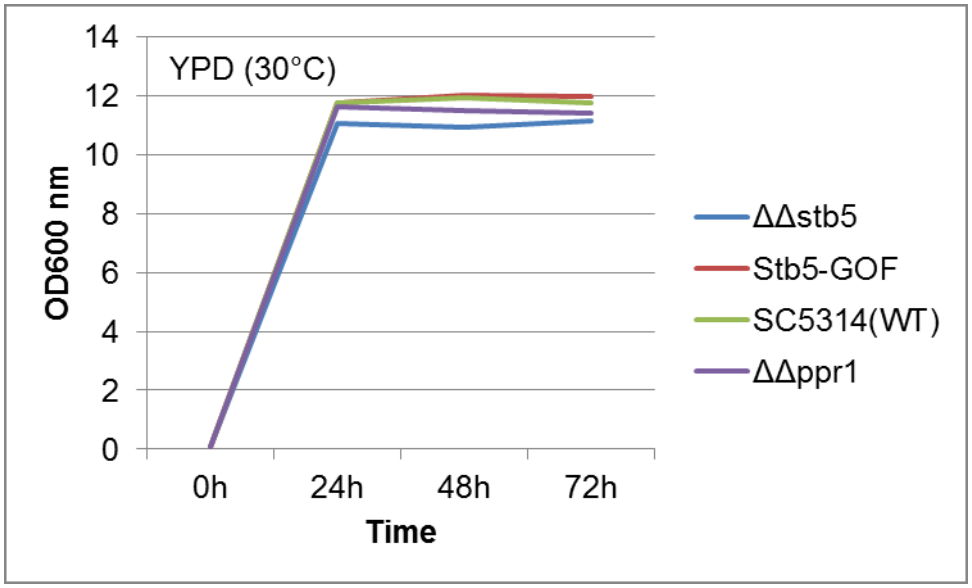
C) Growth curves of ΔΔstb5, SN95 (WT=wild type) and Stb5-GOF in YPD.

The findings were verified by culturing *stb5* null mutants in YNB media with 19.6 mg/ml acetate as the sole carbon source at 30°C. *stb5* showed slow growth relative to the wild type SC5314 and other control strains, SCSTB5GAD1A (Stb5 gain of function strain) and *ppr1* null mutant (Figure 3.3A). All strains showed about the same growth pattern in YPD at 30°C (Figure 3.3B) and the *stb5* null mutant showed better growth compared to Stb5-GOF strain at 30°C in medium with glucose as the sole carbon source (Figure 3.3C). It has been reported [92] that *icl1* and *mls1* null mutant strains each show slow growth at 30°C in medium with acetate as the sole carbon source, and grow even slower in the same media at 37°C. *stb5* null mutants had the same growth pattern as *icl1* and *mls1* in media with acetate as the sole carbon source at both 30°C and 37°C (Figure 3.3A, B and C and Figure 3.4A, B and C), which further supports that Stb5 is a transcriptional regulator of *icl1* and *mls1*.

A


YNB +
Acetate
(30°C)
OD600 nm
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0h
24h
48h
72h
Time
ΔΔstb5
Stb5-GOF
SC5314(WT)
ΔΔppr1

B


YPD (30°C)
OD600 nm
14
12
10
8
6
4
2
0
0h
24h
48h
72h
Time
ΔΔstb5
Stb5-GOF
SC5314(WT)
ΔΔppr1

C

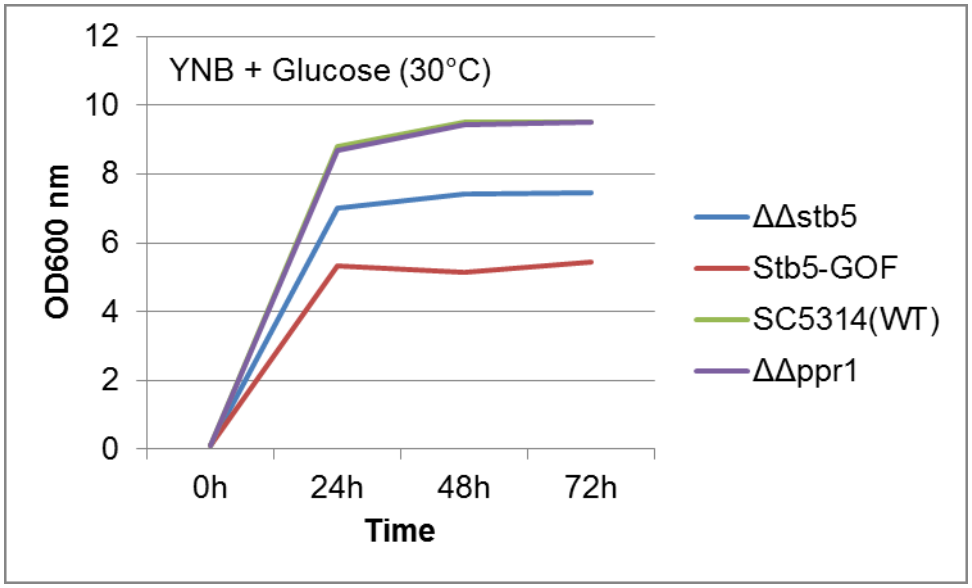


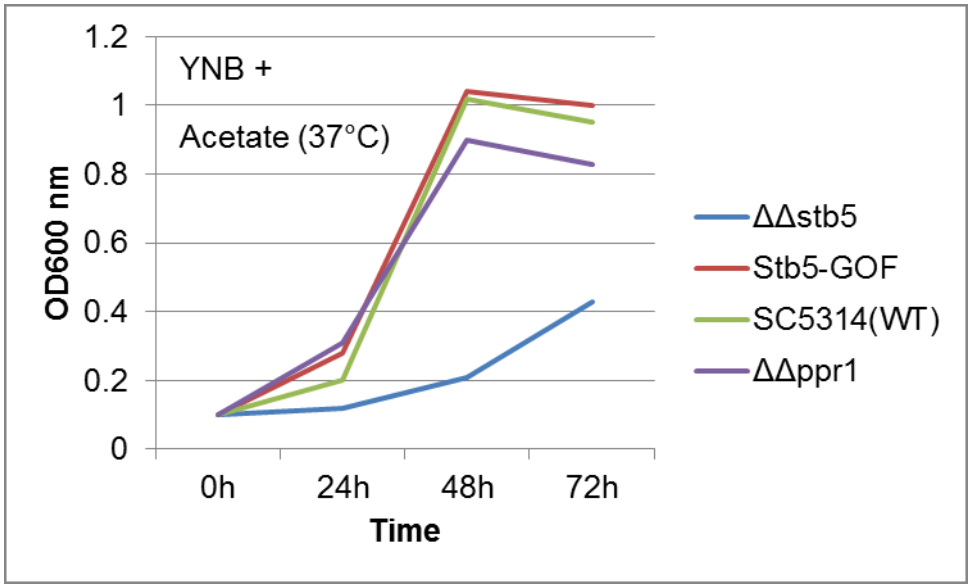
Figure 3.3: Acetate utilization assay of a *C. albicans stb5* null mutant strain, a Stb5 gain of function mutant strain, *ppr1* null mutant (ΔΔppr1) and the wild type strain SC5314 at 30°C. YPD and YNB + glucose media are used as controls at 30°C. Primers used for transformation in this study are presented in Table S1.

A) Growth curve of *stb5* null mutant (ΔΔstb5) strain, Stb5 gain of function mutant (Stb5-GOF) strain, *ppr1* null mutant (ΔΔppr1) and SC5314 (WT=wild type) strain in yeast nitrogen base (YNB) medium with acetate as the sole carbon source at 30°C.

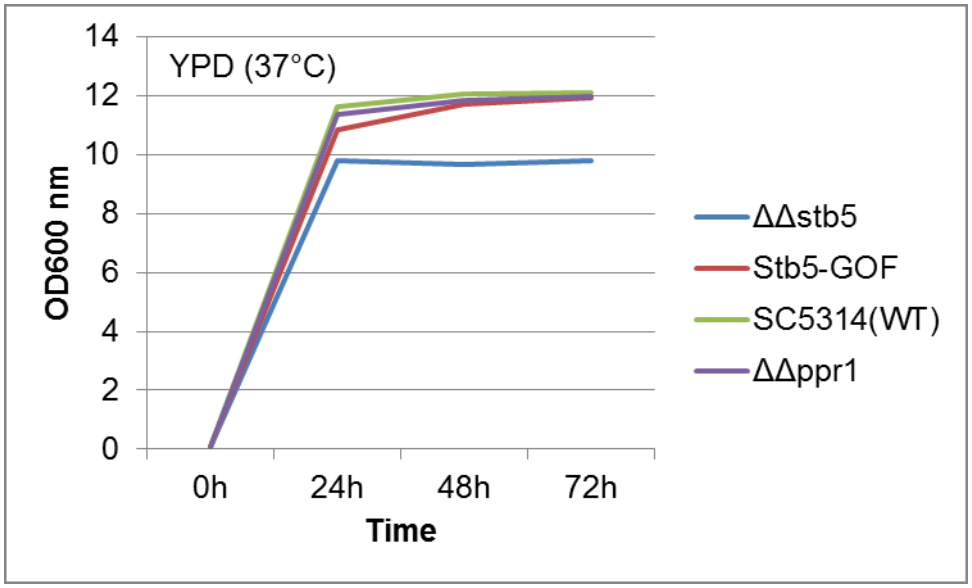
B) Growth curve of *stb5* null mutant (ΔΔstb5) strain, Stb5 gain of function mutant (Stb5-GOF) strain, *ppr1* null mutant (ΔΔppr1) and SC5314 (WT=wild type) strain in YPD at 30°C.

C) Growth curve of *stb5* null mutant (ΔΔstb5) strain, Stb5 gain of function mutant (Stb5-GOF) strain, *ppr1* null mutant (ΔΔppr1) and SC5314 (WT=wild type) strain in yeast nitrogen base medium with glucose as the sole carbon source at 30°C.

A


YNB +
Acetate (37°C)
OD600 nm
1.2
1
0.8
0.6
0.4
0.2
0
0h
24h
48h
72h
Time
ΔΔstb5
Stb5-GOF
SC5314(WT)
ΔΔppr1

B


YPD (37°C)
OD600 nm
14
12
10
8
6
4
2
0
0h
24h
48h
72h
Time
ΔΔstb5
Stb5-GOF
SC5314(WT)
ΔΔppr1

C

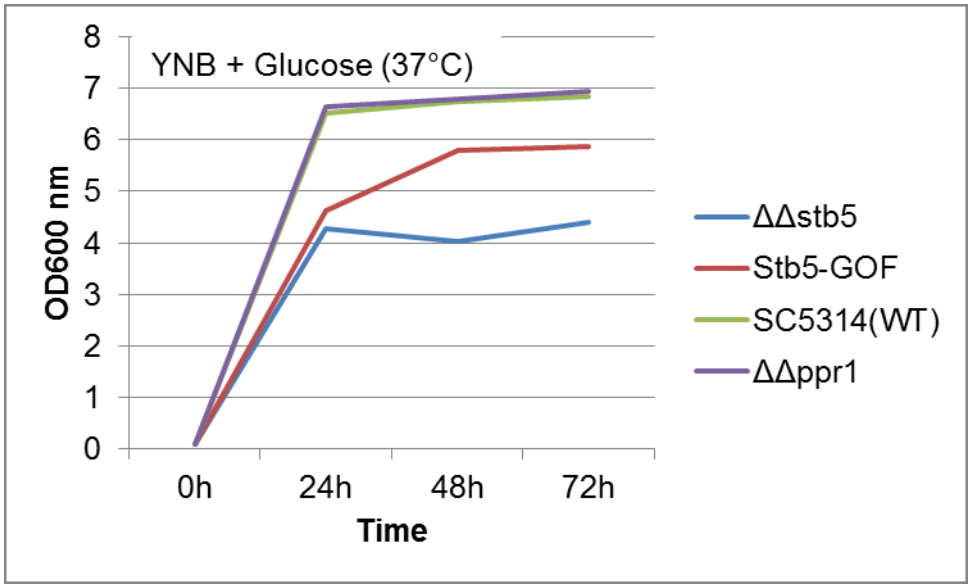


Figure 3.4: Acetate utilization assay of a *C. albicans stb5* null mutant strain, a Stb5 gain of function mutant strain, *ppr1* null mutant (ΔΔppr1) and the wild type strain SC5314 at 30°C. YPD and YNB + glucose media are used as controls at 37°C. Primers used for transformation in this study are presented in Table S1.

A) Growth curve of *stb5* null mutant (ΔΔstb5) strain, Stb5 gain of function mutant (Stb5-GOF) strain, *ppr1* null mutant (ΔΔppr1) and SC5314 (WT=wild type) strain in yeast nitrogen base (YNB) medium with acetate as the sole carbon source at 37°C.

B) Growth curve of *stb5* null mutant (ΔΔstb5) strain, Stb5 gain of function mutant (Stb5-GOF) strain, *ppr1* null mutant (ΔΔppr1) and SC5314 (WT=wild type) strain in YPD at 37°C.

C) Growth curve of *stb5* null mutant (ΔΔstb5) strain, Stb5 gain of function mutant (Stb5-GOF) strain, *ppr1* null mutant (ΔΔppr1) and SC5314 (WT=wild type) strain in yeast nitrogen base medium with glucose as the sole carbon source at 37°C.

3.6 Discussion

Inhibition of the glyoxylate cycle has been shown to render *C. albicans* significantly less virulent [93]. Key enzymes of the glyoxylate cycle include *Icl1* and *Mls1*, and studies have shown that *ICL1* null mutants of *C. albicans* are less virulent [93]. *ICL1* has therefore been proposed as a potential drug target against candidiasis [85]. We report for the first time a zinc cluster transcription factor, Stb5, which positively regulates the glyoxylate cycle and drug efflux pumps in *C. albicans*. s*tb5* deletion mutant strains of *C. albicans* showed sensitivity to caspofungin and anidulafungin which are members of the FDA approved echinocandins antifungal drug class. s*tb5* null mutants also showed poor growth at 30°C and worse growth at

37°C on media with acetate as the sole carbon source. These findings correlate with genes bound by Stb5 as observed through ChIP-Chip, and also correlate with the upregulated genes observed in transcription profiling data from the Stb5 gain of function mutant strain SCSTBGAD1A (Table 3.1). Stb5 binds and upregulates transcription of the genes encoding the glyoxylate cycle enzymes isocitrate lyase (*ICL1*), malate synthase *MLS1*, carnitine acetyl transferase *CTN1* and *orf19.1867* (putative malate permease *MAE1*). These enzymes are required for acetate catabolism and this suggests why *C. albicans* strains with the *STB5* deletion do not grow in medium with acetate as the sole carbon source. The reduced ability for *stb5* null mutants to grow in medium with acetate as the sole carbon source at elevated temperatures makes the inhibition of Stb5 even more clinically relevant since human body temperature is 37°C. Many drug pump genes, including genes encoding Cdr1 and Pdr16, are bound and upregulated by Stb5 as shown in ChIP-Chip and transcription profiling data (Table 3.1). In *S. cerevisiae,* Stb5 binds as a heterodimer with Pdr1 to the promoter of multidrug resistance genes to confer drug resistance. Pdr1 however has no reported ortholog in *C. albicans*, implying that Stb5 binds to the promoter of its target genes in *C. albicans* either as a monomer, a homodimer or a heterodimer with a transcription factor other than Pdr1. What is intriguing however is that we could not find a characteristic zinc cluster binding site at the promoter of *C. albicans* Stb5 target genes, which left us with the open question: Does *C. albicans* Stb5 regulate its target genes using a motif different from the traditional ZCF CGG-type motif? *YOR1*, a reported target of Stb5-Pdr1 heterodimer in *S. cerevisiae* was also found to be upregulated by Stb5 in *C. albicans*. Also intriguing is the fact that, contrary to the Yor1 activator role of Stb5-Pdr1 observed in *S. cerevisiae*, Stb5 appears to interact with Pdr1 to negatively regulate ABC transporter genes including *YOR1* in *C. glabrata*, as well other pleiotropic drug resistance genes such as *CDR1* [94]. Stb5-Pdr1 is also found to negatively regulate azole resistance genes in *Candida glabrata* [94]. However, only the pleiotropic drug response element (PDRE) of *C. glabrata* PDR1 (TCCACGGA/G) was found upstream of seven Stb5-Pdr1 target genes and no motif similar to that reported for Stb5 in *S. cerevisiae* (CGGN(c/g)TA) was found, which suggests the possibility that the negative regulatory role of *C. glabrata* Stb5 could be due to altered promoter occupancy of some transcriptional targets of *C. glabrata* Pdr1p. Stb5 clearly regulates drug efflux pumps in *C. albicans* as shown in this study, and while previous studies showed that Cdr2 is the major echinocandin efflux pump, Cdr1 also identified in this study, among other Stb5 regulated pumps, does pump out

echinocandins in *C. albicans* which could explain the increased sensitivity of *stb5* null mutants to YPD medium containing caspofungin or anidulafungin. We did not observe any improved fitness of the Stb5-GOF relative to the WT strain SC5314 in the presence of echinocandins and this could indicate that a basal expression level of the drug pumps is enough to confer maximum resistance to echinocandins such that overexpression of the pumps does not lead to any significant change in drug resistance. Also, maybe the cell membrane can only accommodate so many drug pumps such that extra transcripts of drug pumps only end up degraded. The null has no Stb5 and therefore insignificant expression of drug pumps, which explains the sensitivity to the drug. There is a possibility that the ability of Stb5 to positively regulate genes encoding acetate catabolic enzymes Icl1 and Mls1 could also lead to the breakdown of the acetate component of caspofungin acetate rendering the drug ineffective; this model, however, may only apply to caspofungin but not anidulafungin which lacks an acetate component, making drug pumps the most likely cause of echinocandin resistance. This, together with the individual roles of drug pumps and the acetate catabolic pathway in the success of *C. albicans* as a pathogen links both pathways to *C. albicans* pathogenicity and could explain why a single transcription factor, Stb5, regulates both processes.

These studies show that inhibiting the ZCF Stb5 shuts down the glyoxylate cycle which is required for *C. albicans* to use alternate carbon sources in order to thrive as a pathogen, makes *C. albicans* less virulent and prevents *C. albicans* from building resistance against the most clinically used drug class against candidiasis, echinocandins. Stb5 could therefore make a suitable drug target against candidiasis and resistance against echinocandins.

Chapter 4: Put3 positively regulates proline utilization and ribosome biogenesis in *Candida albicans*.

4.1 Preface

The work presented in Chapter 4 is from the following manuscript under preparation: Tebung, W. A., Omran, R. P, Morschhäuser, J. & Whiteway, M., (2016). **Put3 positively regulates proline utilization and ribosome biogenesis in *Candida albicans*.** Target journal: Molecular Microbiology.

Author contributions:

Walters Aji Tebung performed experiments, bioinformatic analyses, wrote paper; Raha Parvizi Omran contributed to phenotypic experiments; Joachim Morschhäuser provided activated transcription factor reagents; Malcolm Whiteway designed experiments, wrote paper.

4.2 Abstract of manuscript

Cells have evolved regulatory mechanisms to ensure that they choose nutrient sources efficiently. We have used direct phenotypic studies, transcription profiling and ChIP-Chip analysis to investigate the role of the Put3 ortholog in the fungal pathogen *Candida albicans*. We show, using phenotypic studies, that in contrast to *S. cerevisiae* which can only utilize proline as a nitrogen source, *C. albicans* utilizes proline both as a carbon source and as a nitrogen source. A *C. albicans put3* null mutant strain failed to utilize proline suggesting that Put3 is required for proline degradation in *C. albicans*, as in *S. cerevisiae.* The *C. albicans put3* null mutant strain was able to grow efficiently with glutamate, a product of proline catabolism following the action of the enzymes Put1 and Put2, as the sole carbon or nitrogen source. This suggests that Put3 controls proline catabolism by regulating the first and second enzymes of the pathway, Put1 and Put2 respectively. Our transcription profiling and ChIP-Chip data further confirmed that Put3 transcriptionally regulates Put1 and Put2 in *C. albicans* as both enzymes were upregulated in a strain expressing a hyperactive Put3, and Put3 showed direct binding to the Put1 promoter in our ChIP-Chip data. Our phenotypic findings also show that in this fungal pathogen, Put3 directs the degradation of proline even in the presence of a good nitrogen source such as ammonia, which is in contrast to Put3 regulated proline catabolism in *S. cerevisiae* that can only occur in the absence of a rich nitrogen source. The zinc cluster transcription factor (ZCF) Put3 was initially

characterized in *Saccharomyces cerevisiae* as the transcriptional activator of Put1 and Put2. Although in general, direct binding of ribosome biogenesis genes by Put3 in *C. albicans* was not observed in our ChIP-Chip data, the expression of many ribosome biogenesis genes was shown to be altered in our transcription profiling data using the strain expressing the hyperactive Put3. This suggests that Put3 indirectly regulates ribosome biogenesis. Other *C. albicans* cellular processes identified in this study to be regulated by Put3 include carboxylic acid metabolism, and hyphal growth.

4.3 Introduction

Every organism requires carbon and nitrogen for survival, but nutrient choices and nutrient assimilation mechanisms vary among species. *C. albicans* uses available galactose as a carbon source even in the presence of glucose in contrast to *S. cerevisiae* that shuts down the galactose catabolic pathway and uses glucose as a carbon source when both galactose and glucose are available [95]. Galactose is efficiently utilized as a carbon source only in the absence of glucose [95] in *S. cerevisiae*. Also, galactose catabolism is not regulated by Gal4 in *C. albicans* as it is the case in *S. cerevisiae* [15]. *S. cerevisiae* has the ability to acquire carbon and nitrogen from varying sources, but tends to pick and choose nutrient sources when presented with options [95,96]. This case of Gal4 transcriptional rewiring and metabolic dynamics observed between *C. albicans* and *S. cerevisiae* highlights the need to tread carefully when predicting protein function and the structure of metabolic pathways in related organisms based solely on orthologies with genes in model organisms. We studied the role of Put3 in *C. albicans* and our findings suggest that Put3 function has been fundamentally preserved between *C. albicans* and *S. cerevisiae*. However, we show that in contrast to *S. cerevisiae*, *C. albicans* uses proline both as a carbon and nitrogen source, unlike *S. cerevisiae* where Put3 only activates proline degradation in the absence of a rich nitrogen source. The presence of a rich nitrogen source does not prevent *C. albicans* Put3 from breaking down proline to acquire carbon for cell growth. Nitrogen catabolism has been shown to be hierarchical in *S. cerevisiae*; proline is not used as a nitrogen source in the presence of ammonium sulfate, even though both can serve as nitrogen sources [96]. Also, proline only serves as a sole nitrogen source in *S. cerevisiae* and not a sole carbon source or sole carbon and nitrogen source [97]. In *S. cerevisiae,* proline serves as an inducer of Put3, and available nitrogen sources dictate the phosphorylation of Put3 and hence fine-tune its activation

of Put1 and Put2 [96], an event that seems to be bypassed in *C. albicans* since the pathogen can still catabolise proline in medium containing ammonium sulfate.

4.4 Materials and methods

4.4.1 Strains, media, plasmids and transformation

Standard procedures for *C. albicans* cell growth and transformation [58] were followed. *C. albicans* strains for transformation, ChIP-Chip analyses and transcriptional profiling experiments were cultured in YPD. Yeast nitrogen base medium (YNB) at 6.8 mg/ml supplemented with glucose (20mg/ml), or proline (15.3mg/ml) was used for phenotypic studies testing for the ability of the *put3* null mutant to utilize proline as a carbon source. Yeast carbon base medium (YCB) at 11.7 mg/ml supplemented with ammonium sulfate (5 mg/ml), or proline (8.7 mg/ml) was used for phenotypic studies testing for the ability of the *put3* null mutant to utilize proline as a nitrogen source. The *put3* null mutant strains were also cultured in control media (YNB at 6.7 mg/ml without a carbon supplement, YCB at 11.7 mg/ml without a nitrogen source and YPD). The SCPUT3GAD1A strain was constructed as described [5]. One *PUT3* allele was replaced by transformation with the *HIS1* marker and the other allele with the *ARG4* marker in the SN95 background strain to obtain the *put3* null mutant. Oligonucleotides used for *put3* null mutant construction were amplified from plasmids pFA-HIS1 for the first allele knockout and pFA-ARG4 for the second allele knockout [60] using the respective forward and reverse primers (PUT3_Marker_KO_F that has a 97 bp region that is complimentary to the sequence just before the *PUT3* start codon and PUT3_Marker_KO_R that has a 95 bp region that is complimentary to the sequence just after the *PUT3* stop codon). Deletion of the first allele was confirmed by PCR using the forward primer PUT3_KO_Check_F that binds upstream of the *PUT3* gene and the reverse primer PUT3_KO_Check_R that binds downstream of the gene. Deletion of the second allele was confirmed by PCR using two primer pairs; forward primer PUT3_KO_Check_F that binds upstream of the *PUT3* gene and reverse primer FT-H2 that binds upstream of the *HIS1* gene (*HIS1* promoter region) and forward primer FT-U3 that binds inside the *HIS1* gene and reverse primer PUT3_KO_Check_R that binds downstream *PUT3*. The *put3* null mutants were confirmed using the forward primer PUT3_KO_Check_Internal_F and the reverse primer PUT3_KO_Check_Internal_R that both bind inside the *PUT3* gene; no band is expected for *put3*

null mutants for this primer pair. Oligos used for transformation in this study are presented in Table S1.

4.4.2 Proline Utilization Assay

The *put3* null mutant strain, SCPUT3GAD1A (Put3 gain of function strain), SC5314 (wild type strain) and the unrelated *ppr1* null mutant strain were each cultured in 10 mls of one of the following media in 50 ml falcon tubes for 3 days at 30°C with shaking at 220 rpm (6.8 mg/ml Yeast Nitrogen Base (YNB), YNB + proline (15.3 mg/ml), YNB + glucose (20 mg/ml) and YPD). OD600 data were collected every 24 hours throughout the incubation period.

The *put3* null mutant strain, SCPUT3GAD1A (Put3 gain of function strain), SC5314 (wild type strain) and unrelated *ppr1* null mutant strain were each cultured in 10 mls of the following media in 50 ml falcon tubes for 3 days at 30°C with shaking at 220 rpm (11.7 mg/ml Yeast Carbon Base (YCB) (Sigma-Aldrich), YCB + proline (8.7mg/ml) and YCB + ammonium sulfate (5mg/ml)). OD600 data were collected every 24 hrs throughout the incubation period.

4.4.3 Transcriptional Profiling Experiments

Transcriptional profiling experiments were carried out as described (Chapter 2 of this thesis) with a few exceptions. Briefly, transcriptional profiling experiments were performed for the strain SCPut3GAD1A (Put3 gain of function mutant) versus the background strain SC5314 (wild type). Single colonies of each strain were each inoculated into 50 ml YPD and incubated overnight at 30°C on a 220 rpm shaker. The overnight cultures were diluted to OD600 of 0.1 in YPD and grown to an OD600 of 0.8. Total RNA was extracted using the QIAGEN RNeasy minikit protocol and RNA quantity was determined using a NanoQuant machine. For direct dye incorporation, 20 μg of total RNA from each sample was reverse transcribed using oligo(dT)23VN and Superscript III reverse transcriptase (Invitrogen) in the presence of Cy3 or Cy5; dye swaps were employed for each sample. Template RNA was eliminated from the synthesised cDNA by simultaneously adding RNase A (Sigma) to a final concentration of 0.05 mg/ml and 0.05 units/μl RNase H (New England Biolabs) to each sample, then incubating for 30 minutes at 37°C before purifying the labeled cDNA with a QIAquick PCR Purification Kit (QIAGEN). Hybridization, washing, scanning, and normalization were performed as described [62], with the following exceptions. Scanning was carried out using an Axon GenePix 4000B microarray scanner, and data analyses and normalizations were done using GenePix data analysis

software. Scanning settings were 635 nm for Cy5 and 532 nm for Cy3. The median of ratios mutant-Cy5/non-tagged-Cy3 or Mutant-Cy3/non-tagged-Cy5 values were statistically analyzed in the MultiExperiment Viewer (MeV) software using a *p*-value cut-off at 0.05. Positive significant genes (upregulated genes) were candidates for Put3 regulation.

4.4.4 ChIP-Chip

ChIP-Chip experiments were performed as described (Chapter 2 of this thesis) with minor changes. Briefly, the SCPUT3GAD1A strain containing the chromosomally inserted Put3-HA fusion as well as the background strain SC5314 (untagged) were cultured to an optical density (OD) of 0.6 at 600 nm in 50 ml of YPD. Crosslinking for each 50 ml culture was carried out in 1.5 ml of 37% formaldehyde for 30 minutes, then ChIP performed as described (Chapter 2 of this thesis). ChIP DNA extracted from tagged strains was labeled with Cy5 dye and ChIP DNA from untagged strains SC5314 was labeled with Cy3 dye, and were then co-hybridized to Agilent 8X15K whole genome arrays containing 14490 60-mer intergenic and intragenic oligonucleotide probes. Microarray hybridization, washing, scanning, and normalization were performed as described [62] with the following modifications: Axon GenePix 4000B microarray scanner was used to perform scanning, and GenePix data analysis software and Multiexperiment Viewer (MeV) software were used to analyse and normalise data; a 0.05 P-value cut-off was used for MeV analyses. Scanning settings used were 635 nm for Cy5 and 532 nm for Cy3. The log of ratios Cy5/Cy3 (635 nm/532 nm) with a cut off of at least 1.5 for each spot was considered to be an indicator of significant Put3 binding.

4.4.5 Bioinformatics

The CGD tool "Go Term Finder" (http://www.candidagenome.org/cgi-bin/GO/goTermFinder) [63] was used for Gene Ontology analyses. Fungal blast analysis of *C. albicans* Put3 was performed using the Saccharomyces Genome Database (SGD) fungal blast tool (http://yeastgenome.org/blast-fungal) [66]. Protein sequences were aligned using the SIM Alignment Tool (http://web.expasy.org/sim/) [67], and graphical representation of protein alignment was generated using the LALNVIEW program [68].

4.5 Results

4.5.1 Bioinformatic determination of Put3 ortholog in *C. albicans*

The proposed ortholog of *S. cerevisiae* Put3 in the pathogen *C. albicans* is encoded by the gene C1_07020C_B (*PUT3*). These two orthologs have about 41.7% sequence identity that spans 821 amino acids (over 83% of each protein) at the N-terminal part of each ortholog (Figure 4.1). The next closest *S. cerevisiae* protein in terms of sequence identity to CaPut3 is Asg1 which has a stretch of 214 amino acids with just 24.8% identity and that does not include the zinc cluster domain. These findings provide strong support that C1_07020C_B is the structural ortholog of Put3 in *C. albicans*.

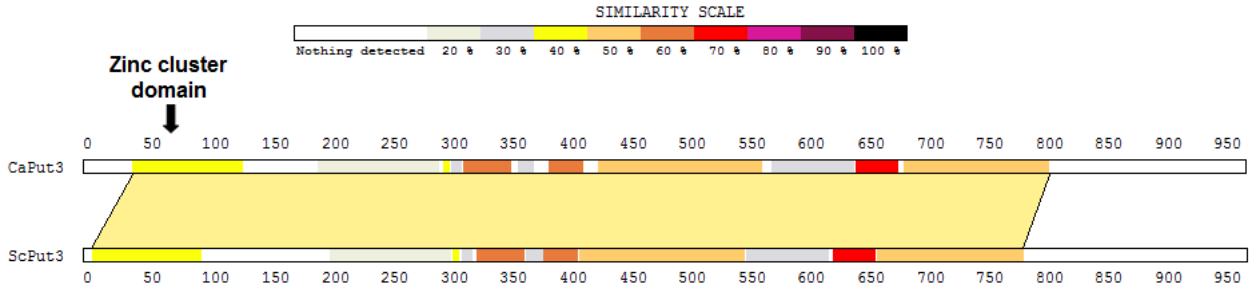


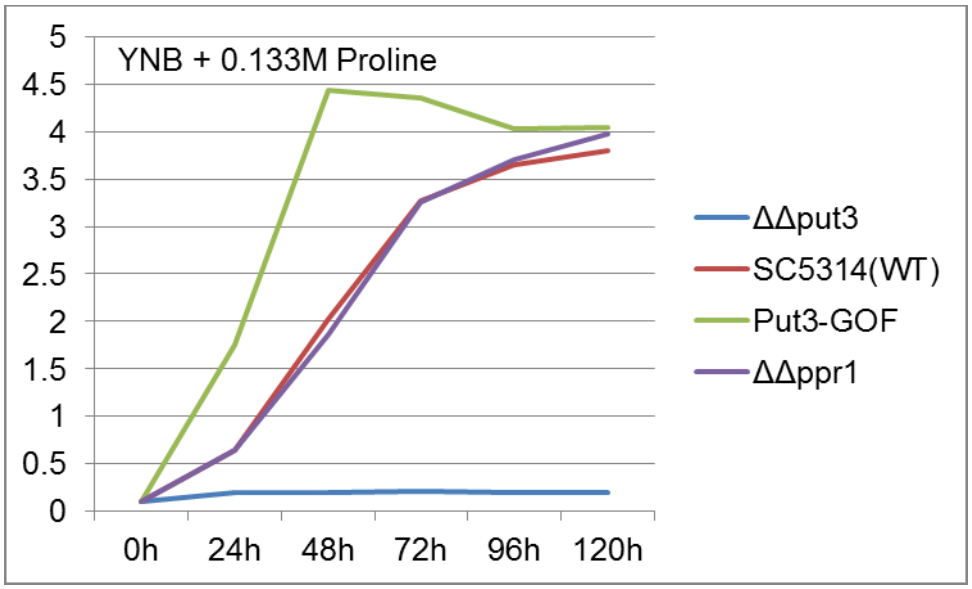
Figure 4.1: *C. albicans* Put3 aligned with *S. cerevisiae* Put3. The shaded alignment area has 41.7% identity and includes the zinc cluster domain.

4.5.2 Proline utilization

Our phenotypic studies show that *C. albicans* is able to metabolize and use proline as a sole carbon source and is able to use proline in the presence or absence of ammonium sulfate, an observation that contrasts with findings that *S. cerevisiae* cannot use proline as a sole carbon [97] and can only use proline as a nitrogen source in the absence of a rich nitrogen source such as ammonium sulfate. Proline degradation is controlled by Put3 in *S. cerevisiae*, and this degradation is not activated when a better source of nitrogen is available [96,98]. This nitrogen sensing regulation of Put1 and Put2 by Put3 is achieved by fine-tuning Put3 activity through phosphorylation patterns that depend on the quality of available nitrogen in the growth medium [96,98]. Rich nitrogen sources induce phosphorylation patterns that inactivate the ability of Put3 to induce proline catabolism genes. The *C. albicans* wild type strain SC5314 and the unrelated disruption *ppr1* null mutant strain (the *ppr1* null mutant strain was constructed similarly to the

put3 null mutant strain) showed significant growth in yeast nitrogen base (YNB) medium (The YNB medium contained ammonium sulfate) that contained 15.3 mg/ml proline as the sole carbon source. This shows a more liberal use of proline by the pathogen as potentially both a carbon and a nitrogen source (Figure 4.2A). To establish if Put3 regulates the proline degradation process in *C. albicans,* we cultured the wild type as well as the *put3* null mutant and Put3 gain of function mutant (Put-GOF) strains for 5 days at 30°C in YNB medium containing 15.3 mg/ml proline as the sole carbon source. The *put3* null mutant strain did not grow in YNB with proline as the sole carbon source after 5 days at 30°C, suggesting the inability of the strain to metabolize proline (Figure 4.2A), while the Put3-GOF showed a better growth rate compared to the wild type strain SC5314 and the control *ppr1* null mutant strain (Figure 4.2A). These findings suggest that Put3 regulates proline catabolism in *C. albicans*. All strains had normal and similar growth patterns in YNB with 20 mg/ml glucose as the sole carbon source (Figure 4.2B), and in YPD; and all strains showed no growth in YNB without a carbon source. We also used 8.7 mg/ml proline as a sole nitrogen source in yeast carbon base (YCB medum) to culture wild type *C. albicans* strain SC5314, the control *ppr1* null mutant strain, the *put3* null mutant strain and the Put3 gain of function strain. All strains were able to grow in the medium except the *put3* null mutant strain (Figure 4.3A), indicating that Put3 regulates the use of proline both as a carbon source and a nitrogen source in *C. albicans*. All strains showed similar growth in YCB medium with 5 mg/ml ammonium sulfate as the sole nitrogen source (Figure 4.3B).

A



B

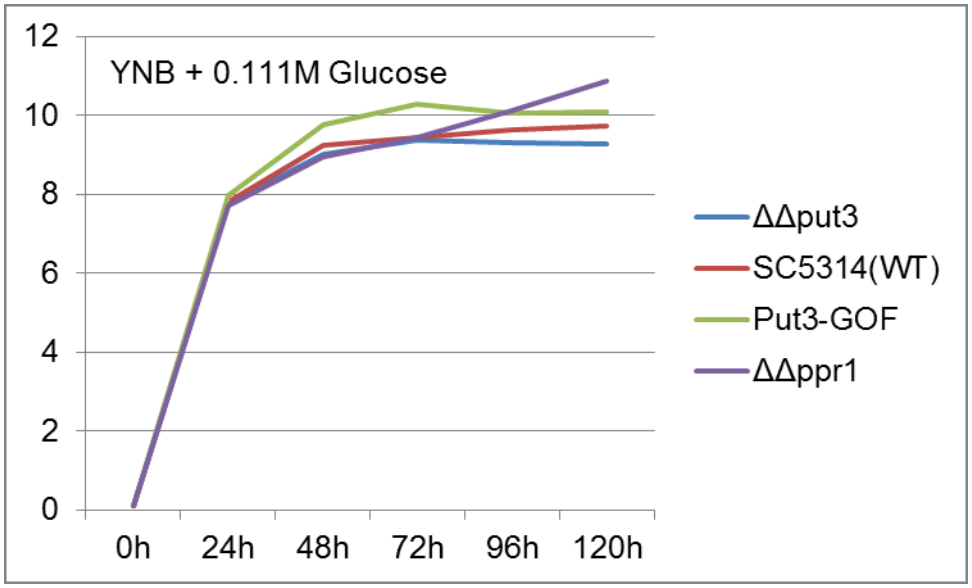
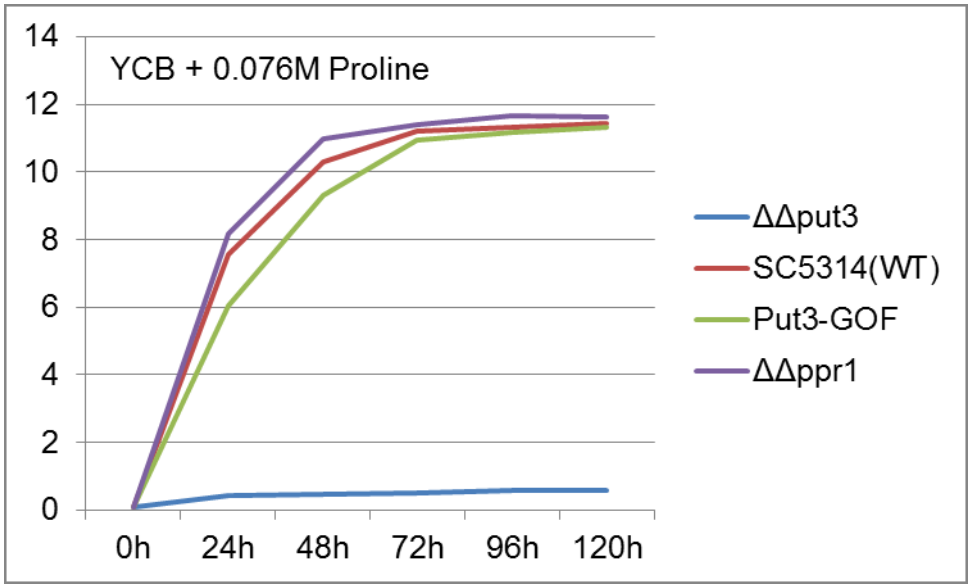


Figure 4.2: Proline utilization as carbon source assay of a *C. albicans put3* null mutant strain, the wild type strain SC5314, a Put3 gain of function mutant strain and *ppr1* null mutant (ΔΔppr1) strain at 30°C. YNB + glucose media is used as a control at 30°C. Primers used for transformation in this study are presented in Table S1. **A)** Growth curve of *put3* null mutant (ΔΔstb5) strain, SC5314 (WT=wild type) strain, Put3 gain of function mutant (Put3-GOF) strain and *ppr1* null mutant (ΔΔppr1) in yeast nitrogen base (YNB) media with proline as the sole carbon source at 30°C.
B) Growth curve of *put3* null mutant (ΔΔstb5) strain, SC5314 (WT=wild type) strain, Put3 gain of function mutant (Put3-GOF) strain and *ppr1* null mutant (ΔΔppr1) in yeast nitrogen base (YNB) media with glucose as the sole carbon source at 30°C.

4.5.3 Identification of Put3 regulated genes using transcriptional profiling experiments

To identify the pathway through which Put3 regulates proline degradation in *C. albicans*, we did transcriptional profiling experiments using the Put3 gain of function mutant strain SCPUT3GAD1A, generated by addition of a Gal4 activation domain to Put3 [5]. Such gain of function (GOF) mutant strains can allow for network up-regulation in the absence of any stimulatory condition. Transcription profiling using hyperactive transcription factors is an efficient way to identify the regulatory pathway of transcription factors.

A



B

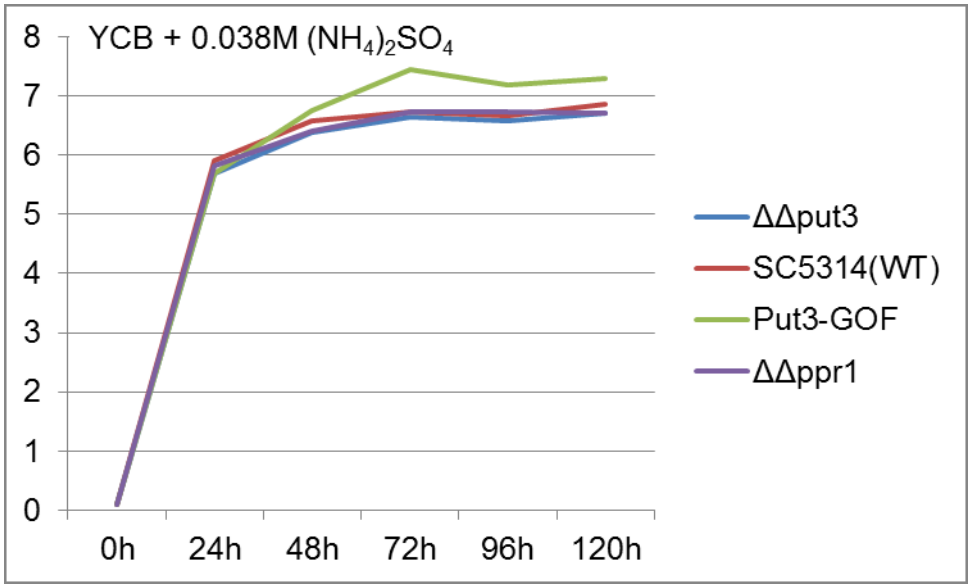


Figure 4.3: Proline utilization as nitrogen source assay of a *C. albicans put3* null mutant strain, the wild type strain SC5314, a Put3 gain of function mutant strain and *ppr1* null mutant (ΔΔppr1) strain at 30°C. YCB + ammonium sulfate media is used as a control at 30°C. Primers used for transformation in this study are presented in Table S1.

A) Growth curve of *put3* null mutant (ΔΔstb5) strain, SC5314 (WT=wild type) strain, Put3 gain of function mutant (Put3-GOF) strain and *ppr1* null mutant (ΔΔppr1) in yeast carbon base (YCB) media with proline as the sole nitrogen source at 30°C.

B) Growth curve of *put3* null mutant (ΔΔstb5) strain, SC5314 (WT=wild type) strain, Put3 gain of function mutant (Put3-GOF) strain and *ppr1* null mutant (ΔΔppr1) in yeast carbon base (YCB) media with ammonium sulfate as the sole nitrogen source at 30°C.

The key proline catabolism enzymes Put1, and Put2 (median of ratio = 2.20325, P-value = 0.022913117), were among the genes upregulated in strains expressing the hyperactive Put3 (Table 4.1). This suggests that Put3 has preserved its proline catabolism regulation role between *C. albicans* and *S. cerevisiae*, and it regulates this pathway by transcriptionally activating Put1 and Put2. Other upregulated genes observed in the transcription profiling data using the Put3 gain of function mutant strain are genes involved in ribosome biogenesis, carboxylic acid metabolic processes, filamentous growth, response to stress, RNA metabolic process, and cellular protein modification among others (Table 4.1).

Table 4.1: *C. albicans* Put3 regulated genes from ChIP-Chip and transcription profiling. Both ChIP-Chip and transcriptional profiling data were analyzed using GenePix and MeV. Put3 upregulated genes from transcriptional profiling that also show Put3 binding in ChIP-Chip data are presented.

		Transcription profiling		**ChIP-Chip**	
Gene	**orf**	**Median of ratio**	**P-values**	**log of ratios**	**P-values**
	orf19.670.2	28.522	0	1.557	0.014703244
YHB1	orf19.3707	8.709	1.08409E-05	2.465	0.19247757
GLN1	orf19.646	7.447	5.2656E-05	1.588	0.006385783
PTC8	orf19.4698	6.080	8.82763E-05	2.519	1.13E-04
PUT1	orf19.4274	4.452	0.000717051	2.308	9.52E-04
CYB2	orf19.5000	3.882	0.001373703	1.512	0.08496411
ATP1	orf19.6854	3.727	0.001810438	1.537	0.016077295
IDP2	orf19.3733	3.300	0.003476847	3.738	0.019314699
	orf19.1486	2.819	0.00730525	1.921	0.027921325
	orf19.1549	2.753	0.009056837	2.148	0.1706846
	orf19.4405	2.727	0.020243147	1.667	0.004911663
	orf19.6853	2.668	0.009055289	1.537	0.016077295
	orf19.5026	2.368	0.016391512	1.691	0.016881298

RGT1	orf19.2747	2.351	0.017607247	1.644	0.048810214
FAD2	orf19.118	2.236	0.023882609	2.038	0.004570739
	orf19.1109	2.162	0.041091837	1.559	0.014572118
	orf19.4273	2.144	0.038977854	2.222	0.00168116
UTP4	orf19.1633	2.113	0.029476538	1.847	0.003489993
	orf19.2782	2.037	0.033253834	1.920	0.17685506
	orf19.3585	2.034	0.033969335	1.693	0.20605038
	orf19.6828	2.009	0.04790305	1.563	0.10103865
RPL35	orf19.5964.2	1.993	0.016881298	1.633	0.005551415
GIN4	orf19.663	1.967	0.040771257	1.869	0.051601104
CAS1	orf19.1135	1.947	0.044322442	1.555	0.04001725

4.5.4 Direct identification of genes bound by *Candida albicans* Put3 using ChIP-Chip

Phenotypic studies revealed that Put3 regulates proline degradation in *C. albicans* and transcription profiling experiments revealed that *C. albicans* Put3 regulates the proline catabolism genes Put1 and Put2, as well as genes involved in ribosome biogenesis and other cellular pathways. Chromatin immunoprecipitation followed by microarray analysis (ChIP-Chip) effectively identifies the direct binding targets of specific transcription factors. This analysis was carried out using the *C. albicans* strain SCPUT3GAD1A [5], constructed from SC5314 and has an HA epitope sequence fused to *PUT3* gene. After chromatin cross-linking, target binding sequences were identified by amplifying and labeling immunoprecipitated DNA sequences and hybridizing these labeled sequences to Agilent 8X15K whole genome tiling arrays representing the *C. albicans* genome. Put3 target genes were ranked based on their log of ratios (PUT3-HA-Cy5/non-tagged-Cy3) values, then target genes with log of ratios of at least 1.5 that are also upregulated in the SCPUT3GAD1A strain expressing a hyperactive Put3 [5] (Table 4.1) were further analyzed for function using the Candida Genome Database Gene Ontology tool. Consistent with transcription profiling data for the SCPUT3GAD1A strain, Put3 targets identified by ChIP-Chip include Put1, an essential enzyme of the proline catabolism pathway, as well as genes involved in ribosome biogenesis, carboxylic acid metabolic processes, filamentous growth, response to stress, RNA metabolic processes, cellular protein modification among others (Table 4.1). Put2, however, was not bound by *C. albicans* Put3 as shown in our ChIP-Chip data and this

could imply that Put3 regulates Put2 indirectly. Binding of *C. albicans* Put3 to the promoter of Put1 further confirms that *C. albicans* Put3 regulates proline catabolism in the pathogen, a role that is also played by its ortholog in *S. cerevisiae*.

4.6 Discussion

We did phenotypic studies using a *C. albicans put3* null mutant and found that this mutant showed no growth in YNB medium containing proline as the sole carbon source (Figure 4.2A), and no growth in YCB medium containing proline as the sole nitrogen source (Figure 4.3A). In contrast, the wild type strain SC5314, the Put3-GOF strain and *ppr1* null mutant strain used as controls all showed significant growth in both YNB medium containing proline as the sole carbon source (Figure 4.2A) and YCB medium containing proline as the sole nitrogen source (Figure 4.3A). Intriguingly, the Put3-GOF strain had a shorter lag phase in YNB medium with proline as the sole carbon source, which further highlights the importance of Put3 in proline catabolism (Figure 4.2A). All strains showed similar growth in YNB medium containing glucose as the sole carbon source (Figure 4.2B), YCB media containing ammonium sulfate as the sole nitrogen source (Figure 4.3B), YNB with glucose as the carbon source, and YPD.
Put3 regulates proline use as a nitrogen source in *S. cerevisiae* by controlling expression of *PUT1* and *PUT2*, which encode the steps required to process proline to glutamate. We found that glutamate complements the *C. albicans put3* null mutant for both carbon and nitrogen needs, suggesting a requirement of Put3 to express the *PUT1* and *PUT2* genes for proline utilization. We further present transcription profiling and ChIP-Chip data to support the role of Put3 in the positive regulation of *PUT1* and *PUT2*.

We therefore report here that the zinc cluster transcription factor Put3 has conserved its proline catabolism role between the related species *C. albicans and S. cerevisiae*, however, *C. albicans* is able to use proline both as its sole carbon source (Figure 4.2A) and as its sole nitrogen source (Figure 4.3A) which is contrary to *S. cerevisiae* that only uses proline as its sole nitrogen source but not as its sole carbon source or sole carbon and nitrogen source [97]. We also show that *C. albicans* degrades proline even in the presence of ammonium sulfate (Figure 4.2A), something *S. cerevisiae* cannot do. We further confirmed the role of Put3 in *C. albicans* using transcriptional profiling and ChIP-Chip experiments. Consistent with our observations from phenotypic studies, in *C. albicans,* Put1 and Put2 are under the regulation of Put3 as shown by our transcriptional profiling data, and our ChIP-Chip data suggest that Put3 binds *PUT1* (Table

4.1); *PUT2* does not appear to be bound by Put3 in *C. albicans* as suggested by our ChIP-Chip data. These findings suggest that *PUT1* and *PUT2* are under direct and indirect regulation of Put3 respectively in *C. albicans*, unlike in *S. cerevisiae* where Put3 regulates both genes through direct binding [96]. Put3 also regulates ribosome biogenesis genes, carboxylic acid metabolic process, filamentous growth, the response to stress, RNA metabolic processes, and cellular protein modification among others as revealed by our transcription profiling and ChIP-Chip data (Table 4.1). These other roles have so far not been reported for *S. cerevisiae* Put3, therefore, it will be neccesary to carry out phenotypic studies in order to finalyze the role of *C. albicans* Put3 in the pathways, and to establish whether or not *S. cerevisiae* Put3 plays a role in the processes. Such findings would either further strenghthen the proposed conservation of Put3 function, or reveal hidden cases of rewiring in Put3 function.

Our findings show that *C. albicans* uses proline both as a carbon source and nitrogen source, and that *C. albicans* Put3 regulates proline catabolism even in the presence of ammonium sulfate to provide the cells with carbon and nitrogen, events that are contrary in *S. cerevisiae* which can only use proline as a nitrogen source and whose Put3 does not possess the ability to effectively activate the proline catabolic pathway in the presence of a rich nitrogen source such as ammonium sulfate. Put3 is therefore not rewired, at a general level, between *C. albicans* and *S. cerevisiae* because its function remains the same in both *C. albicans and S. cerevisiae*, however, the ZCF is less tightly controlled in *C. albicans* and also plays a role in the breakdown of the proline carbon skeleton to usable carbon.

Chapter 5: General conclusions and future work

Our research aims were to identify zinc cluster transcription factor rewiring between *C. albicans* and *S. cerevisiae*. We identified several scenarios regarding transcriptional rewiring presented in Chapters 2, 3 and 4 of this thesis.

Chapter 2 of this thesis is about Ppr1, a ZCF initially characterized in *S. cerevisiae*. Ppr1 binds CGG(N6)CCG in *S. cerevisae* to regulate transcription of genes involved in *de novo* pyrimidine biosynthesis. A clear structural otholog of Ppr1 exists in the human pathogen *C. albicans* and one would expect, considering the structural similarity, that the two orthologous ZCFs will play the same role in the two species. We found otherwise. In *C. albicans,* Ppr1 regulates Dal genes involved in purine catabolism using the same DNA binding motif, CGG(N6)CCG, as in *S. cerevisiae*. Among other transcriptional rewiring cases is the example of the ZCF Gal4 that regulates glucose utilization in the pathogen but drives galactose catabolism in the bread yeast. Transcriptional rewiring helps explain, at least in part, how species with similar genomes survive in their different home environments. However, little is known about the forces that drive transcriptional rewiring. We identified that Ppr1 rewired at the lineage leading to *N. castellii* under pressures to minimize oxygen use in *S. cerevisiae* and related species. How cellular systems copy and paste transcription factor binding sequences from the promoters of multiple genes in their old regulon to the promoter of a completely new gene set in their new regulon remains a mystery. More work would be required to address this question.

In Chapter 3, we discuss a case of partial transcriptional rewiring. Stb5 initially identified *in S. cerevisiae* as the regulator of drug efflux pumps, is found in our *C. albicans* studies to regulate not only drug efflux pumps, but also glyoxylate enzymes. Glyoxylate enzymes appear to be under the regulation of Sip4 in *S. cerevisiae*. The glyoxylate cycle has been proven important for *C. albicans* survival during pathogenesis. Enzymes of the glyoxylate cycle such as *ICL1* allow the pathogen to freely use non-fementable carbon sources such as acetate when glucose is limited during pathogenesis. Researchers have therefore proposed *ICL1* as a drug target against candidiasis. Stb5 transcriptionally regulates *ICL1* and drug efflux pumps and could potentially serve as a drug target. Stb5 will make a drug target because it controls two pathways, the glyoxylate and drug resistance pathways, that are both important for *C. albicans* virulence, and it is a member of the fungal specific class of transcription factors called zinc cluster transcription factors which should help in the development of antifungal drugs with minimal off targets in the

human host. Transcription factors would make excellent drug targets for cell destruction due to their involvement in multiple pathways. However, this promiscuous nature of transcription factors makes it challenging to use them as drug targets in the treatment of diseases that require cell healing due to extensive side effects. Other challenges faced when considering TFs as drug targets is the difficulty of finding drugs that would specifically disrupt protein-DNA interaction or protein-protein interaction, rather than a drug that simply binds the active site of an enzyme to inhibit its function [99]. Advances in Protein-DNA and protein-protein interaction studies have, however, led to the discovery of TF-targeting small molecules such as nutlins, which are specific MDM2-p53 antagonists currently in the early clinical trials for cancer treatment [100]. These advances in the understanding of protein-DNA and protein-protein interactions will allow for the exploitation of transcription factors as drug targets. *C. albicans* transcription factors such as Lys144, Sut1, Efg1, Crz1, Nrg1and many others have been identified as potential antifungal drug targets but have not actually been targeted [99]. Research on the identification of drug targets that are TFs, as well as drugs against these transcription factors would therefore improve the efficiency to fight candidiasis and diseases overall.

Chapter 4 is about Put3 that has conserved its function between *S. cerevisiae* and *C. albicans*. Put3 transcriptionally regulates proline catabolism in both species. We, however, discovered that *C. albicans* uses proline both as a carbon source and nitrogen source and also degrades proline even in the presence of ammonium sulfate. These findings are in contrast to *S. cerevisiae* that does not use proline as its sole carbon source or sole carbon and nitrogen source [97], but only uses proline as its sole nitrogen source. Also, the *S. cerevisiae* Put3 ortholog only degrades proline if no rich nitrogen source is available, contrary to *C. albicans* Put3. These differences could be a result of post-transcriptional regulation as we find the transcriptional arrangement in both species to be identical. It would be interesting to identify how *C. albicans* escapes the inhibition of Put3 that occurs in the presence of a rich nitrogen source such as ammonia, as observed in *S. cerevisiae*.

We would still need to screen the remaining of the 82 ZCFs in *C. albicans* in order to have an overall picture of how different organisms use rewiring to deal with challenges encountered in their lifestyle. Other points requiring more study to fully address transcriptional rewiring include the identification of the driving force or forces of rewiring, the importance of the rewiring process for evolution, whether or not transcription factors rewire randomly, and how

cells coordinate the seemingly complicated transfer of transcription factor binding sites from one gene set involved in a specific cellular process to a completely different gene set involved in a different cellular process. Is it coincidence that transcription factors seem to rewire between related cellular processes (Gal4 for the sugars and Ppr1 for nucleotides)?

Current knowledge of transcriptional rewiring can only partially address questions such as how and why transcription factors rewire, and whether or not rewiring is random. As evidenced in Gal4 and Ppr1 studies, it is unlikely that transcriptional rewiring is a random process. The respective transfer of Ppr1 and Gal4 DNA binding sites from one related gene set to another supports a more ordered process and dismisses the likelihood that transcriptional rewiring is random. Several transcriptional rewiring driving forces have also been identified. Some examples of rewiring drivers include horizontal gene transfer [41], gene duplication [42] and change of inducers (as reported for Ppr1 in Chapter 2 of this thesis). A fungal transcription factor, SB5, originating from a virus through horizontal gene transfer took over the transcriptional regulation of the cell cycle from the original transcriptional regulator E2F [41]. This highlights how horizontal gene transfer could drive transcriptional rewiring. A transcriptional rewiring event driven by gene duplication involves MALR that ancestrally binds CGC and CGG to regulate maltose utilization as well as palatinose utilization, but has evolved two paralogues following gene duplication that regulate distinct pathways. One of the MalR paralogues binds CGC to regulate maltose utilization, meanwhile the other paralogue binds CGG to regulate palatinose utilization [42]. Finally, we also report for Ppr1 in Chapter 2 of this thesis how changing the inducer of a transcription factor could drive rewiring. These findings indicate that many factors could induce transcriptional rewiring and further research would uncover many more driving forces that could help in the understanding of how and why transcription factors rewire.

In many rewiring cases, a transitional stage where two transcription factors compete for the same function often seem to precede the rewiring event. We identified a fairly equal distribution of the Ppr1 and Dal82 motifs at the promoter of Dal genes in species just before Ppr1 rewiring which could suggest that both transcription factors compete for the regulation of Dal genes in those intemediary species. Also, Before SBF and Whi5 took over cell cycle regulation from E2F and Rb in fungi, they all co-existed in many basal fungi and the lineages formerly known as 'zygomycetes' and could suggest that both competed for cell cycle regulation before the

rewiring event [41]; in fact, SBF and E2F have the same DNA binding specificity which further strenghthens this hypothesis [41]. Finally, the two paralogues of MalR covered for each other in the transitional species around *Saccharomyces bayanus* just before the rewiring event that led to the pathway specialization of each of the paralogues [42]. These findings suggest that many pressures can initiate transcriptional rewiring, but a transcription factor can only rewire if another transcription factor takes over its ancestral role. This conclusion, however, may not apply to all or most cases of rewiring and it is only through extensive research covering most rewiring cases that a more reliable general conclusion can be made.

Reference List

1. Ledent, V., and Vervoort, M. (2001). The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Res. *5,* 754-770.

2. Laity, J.H., Lee, B.M., and Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. Curr. Opin. Struct. Biol. *1,* 39-46.

3. Pelttari, K., Barbero, A., and Martin, I. (2015). A potential role of homeobox transcription factors in osteoarthritis. Ann. Transl. Med. *17,* 254-5839.2015.09.44.

4. Nikolaev, Y., Deillon, C., Hoffmann, S.R., Bigler, L., Friess, S., Zenobi, R., Pervushin, K., Hunziker, P., and Gutte, B. (2010). The leucine zipper domains of the transcription factors GCN4 and c-Jun have ribonuclease activity. PLoS One *5,* e10765.

5. Schillig, R., and Morschhauser, J. (2013). Analysis of a fungus-specific transcription factor family, the *Candida albicans* zinc cluster proteins, by artificial activation. Mol. Microbiol. *5,* 1003-1017.

6. Vandeputte, P., Ischer, F., Sanglard, D., and Coste, A.T. (2011). In vivo systematic analysis of *Candida albicans* Zn2-Cys6 transcription factors mutants for mice organ colonization. PLoS One *10,* e26962.

7. MacPherson, S., Larochelle, M., and Turcotte, B. (2006). A fungal family of transcriptional regulators: the zinc cluster proteins. Microbiol. Mol. Biol. Rev. *3,* 583-604.

8. Marmorstein, R., and Harrison, S.C. (1994). Crystal structure of a PPR1-DNA complex: DNA recognition by proteins containing a Zn2Cys6 binuclear cluster. Genes Dev. *20,* 2504-2512.

9. Mamane, Y., Hellauer, K., Rochon, M.H., and Turcotte, B. (1998). A linker region of the yeast zinc cluster protein leu3p specifies binding to everted repeat DNA. J. Biol. Chem. *29,* 18556-18561.

10. Remboutsika, E., and Kohlhaw, G.B. (1994). Molecular architecture of a Leu3p-DNA complex in solution: a biochemical approach. Mol. Cell. Biol. *8,* 5547-5557.

11. Liang, S.D., Marmorstein, R., Harrison, S.C., and Ptashne, M. (1996). DNA sequence preferences of GAL4 and PPR1: how a subset of Zn2 Cys6 binuclear cluster proteins recognizes DNA. Mol. Cell. Biol. *7,* 3773-3780.

12. Pfeifer, K., Kim, K.S., Kogan, S., and Guarente, L. (1989). Functional dissection and sequence of yeast HAP1 activator. Cell *2,* 291-301.

13. Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S.C. (1992). DNA recognition by GAL4: structure of a protein-DNA complex. Nature *6368,* 408-414.

14. Johnston, M. (1987). A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. Microbiol. Rev. *4,* 458-476.

15. Martchenko, M., Levitin, A., Hogues, H., Nantel, A., and Whiteway, M. (2007). Transcriptional rewiring of fungal galactose-metabolism circuitry. Curr. Biol. *12,* 1007-1013.

16. Askew, C., Sellam, A., Epp, E., Hogues, H., Mullick, A., Nantel, A., and Whiteway, M. (2009). Transcriptional regulation of carbohydrate metabolism in the human pathogen *Candida albicans*. PLoS Pathog. *10,* e1000612.

17. Lavoie, H., Hogues, H., and Whiteway, M. (2009). Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. Curr. Opin. Microbiol. *6,* 655-663.

18. Lohr, D., Venkov, P., and Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. *9,* 777-787.

19. Traven, A., Jelicic, B., and Sopta, M. (2006). Yeast Gal4: a transcriptional paradigm revisited. EMBO Rep. *5,* 496-499.

20. de Boer, C.G., and Hughes, T.R. (2012). YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities. Nucleic Acids Res. *Database issue,* D169-79.

21. Whiteway, M., Tebung, W.A., Choudhury, B.I., and Rodriguez-Ortiz, R. (2015). Metabolic regulation in model ascomycetes - adjusting similar genomes to different lifestyles. Trends Genet. *8,* 445-453.

22. Scannell, D.R., and Wolfe, K. (2004). Rewiring the transcriptional regulatory circuits of cells. Genome Biol. *2,* 206.

23. Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., and Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. PLoS Biol. *3,* e1000329.

24. Kelly, M.T., MacCallum, D.M., Clancy, S.D., Odds, F.C., Brown, A.J., and Butler, G. (2004). The *Candida albicans* CaACE2 gene affects morphogenesis, adherence and virulence. Mol. Microbiol. *3,* 969-983.

25. Durrenberger, F., Wong, K., and Kronstad, J.W. (1998). Identification of a cAMP-dependent protein kinase catalytic subunit required for virulence and morphogenesis in *Ustilago maydis*. Proc. Natl. Acad. Sci. U. S. A. *10,* 5684-5689.

26. Ramirez, M.A., and Lorenz, M.C. (2009). The transcription factor homolog CTF1 regulates {beta}-oxidation in *Candida albicans*. Eukaryot. Cell. *10,* 1604-1614.

27. Pennacchio, L.A., and Visel, A. (2010). Limits of sequence and functional conservation. Nat. Genet. *7,* 557-558.

28. Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S. *et al*. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science *5981,* 1036-1040.

29. Johnston, M. (1999). Feasting, fasting and fermenting. Glucose sensing in yeast and other cells. Trends Genet. *1,* 29-33.

30. Moehle, C.M., and Hinnebusch, A.G. (1991). Association of RAP1 binding sites with stringent control of ribosomal protein gene transcription in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *5,* 2723-2735.

31. Hogues, H., Lavoie, H., Sellam, A., Mangos, M., Roemer, T., Purisima, E., Nantel, A., and Whiteway, M. (2008). Transcription factor substitution during the evolution of fungal ribosome regulation. Mol. Cell *5,* 552-562.

32. Mallick, J., and Whiteway, M. (2013). The evolutionary rewiring of the ribosomal protein transcription pathway modifies the interaction of transcription factor heteromer Ifh1-Fhl1 (interacts with forkhead 1-forkhead-like 1) with the DNA-binding specificity element. J. Biol. Chem. *24,* 17508-17519.

33. Tuch, B.B., Galgoczy, D.J., Hernday, A.D., Li, H., and Johnson, A.D. (2008). The evolution of combinatorial gene regulation in fungi. PLoS Biol. *2,* e38.

34. Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yanai, I., McClellan, M., Berman, J., and Barkai, N. (2005). Rewiring of the yeast transcriptional network through the evolution of motif usage. Science *5736,* 938-940.

35. Booth, L.N., Tuch, B.B., and Johnson, A.D. (2010). Intercalation of a new tier of transcription regulation into an ancient circuit. Nature *7326,* 959-963.

36. Cote, P., Hogues, H., and Whiteway, M. (2009). Transcriptional analysis of the *Candida albicans* cell cycle. Mol. Biol. Cell *14,* 3363-3373.

37. Hussein, B., Huang, H., Glory, A., Osmani, A., Kaminskyj, S., Nantel, A., and Bachewich, C. (2011). G1/S transcription factor orthologues Swi4p and Swi6p are important but not essential for cell proliferation and influence hyphal development in the fungal pathogen *Candida albicans*. Eukaryot. Cell. *3,* 384-397.

38. Moran, G.P. (2012). Transcript profiling reveals rewiring of iron assimilation gene expression in *Candida albicans* and *C. dubliniensis*. FEMS Yeast Res. *8,* 918-923.

39. Maguire, S.L., Wang, C., Holland, L.M., Brunel, F., Neuveglise, C., Nicaud, J.M., Zavrel, M., White, T.C., Wolfe, K.H., and Butler, G. (2014). Zinc finger transcription factors displaced SREBP proteins as the major Sterol regulators during Saccharomycotina evolution. PLoS Genet. *1,* e1004076.

40. Sherwood, R.K., Scaduto, C.M., Torres, S.E., and Bennett, R.J. (2014). Convergent evolution of a fused sexual cycle promotes the haploid lifestyle. Nature *7488,* 387-390.

41. Medina, E.M., Turner, J.J., Gordan, R., Skotheim, J.M., and Buchler, N.E. (2016). Punctuated evolution and transitional hybrid network in an ancestral cell cycle of fungi. Elife 10.7554/eLife.09492.

42. Pougach, K., Voet, A., Kondrashov, F.A., Voordeckers, K., Christiaens, J.F., Baying, B., Benes, V., Sakai, R., Aerts, J., Zhu, B. *et al*. (2014). Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. Nat. Commun. 4868.

43. Lorenz, M.C. (2013). Carbon catabolite control in *Candida albicans*: new wrinkles in metabolism. MBio *1,* e00034-13.

44. Sandai, D., Yin, Z., Selway, L., Stead, D., Walker, J., Leach, M.D., Bohovych, I., Ene, I.V., Kastora, S., Budge, S. *et al*. (2012). The evolutionary rewiring of ubiquitination targets has reprogrammed the regulation of carbon assimilation in the pathogenic yeast *Candida albicans*. MBio *6,* 10.1128/mBio.00495-12.

45. Leach, M.D., Stead, D.A., Argo, E., MacCallum, D.M., and Brown, A.J. (2011). Molecular and proteomic analyses highlight the importance of ubiquitination for the stress resistance, metabolic adaptation, morphogenetic regulation and virulence of *Candida albicans*. Mol. Microbiol. *6,* 1574-1593.

46. Harcus, D., Dignard, D., Lepine, G., Askew, C., Raymond, M., Whiteway, M., and Wu, C. (2013). Comparative xylose metabolism among the Ascomycetes *C. albicans*, *S. stipitis* and *S. cerevisiae*. PLoS One *11,* e80733.

47. Wong, B., Leeson, S., Grindle, S., Magee, B., Brooks, E., and Magee, P.T. (1995). D-arabitol metabolism in *Candida albicans*: construction and analysis of mutants lacking D-arabitol dehydrogenase. J. Bacteriol. *11,* 2971-2976.

48. Hahn-Hagerdal, B., Karhumaa, K., Jeppsson, M., and Gorwa-Grauslund, M.F. (2007). Metabolic engineering for pentose utilization in *Saccharomyces cerevisiae*. Adv. Biochem. Eng. Biotechnol. 147-177.

49. Barelle, C.J., Priest, C.L., Maccallum, D.M., Gow, N.A., Odds, F.C., and Brown, A.J. (2006). Niche-specific regulation of central metabolic pathways in a fungal pathogen. Cell. Microbiol. *6,* 961-971.

50. Ramirez, M.A., and Lorenz, M.C. (2007). Mutations in alternative carbon utilization pathways in *Candida albicans* attenuate virulence and confer pleiotropic phenotypes. Eukaryot. Cell. *2,* 280-290.

51. Moran, G.P., Coleman, D.C., and Sullivan, D.J. (2011). Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. Eukaryot. Cell. *1,* 34-42.

52. Askew, C., Sellam, A., Epp, E., Mallick, J., Hogues, H., Mullick, A., Nantel, A., and Whiteway, M. (2011). The zinc cluster transcription factor Ahr1p directs Mcm1p regulation of *Candida albicans* adhesion. Mol. Microbiol. *4,* 940-953.

53. Pan, T., and Coleman, J.E. (1990). *GAL4* transcription factor is not a "zinc finger" but forms a Zn(II)2Cys6 binuclear cluster. Proc. Natl. Acad. Sci. U. S. A. *6,* 2077-2081.

54. Bruckner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. Int. J. Mol. Sci. *6,* 2763-2788.

55. Hong, M., Fitzgerald, M.X., Harper, S., Luo, C., Speicher, D.W., and Marmorstein, R. (2008). Structural basis for dimerization in DNA recognition by Gal4. Structure *7,* 1019-1026.

56. Flynn, P.J., and Reece, R.J. (1999). Activation of transcription by metabolic intermediates of the pyrimidine biosynthetic pathway. Mol. Cell. Biol. *1,* 882-888.

57. Roy, A., Exinger, F., and Losson, R. (1990). cis- and trans-acting regulatory elements of the yeast *URA3* promoter. Mol. Cell. Biol. *10,* 5257-5270.

58. Gola, S., Martin, R., Walther, A., Dunkler, A., and Wendland, J. (2003). New modules for PCR-based gene targeting in *Candida albicans*: rapid and efficient gene targeting using 100 bp of flanking homology region. Yeast *16,* 1339-1347.

59. Noble, S.M., and Johnson, A.D. (2005). Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. Eukaryot. Cell. *2,* 298-309.

60. Lavoie, H., Sellam, A., Askew, C., Nantel, A., and Whiteway, M. (2008). A toolbox for epitope-tagging and genome-wide location analysis in *Candida albicans*. BMC Genomics 578-2164-9-578.

61. Bataille, A.R., and Robert, F. (2009). Profiling genome-wide histone modifications and variants by ChIP-chip on tiling microarrays in *S. cerevisiae*. Methods Mol. Biol. 267-279.

62. Nantel,André, Rigby,Tracey, Hogues,Hervé, and Whiteway,Malcolm. Microarrays for Studying Pathogenicity in *Candida albicans*.

63. Inglis, D.O., Arnaud, M.B., Binkley, J., Shah, P., Skrzypek, M.S., Wymore, F., Binkley, G., Miyasato, S.R., Simison, M., and Sherlock, G. (2012). The Candida genome database incorporates multiple Candida species: multispecies search and analysis tools with curated gene and protein information for *Candida albicans* and *Candida glabrata*. Nucleic Acids Res. *Database issue,* D667-74.

64. Klepper, K., and Drablos, F. (2013). MotifLab: a tools and data integration workbench for motif discovery and regulatory sequence analysis. BMC Bioinformatics 9-2105-14-9.

65. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *Web Server issue,* W202-8.

66. Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R. *et al*. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *Database issue,* D700-5.

67. Huang, X., and Miller, W. (1991). A time-efficient, linear-space local similarity algorithm. Advances in Applied Mathematics *3,* 337-357.

68. Duret, L., Gasteiger, E., and Perriere, G. (1996). LALNVIEW: a graphical viewer for pairwise sequence alignments. Comput. Appl. Biosci. *6,* 507-510.

69. Morozov, A.V., and Siggia, E.D. (2007). Connecting protein structure with predictions of regulatory sites. Proc. Natl. Acad. Sci. U. S. A. *17,* 7068-7073.

70. Wang, H., Xu, Z., Gao, L., and Hao, B. (2009). A fungal phylogeny based on 82 complete genomes using the composition vector method. BMC Evol. Biol. 195-2148-9-195.

71. Lee, I.R., Yang, L., Sebetso, G., Allen, R., Doan, T.H., Blundell, R., Lui, E.Y., Morrow, C.A., and Fraser, J.A. (2013). Characterization of the complete uric acid degradation pathway in the fungal pathogen *Cryptococcus neoformans*. PLoS One *5,* e64292.

72. Wong, S., and Wolfe, K.H. (2005). Birth of a metabolic gene cluster in yeast by adaptive gene relocation. Nat. Genet. *7,* 777-782.

73. Scott, S., Abul-Hamd, A.T., and Cooper, T.G. (2000). Roles of the Dal82p domains in allophanate/oxalurate-dependent gene expression in *Saccharomyces cerevisiae*. J. Biol. Chem. *40,* 30886-30893.

74. van Vuuren, H.J., Daugherty, J.R., Rai, R., and Cooper, T.G. (1991). Upstream induction sequence, the cis-acting element required for response to the allantoin pathway inducer and enhancement of operation of the nitrogen-regulated upstream activation sequence in *Saccharomyces cerevisiae*. J. Bacteriol. *22,* 7186-7195.

75. Dorrington, R.A., and Cooper, T.G. (1993). The DAL82 protein of *Saccharomyces cerevisiae* binds to the DAL upstream induction sequence (UIS). Nucleic Acids Res. *16,* 3777-3784.

76. Finkel, J.S., Xu, W., Huang, D., Hill, E.M., Desai, J.V., Woolford, C.A., Nett, J.E., Taff, H., Norice, C.T., Andes, D.R. *et al*. (2012). Portrait of *Candida albicans* adherence regulators. PLoS Pathog. *2,* e1002525.

77. Hofman-Bang, J. (1999). Nitrogen catabolite repression in *Saccharomyces cerevisiae*. Mol. Biotechnol. *1,* 35-73.

78. Suarez, T., de Queiroz, M.V., Oestreicher, N., and Scazzocchio, C. (1995). The sequence and binding specificity of UaY, the specific regulator of the purine utilization pathway in *Aspergillus nidulans*, suggest an evolutionary relationship with the PPR1 protein of *Saccharomyces cerevisiae*. EMBO J. *7,* 1453-1467.

79. Liu, T.D., and Marzluf, G.A. (2004). Characterization of pco-1, a newly identified gene which regulates purine catabolism in Neurospora. Curr. Genet. *4,* 213-227.

80. Peman, J., and Quindos, G. (2016). Current aspects of invasive diseases caused by Candida and other yeast fungi. Rev. Iberoam. Micol.

81. Patterson, T.F. (2006). The role of echinocandins, extended-spectrum azoles, and polyenes to treat opportunistic moulds and Candida. Curr. Infect. Dis. Rep. *6,* 442-448.

82. Pfaller, M.A. (2012). Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment. Am. J. Med. *1 Suppl,* S3-13.

83. Cappelletty, D., and Eiselstein-McKitrick, K. (2007). The echinocandins. Pharmacotherapy *3,* 369-388.

84. Schuetzer-Muehlbauer, M., Willinger, B., Krapf, G., Enzinger, S., Presterl, E., and Kuchler, K. (2003). The *Candida albicans* Cdr2p ATP-binding cassette (ABC) transporter confers resistance to caspofungin. Mol. Microbiol. *1,* 225-235.

85. Cheah, H.L., Lim, V., and Sandai, D. (2014). Inhibitors of the glyoxylate cycle enzyme *ICL1* in *Candida albicans* for potential use as antifungal agents. PLoS One *4,* e95951.

86. Qin, Y., Zhang, L., Xu, Z., Zhang, J., Jiang, Y.Y., Cao, Y., and Yan, T. (2016). Innate immune cell response upon *Candida albicans* infection. Virulence 1-15.

87. Dunn, M.F., Ramirez-Trujillo, J.A., and Hernandez-Lucas, I. (2009). Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. Microbiology *Pt 10,* 3166-3175.

88. Lorenz, M.C., and Fink, G.R. (2002). Life and death in a macrophage: role of the glyoxylate cycle in virulence. Eukaryot. Cell. *5,* 657-662.

89. Kasten, M.M., and Stillman, D.J. (1997). Identification of the *Saccharomyces cerevisiae* genes *STB1-STB5* encoding Sin3p binding proteins. Mol. Gen. Genet. *4,* 376-386.

90. Akache, B., and Turcotte, B. (2002). New regulators of drug sensitivity in the family of yeast zinc cluster proteins. J. Biol. Chem. *24,* 21254-21260.

91. Larochelle, M., Drouin, S., Robert, F., and Turcotte, B. (2006). Oxidative stress-activated zinc cluster protein Stb5 has dual activator/repressor functions required for pentose phosphate pathway regulation and NADPH production. Mol. Cell. Biol. *17,* 6690-6701.

92. Idnurm, A., Giles, S.S., Perfect, J.R., and Heitman, J. (2007). Peroxisome function regulates growth on glucose in the basidiomycete fungus *Cryptococcus neoformans*. Eukaryot. Cell. *1,* 60-72.

93. Lorenz, M.C., and Fink, G.R. (2001). The glyoxylate cycle is required for fungal virulence. Nature *6842,* 83-86.

94. Noble, J.A., Tsai, H.F., Suffis, S.D., Su, Q., Myers, T.G., and Bennett, J.E. (2013). *STB5* is a negative regulator of azole resistance in *Candida glabrata*. Antimicrob. Agents Chemother. *2,* 959-967.

95. Sellick, C.A., Campbell, R.N., and Reece, R.J. (2008). Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them. Int. Rev. Cell. Mol. Biol. 111-150.

96. Huang, H.L., and Brandriss, M.C. (2000). The regulator of the yeast proline utilization pathway is differentially phosphorylated in response to the quality of the nitrogen source. Mol. Cell. Biol. *3,* 892-899.

97. Brandriss, M.C., and Magasanik, B. (1979). Genetics and physiology of proline utilization in *Saccharomyces cerevisiae*: enzyme induction by proline. J. Bacteriol. *2,* 498-503.

98. Leverentz, M.K., and Reece, R.J. (2006). Phosphorylation of Zn(II)2Cys6 proteins: a cause or effect of transcriptional activation? Biochem. Soc. Trans. *Pt 5,* 794-797.

99. Bahn, Y.S. (2015). Exploiting Fungal Virulence-Regulating Transcription Factors As Novel Antifungal Drug Targets. PLoS Pathog. *7,* e1004936.

100. Duffy, M.J., Synnott, N.C., McGowan, P.M., Crown, J., O'Connor, D., and Gallagher, W.M. (2014). P53 as a Target for the Treatment of Cancer. Cancer Treat. Rev. *10,* 1153-1160.

Appendix: Supplemental material

Table S1: Oligos used for DNA amplification in these studies.

Related to Table 2.1, Figure 2.5, Table 3.1, Figue 3.2, Figure 3.3, Figure 3.4, Figure 4.2, Figure 4.3 and Figure 4.4.

Name	Sequence (5' - 3')
PPR1_TAP_URA3_F	AATAATAATTCATCAATTCGTGATGAAAAGAAAACTTTTGAATTGATTCATGATATGCCTAATGAAATTATTTGGGATGAATTTTTTGCCAACAAACAAGGTCGACGGATCCCCGGGTT
PPR1_TAP_URA3_R	AATATAACAACTACGCTGTTCGTAGTATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAACCCGCTACCCTTTACCACTCCTTCTCGATGAATTCGAGCTCGTT
PPR1_Marker_KO_F	AATTGATAAACAAATTAACCAATTAATTAATTAATACTCTCTCTACTTTTCTTATCATTACTTCAGAAAATCATTCCTCTTTTGCATATACACATACGAAGCTTCGTACGCTGCAGGTC
PPR1_Marker_KO_R	TAACAACTACGCTGTTCGTAGTATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAACCCGCTACCCTTTACCACTCCTTCTCTGATATCATCGATGAATTCGAG
PPR1_TAP_URA3_check_F	CCAGCCAATAGCATTCCTTC
PPR1_TAP_URA3_check_R	GACCCGGTCAAATTGAAATAAA
FT-U1	TTGAAGGATTAAAACAGGGAGC
FT-U3	TATAGGTCTTAGTGTTGACTGT
PPR1_KO_Check_F	AAATTCCCATTCCCATTTGC
PPR1_KO_Check_R	CAATTCCGGGAGTTTATCACA
FT-H3	GGACGAATTGAAGAAAGCTGGTGCAACCG
FT-H2	CAACGAAATGGCCTCCCCTACCACAG
PPR1_KO_Check_Internal_F	GTCGTTGTGATGCTGCAAAT
PPR1_KO_Check_Internal_R	TTTTCGATGATGACGTGGAA
STB5_TAP_URA3_F	TTGTCCAAAGAACTATTAGGTGCAAGCTCATCATATCATGAGATATTGAATGTTCATTCAGTTGATCTTTGTATCGATACTGATTTCATGTACAATTTTGGTCGACGGATCCCCGGGTT
STB5_TAP_URA3_R	ATATATAAGGCACACAATCAAGGCATTTGCAATTCTATTTGTGGATACTACAATAAAATAAATAAATTAATTAGACAGTACAT

	ACTAAATAACAAGGAATCGATGAATTCGAGCTCGTT
STB5_Marker_KO_F	AAACATAAATCTCCAGAAGTTTTATTTCTCAACACGGGGTTTTTAGAATTAGTCATACGGTGTTCCCTACAACCTTTGCTCAGCGCCAAAAACACACGAAGCTTCGTACGCTGCAGGTC
STB5_Marker_KO_R	ATAAGGCACACAATCAAGGCATTTGCAATTCTATTTGTGGATACTACAATAAAATAAATAAATTAATTAGACAGTACATACTAAATAACAAGGAATCTGATATCATCGATGAATTCGAG
STB5_TAP_URA3_check_F	TCCAAGACCATCACAAGACCT
STB5_TAP_URA3_check_R	GATGATGCGTTGGTTTGGTT
STB5_KO_Check_F	AATGGCGCGACTGAAAATAG
STB5_KO_Check_R	TCTGATGATGCGTTGGTTTG
STB5_KO_Check_Internal_F	CTTGCTGATAATCCGGCTGT
STB5_KO_Check_Internal_R	CCCATGATTTCCCAAGATGT
PUT3_Marker_KO_F	CATTCCGAACAAGGTATTAGAAATAATTATTTTCTATCAACCAACACCCACATAAATCATTCCTTCATTTACTTATATATAATCCGATTCTTGTACAGAAGCTTCGTACGCTGCAGGTC
PUT3_Marker_KO_R	GTTTTTGTAATATATTGTATTATATAGAAAATTTTATTACCATCACAGAATAAATGTACAGACATAAATATATATTTGCCTCACTCCCGCACAATTCTGATATCATCGATGAATTCGAG
PUT3_KO_Check_F	CGTTGTGCGAAACTACCAAA
PUT3_KO_Check_R	ACAGGCCAATGAGAATACGC
PUT3_KO_Check_Internal_F	AAGAGCCTTCGGAGGAAAAG
PUT3_KO_Check_Internal_R	TCGGGACTTTTTGTTTGGAG

Table S2: Top 50 Ppr1 ChIP-Chip Sequences submitted to MEME to identify Ppr1 DNA binding motif.
Related to Figure 2.4A, Chapter 2.

ChIP-Chip identified top 50 sequences corresponding to C. albicans Ppr1 target genes used used to identify C. albicans Ppr1 DNA binding motif
>orf19.230
AAATAATATTGTTGATGATTAATTACAGAAAAGAATTCATTCTATACATAAATATATATATAT ATATACATATATTTATATTTATGTACATATATATATCTATATGAGAAAAATTAAACTATTTAC AAAAAAAAAGTTCATAAAATTCAATCTCAATGGTAAGCAGTACTACCTATTATTATTATCTTT CCATTGATTTAAAATATAAATTGTTTTATCAATTAATTCCATTATTTCATTTTCATCATTTTCC AACATTAATTTAACATCACCTATAAAATGACGTAATTTCATTAAAGATTTTTTACCTTCCCCC AATAATAATAATAATGATGATGATGATGATGATTCTTGATCATCAACATCATCTAATGAGTTA TTATTATTATTAGTAGTAGTAGTAGTAGTGCTACTCTTCTGTATTGATCGTAATTTCTTGAATA ACATATTTTCAATTAATAAAGTTTGATCTAATAAATATGGAGTTAAATAAATTTGTT
>orf19.4691
TTTGATATGTACAGTGAATGATCCCCCCTCACATTATGGTACATTCCATCATCATTCAAAAAG AGATTTTTTTTAATAATAATGAATTATATTATGTAATCAAAAACCAATCAAATTTCAAGAGAC TTCTACAAGAATATAATATGTGGAGTAATTTTAAGGTCTTTTATATTCTAACAAATTAGTAAC TAATAATTATACCCAAATGAAAGTAATTCACTTGGAATAACACTCAAACTATAGTTAATTGTT TAAGAAAAGGAAAATAATATATTTCCGTAAATAGATCATATATTTAACAAACTAAGATTTAT TGATGAGCGAATATGAAAAGGAGAGAATCTTAATGTTTTGTTTCACAACTTTTACGAACTTTT TTTTTTTTATTTAATGTGAATCACCCAAAACTTAACATTAATGAGTATGGAGTTCTTTCTTCTT GAGCCTGTTTTGTCAACCACTGTACTGGTTGAATCAAGGGTAATGAGGATGTAGTCTGT
>orf19.1358
AACTACTGATTTTTTTGTGTTAAATTAATTTATTCACCACAAGGACACGACTATCCTACACTC ACATACAAAAAAATTTATTAACAACAACAACAACAACAACAGCAGCAGCAGGGAGAGAGAG AGATAAAAAAAAATTTTTCATTATAAAATCAACAAGAAATAACCTTGCGCTTAAAAAAAGTT GAATAAAATTTCCTCTTTTTCTTCTTCTTCTTTCTCTTACTAATTACACCAATTATATAATT AAGAACAAAGTAACATAACAACAACAACAACTACTACTACTACTATTACTATTACTACTACG TCTTAAATCTGCTATAAACTTATCATAAATTATTATTATTTATTATTTACCCCTTTTTTATTCAT TATATTGTCCAAATTTTTTCCAAGATTATATACATTAGAAAAAATGTGTGCCTAAATTTATTT ACCCTTGTAATACAATAAATTTGTTTATTAATTAATTAGAAAAAAAAAGTTGTTTTGTT
>orf19.2065:orf19.2064
TGTATATTGTGAAACAATTTGCTTTTGAAAGTCGTCTTGAGAAAGAATTTTGGTCATCGAAAA TGCTTTAAATAATGTTTAAAAAAAGTTATGTGTGATTATATAAAATATCATATATGCAATTGT

AGAAAGGAGATCTCCACCAATTTATAGTTGAAAGAGTTTTGAAAAATCCAATCTTTGATTCG
GGGTTGTCCGAGACAAACTTCCTATTGATAGTCGCACAGTTTGCCGAATTAGATGTTAGATTG
GTAGCGCCAGGTACACCACAATTGGAAATAAATCCCGCCAACAATTGTTTGGCAATTAAGTG
ATAGCAGCTTTCAAATACGATAGTATCGTACATATTACGAAGATCTTTTAATAGTTAACGAA
ATACAAGTGTACATAAAAATATATGCGAGCCCATTTGAGCCAAATCAAGTAAACTACATCGT
GAATAATCATGTATAAAGCGCCCATCGGTTTTATCTTTGACTTTTATAGCCGATCAGTACCCC

>orf19.2928

ATCACTCCTCTTCCTCCTATTATCAGACCACCCACCACCACCACCAATAACTAACAAACCTAC
AATTACATAACAAATAAAACAAACAAACAAAAAATACTGTATACATATAGATAGATAGATA
GATAGATAGATAAAAACAAGTAATATATTTATGTATATATATAAAAACAACTGTGTAATAAT
AATAATAATGATAATATAATAATATAATATGAAAAAAAAAGAAAAAACAAAAACTAATATA
TAAAGTTACATTTGGAAATAATCCTTTCCCCCGTTTTACTGTTTTTTATTTTCTTTTCTTCTTTT
CAAAAAAAAAATTCTTAAAAAAAAAACAAAAACCAAAATTTATGAATAAATGAAATAATTA
TAAATGAATAAACCACTTCCGCTCCCCCACTTAATCTCTTAAACTCTTAAACTCTGAATGGAT
TTGTAGAATAAGGGGTGGTACTTCCAGTAAAACGAGTACTCAAATAACTCCAAGTGTAGGTT
TT

>orf19.77.1

TAGTAGTGCAATTTAGTAAGCAAGTTGGCTAAAAAAGAAAAAAAAGAAGAATTTTTTTTAAA
TTACACAAAACCAAATAATAATAATAATTTTTTTTTTTTACGTGTCTTATTATTTAAATCA
ATTTAAAGTATCATGCACCCACTTCATCCACCCATCCATTTTTTAATGATATATCAAACATTC
CCTCTCTTCTTCAACTTCAACTTCTACTTGTTGTAATTGAAAGCTTCTACTTATAGATTAAAGA
TAGTTCTTTATGCAAATAATTTAACTAAAATCTATTTACTTTCACTCTAAATTATATAAATTGA
AATCAATAAATAAATAAAAGAACAGTAGGACTTGTTAGTTAATAAGACTTCATATAATTAAA
CAATAAATGAATTTAGAAAAAAAAAAAAAAAACAGGTGGAAGGTTGTTGTTGTTTATTGTATT
CTATTAAAGGAAAAGGAGAGATGGGATTAGAGGTTTCTAGTTTCCATTTTGTATGTTCA

>orf19.5450

CAACACTCTAGTCGATGATAATGCTCGTGACATGGTTGCCTGATAAGAGAAAGAAAACTACT
AAGGTCGACGAAATTTTTACAATATATATATTTTTTGAGATAAACTTTTTTTTCTCTTTGTGTA
ACGATTCGGTTTTGGCCGAAACACTCATCGGGATGCGAGGCCCACTAGTATGTGATCAATTG
TTTACCCGATTTAGAATTGGCTACCCCATCTTTATACAACTCATGCAATTCCTTGCTACCATCA
AATCTCGTCTCGGTCGATTTCGCATCTTTTAATTCACCAGACTCGTACCATTCAATGATTTGAT
TCAATGTCTTCAACTTCAATTCCCTATTGTGTTTCAACAACTCAGTCACCCAGAACCCAGCAC
TAGTAAAGTTCTTGAAAATGTGTAATGACGTAGGAATGGTTACTGGCTGGAATGACATCCCA
CCATAAGTCAACATCAACCCGTTATTGTTCAATTTTCTGGCAATACCAGTAGAACTTTT

>orf19.6272
ACCTCAAGATTTAATCATTTCAAATTCTAAATTATCACAATTAAGAAATAATTTGACCAAAGT TAAGGAATTTTCTATTGATGATCGAGCAAAAATTATTGAAAATTTAGAAAATGATATTAAAT TAGATGAAAATTTTGGTGAAAAAGGTGATACTAAAGCTCCTGATATTAATGAATCAGTTGAT TTACATTTTATTTCATTTATTAAAAGTACGAAAAATGGTCATTTATATGAATTGGATGGTAGA AGAACTGGTCCTATTGATTTAGGTGAGTCGAATAACAAGGACAATAATAATAATAATAATAA TATTATTAATGATCCTTTATTGATTGAAAAAATTCAATTTTATATTGATAATGCTGATGAAGC AAATAAAAATAATTTTGCAATTATGGCTATTGGTCCTTCCTTTGATTGATAAAGGAGGGGGG GAAGGGGTTGAAAGTTGATAAGTATAGTTTATGAATTTGAGACACAAGAAAAAGGTAATGCT A
>orf19.6689
ACTTTACTGATTGATAAACAAAATATGGAAAAAGCTTTAACTATGGATATGTTGGCTACTGA TTTAGCTGATTATTTAGTTAGAAAAGGAGTTCCATTTAGAGAAACTCATCATATATCTGGTGA ATGTGTTAGAAAAGCTGAAGAAGAAAAATTATCTGGAATTGATCAATTATCTTTTGAACAAT TTCAACAAATTGATTCTCGTTTTGAAAAAGACGTTATGGAAACTTTTGATTTTGAAGCTAGTG TGGAAAGAAGAGATGCTATTGGTGGTACTGCTAAAAGTGCCGTTTTAAAACAATTAGAAAAT TTAAAATCAATTTTAAGTTAATTTATAAATAGTCATATAATAATCACAGTATTGTTTAACTTA CATAAAATATAATCAATTAACACAGAGATACCTTGTACTCCAATACTAGTGTTATAATGTTTA TAACAGCAGCTTTCAAGAATAAGGAACTAAATAAAAGCTTCTTCCATCACAACACACAACAC
>orf19.6480,orf19.6479
TTAATCGGATTGGATTGGATTGAATCGGATTGGTATGAATTGGTCAGAAAAAGAGAGAAAGA AAGAAAAAAAGAATTTTTTTTTTCCTTGGTGGGTGGGTGGTGTCTTGTTTTGGATCAACAACA ACAAGCAATTTCATTATATACAACACGTATTCAAAATAACACGTATTCATTATATATATATAT ATATATATATAGTCAACTACTAAAGATAAGAAAGATATTAACAAGATAAAAAAGGAATAAG ATAGAAAGAACTATATAGGATATATAGATATACTAACACGCTTTTATTGTGTTATTTGTAAAG CAGATTTTTAATGCAGGGATTGGGGTAGATGAAACAATTTCATATTACACACCAATTAAACT CTTGAAGAAGGGAGTTGTTTCAAATGAGCAACTTTAATAAGCAATACGGGGCATTCACAAGT AGATCGAAATAACACTTTCTTACCTTTAAACCATTAGTAATAGTAATAGTAGTAGTAGTAGTA G
>orf19.1043
TTATTCTAATTCTAATTCTAATTTTTCTAGTTCTGGGAAATTAGTGGAAGATGTATCAAAATT ACAAAATAAAATTGTGTTTTTATTTTTAGAAGGGACTTTATCAAATAATAAATCTATTTTACC ATTTATTAAATTGAATTCTAAATATTCATTTGATGATTTTAATATTAAATCATTAGTTGTGAAA TTAATACCAAATTATTTTACATTACCAATTCCTCATGTTAGTAAATTACAATATTTTTTCGAAC

TTTTAACAAATTTATCTCAAAAACTGGTTCGTTATAAAATGTATAAATTTGATGAATTTGATA TTGCCAAGATTAGAAATTCATTTGAATTGAATTCTTTAACTTCAATGAGTCAAGATTTGAATA TTGATTCCAAGAATAAATTTATAGACTATTATTTTGATCATAGTATAAAGAAGAATTAAAAAT GTTGTATATATAAATAAATATATATATATATATATGTATTATATGAAGCTACCCTTT
>orf19.155
ATCATATAATTGTTAATAATTTCGAATATTAAAAGAAAAAAGGGGAAACTGAGTTATATATC ACAATTGTTGATTGTATCTGATTTAATGTCAAAATCTATAGATTATATCAATAGGAAATGGGG TCTGCAGATAGATAAATAAATAAATAAGCAATTAAGTTAATGAATAAATGATTGAAAACCCA AACTTATAATCAAAAAAAAAAAGCCAATATAGTAAATTTATTATAATCTGTGATAATTGATG GAATAATGGGAATTTGATATATTTATGGAGGAGTAAATAAACCTCTTAATAATAGAAGATAA GTATATAAATTAATATTAAATCAATTGAAATAAAACTAGTTGTGATTTATAATTAAGAACTA GTAGTAGTAGTAGTAGTGGTGGTGATGGAGGTGATGAAGGTGGTGGTGATATGGGTGAGTTA GTTAGTTAATTAGTTTGCTTATTTGTTTCAAGTTGAAAGAGAAAGAGAAAGAAGGAAAAAAT TTT
>orf19.4640
AATTATAGAACATTCCCTATGGATGGAGGAATTGATAAAAAGACAAAAGAAGTTGTCGAAG ATCAAACAAGTGGTATAGGTGTAGTTTTATTATAATTAGAGTGAAACTTTTATATATATATAT ATATATATAATATAAAGCATTCCATGTTTTGGTTATGATTGCTAGTATAAGTATAAGTATAAG CATAAGTATAAGTATAAGTATAAGTATAAGTAGAGTGGTTGTTGTTCATTTATTGATCATCTC TGTCTTCACTTAAATTCTTTTTGTCAATCTTTTGGTAAAGAATTTTTTTTTTTTTTGTTTTTTTG TTTTTCAATCTCATCCCATTTTTTTTTTATTTTTTTTTGAAAAATAATTTCTATAATTCAAACCT ACTCCCTTATTCCCTTCAATAATAGTAATAAATAAATAAATAGAAATAATGATTTCTTCAAGT ACTTGGGTTCCAAGGGGTTTTGCTGCAGAATTCCCAGAGAAATATGAATTAGATGA
>orf19.5454
AACCCTTCCCATTCAGTTCTACCGGGCTCGTTCAAGTGGACATGTGCATCAACTAGCCCAGGC ATAATCACTGCTGGCGTAACATTTCTATAGTCTATAGGAAAAACATTATAGCGTGCCAAAAT CGGATCATTGGGCGGTAAAACCCTATCAACAATCTCTAAAATCGTGCCTGACTCTACCGAAA AAATGATGGTGGCTGGTAAAAGTGTGTCATTAACCAACACTCTAGTCGATGATAATGCTCGT GACATGGTTGCCTGATAAGAGAAAGAAAACTACTAAGGTCGACGAAATTTTTACAATATATA TATTTTTTGAGATAAACTTTTTTTTCTCTTTGTGTAACGATTCGGTTTTGGCCGAAACACTCAT CGGGATGCGAGGCCCACTAGTATGTGATCAATTGTTTACCCGATTTAGAATTGGCTACCCCAT CTTTATACAACTCATGCAATTCCTTGCTACCATCAAATCTCGTCTCGGTCGATTTCGCATCT
>orf19.915
TTCATTTAATCCAATAAATGATGATGCTGATGATGTATGATTATTATTATTATTAATAAT

AATTAAATCGGTACTTTTATTGAAAAAATATCTTTTGGAAGATTCATTTAATGCAAGGAAAAC
TAATGGATTAATTTCTAATTTATTGATTTGAATATTTATCAAATGATAATCAATTGGATAATT
AGTATTATGTAAATGTCCATGCATGCTCCCATCTAATAATCTTGCCGAAACATTAATTTGTTT
ATTTTGTAATAATCTTGTAATAATTGTCGACGTTTTTGTTTTCATTGATAATATTATTAAATCA
CCAATATTTTCAAAATTTAATATATGAGGTGATATTAATCCTTGTGAATTTTTACATTTATCAA
TAAATTCAAAACAATCTTGGAAAAATGGGATTTGATCATCTTCATGTCCATCAAACTTATCAT
ATTCATCATCTATATTATTCCCACAATTCATATCAAAATATTCATTATCATTACCCT

>orf19.1587:orf19.1588

CCAACCTAGAAGAGGGGTGATGCTATGTCTTTTATCAAGAAAGTTATGTATCGAAGTAGGAT
TAGGCATGGTTTAAAAAATGTTGACTTTCCAAAATAAAAAATTAACACAATCAACAAATCCA
CCTTTTTCTTTGCTTTGGGTTCTTGATATAATTTATATACCTCCTCAAAATCGGGTTTGTTTGC
AAAAGGGCCGCGACAATGGCACTGGGATCGCGAACAATTTTCATGGCTATCTATTGATTTAA
AATTTGATCTGACTCCGGCCATTTGCGAAATGTAAGGGCGCGTGCGACCTTACTTAGCAACA
GCAAACTAAGCCGGGTACCCGAATCGTTTTGTTTTATTAGAATTTATGGCACATGTATTCTAT
CAAATCATTCATGACGTCAGTAGCAAACTTCTATTACAGACCACAATAGCTTTTATGGGAAA
CTAAGACAATCACGTGGACATGTACTAACAACTAAGGAAAGTTAATTATGGTGACTCGTATC
T

>orf19.1078

AAATTTCAAATATTTAAAAGAAGAAGAAGAAGGAGGAGGAAGAGAAGGAGAGACAATCAT
TTTTTATTGTTGTTGTTGTTGTTGTTTGTTGTTGATTAATAATATTATTAGTTAGGATGGA
GAGAGAGAGAGAGAGAGAGAGTCATATAAGACCAATGGAAAGTCTGGAGAAACGGAACTTT
TCCTTTTTTTCTTTCTTTCTTTCTTTCTTTTTGGTAGCTCTTT
TGATTGCTTCAAAAATTCTCTCTCTCTCTCTTCTCCCCCCTCCCCGCCACCACGACTACTAC
TACTACTACTATTACTACTATTACTGCCCACTACCACCACCACCACCACTACCACTTGTTATC
AATTTTTTGTTCGACAAAAACTTTTTTTTTTCAGGTTTTTTTTTATCGTCGGAGATAACCGGAA
TCACGTGGAGAAGGAGTTGGCCTTCCGAAATGAGTAATATTACGCGGCTCTCACATACAAGA
ACTATCAATAGTTACTACA

>orf19.1078(2ndprobe)

TACTACTATTACTGCCCACTACCACCACCACCACCACTACCACTTGTTATCAATTTTTTGTTCG
ACAAAAACTTTTTTTTTTCAGGTTTTTTTTTATCGTCGGAGATAACCGGAATCACGTGGAGAA
GGAGTTGGCCTTCCGAAATGAGTAATATTACGCGGCTCTCACATACAAGAACTATCAATAGT
TACTACAATCTTTGATTAATTTTTTCGATAGCCGTTATTTTGGCAATCAACCTTAACTCTCTTT
AGATTAGCATTTTTGATTAGTTGTTCAGTTTGGCCTAAATATTTTTATGTGGTCTTTTTCTTTTT
TGACTCTGAAAAAATTATGGGGACAGCACAATAATACTGGCTTAGCAATGTCTTGATAATTG

CTAATTAGTCTCGCGACAGGTTAA ATAATACGTTGAAAAGCACTTATAACTTCTGAATGATAAACTCCTTTATTACTATCAGTTGGA GTACAAGTGTTATAGTTTTGGTTGTATATTTTG
>orf19.1681
GATGATTTGACTAGATTACCAGATGTGTTATTCAGTTACGTCAAACAATGGAGACATCCATTA ATTGCTCAAAAGGGGGCTCCAGGTAAGAAAGAATTAGAAAGATTAAGAGACGTCCAACATT TCACTCCATAGATGATATGATTATTTATAGTTTAAGTGTAAAGAGTTTATTTTTAAAAAGAAT ATATTCAGTAAACTAATAGTTAAAAAAAGGGAATAAAAGTAGAACAATGCAGCAATGTTTTT GAGGCTGAAATATAAATTCTGGAGATCTCCGAAGTTTAGACAACACCAAAAAAAAAAAGGA AAGGAACTATATAAAGGCAGAAAGATAATCTCCATTCTACCAGTCTTAGTTTTTTTAAAGGT GAAGAATATTTTTTGTACATCTAGTTAATATGGCATTACCAACAATCGAATCGTTCAAAAGAC TATCATCAAAAGATAAAAGACAAGTGCTTGATCATTTATTTGAGCCATGCAATACATTATCC AAT
>orf19.6470,orf19.6472
GATCTTAATTGCTATAGTTTCTTTACTGAGCACTGCCATGAAAAAGCTTTTTCTAACCAAAAG GTTATGTTAATTTTGTGGAGAAGAGATTTTGATATTGTATTATAAGCTCTACGTCATTGAAAC ATGTCACACATCACTAACTCTTGTAAACTTGTAAATTTTCGTCAATTCTTGAAATTACAACTC TATTCTTCTAGTTGGATCAACCAATTGTCTAATTGACAGGAATTGTCGTCAAGTTGACGTGGT TTTAGACCTAACTTTATTAGGTGGAGTAAAATTGAAATGGTGCGCCGGATATTACCGAGAAG CCACACCCTCCCTCCCTCCAGTCTCTACAACCAAGATAAGACGAACAGCCAAAAAAACAAAA AAATATCGATGCTATTTACCAACCCTCTTTCTTCAATTATCTATACTCCTCGATGTCATTTAAA GAAGGTTCCAAGTTCCCAGCAGGTGTTGAATTCTCCTATATTCCAATTCAATTGGATGAT
>orf19.2003
ATCAGGTGGATTTATGGTTAAACTATGTGTGGCCAATTCTACGAGATATCTTCATGCGGCATC CGCCGAAAGTTTATACAACAACAATAAGAAAAATATAAATACCTCATGCCAAATTATTATGT GGGGCTAAAATTAGTGTTTATGTATTCTTGAGCAATTGATTGGTTATTGATTAAACTAGATTT AGTAAATATGATCTTACTTCCAAACAACGGCAATCAATGCCTTTTTTTTTCGTAATTCAATTT ATTTTATTGTTGGGAATGAAGATATTATACATACATACATCCTCAATAATTATGTCATTTCTA CACAATTATAACGACTCTTTGGCTAAATAATCCCTACATTTGACTAATGCAGAAATCGGCTAT ATCAAATCATTATCTGTAATAGAACTAATAAGTAATTTTAAAAATATTTGCAGTTAATAATCG TAAATATTTCAAGAAGATCTTGATTCAACATTGGTTAAAAAAATTTAAAGGACGTTTCTT
>orf19.6259
CTGGTGAAAATGTAACTTGTTTCAATTCAGCCATGCTAAATGTTTGGTTTTTGGAGTTTGGAG ATTGTTTGTATTTTAATGTTTGATGCTCATCACCAAAACGCATTAAACAGGTTGGAAAAAAAA

AATCGAGACCATAAATGTCATCCGGGTAAATAAATTATTTGTACTGTATATAAAAAGTCACG
CTTAACTCACGTGCGTCGGAACTTTCCGGAAAAGTGTTCCCTAATTTGAAAGTTTATTAAGCA
ATGGTTGATTCCATTTTCGTTTATTTTTGGGGCCAAAATATTTACCAATTTTTACAAGAAAAA
AAAGTGCTTAAGGATTTAATCAAAATCATATGGGACAAATTGAATTTCTTGTTAATGCCGGA
ATCAATTTTACAAAAATCACATGATGGTCAAACCAATACACTCATTTTATTAATACAATAATA
TTATTTAATGATTGTAGTTTGCTTTCAATGGTTGGCCAAAGTATACTTGGTTTTCAGGCCA

>orf19.7473

TGTTGTTATTCTCATAGTAGTAGTCAGAGTCCTCACTAATGAAACTATTCATTACCAACCACT
AAAATTCTTTTGCTCTCTCTCTATTTTTGCATTCAATGTTGAGAATTAAATTGAATTTGAACCT
GTTTGAATGATCTGGTTTTTTTAATAAAGTGTGTAAATGTGAGTGAACAAGAAATCCGAAAA
TTCGCTCAACGCCGAAAAAAAAGCTTTAAAAAAATCAAAGTTATCTACGGAAAATAACGCA
AACACTCACCCGTAAATTACCCCCACCTACAACAACCAATATTAATTCACAAAATAAATATT
TTCAGTTTATTATTGCTTATATATACAATCAAATTGTCTGGTTTATTAAGTAATAGAGATTGTT
AAGGTAAAATTGAATTAAAGTACAACCAACAATATAACCAGATTGCAAATGTAGAACGATA
TATTAAAGCGTCGGAGTACAACAATCATTTCAAAGTTAATTGATTGGTTCATCTTTGGTATTT

>orf19.2114

GTCGGCTTTGGTGTACGATTCATCAAATTTACCTCTCAACAAAACCTGGACATTAGCTTCAAG
AACATCTTGCACAGTTGGGTTTGATGGATCTTTTTTGACCTTTAAAAATTTGACATTACCTTTA
CCGTAGGTTGAGGAGTATAATTCGGATTGCATGATTAGTTGTTGATTGAGAGAAAACTAGAA
AAAAAGGAAGAGGAATAAACCAGTCTATTTATACTTTTTTGTCCATATCTATTGGTTTGCTTT
AATTTCGGAGATCTCCACAGAAAGCCGCACACACGTCTACTTCAAGTAATCTCTGATAACAA
AAAAAAAAAAAAATCGGGGTGCCTATGCGGAGGACCGCAAACATATATTAACACTATTTAA
TCATTTTTCTTATCAGTATCAGCTGATTCTTCAACATTGCTTTCATCTTCGGCAAATCGTAAGC
TTTTTACTTTCTTGGTGGTTGTAGCGTCCTGAACACTTACCCCAGTTGCTTTAACATTTTC

>orf19.3038

ATTAGAAGAAACGTAAAATATATTGTTTCATTTCATGTGTGTATAATGTACTAGAGAAACTTC
ATAGAAAATGTATGAAGAATAAATTGCAAGTCCATTGGAGTAAGGATAATATCTGAGAAACT
TATAGGGATCAAATTTATTGTCAAAATGGTATCTTTATACTCTTATTCTTTGGTTGAATAGGTC
TTCAGTATAGTTTTATAGATATGATACTGTGTGTTTTCTTTTTGTCATATATCTTTACAACTAT
AATCTAGGTCTATGTGGTTCGCCTATATTCTCACCTGAGGCATCCACATACTTACATTCATCC
CCCAGCAGCTTTTTGCCTTGTCAACAAAAGAACTTTACGGGCAAAAGCGGATTCTGGTGGTG
ATCATATTTTTATTTTTGGTTCTTCGTTTTTGGTTTTTCTCAGCCGTCGAGGAAATTAGTTCTG
GAGTTGCCGGGAATACATACGTGATTTAAGTTGTGGTTGGTGGTATGATCATTGCTGC

>orf19.1727

TTCATATCAGTATAATTCTTGTGGTGTTGTTTATCTTTAATTCCTTTCAAAAAAATTTTTAGTT
CTCAAATTGGATAAAATTTTTAAAACTTGCATTAAATAAATACTTACGTATATTAATCATCCA
TCTATCAAAATACTTGCATATATATATATATATATAACCAACAACTAATTTCCCACCAGGATA
AATTTTTTTTTTTTTTTATTTTCAGAAAGAAAACCCTTTCTTTCAAATTATCAACACTTACTTTAT
ACTGTTTTATTATAATTATCATTTATTAATTAATTATTATTTAATTTAATCTACTATTTATCTAG
AAGAGAGAGTTATTATTATTTCTTTCAACACATTTTCCTTTCCAAAGCAAAGCAAAATTTTGA
GTTCATATATACAAATTAAACATTATTACTTATCCTCATATTTATATTTATATTCCGTACTATG
GCTCCAAGAACACCAGCCGATAAGCTTGGTCATGCATCTACATTACAACTCCC

>snR74

AGGAATTAACAGTCCCTAAACAACAGTAGACCGTCGTGAAAAAATGTGGTATAATCATCATA
AACATAGAAAAAACCACTTAATTATGTCTCCTTAAAACCAAACATATATATATATATATATAT
ACGAAAAACAACTAGAATCGGTTTTTATTGGAAATCTTCAATATAAATCATCTTTTTCTATTT
TTTTTTTTAAAAAAAAAACAATTTAAAAAATCAGACAATATACTTTATATATCTTCATTAATT
CAGACATATGCTTGTCAAAAAAAATTGTGAGGATAATTTCATCATATTTTAAAAATTTTGTTT
TTTTTTTATTTCATTTCATTTCATTTCAAATAATGACTATTGAAGACCCACTAAAATCCTGATT
TATAAATCTATTGTCTTCCCCCTTTATTAATTTAGTTAGTATAACAATTTCAATTTGGAAAAAA
ATAAAACAACAATAAAATTGACTTAGATCAGAAATGCATGAATGGTAATTAAATAAAA

>orf19.4889

ATCGTGTGACTATAAAAATATAAAGGATGCCAAATAGTAGTAGTAAAATAGTAAATAGAATT
GCAAAACAAAAATGATTAAATAGCCCTTTATCAATAAATTTTTAAAGTTAGTTTCTTCTGGAA
CCCTACCCTCTTGGTGTTAATCTTTTAAGTTAATATTTATAGTTAATAAAGTAGAAGTGTCTAT
TTATTGATTGTTGTTGTTGTTGATTAAGAATATAAAGAAAAACAGAAAAGAAAAAAAGAAGG
TTTAAAAAAGTTAATTGTGAAGTAAAAGGGTTGAAAAATTTTTTTTTTTTCTGTTTCTCTCTTT
GAGATTCTTTGACATATTTATTATTATAACACTATGCTATACTAAAAACAGTACTACCAATTG
AATTAAATTAAATTAAATTAAATTAAATTATTAGACCAATTTCAATAAAGATAAGCAATTTA
AGAATAGATGGATAGAT
AAATAAACAGTACTCCACTCTTTTCTCTTCAAAAAAGATAAAA

>orf19.3699

ATTCATAATAAACAATTACGTGAATCATATATAAATTGGTTAAATATTATGAAATTTATAAAT
TTAAAAATCATTCGAATAGGTGGATTGAAATTATCTAATATTGATTATTTAAATATATCAATT
CTGACATATATTACAAAAATAGAAACTGACACTAGTGGTAATAGTAATAATAGATCAAATAC
AATTGTTAAAGAATTAATACAATTAACAAATGAGAATTGTGATATTGATAACCATGATGATG
GCCACGATGATTATTGGATAACTTATAATATTCATGATTTGAAGATTTTCCAATTGGAAGATA
TAATGATTGGAATTCAATCATGGAGTTATATTTGGTTTAATATTTTCTTTGAATCATTTAAATC

AAATGGTCATGGTCATGGTCATCACAATAATAATAAGGTTTTATTCAATTGGGATGATGATAT
AGATGGATTTAAAGGATTTAAACAAAAGGGAATAAAATTATTTAATGAAATAGAATTACA

>orf19.4656

ATTTACAAAAATTATTATTGATAATGGGAAATATTATCGAGAAATATTGATTTTATATTTAAA
ATTTATTAAAAAATTGATTAAATTGACGAAACACAAAGATTCACCACCATATAAAGTGCTTG
AATATATATTTCAAAATCTTGATAAAGATCATAATAGCAACGATATTTTGGATATTAAATATT
TACTTTTACAGCAGGGGAGAAGATATACTAAAGAAATGTTTATGGAATTATTGAAAAGTTTA
GTACATGGAGTCAATACTACTAATGTTCTTGAACAATTGAAAACTATTGCTATAGTTAATCAA
CAAATATTAAATGGAGAGTTAATAGTTAATAATAATGATGTCGTTGATGAATTATTAGTGAA
AAATTATTTTCATCCGTTAGAAAATTTATTAAATCAATTCATATCATTATTAGAAGATTCAAA
ACTACAAGATAATAATAATAATAATCATAATAATCCTAATTGGGAAATCAAAATTTTCCAAA

>orf19.4195.1

AAGATCATTGAGCTATATTTATCTTTGATTTTTGGGTAAACCTCTTGCTAAAAGAAATTGCCA
TGATCTTCCATATTTAATGTGTATTATCTTTTCCATACACCTCACAGAACTTCCAGTAATACAA
TAAGCAAACGAAAGAGTAAAATAATAGACTCACATATTTATTTATCACTCCAACGACCATAC
CACAATTAACTTAAATATGTCCACTCAAACGTTTATAAGTTCACGGCACTTTAATTCCGCTTG
CGTTATCACGGCATTTTCCGTCTATTCAATTCCGTCTTGTGCACTTACGGGCCAAAATTTTTTT
TAAGGTGATGCTTTTAGCAACTTAGTTAACATCGCAAAAAAGCTTGTAAAATTCTCTTTCAAT
TCTTCTTGCTTTCAACAGTCACTTCAAAGTACCAATCTTTACTTTGTTTATTGCATTTATTACG
ATTTTAACATGACGTTTGACGACAAAAAAGGTTTACAAATCGCTCTTGATCAAGCCA

>orf19.3799

AACATTGAGAAACTTGTGAGTACCTTAGATATTGATTGAGCAATAAAAGAGTATCTATGTAA
TATATATTACAATTGCTCTTGTAGAAGAAATTGATTGAAAATTGCTAAATCTGAGTTATACTT
TTTTTTTGTTTTTATTTTTGTACATTAAACTTAGCAAAGATAAAAGCAAGAATTCCCTATTCAA
GGCAATTTAAAAAAAAGAACAAAAGAACAAACCCAAAGCAAATATTTAAAACTAATAAAGT
GTATTTGATAAAAGGTAGGTGATCAATAATAGATCAATTCAATTGAAAACAGAATAAATATA
TAGAAATGTAAAGAATTAAAATTAAATAGTACCAAGACTCAGAGGTTATAACCCTTTAAAAT
GCATTTATTATCGTCTCACTATTTAGTTTATAAAGATTGATTGAAAAAAATTTTTTTTTTTCTG
ATTTGAAATTTTTCCGTCAAATTTAGCACGACTAAAAAAAGAAGCATTTATTATTATTCAAC

>orf19.3469

GTAGTAGTGAGAGAGAGAGTGTGTGTGTGACACAAAATACAACTGACGCGATCAATTTAA
CTTAATTAAGAAAAAGAAAAGTAAAAAGTAAAAAGGAGGAGGAGGAACAAAAGAATTTTTA
ATTAATTAAGTTAATTAAGTTAATTAGTGTAGTAGTAGTGTAATAGTTGTTGAAGCAAGTCAA
AAAAAAAAGAAGAAATTAACAACGATCTTCTCAACTTCATCTTCCCACTCTTTTCTTACTCCT

TTCAAATTGTTGAAATTGCTTATATCTGTTTCTTTTTTTGAATTGCTATCACTTTACTATATATA
TATATATATATGTATATATAATATCTATCAGTTTTTTATATTACTTTCCAAAAAAGAAATAGTT
TTTTTTTTTTACAGGACATCTTTCAATTCACTACAGATAATTTTCACTCCCTAATCATCAACCA
GTATCCAATATGTCAGCAACGATAATTGATAATGGACAACCAAAGGCAAAGACTCCAC

>orf19.5701

AATCTCATGATTATACAAATTCAACTCATTATCGAGATTTACAAAATTCTTATATTATGGATA
CAATTAATTTCTTATGGAAAGAACGATTTTTAAAACAAGATGCAAATTTAACTAATATTAGTT
CAAATAAAGGATTTTTTTTACATCCTGATTTTGAAAATAAATTAAATTCATTACATACTTTTG
ATTGTTCTTATTATTTATCTTCTTCACCTTCAACTATTGGAGAATTTTTTTATAATCCAGCTTTT
TCATATATCGTGACAAAAATCATTTGGGATATTGAAGATAATCATAATGGTATTAATACTAG
ACATGAAGGTCCTATATCAAGAGAAAGTTTAATACGATTAAGAAATGATAGTGATAAAATTT
GGTTATCATTATCATTTGATGAATTGAAATTGTCAATACTTCAAAATTTGGAAATTAAAAAAT
TTAATGGTATTGCTGATTTATTATTTAATTCTTTGAAAAGTTTAGCTGATAAAAGAAGT

>orf19.1422:orf19.1421

ATCGGAATGTGAATTTTGATCAATTCTCTTGTCAGGTTCTAGATTTGGAAAATCTCCCATTTCT
TTAAAAAGTTTTGCCTTAATAAATATGAAAAATGATCAGATATACTGTTGAGCTTTAACACA
GTTAGTTTGAAAAAATGATTATACTGCTATTAAGTGTAACGCTGTTTGAATACGTCTTTCGAT
AGTGGCACTGGATTTATAATTTTTTTTTTTCTCTTACTAACTCTTCCTCAGTTTATATGACAAA
TAGTGTATCCGCAAATTGGGTGATTAGTTTCTTTTTTTTTTCCCTTTGTTTTGTTTTGGTTTTTT
TTTCCCCGAATTTTCACGTGATCAGTAAATAATGCGTGTACGTATTCCAGGTGATTACGTAAG
ATTTACCTTCGGTTATCTCCATTCTTATCACTAGAATTTTTCAGCATAAAAATAGACAACAAG
TAAGTACTTTTTATTCTTCATTAAACAATGGTATTGAAAACATTCAAAGTTGATGG

>orf19.193

TCAAAATATTCATCACAAACAATTCAAAAGGCGCAATCTAAATTTGAAAAGAATTTTGATAT
TAAATTCAATTGTAATTGAAGGTTAACTATATATATATAGATATATATACATACATACATACA
TAATTAGTCTTGTTCTTTTATCTTAATTTATAAATTATTTAATTGATTGAGATTTTCAATTAAT
GAATCAATTTCCGGAAAATCATTAATGAAGAATTGTGATTTTTCTTGAATTGATTTACCTCCT
CCACTATTATCAATTTCTTCTGATAAATAAAGCATTAAAAATTCAATAATTTTGAAATTCGAA
TTATTATTTAATTCTTCATCTTCATTATTGTTATTATTACTGTTAGTGTTAGTGTTGGTGTTGTT
TTGATCATTAATATTTTCAAATGAACTCAATTTAAAATGTTTAATTAATATACTACAACCATT
TAAATCTTCAAATATTCGATAATTATCATAATTCTTCAATAAAATATGAATTAAAGT

>orf19.2827

AAATTAAAGATCAACAATCAATTATTGAAGAATTAAAAATTCAAGATTCTTTAATGAAACGG
AAACTGGATGATTTAAAACAAACAATTTTTCAAATACAAAATGAATTTAATAGTCAAATTGA

ATCATTAGATCAACAAATTAGAGAAACTAAATCAAATATTCAATCGGATATTTTACAATTGG
AAAGAAAATTACGTGATGTCAATTTGAATAAAAGTTTGATTCAAGAACAATTGATTGGTAAA
CGACAAAAAATTGATGAAGATGTCATTAATTTAACGTCATTTATATTGAAATTTAAAACTAAT
ATTCAAGAAAAATTGATTAATATTGCTGATATTGTTCAAGATCAATTGAAAAAAGAACAACA
ACAACAACAACAACAACAAGCATCACAACAGCTAGAGTGAAAAAAGAGTTTCATTAATGTT
CAGTTAAATCAAATAAATTCTTTAACTTAATCTTATTATCAATAATTCGTTTTCTAAACATGC
ACT

>orf19.1707

TCAATAGTCAATTTTCTGACAACTCTTTTCTTATCATCTAAAGCAGGTAATAAATCATTTAAT
AATTTTTGTTTATTAATAATATCATCAAATTTTTCAATTATTAATTTCAAATTCTTTAAACTTA
ATACTTTAATATCAATATTCGATTTTGATTTCCCTTTTGTTGTTGCGGATGTAATTAATTCAAT
TAAAATTGGGATTAATTCAGTAATATTCTTCTTAGTCAATAATGATGAATTTGTATTGATGGT
GGCGATGGTATCTTTTTCTTGATCAGTTTCCATCTCTTTCTATAATTCTCAATCTGTATCTATA
TCGTTTTCTTCTTCTTTTAAAATAATATTCAATATTATCAATGATGCCAATAAATTTGAATTTG
AATTAGCATCTATTGGAATTTTAATTAATGAATATAATGTAATTGGTAGAATTTCTGATAATT
GATTAATTAAAATATTATTATTACCATTAATTGATAAATTTTCAATAATCAATGA

>orf19.3633

TATGAAACAATTGAACAATGACACCAGAAGCTGGTGTTCCTAATGTGGTATCTAATATATGA
CAAGTAATTGGATCAACAGACATTATGGAAAAAAAAAAAGGATATGTATGTTTTCAAGGTGC
TATTATGTTAACTGGTGAATATATGGTTTGCCTTGTAGTTGGCTATAAAAATGATAATACTGA
AAGAAAGAATTGAAAGAAGTATTGCTGATATTTTTTTCTATATATATGGTGGATTCAATTTAG
TTTTATATAGTTTGATAAGATATTTCTCTCCCACCGCAGATAACCGAAGATCAAAATTGCTCT
CTCCGCAAAATCACCAATTACAAATAAACCCACACAAATGGTAATAAATATACTATATAAAT
AAATAAATAAATAAGGTAAACATAAAGGTTCTATAATAATTATCCTATATATTTCCATGATAT
GTTTTTTGTATCATATTTCCATAAACGATCTTGTTTTAAAACCATTGTCAAAATGTAATCAA

>orf19.4231

TTGACTTTTAATCGCTTCTAACCATATCAACACGTTTAAAACTTTTTTCTTTCTTTCTTTTTTTC
TTTCTTTTTTTTTTTTTTAGTTTTTTAAATTCACTGGAGAATTATATAAATATCAAAGCAAATAA
CCTCTATTTCCAATTCTAGTTTCAACCAACCTTTTCCTTTTTCAATCTTTGAATATCATTATTGG
GATCTTGAATTTTTCAAATATTGCTTTTATTTTTTCATTGCATTTTTAACGATTTCACTTACTTT
TCTTTGTTTTTGTCACTTATTTCTTCTTAACTATTCAACAATAACAACAACAACAGTTCTTGAT
TCAAATTTATTCACTATATTTAAATACGATTTCACTATTCCTCTACTGTTGCTTGCTATTACGT
TATTAACATAATCTATTTATTTTTTCATTTGCTACGATCTACCACCAAAAAAAAAAATGTTAA
TCATCTCAGAATTCAATCCTCAATTAATGCATTCGGACGATTCCGAAGAC

>orf19.35

TCTGCAAATTATGTTTAAAAATATTGATCAATTGAAAATTTGGAAATTAAAAGATGTTTTAAT
TGAAAAATATAAATTTAATAAAAATGATTATGAAGTTGATTTAATTTGTGATTTGATTTTAAA
ATGTTTAAAATATGATTTAAATGAAAGATACGATGCTAATTCTTTATTAAAACATCCTTGGTT
TGATAGTGATAACAATAGCACTATTAGTCCTGGATTGGAAGATAGATTGAAAAATTTACCAA
ATTGTCATGATGATTTACCGGGTTATACATGTGAATGTAAAATATAAAGAAACTACAAATTA
ATTAACAATTAAGATTATATAAATATATTACATACATACATATATATATGTTTAGTTGTTT
TTTCTGGTTTACATATAGCACATATGAAAGAGATGTTTTGATTTATTAACAATAATGGGAATA
ATAATGGGGTATTAATGCTCTCTACACATTAAAAAATACTATCAGTACCAATTCAATCCAC

>orf19.4152

TAGTGCTCACCAAAGACAAAAAAAAAAAAGATTAAAGTGATGAGAACAAGAGAATCACAAC
TGAAAAAATTTTTTTTTTTTTTTTTGCTTCCTTTGCCTTTCTCTCTCTCTCTCTCTGGCCACTCT
TTCTATATCTTAATTAAACAGAATAGAACATAAAAAAAAAGAAACTGCTATTAATATAATAA
CACACACAATATACACTAAGAATTTATATGTTCCTGTATATTAGTTGCCTATCTTAACACATT
ACCAGAGATTAACAATTACCCTGCTAGATTGTAAGTGTTTCAAAGATATGATTTGATATACA
ACAACATTATCAACAACAACAACAACCTACATGTTATAGATAGAGAAATAGAAATTAGAAG
AGAGTGAAGAAAAAAGATTACATATTCACTTGGGTTGGGTGATTACTTATCAAGTTTAGTGC
ATGAGCTACTTTTCTTTTTTTTTGACTTCTTTCTTCTGTCGTTGGTTGATGTGTCTTCTCAAG

>orf19.1041

TTGGTTGTAACTTTTTTTTTTTTCCACCATCTGGTTTGTTTGTTTGTTTGTTGATGTTGTGTTGG
ACTGCTAAATTTTTTTTTTTTTTCCTTACTTGTACGAACAGAGGATGGATGGATACTTTTGAGG
ATGGATGGATACTTTTGAGAAAGCATAACTTTGAGTTATTATATCTATTTAGATAATCCGGAT
AAACTCTACAGTATGTATTTGTTTTCTAATTTGATTATAGTAATCGTATACTAAATAATAAAT
ACTTATAAAGACAAATAAATTTAGAATAATTCCATTTTGGTTTTTTTTTTTGTTTTTTTTTTTTG
GGTTCTGGTATAGTTTTATATATATATATATATATATTCATTGACCTGAAATCATTTTGTTTTT
AAGCAGCAAAAGACAAAGAGAAACCTAATTTATGAACTGGTTCAGCCAATTTCAAGGCATC
AAAGGAAGCACCCAAACCTAATTTAACCCCTGGTCTTAATTCTTGAGTGTAAGAC

>orf19.7038

AAACAATTATCAATTAATCAAATTGATTTATTAGATAAACGTATAAAAGATCAAGAAATTAA
ATTTAAAAAAATTAGTGAAGAAAATCCTGATATCAAAGGTGGTGAATTAATTAAATTAAGAC
AATCAATTATAAATGATAAACAAGAAATTTTTCAACAATTAAATAAAGATTGGTTAATTAAA
CAATGTTGTTTACAAGAATTTATTATTTTCCAAGAAACTCAATTTTTAATTACGGAACTTTGG
GTTGAATGGTGTAAAGATCGTTTGAAATGTCAAGAGAAATTAGTTGGTTTATATGATAATTTA
AATCAAGAAATTATTCATGATATGCCTCTAGAGAGATGATCCCCTTGAGAGATGATCCTAGA

AAGATAAAGCTTTTAGTAGTAATTCCTTCAACTTCTATATCTATACATACTATTTTAATTTACT
GTCATTTCATTTCATTTCATTTCATTTCATTCTCCCATTTATTAATTAATTCATGATATGAT

>orf19.6570

CCTTTATCTTTTTAGTAAATAAATTACTGTAACTTTAAACTCGTTGATAATGATAAATTAAAA
CAAATCGTAATAGCTTTGTTCCCAAATAGGAAAACTGATAAACTATTGTTGTTTTTGTTGCTG
CTGAAAGAATGTTCCGATTTTCCAATGGAACGGCAAAACTTCTCACTTTCGGGGAAATGCAT
CGGAAGTTTGTCACCTATTATTAAAAAGCAGGCACGGCATCAAATAGATAAAGAAATTAAAT
ATTATTAGGAACAAGATAAAAGAAGTCAACTTAGTATACATTTAAAATGTAACGGTATTGTA
CTGTTCTTAGTGTAATTGTGATGATTTGTAGGACCATTTTAGTTTGTGAAACTTTTTAGGTCTC
ATCTCAATTTGTTTACTAAATGAACCGTTTGAGTAAACGATACAAAAGTAGGTTTCATTATAT
TCTGTTCTGTTTAATTAATCAATATATGAAAAATTAGACAATTTTTTTGACTTCTATAATT

>orf19.1080

TGCTAATTAGTCTCGCGACAGGTTAAATAATACGTTGAAAAGCACTTATAACTTCTGAATGAT
AAACTCCTTTATTACTATCAGTTGGAGTACAAGTGTTATAGTTTTGGTTGTATATTTTGTGTGG
GTCTAAGACAATTTCTATTTTAACACTATTTTTTCGCAATTATTTGCAGTTAAATATTGTATGA
TAAGTATTTTCATAAAACTGTTAACCACCACAAATATTATCAATCTACTATTTAATCAGCTTC
TGGAAACAACAAGCAAAAACATTGTTAACCTATTTTTTACTTGCCCCCCCCCCCAAATATTAT
AGTATAACATAATATTTAATAAACAAGATGTATAAATGTGTAACCACTTTAATTTTTCTTTAA
TCCTAAACTTTCTTTAAATGAATTGGCTTTTGTGATTCATTAATACTTAACATAACTATTTTGT
TATGAAAAAGGTGTCCAAAAACATTATTGGAAACTGAACAACAACCACCTTAACCT

>orf19.6972

AAGCGGGGGGGATGAATAATAATCTTAAAAAGAAAGAACAAATGAATCAATGAATCAATGA
ATGAATGAATAATTAAACGGGAATTCTTAAGGGGAATTGAAATCAAATGTGAAATTGGTGGT
AATCAGAATTTTGTTTTAGTTGATAATTTTAAAGAAAAGAAAAAAAAAAAAGGGGAGGTAGT
GTTAATTGGTTCAAAACAAATTTAAACTTTATAAATTCTGAAAACGCGTCTGCGTCTGAAATT
ATAACAAAACATTTTTTTTTAAAGTGTAGCATCACGTGACTCGGAATTATAACTTTATTATTA
CTACTTAACTGATACATTTTTATATTCATTACAAATCAAACCATTCGAAAACTATAAACATCT
TCTTAATTGCCCAATCTAAATATGATATTCAAGTAATCTTGAATTATTTTCTTTCAATAACACT
TCATCTTCATTCCAAACAATTTTAATAATATCATAAAATTCTTTAGCATAAAACAAGGAAAT

>orf19.5026

TGCTTTTTTTTTTCTCTTCTTGTTCTCTGTTCTCTTATGTTGAGTGTATGTTAATAAATATTTAA
TGTATGCTTTGTGTGTATGGATGAGTGTTTCTTTCTTTTTTTTTTCCTCTTTATTTTATAGCTTC
TTTAGTTAAGACACTTTTTTCAAAAGATTTCCACCAAAAAGGGTAAATAATAATAGGAATAA
TAACAAAATTATTAGAACTTATGGGACGATGTAATATAAAGTACGGTCGTTGAATCAATTAT

AATAATATGTGTAGATGTGTGAAGATGTGTGAGTTTATGATGATTTATAAAATGAAGTAACT TTTTTATTGAGAAATTTGTTTGTAATATGGGGTAAAGGAAGTCAAATATAGTTGGAAATTAGT AGTTATTCTTCTTCTTCTCGAGTTCCCAAATTAAATATTATAAGGATATTGATTGTCGTTTAGA TATTATACAAAGACACCTTAATGAAATTGAAATCAACCTTCAAAACAAAAGGAAAAG
>orf19.1716
TAATTTTGAAGATTATAATGATGTTCTTGAAGCGTAAAACATTTGACCCCCCTGTCATGATTT CTAGAAGGACCACCTTTGATTGTAAATAGTAATAATTACCACCCTTATTTAATTATTTATTTA ACTTATTTATTTATTTATTATACATATATACAAATCTAATAAAGTGAAAATCTCCCCCTTCAC ATTTATAATTGGCCAGTTTTTTTCAAATAAGCATTCCAACCAGCATCTCTATACCTTTTACCTT CAATATCTGGATCTCTTCCTTTACCAAACAATCCTCTACCAACAATGATAATATCAGTTCCAG TGCTAACAACTTCATCAACAGTTCTATATTGTTGTCCTAATCCATCACCTTTATCATCTAATCC AACTCCAGGTGTCATAATAAGCCAATCAAATCCTTCTTCTTGTCCACCCATATCACGTTGGGC AATAAATCCAATAACAAATTCCTTATCGGATTTAGCAATTTCAACAGTTTTTTGAGA
>orf19.1716(2ndprobe)
TCTTAGTGTTGACTGTCATATCTTTTGTTTGCTATTTCTATTTCTATTTCTTTCTTCAACTGTAA AAGGGGCGATATATAAACCAGTGTGTATGGGGTTGTTGCTCAGGAATTGATCAATGTTTAAA AGAATGGTTGAAATCATCATTGAAGAACAAAAAAAAAGATTTTTTCTTTCATTATGTTATCTG TATAATATGCAGGAAATAAGATTGCAAAAAAATAGCCGTTTTTTTGGATACTATCAAACAAG AGGAATGACTAATACAGAATTAGTGTTACGAATCAATGGCACTACAGCAACTTTCAACAAAA TAATCTACGTTCAAAATATAATATATACTATTAAATAAAATCAAATAATAAAATAATTTTAA AGAAAGTAGTACCAGCTCTTTTTTTTGTTTCCGTTTATACCATCCAAATCAATTCCTATTAGTA CTTTGTAATTACTCTCAGTTCAAACTGTTTAAATAATTCATAAGTTTTTTTTGAAATGGA