Postgenomic studies of *Candida albicans*.

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor in Philosophy, Biology Department

Mikhail Martchenko

Department of Biology

McGill University, Montreal

February, 2006

© Mikhail Martchenko, 2007

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-32310-6
Our file Notre référence
ISBN: 978-0-494-32310-6

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

ACKNOWLEDGMENTS

First and foremost, I want to express my gratitude to Dr. Malcolm S. Whiteway, my supervisor, for his guidance, advice and support in all stages of my studies.

I am also very grateful to each and everyone at Dr. Malcolm Whiteway's laboratory for their support during the past five years.

As well, I want to thank my supervisory committee members, Dr. Howard Bussey and Dr. Richard Roy for their encouragements, suggestions, and positive criticism; the Biology Department, especially Ms. Susan Bocti, for her warm welcoming and help.

I would also like to thank Dr. Aaron Mitchell for the *rim101* and *mds3* mutants and their wild type parental strains, and for the for the *C. albicans* mutant libraries, Anne Marcil for assistance with the end point dilution survival assay, Dr. Ursula Oberhozer for donating the *ACT1* primers, Dr. Catherine Bachewich for the CB20R primer, and Daniel Dignard for bioinformatics assistance. We thank Dr. Alistair Brown and Clair Russel for CRC103 and CRC106 strains, and for the CIp-*lexA* and CIp-*lexA-GCN4* DNA constructs.

I gratefully acknowledge FRSQ-FCAR-Sante for their M.Sc. and Ph.D. Scholarships and NRC for the Graduate Student Scholarship Supplement.

To my wife

COTRIBUTIONS of AUTHORS

1. Assembly of the *Candida albicans* Genome into 16 supercontigs Aligned into 8 Chromosomes.
Mikhail Martchenko – closed the GAPs by PCRs; identified *TLO* genes as telomeric (second authorship).

2. A Human – Curated Annotation of the *Candida albicans* Genome.
Mikhail Martchenko – annotated 1200 *C. albicans* genes (fourth authorship).

3. Superoxide Dismutases in *Candida albicans*: Transcriptional regulation and Functional Characterization of the Hyphal Induced *SOD5* Gene.
Mikhail Martchenko – the principal investigator: performed Northern blot analysis, *SOD5* deletion, Superoxide dismutase activity staining, characterization of *sod5* strain and sensitivity to hydrogen peroxide, macrophage and virulence assays, and fungal burden;
Anne-Marie Alarco – assisted with the animal work;
Doreen Harcus – assisted with Southern blot;
Malcolm Whiteway – project supervisor.

4. Transcriptional Activation Domains of the *Candida albicans* Gcn4p and Gal4p homologs.
Mikhail Martchenko – the principal investigator: performed the genome-wide comparison of transcription factors *in silico*, strains construction and β-galactosidase assays to find activation domains of Gal4p and Gcn4p;
Anastasia Levitin – assisted with cloning;
Malcolm Whiteway – project supervisor.

5. Transcription rewiring within the galactose metabolism circuitry of the *Candida albicans*.
Mikhail Martchenko – the principal investigator: performed microarray analysis, *GAL4* deletion, promoters analysis and β-galactosidase assays;
Anastasia Levitin – assisted with microarray experiments;
Wiriya Chiranand - performed the magnetic bead pull-down assay;
Herve Hogues and **Andre Nantel** – assisted with the bioinformatic analysis;
Michael Gustin – the collaborator and the supervisor of Wiriya Chiranand;
Malcolm Whiteway – project supervisor.

ABSTRACT

We assembled the genome of the human fungal pathogen *Candida albicans* into eight chromosomes, and annotated each of its genes. A genome comparison with *Saccharomyces cerevisiae* revealed an increased number of *C. albicans* superoxide dismutase genes. We analyzed the expression patterns and the function of one of these genes, *SOD5*, whose role is to protect the pathogen against extracellularly produced, neutrophil-generated superoxide radicals. Comparative genomics also showed that although many of the *C. albicans* transcription factors, such as Gal4p and Gcn4p, have homologues in *S. cerevisiae*, the sequence similarities occur only in the DNA binding motifs of those proteins. Deletion analysis of CaGcn4 and CaGal4 proteins show that the N' and C' termini respectively are needed for their transactivation ability. These two transactivation regions show no sequence similarity to the equivalent domains in their *S. cerevisiae* homologues, and the two *C. albicans* transactivating domains themselves show little similarity. A comparative analysis of the transcriptional machinery between *C. albicans* and *S. cerevisiae* showed low sequence similarity of the mediator complex that bridges activation domains of transcription factors to the RNA polymerase II complex. We performed a comparison of intergenic DNA regions to identify the *cis*-regulatory elements from *Candida* and *Saccharomyces* species to examine the organization of the transcriptional regulatory networks between these two organisms. We observed that the *C. albicans GAL* genes lack Gal4p binding sites, but that such sites are found upstream of telomeric genes and genes involved in glycolysis, and we show that CaGal4p regulates the expression of those genes. We identified the regulatory DNA sequences in the promoters of *GAL* genes, including a *GAL* – specific palindrome necessary for *GAL10*

expression. Cph1p, the *C. albicans* homolog of the Ste12p transcription factor controlling pheromone-induced gene expression in yeast, acts through this *GAL*-specific palindrome, functioning as an activator in the presence of galactose. This shows *C. albicans* and *S. cerevisiae* can regulate the same process by different regulatory circuits.

RÉSUMÉ

Nous avons assemblé le génome du pathogène fongique humain *Candida albicans* en huit chromosomes, et avons annoté chacun de ses gènes. Une comparaison de génome avec *Saccharomyces cerevisiae* a indiqué un nombre accru de gènes superoxydes dismutases de *C. albicans*. Nous avons analysé les modèles d'expression et la fonction d'un de ces gènes, *SOD5*, dont le rôle est de protéger le pathogène contre des radicaux superoxydes produits extracellulairement par les neutrophiles. La génomique comparative a également démontré que bien que plusieurs des facteurs de transcription de *C. albicans*, tels que Gal4p et Gcn4p, aient des homologues dans le *S. cerevisiae*, les similitudes de séquence se trouvent seulement dans les motifs de liaison de ces protéines à l'ADN. L'analyse de suppression des protéines CaGcn4 et CaGal4 démontre que les termini N' et C' sont respectivement nécessaires pour leur capacité de transactivation. Ces deux régions de transactivation ne montrent aucune similitude de séquence avec les domaines équivalents dans leurs homologues de *S. cerevisiae*, et les deux domaines transactivants *C. albicans* eux-mêmes montrent peu de similitude. Une analyse comparative de la machinerie de transcription entre *C. albicans* et *S. cerevisiae* a montré une faible similitude de séquence du complexe médiateur qui relie les domaines d'activation des facteurs de transcription au complexe de la polymérase II de l'ARN. Nous avons effectué une comparaison des régions intergèniques de l'ADN pour identifier les éléments de la *cis*-régulation des espèces *Candida* et *Saccharomyces* afin d'examiner l'organisation des réseaux de régulation de transcription entre ces deux organismes. Nous avons observé que les gènes *GAL* de *C. albicans* manquent d'emplacements de liaison Gal4p, mais que de tels emplacements sont trouvés dans les promoteurs des gènes

téloměriques et des gènes impliqués dans la glycolyse, et nous avons montré que CaGal4p règle l'expression de ces gènes. Nous avons identifié les séquences de l'ADN de régulation dans les promoteurs des gènes *GAL*, incluant un palindrome spécifique pour le *GAL* nécessaire pour l'expression du *GAL10*. Cph1p de *C. albicans,* l'homologue du facteur de transcription Ste12p contrôlant l'expression du gène en levure induit par la phéromone, agit à travers ce palindrome spécifique à *GAL* et fonctionne comme activateur en présence du galactose. Ceci montre que *C. albicans* et *S. cerevisiae* peuvent régler le même processus par différents circuits de régulation.

CONTENTS

CHAPTER 1: INTRODUCTION
1.1 History, Rationale, and Objectives: 1
1.2 Introduction to yeasts and their genomes 2
1.3 *Candida albicans*
1.3.1 A human fungal pathogen 6
1.3.2 Yeast to hyphal transition of *C. albicans* 7
1.3.3 *C. albicans* genome 9
1.4 Introduction to transcription
1.4.1 General overview of eukaryotic initiation of transcription 10
1.4.2 A model case: the regulation of *GAL* genes 14
1.4.3 Another model case: Gcn4p 21
1.4.4 Types of activation domains 22

CHAPTER 2. THE COMPLETION OF *C. ALBICANS* GENOME 24
Section 2.1:
Assembly of the *Candida albicans* genome into 8 chromosomes 25
2.1.1 Abstract 25
2.1.2 Introduction 26
2.1.3 Materials and Methods 29
2.1.4 Results 33
2.1.5 Discussion 39

Section 2.2:
The annotation of the *Candida albicans* genome 46
2.2.1 Abstract 48
2.2.2 Introduction 49
2.2.3 Materials and Methods 51
2.2.4 Results and Discussion 53

CHAPTER 3. POSTGENOMIC STUDIES OF *CANDIDA ALBICANS*: INTRAGENIC COMPARATIVE GENOMIC STUDIES 65
Section 3.1:
Superoxide dismutases in *C. albicans* 67
3.1.1 Abstract 67
3.1.2 Introduction 68
3.1.3 Materials and Methods 71
3.1.4 Results 77
3.1.5 Discussion 89

Section 3.2:
Transcription factors in *C. albicans* 106
3.2.1 Abstract 108
3.2.2 Introduction 109
3.2.3 Materials and Methods 113

3.2.4 Results......115
3.2.5 Discussion......121

CHAPTER 4. POSTGENOMIC STUDIES OF *CANDIDA ALBICANS*: INTERGENIC COMPARATIVE GENOMICS STUDIES
Section 4.1:
Transcriptional rewiring in *C. albicans*......135
4.1.1 Abstract......139
4.1.2 Introduction......140
4.1.3 Materials and Methods......143
4.1.4 Results......144
4.1.5 Discussion......151

CHAPTER 5. FINAL CONCLUSION AND SUMMARY......166

APPENDIX A: Tables......169

APPENDIX B: Animal research and radioactive certificates......173

BIBLIOGRAPHY......176

LIST OF TABLES

Table 1.4.1. Transcriptional machinery: *S. cerevisiae* vs *C. albicans*...........................12

Table 2.1.1. Chromosome size and features...43

Table 3.2.1. Genome-wide comparison of transcriptional activators between *C. albicans* and *S. cerevisiae*...124

Table 4.1.1. Genes whose expression was up-regulated in response to galactose.......154

Table 4.1.2. Genes whose expression was up-regulated in response dextrose............155

Table 4.1.3. Genes whose expression was down-regulated in *gal4* strain...................156

Table 4.1.4. Genes whose expression was up-regulated in *gal4* strain........................158

Table 4.1.5. β-galactosidase activities in Miller units of *C. albicans* promoters in *C. albicans* strains...159

Table 1: Strains used in this study...169

Table 2: Plasmids used in this study..170

Table 3: Oligonicleotides used in this study..171

LIST OF FIGURES

Figure 2.1.1. Schematic representation of the chromosomes in Assembly 20................44

Figure 2.1.2. Highly related ORFs in the *C. albicans* genome......................................45

Figure 2.2.1. Visualization of proteins sequence similarities...64

Figure 3.1.1. Similarity of the CuZn-containing and Mn-containing superoxide dismutases in *C. albicans*..94

Figure 3.1.2. Hyphal-inducing conditions induce *SOD5* expression.............................96

Figure 3.1.3. Alkaline conditions, Osmotic and Oxidative stresses affect SOD5 transcription..97

Figure 3.1.4. *SOD5* transcription is affected by non-fermentable carbon source...........99

Figure 3.1.5. Transcription factors involved in the transcription of *SOD5*...................100

Figure 3.1.6. *SOD5* gene is deleted from UH16 strain..101

Figure 3.1.7. Hydrogen peroxide and limited nutrients inhibit growth of UH16-2 strain..104

Figure 3.1.8. Sod5p activity is required for virulence in animal model........................105

Figure 3.2.1. Genome-wide comparison of transactivators between *C. albicans* and *S. cerevisiae*..128

Figure 3.2.2. Gal4p and Gcn4p act as transcriptional activators in *C. albicans*...........129

Figure 3.2.3. Identification of CaGcn4p Transcriptional Activation Domain..............130

Figure 3.2.4. Identification of CaGal4p Transcriptional Activation Domain...............131

Figure 3.2.5. The phylogenetic comparison of Gcn4p and Gal4p orthologs across the Ascomycota..132

Figure 3.2.6. The comparison of the transcriptional machinery between *C. albicans* and *S. cerevisiae*....134

Figure 4.1.1. *Candida albicans GAL* regulon....160

Figure 4.1.2. Galactose leads to elevated expression of *CaGAL* genes....161

Figure 4.1.3. The function of *C. albicans* Gal4p....162

Figure 4.1.4. Gal4 regulates expression through Gal4 DNA binding site....163

Figure 4.1.5. Cph1p regulates the expression of *GAL10*....164

Figure 4.1.6. Analysis of *cis-GAL* regulation across Ascomycota....165

LIST OF ABBREVIATIONS

a. a.	Amino acid, amino acids
AD	Activation domain
ALS	Agglutinin-like substances
b. p.	Base pairs
CGD	Candida genome database
CuZn	Copper-Zinc
DBD	DNA-binding domain
GAL	Galactose metabolizing genes
H_2O_2	Hydrogen peroxide
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
HOK	MRS subrepeat
k. b.	Kilobase pairs
LTR	Long terminal repeat
M. b.	Megabase pairs
Mn	Manganeese
mRNA	Messenger ribonucleic acid
MRS	Major repeat sequence
N_{11}	Eleven nucleotides
O_2-	Superoxide radical
OD_{600}	Optical density at 600 nanometers
OH-	Hydroxyl radical

ORF	Open reading frame
PCR	Polymerase chain reaction
RB2	MRS subrepeat
rDNA	Ribosomal deoxyribonucleic acid
ROS	Reactive oxygen species
RPS	MRS subrepeat
SAP	Secreted aspartyl proteases
SOD	Superoxide dismutase
TLO	TeLOmere associated genes
UAU	Ura3-Arg4-Ura3 knock out cassette

CHAPTER 1: INTRODUCTION

1.1 History, Rationale, and Objectives:

An important diploid fungal pathogen, *Candida albicans*, which diverged from the baker's yeast *S. cerevisiae* approximately 250 million years ago, causes the majority of human fungal infections (116). Genetic manipulations that are carried out easily in *S. cerevisiae* are more difficult in *C. albicans* due to the fact that *C. albicans* is diploid, lacks a useful sexual cycle, and does not have the abundance of molecular tools that are available for studies in yeast. Currently, reverse genetic approaches are commonly used for functional studies in *C. albicans*: genes are first identified by their sequence and then both genomic copies are sequentially mutated. Generally this approach is made efficient through a previous knowledge of the biological process of interest, emphasizing a usefulness of a model organism like *S. cerevisiae.* In addition, the unconventional *C. albicans* codon usage poses another significant challenge: *C. albicans* translates the universal leucine codon CUG as serine (263). Therefore, many heterologous markers do not fully function in *C. albicans* unless they are modified by codon optimization. Due to those difficulties in *C. albicans* genetics, genomic approaches are very useful in the study of this pathogen.

The genome of the human fungal pathogen *C. albicans* had been sequenced by the Stanford Genome Technology Center and assembled into 412 contigs (148). We completed the assembly of the genome of this fungus into eight linear DNA sequences representing the 8 chromosomes (Van het Hoog et al, in revision). In addition we established a detailed annotation of the genes (33). The assembly and annotation allowed us to perform comparative genomics of inter- and intragenic regions of *C. albicans* with

those of other pathogenic and nonpathogenic fungi to better understand the biology of this organism. In particular, the comparative genomics of non-coding DNA sequences of two organisms with similar gene contents may explain species-specific characteristics that distinguish, for example, *C. albicans* from *S. cerevisiae* and humans from chimpanzees. Our studies show that a genome sequence assembled into chromosomes, coupled with the annotation of its genes:

1) helps to direct understanding of the biology of the organism;
2) helps to underline the differences in the human fungal pathogen *C. albicans* from the model yeast such as its virulence factors (*SOD5*) and the differences in the structure of its transcriptional machinery (Gal4p, Gcn4p, and mediator complex);
3) helps to define the role of transcriptional regulators and show that *C. albicans* and *S. cerevisiae* can regulate the same process (galactolysis) by different regulatory circuits.

1.2 Introduction to yeasts and their genomes.

The increasing number of fungal genomes whose sequences have been completed allows their comparison at the nucleotide and protein levels and improves our knowledge of evolutionary relationships among fungi. Yeasts include important industrial organisms that have been used for many centuries, pathogens, and popular laboratory organisms that serve as general models to understand the eukaryotic cell. For decades, *Saccharomyces cerevisiae*, baker's yeast, has been one of the best-characterized organisms from the genetics and physiological point of view. Sequencing of yeast genomes tells us directly

about the coding potential and regulatory sequences that are necessary for yeasts as well as eukaryotic cells, but also about genes that are specific for a certain species. However, yeasts could also be used as a general model in comparative genomics because of the small genome size(s) and the well-defined evolutionary relationship among different lineages.

Overall, the information on the full sequence of more than 20 fungal genomes is accessible. They include members of the Archiascomycetae, Hemiascomycetae and Basidiomycetae classes, and therefore they cover an evolutionary range that could extend more than 1 billion years (117). This amount of information allows the comparison of fungal genomes covering short phylogenetic distances, for instance, within Hemiascomycetes, or larger distances that include the other two classes. Comparative genomics has emerged as a new discipline; in the case of fungi it allows a better understanding of species evolution and helps to explain the different life styles that can be found among these organisms, and it introduces new information useful from a medical and technological point of view. Comparison of the genomes of two species such as *S. cerevisiae* and *Cryptococcus neoformans* demonstrates that they share at least 65% of their genetic information, which is noteworthy for two species that diverged in the evolutionary tree about 1 billion years ago (117, 191). Obviously, some genetic traits such as those coding for capsule synthesis and other possible virulence factors, such as melanin formation, are specific to *C. neoformans.*

Analysis of the sequence of the 16 chromosomes of the *S. cerevisiae* genome revealed the existence of numerous pairs of homologous regions. On this basis it was postulated that whole-genome duplication had occurred in an ancestor of *S. cerevisiae*

(307). This duplication event was followed by extensive gene loss, which in general would have affected only one of the two members of the generated gene pairs, and was followed subsequently, among other events, by gene inversions and transpositions affecting individual genes. As a result, *S. cerevisiae* individual genes that lack homologues coexist with families formed by two paralogous genes located in sister chromosomal regions (large blocks of homologous genes) on two different chromosomes (e.g. *RCR1* and *RCR2*) (308). A consequence of gene duplication is the possibility for functional redundancy, with its implications on phenotypic stability. Enzymes involved in the defense against oxidative stress in *S. cerevisiae* offer a number of examples of this situation (276, 289). Thus, *S. cerevisiae* contains two cytosolic thioredoxins Trx1, Trx2, and one mitochondrial Trx3; and all three proteins show significant similarity in their amino acid sequences (236, 294, 295). *S. cerevisiae* also possesses two dithiol glutaredoxins: Grx1 is exclusively cytosolic while its homologue Grx2 shares a cytosolic and a mitochondrial location (196, 198). Interestingly, *S. cerevisiae* has a peroxisomal omega class glutathione transferase Gto1 that is induced under oxidative stress conditions, and two homologues, Gto2 and Gto3, located in the cytosol (11, 95). Among the fungal genomes sequenced up to now, only the *S. cerevisiae*-closely related species *Saccharomyces paradoxus* has a predicted peroxisomal Gto1 orthologue; other fungal species contain a single Gto homologue, probably located in the cytosol. In contrast to other fungi therefore, the genome duplication in the *Saccharomyces* evolutionary line has apparently led to a new enzyme activity in an organelle such as the peroxisome as a distinctive trait of *S. cerevisiae* and close relatives. A parallel loss of peroxisomal glutathione transferases in different evolutionary lines from a fungal ancestor that

contained such enzymes could also have led to the present situation, but this seems a less plausible hypothesis, as it would require a larger number of independent genetic changes. The presence of a glutathione transferase protecting against reactive oxygen species generated at the peroxisome could explain the acquisition of some metabolic traits, such as the participation in lysine metabolism by *S. cerevisiae* peroxisomes compared to other fungal species (34).

The above considerations concern evolution within fungi over a long-time scale, that is, a period of about 1 billion years (117). How is evolution operating over a shorter time scale, for instance within the genus *Saccharomyces*? Two recent studies (62, 153) address this question within the *Saccharomyces sensu stricto* group that has accumulated an estimated 5-20 million years of separate evolution (63, 161). The partial sequencing of the genomes of the *S. paradoxus, S. mikatae, S. kudriavzevii* and *S. bayanus* species and the comparison with the *S. cerevisiae* genome sequence (62, 153) permitted the observation that synteny, namely gene position and orientation relative to neighbor genes, is conserved for most of the observed *Saccharomyces* genomes. Nucleotide changes are about twice as frequent in intergenic regions than inside genes, as expected from the fact that these intergenic regions have fewer restrictions for maintaining sequences than coding regions where many nucleotide changes would lead to non-functional amino acid substitutions. Even in intergenic regions, nucleotide changes have not occurred homogeneously. This has allowed the characterization of conserved intergenic sequences that may correspond to promoter regulatory motifs less prone to evolutionary changes than other sequences, for instance those downstream of the 3' ends of genes. This observation demonstrates the existence of a selective pressure that acts conservatively at

gene promoters. With respect to coding sequences, there is a high level of conservation among the *sensu stricto* yeast species and most changes at the nucleotide level have occurred at the third position among synonymous codons, that is, they are conservative (153).

In spite of the high level of genome conservation among the *sensu stricto* species, there exists a small number of genes that are species-specific (63, 153). These may have resulted from recombination events involving more distant species. A large proportion of these species-specific genes have a metabolic function, particularly in sugar metabolism. Although some of the above genome changes are still difficult to explain on a molecular basis, they shed new light on the mechanisms of genome evolution that may be applicable to other closely related organisms, such as humans and chimpanzees. Again, genomic studies on *S. cerevisiae* and related species may open the way for analysis of the evolution of biological systems.

1.3 *Candida albicans*

1.3.1 A human fungal pathogen.

Almost all of us carry *C. albicans* in our gastrointestinal and genitourinary tracts, and on our skin. *C. albicans* can colonize and invade the host tissues when the immune systems are weakened either as a result of cancer chemotherapy or HIV infections, or as a result of elimination of competing flora after antibiotic treatment. These infections can range from thrush in immunocompetent colonized hosts, to life-threatening systemic infections in immunocompromised individuals such as patients with cancer (64). *C. albicans* is the fourth most common hospital acquired infection in the United States, and

its treatment costs more than US $1 billion each year (14, 216). Because this fungal pathogen is a eukaryote and shares much of its biological processes with humans, many antifungal drugs cause deleterious side effects. It is therefore important for *Candida* research to identify targets of antifungal technologies.

1.3.2 Yeast to Hyphal transition of *C. albicans.*

Because the biological processes of *C. albicans* are similar to those of *S. cerevisiae*, rapid advances have been made in understanding *Candida* molecular and cellular biology. *S. cerevisiae* diverged from *C. albicans* 250 million years ago (116), and is an important guide for studying aspects of signal transduction, mating, cell-cycle progression, cell-wall biosynthesis, and metabolism in *C. albicans*. Despite the fact that many processes are conserved between the two organisms, there are also significant differences between the two cousins. For example, while *S. cerevisiae* grows exclusively as budding yeast, *C. albicans* can grow in several morphological forms: it can proliferate as budding yeast, and it can also form filaments, such as true hyphae and pseudohyphae. Hyphal formation of this fungus can be induced by high temperature, basic pH, and by the presence of serum, and this morphological transition is implicated in the pathogenesis of *C. albicans* (35).

Because the virulence of *C. albicans* has been associated with its ability to switch between yeast and hyphal forms, morphogenesis has been the focus of the research of this fungus. *C. albicans* cells in a hyphal state are often found at sites of tissue invasion, and cells that do not form hyphae are often less virulent (105). In addition, other *Candida* species that do not form true hyphae are less frequently isolated from the human host,

which suggests that they are less virulent. At the same time, strains that are unable to grow in the yeast form are also less virulent (30, 105, 170, 190). It is currently thought that the hyphal cells express cell wall proteins that both facilitate adhesion to human tissues, and are important for tissue invasion, as well for escape from the white blood cell-mediated phagocytosis. In contrast, it is believed that the yeast form of *C. albicans* disseminates through the blood stream (105)

Much effort has been made by the scientific community to understand the environmental factors that induce hyphal formation through signaling pathways. Different hyphal inducing factors probably reflect the variety of microenvironments in which the pathogen survives *in vivo* (105, 187). Just as in *S. cerevisiae*, cAMP and MAP kinase pathways target transcription factors (in *C. albicans* the Efg1p and Cph1p factors respectively) that promote morphogenesis. The deletion of *efg1* blocks filamentation in most conditions, while the deletion of *cph1* blocks the filament formation only in response to a limited set of conditions (157, 173, 188). This shows that the cAMP pathway plays a more important role in *C. albicans* morphogenesis than in *S. cerevisiae*. A *cph1 efg1* double mutant fails to form filaments in most *in vitro* conditions, and is avirulent in a systemic mouse model of candidiasis (190). This observation is often viewed as showing that the ability to form hyphae or pseudohyphae is an essential virulence factor. However, these mutations block the expression of hyphal-specific genes, many of which are also required for virulence. In addition, *cph1 efg1* mutants were found to produce filaments under some *in vivo* and *in vitro* conditions (251), possibly due to the action of other hyphal inducing pathways, such as the Rim101p pathway, which is activated by alkaline pH (69, 79) and the Czf1p pathway, which is activated by growth in

a solid matrix (36). Hyphal induction is repressed by transcriptional inhibitors such as Tup1p (31), which associates with its DNA-binding partners Nrg1p (32, 220) and Rfg1p (149).

1.3.3 *C. albicans* genome.

An important tool for the scientific community working with *C. albicans* is the whole genome sequence of this fungal pathogen; this sequence is needed for the application of knowledge acquired from the biology of other model organisms such as *S. cerevisiae*, for the proper interpretation of functional genomic studies, and ultimately for the identification of novel antifungal targets that may be unique to the pathogen and whose inhibition won't have deleterious effects on the host.

The Candida Genome Sequencing Project started in 1996, and in 2004 it produced a diploid assembly constructed from 10.9X coverage, called Assembly 19. It provided single contigs where heterozygosity was not obvious and allelic contigs where there was significant heterozygosity (148). Several important steps along the way to this release are detailed in a review by Nantel (223). The first was the construction of a physical map of chromosome 7 (54). Chibana *et al.* (56) completed the sequence of chromosome 7, identified 404 genes, and compared the synteny to the *S. cerevisiae* genome. The authors sequenced the Major Repeat Sequences and the gaps left in Assembly 19. They then aligned the sequence on the chromosome as determined by the physical map (54). Then there were the two early releases of the emerging genome sequence data, called Assembly 4 and Assembly 6, which were the lower density assemblies that facilitated a great deal of gene analysis. These assemblies provided the framework for the

construction of several microarrays (168, 224), an analysis of genes haploinsufficient for filamentation (300), and the elucidation of several gene families, including a number important in pathogenesis, such as the secreted aspartyl proteinases (*SAP*s) (131), the agglutinin-like substances (*ALS*s) (127), and the phospholipases (*PLB* and *PLC*) (160, 259). Two relatively comprehensive *C. albicans* disruption libraries have been described. One library was constructed systematically by targeted disruption of one allele followed by insertion of a regulated promoter at the other allele (254). The other disruption library was constructed randomly by transposon mutagenesis, inserting a UAU cassette into one allele, which facilitates disruption of the second allele via two spontaneously occurring steps of mitotic recombination (39). One of the very important achievements was Assembly 19, the diploid assembly of the genome of *C. albicans* strain SC5314. The genome sequence provided a great deal of insight into many aspects of genomic organization, such as the large amounts of heterozygosity (148).

1.4 Introduction to transcription

1.4.1. General overview of eukaryotic initiation of transcription.

In general, a genome consists of genes, which are transcribed into mRNA and then translated into proteins. Expression of a typical eukaryotic gene is a complicated process, which involves transcription, RNA splicing, capping, and transport out of the nucleus into the cytoplasm where the processed message can be translated into protein. In this work I will mostly concentrate on the initiation of transcription of genes.

Much of what we know about the regulation of gene expression comes from studies of the yeast *Saccharomyces cerevisiae*. Yeasts, like all eukaryotes, have three

RNA polymerases (253). PolI transcribes genes encoding RNA that functions as a part of ribosomes (rRNA) (246). PolII transcribes genes that code for proteins (mRNA) (113), while the tRNA genes are transcribed by PolIII (98).

The transcriptional machinery of yeast is very complex: in addition to the core polymerase, it can contain at least 100 proteins (178) (Table 1.4.1). The function of many of these proteins is not known, but we can nevertheless make the following general points:

- Some of the protein complexes, such as TBP, TFIIB, TAF17, and TFIIA bind DNA directly to form a platform for RNA polymerase II (41, 106, 177, 233, 277).
- Some of the protein complexes contain enzymes, such as TFIIH, which contains a kinase and two helicases (280). These activities help the polymerase initiate transcription and then exit the promoter. Other factors like histone acetyl transferases (HATs), hidtone deacetylates (HDACs), and histone kinases regulate transcription by chemically modifying histones in nucleosomes (19, 37).
- Some of these proteins are required globally for transcription – when any one of the proteins like TBP, TFIIB and the helicases in TFIIH is depleted from cells - transcription of all genes by PolII stops. In contrast, other proteins, like some members of the nucleosome modifying enzymes and some members of the mediator complex, are needed for the expression of only certain genes under certain conditions (see below) (106, 113, 177, 277, 280).
- Most proteins of the transcriptional machinery can be isolated from cells as parts of complexes. The mediator complex, which consists of more than 20 proteins, can associate with polymerase to form a "holoenzyme" (27, 206, 221). In

addition, proteins found in one complex are sometimes found in another complex as well. For example, certain TBP-associated factors also associate with the HAT Gcn5 in the SAGA (Spt, Ada, Gcn5 Acetyltransferase) complex (175).

The model of gene activation in eukaryotes is consistent with predictions of the "regulated recruitment" mechanism (10, 152, 177, 201, 202, 210, 245, 303):

- The transcriptional machinery is not bound to the promoter of a given gene prior to activation.
- Activators are modular proteins with separable activating and DNA-binding domains.
- The activating domain must be attached to the DNA promoter to activate transcription of the downstream gene.
- A given activating region will transactivate when attached to a heterologous DNA binding domain, even one taken from bacteria.
- Any number of genes can be controlled by a given activator if their promoters contain the appropriate activator binding sequence.

Table 1.4.1. Transcriptional machinery: *S. cerevisiae* vs *C. albicans.*
S. cerevisiae and *C. albicans* transcriptional machinery are compared by calculating the average e values of its individual components.A pair-wise sequence comparison was done on the full length amino acid sequences using a standard pair-wise blastp setting on http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi. P (-LOG10(e-values)) is a statistical measure of the similarity between the two proteins: the higher the value of P – the more homologous the two proteins are.

S. c. NAME	*C. a.* orf19.####	P (-LOG10(e-values))	AVERAGE	Complex
MED1	3434	1		MED-Mediator
MED2	7700	1		MED-Mediator
MED3	5929	2		MED-Mediator
MED4	1878	20		MED-Mediator
MED5	632	2		MED-Mediator
MED6	7420	35		MED-Mediator
MED7	232	34		MED-Mediator
MED8	4497	23		MED-Mediator
MED9	5772	3		MED-Mediator
MED10	5268	32		MED-Mediator

MED11	6909	8		MED-Mediator
MED12	736	22		MED-Mediator
MED13	1451	37		MED-Mediator
MED14	4348	46		MED-Mediator
MED15	5105	23		MED-Mediator
MED16	1343	23		MED-Mediator
MED17	3055	26		MED-Mediator
MED18	4091	10		MED-Mediator
MED19	6895	1		MED-Mediator
MED20	3799	1		MED-Mediator
MED21	7548	23		MED-Mediator
MED22	3317	7		MED-Mediator
MED31	1429	23	17	MED-Mediator
HAT1	779	79		Histone Acetylation
GCN5	705	152		Histone Acetylation
HPA2	6323	32		Histone Acetylation
ESA1	5416	140		Histone Acetylation
SAS3	2540	94		Histone Acetylation
ELP3	7387	266		Histone Acetylation
TAFII145	735	33	114	Histone Acetylation
GCN5	705	152		SAGA
ADA1	307	36		SAGA
ADA2	2331	126		SAGA
ADA3	3023	49		SAGA
STP3	7622	91		SAGA
SPT7	7572	143		SAGA
SPT8	4312	86		SAGA
SPT20	8052	28		SAGA
TRA1	139	999	190	SAGA
RPD3	6801	200		Histone Deacetylation
HDA1	2606	208		Histone Deacetylation
HOS3	2772	147		Histone Deacetylation
SIR2	4761	97	163	Histone Deacetylation
RPB1	7655	999		RNA Polymerase II
RPB2	3349	999		RNA Polymerase II
RPB3	1248	91		RNA Polymerase II
RPB4	7785	17		RNA Polymerase II
RPB5	6340	72		RNA Polymerase II
RPB6	2643	53		RNA Polymerase II
RPB7	3347	57		RNA Polymerase II
RPB8	6314	45		RNA Polymerase II
RPB9	2276	45		RNA Polymerase II
RPB10	7474	1		RNA Polymerase II
RPB11	7805	9		RNA Polymerase II
RPB12	7255	17		RNA Polymerase II
ELP2	2711	192		RNA Polymerase II
CDK8	794	121		RNA Polymerase II
CycC	7355	68	186	RNA Polymerase II
TBP	1837	84		TFIID
TAF2	2135	151		TFIID
TAF5	536	176		TFIID
TAF6	7454	105		TFIID
TAF9	1111	41		TFIID
TAF12	470	46		TFIID
TAF11	6923	14		TFIID
TAF7	1574	45		TFIID
TAF10	3242	34		TFIID
TAF3	4388	17		TFIID
TAF14	798	36		TFIID
TAF13	5174	19	64	TFIID
TOA1	2682	33		TFIIA
TOA2	4625	32		TFIIA
SUA7	3059	94		TFIIB
SSU1	7313	41		TFIIF

TFG2	2111	73		TFIIF
TFA1	4851	61		TFIIE
TFA2	4882	52	55	TFIIE
TFB1	5297	56		TFIIH
TFB2	5846	129		TFIIH
TFB3	567	63		TFIIH
TFB4	4194	46		TFIIH
RAD3	7119	999		TFIIH
SSL1	1457	109		TFIIH
SSL2	2857	318		TFIIH
KIN28	6239	118		TFIIH
CCL1	4542	47	209	TFIIH
SWI1	5657	36		SWI/SNF
SWI2	1526	283		SWI/SNF
SWI3	4488	88		SWI/SNF
SNF5	5871	57		SWI/SNF
SNF6	2827	1		SWI/SNF
SNF12	2265	41	84	SWI/SNF
STH1	7869	312		RSC
RSC1	2964	107		RSC
RSC2	2964	113		RSC
RSC3	5133	4		RSC
RSC4	2041	19		RSC
RSC6	2265	24		RSC
RSC8	7234	68		RSC
RSC9	2473	30		RSC
SFH1	5093	28	78	RSC
ISW1	4437	272		ISW1
IOC2	7506	5		ISW1
IOC3	2266	12		ISW1
IOC4	7023	14		ISW1
ISW2	7401	999		ISW1
DLS1	3063	4		ISW1
DPB4	2088	28		ISW1
ITC1	5510	34	171	ISW1
PAF1	3613	23		PAF-Mediator
CDC73	2983	44		PAF-Mediator
CTR9	7067	109		PAF-Mediator
RTF1	7047	66		PAF-Mediator
LEO1	7116	24		PAF-Mediator
CCR4	5101	149		PAF-Mediator
HPR1	none	0		PAF-Mediator
MED15	5105	23	55	PAF-Mediator

1.4.2 A model case: the regulation of *GAL* genes.

Galactose is utilized by almost all organisms through its conversion to glucose-6-phosphate in a reaction catalyzed by the enzymes of the Leloir pathway (179). In *Saccharomyces cerevisiae* these enzymes are encoded by *GAL1* (galactokinase), *GAL7* (galactose-1-phosphate uridyl transferase), *GAL10* (UDP-glucose-4-epimerase), and *GAL5* (phosphoglucomutase) (75-77). Galactose enters *S. cerevisiae* cells through a specific permease, encoded by *GAL2* (297). Because eukaryotic genes are not organized

in operons, it is intriguing that the *GAL1*, *GAL7*, and *GAL10* genes are clustered near the centromere of chromosome II, and the expression of those genes is induced by galactose (192). The level of induction of *GAL* genes is unusually high for yeast genes. When cells grow in the absence of galactose the level of expression of *GAL* genes is very low. The addition of galactose induces *GAL* genes more than 1000-fold, which makes them among the most highly expressed genes in the cell. This high level of expression is attributed to the action of a single transactivator, Gal4p, which binds specific DNA sites upstream of *GAL* genes (147). In addition, Gal4p will activate a wide array of genes in yeast and many other higher eukaryotes if Gal4p binding sites are introduced nearby (12). There is also one additional regulator of *GAL* genes: the repressor Mig1p, which only works in the presence of glucose (43). This ensures that if both galactose and glucose are present, glucose will be utilized first. A Mig1p site is also found upstream of *GAL4* gene, so that glucose reduces the level of Gal4p in addition to repressing *GAL* genes directly (227). Mig1p binding sites are also found upstream of many other glucose-repressed yeast genes (107, 192).

The four sites on the *GAL1* promoter to which Gal4p binds comprise the galactose upstream activating sequence (UAS_G), and lie about 275 base pairs upstream of the *GAL1* transcriptional start site. A binding site for the Mig1p is found between the UAS_G and the *GAL1* transcriptional start site. Each *GAL* gene, like other typical eukaryotic genes, contains a TATA element, which binds TBP along with many other proteins including polymerase. The *GAL* genes can exist in one of the three states (147, 192):

- In the absence of galactose and glucose Gal4p is bound to UAS_G but does not activate transcription because it is complexed with an inhibitor Gal80p.

- In the presence of galactose Gal4p is freed from Gal80p, and the specific *GAL* gene is transcribed.
- In the presence of both glucose and galactose the gene is kept off by the repressor Mig1p. This protein recruits the repressing complex, a key component of which is Tup1p.

The specificity of any transcription factor comes from its DNA binding domain (DBD). Gal4p homodimerizes and recognizes each of the 17 base pair sites ($CGGN_{11}CCG$) using a zinc cluster type domain (176, 303). This motif inserts an α-helix into the major groove of DNA. Mig1p binds CCCCRNNWWWWW DNA using a related motif called a zinc finger (199).

The protein Gal3p senses galactose in the environment, binds it along with ATP, and interacts strongly with Gal80p (240). This interaction frees Gal4p's activation domain (AD) (272). In the absence of glucose, Mig1p is held in the cytoplasm in a phosphorylated form (274). A kinase, Snf1p, which is responsible for this phosphorylation, is inhibited by glucose: this inhibition allows Mig1p to enter into and remain in the nucleus (87).

Gal4p has distinct DNA-binding and activating functions, which are located on separate domains of the protein, and the specific mode of DNA binding is not critical for activation (217, 271). The Gal4p DBD is found within the first 100 amino acids of the protein, and this region also contains a nuclear localization signal and the Gal4p dimerization sequences. Gal4p DBD is able to bind the DNA but does not activate transcription. The residues 100-881 of Gal4p can be attached to a different DBD, for example the DBD of the bacterial *lexA*, and the resulting hybrid protein can activate the

transcription of a gene that bears binding sites recognized by the new DBD (152, 201, 202, 210). This experiment shows three important features of Gal4p:

- The DBD does not suffice for transactivation – an AD is also required.
- The AD must be tethered to DNA to work.
- The same AD will function when attached to any type of DBD, even one from bacteria.

In fact, the link between the AD and the DBD does not have to be covalent. This was shown by attaching the AD to Gal80p in a strain where Gal4p bears a weakened AD, but nevertheless binds Gal80p (200). Interestingly, the Gal80-AD hybrid activates the transcription of the DNA bound Gal4p. The Gal4p DBD, in fact, tethers the Gal80-AD to DNA so that it activates the transcription of the downstream gene. This implies that any protein-protein interaction that brings the AD in close proximity to the DBD will elicit transcription; this is the basis of the "two-hybrid" system, which detects pairs of interacting proteins.

S. cerevisiae Gal4p contains two "acidic" ADs (ADI and ADII): they bear an excess of negatively charged glutamic and aspartic acids (201). Each of the ADs is about 100 amino acids long, and fusing either one to a DBD creates an activator. The activation strength of Gal4p is proportional to the length of ADs: the sequential deletion of the ADII proportionally reduces the transactivation strength of this protein. In addition, mutants in which Gal4p ADI bears more acidic and less basic residues show an increased Gal4p activation ability. Finally, ADII, one of the two Gal4p activating regions, also binds Gal80p. These two functions can be distinguished by mutation: some single amino acid substitutions in ADII decrease or even abolish Gal80p interaction, while having little or

no effect on ADII activation ability. This suggests that the same stretch of peptides can form two alternative structures: one somewhat defined structure needed to bind Gal80p, and another less-defined structure required for activation.

When attached to a DBD many different peptides work as ADs (202). Most of those new ADs resemble natural ADs in that they are acidic and include hydrophobic amino acids. For these new ADs, as for the natural ones, scrambling the order of amino acids destroys the function: the arrangement of acidic and hydrophobic residues in the sequence is important. These properties of ADs suggest that they contain reiterated motifs that are difficult to recognize at the primary amino acid sequence level.

Activators that work by recruitment cannot activate without binding to DNA (200-202). If not bound to DNA, a transactivator would not be able to direct the transcriptional machinery to the promoter of a gene. In fact, if an unbound transactivator had any effect, it would be predicted to be negative: at high concentrations unbound "AD A" would titrate components of the transcriptional machinery and make them unavailable for fruitful interaction with a DNA-bound activator "B". The observation is that the stronger the AD and the more highly it is expressed – the tighter it is bound to its transcriptional machinery target. Overexpression of a Gal4p fragment lacking the DBD does not activate transcription, and if *GAL* genes are being activated by a DNA-bound activator, their expression is inhibited. Similar results have been observed with other activators. In general, the stronger the AD and the more highly it is expressed, the greater the degree of inhibition, a phenomenon called "squelching". It is believed that it is for those reasons that one can rarely find activators free of inhibitors in the nucleus, except when they are working appropriately.

A protein-protein interaction, created by mutation, between a DNA-bound protein and the transcriptional machinery can suffice for activation (120). As mentioned before, the Gal4p (1-100) fragment dimerizes and binds DNA efficiently, but does not transactivate because it lacks an AD. A mutant yeast strain was isolated in which this Gal4p fragment activated transcription as well as the intact Gal4p. The mutant strain bears a single amino acid change in the Gal11 protein, which is a component of the transcription mediator. The mutation in Gal11p created an adhesive protein-protein interaction between Gal11p and Gal4p (1-100). The direct tethering of the transcriptional machinery, as in the case of *lexA*-Gal11p or Gal4(DBD)-Gal11p, leads to the expression of the *GAL1* gene, provided the *lexA* binding site has been inserted nearby.

In bacteria, polymerase is the only relevant complex needed for the activator to recruit to start transcription. Yeast, as all other eukaryotes, has several additional complexities. The transcriptional machinery comes in parts: the DNA is wrapped around histones to form nucleosomes, and various complexes chemically modify nucleosomes or change their disposition on DNA (234, 285). Each nucleosome contains 146 base pairs of DNA wrapped around the histones. Each nucleosome is separated by linker DNA, which contains 15 – 20 base pairs. The nucleosomal DNA is less accessible than naked and linker DNA. The nucleosomes of the actively transcribed genes are often modified. The tail of histones H3 and H4 are typically acetylated and deacetylated at specific lysine residues (16, 19, 37, 108, 175). Histone acetyltransferases or HATs, enzymes that acetylate histones, have been shown to be required for the full induction of the transcription of certain genes. HDACs, enzymes that deacetylate histones, have been seen to repress certain genes.

These facts could be interpreted as follows. Nucleosomes present a barrier to the transcriptional machinery; acetylation of the histones weakens that barrier, while deacetylation strengthens it. It is possible that the acetylation of lysines in the histone tails eliminates positive charge and therefore weakens interaction of those tails with the negatively charged DNA. The expression of the genes can also be affected by enzymes, such as the Swi/Snf complex (65, 279), that alter the mobility or configuration of histones on DNA. In light of these observations it is possible to consider some scenarios of how activators might work. On the one hand, direct recruitment of transcriptional machinery by an activator would suffice for the initiation of transcription. On the other hand, recruitment of nucleosome modifiers might serve to initiate the event, with the transcribing complex then forming without additional help. Yet another possibility would be that both kinds of complexes are directly recruited.

The induction of the *GAL1* gene results in the appearance of various components of the transcriptional machinery at the promoter, such as polymerase, TFIIH, TFIIE, TBP, mediator components, and SAGA (178, 180, 241). Components like TFIIE and mediator are only found at the promoter, while others are found both at the promoter and downstream in the gene itself along with polymerase. In total, SAGA and mediator complexes contribute to activation of *GAL1* by Gal4p; both of those complexes appear at the promoter along with polymerase upon activation, and may be recruited directly, or some might bind cooperatively along with those recruited directly. The following experiment showed that to activate efficiently, Gal4p had to touch both SAGA and the mediator. When a fusion protein comprising either a component of SAGA or a component of mediator complexes is attached to the DBD it elicits very weak activation.

But when binding sites for both of these fusions are placed on a promoter, strong activation is seen (40).

If Gal4p binds its sites sufficiently tightly, interacts with the transcriptional machinery sufficiently strongly, and that machinery binds the *GAL1* promoter sufficiently avidly, the help from nucleosomal modifiers is unnecessary. On the other hand, weakening one or another of the interactions can impose a requirement for the nucleosomal modifiers. But there is a cell cycle-dependent state in which the induction of *GAL* genes depends on nucleosomal modifiers, such as Gcn5p and Swi/Snf, even in the wild-type configuration (178). The state of DNA differs at different stages of the cell cycle: chromosomes are condensed during mitosis, for example. It was observed that during mitosis the induction of *GAL1* would proceed normally only if both Gcn5p and Swi/Snf are present, which suggests that at this stage the chromatin is in such a configuration that those nucleosome-modifying enzymes are required (158).

1.4.3 Another model case: Gcn4p.

In *S. cerevisiae GCN4* encodes a transcriptional activator of amino acid biosynthetic genes that responds to amino acid starvation (123, 124). ScGcn4p is tightly regulated at both the transcriptional and translational levels. The 5' leader region of *GCN4*, which codes for a transcriptional activator of amino acid biosynthetic genes in response to amino acid starvation, contains four small upstream ORFs (uORF1-4). These uORFs act as negative regulators of translation: the ribosome initiates translation at uORF1 and becomes re-activated for translation at subsequent uORFs. Under environmental stresses, such as amino acid starvation, the translation of *GCN4* is

induced: Gcn2p kinase phosphorylates eIF2α, the α subunit of the translation initiation factor, and the scanning ribosome is not re-activated until it bypasses the uORFs and initiates translation at the *GCN4* open reading frame (122).

Gcn4p binds the 5'-TGACTC-3' DNA binding site, located upstream of many genes induced during amino acid starvation, as a homodimer with its basic leucine-zipper (bZIP) found within its carboxy terminus (7, 126). In *S. cerevisiae* Gcn4p there are two transcription activation domains: one resides in an acidic segment in the center of the protein between residues 107 and 144 (125), the second Gcn4p activation domain is located in the N-terminal 100 amino acids. The two activation domains are functionally redundant and can work independently to produce high-level activation (78, 141).

1.4.4 Types of activation domains.

The transcriptional activation domains are classified according to their amino acid composition: rich in acidic (e.g. *S. cerevisiae* Gal4p and Gcn4p) or basic residues (tobacco BBC1), or rich in glutamine (*S. cerevisiae* Mcm1p), threonine-serine (human OCT2), or isoleucine (NTF1) (8, 38, 61, 83, 125, 146, 201, 202). The DNA binding modules also fall into several classes, such as zinc-finger, leucine-zipper, and helix-loop-helix motifs (81, 84, 306). Although the activation domain is critical for function, and can provide a level of regulation, the gene specificity of such transcriptional activators is ultimately determined by the DNA-binding address of the protein.

The role of the transcription activator is to recruit the RNA Polymerase II machinery to the promoter to which it is bound. Thus, the transcriptional activation domains of both Gcn4p and Gal4p interact with Gal11p, which is a component of the

mediator complex that binds the RNA Polymerase II machinery (10, 114, 142, 235, 248). Three general models have been proposed to characterize the structure of the transactivation domain. In the first hypothesis, it was proposed that activation domains are unstructured "acidic blobs" that interact with their targets via ionic interactions. This model is supported by the observation that the removal of the residues of activation domains decreases the activity gradually, rather than abruptly (271). A second model proposes that acidic activation domains form amphipathic α helices, in which acidic residues are aligned on one face of the helix. This model is supported by the observation that an artificial 15 residue peptide, designed to fold into an amphipathic α helix, shows transactivating abilities when fused to the Gal4p DNA binding domain (202). The authors of the third model argue that the most likely secondary structure is an antiparallel β sheet (301).

CHAPTER 2. THE COMPLETION OF *C. ALBICANS* GENOME

The previously discussed achievements have greatly advanced the pace of molecular analysis of the pathogenesis and life style of *C. albicans*, but Assembly 19 was not a finished sequence, since it contained a total of 412 contigs, of which 266 were allelic. In order to provide a finished sequence, called Assembly 21, we used a synteny analysis with preliminary versions of the *Candida dubliniensis* genome assembly, bioinformatics, a Sequence Tagged Site (STS) map of overlapping fosmid clones, and an optical map to determine the sequence of each of the eight chromosomes. This assembly has eight linear DNA sequences including 9 MRSs (3 completely sequenced), 14 MRS subrepeats RB2 sequences, 2 MRS subrepeats HOK sequences, and the rDNA. In addition to its usefulness for gene mapping, Assembly 21 reveals some interesting biological features, including a putative transcription factor gene family with members proximal to at least 14 of the 16 telomeres, a telomere-like sequence in the middle of chromosome 1, information on the relationships of chromosome location to similarity of gene families, and a revised ORF list. Assembly 21 stimulates experiments in such areas as evolution and genome dynamics.

Section 2.1 is the second version of the manuscript that has been resubmitted to Genome Biology, and it describes the assembly of the *C. albicans* genome. Section 2.2 is an article that describes the annotation of the *C. albicans* genome.

Section 2.1: Assembly of the *Candida albicans* Genome into 8 Chromosomes.

Assembly of the *Candida albicans* Genome into 16 Supercontigs Aligned on the 8 Chromosomes.

Marco van het Hoog[*1], Timothy J. Rast[*2], Mikhail Martchenko[1], Suzanne Grindle[2], Daniel Dignard[1], Hervé Hogues[1], Christine Cuomo [3], Matthew Berriman[4], Stewart Scherer[5], B. B. Magee[2], Malcolm Whiteway[1], Hiroji Chibana[t6], André Nantel[1], and P.T. Magee[$2]

[1]Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec, Canada; [2]University of Minnesota, Minneapolis, MN; [3]Broad Institute of MIT and Harvard, Cambridge, MA; [4]Wellcome Trust Sanger Institute, Hinxton, United Kingdom; [5]3938 Paseo Grande, Moraga, CA 94556; [6]Chiba University, Chiba, Japan.

* These authors contributed equally to this manuscript.

2.1.1 ABSTRACT

The 10.9x genomic sequence of *Candida albicans*, the most important human fungal pathogen, was published in 2004. Assembly19 consisted of 412 supercontigs, of which 266 were a haploid set, since this fungus is diploid and contains an extensive degree of heterozygosity but lacks a complete sexual cycle. However, sequences of specific

chromosomes were not determined. Supercontigs from Assembly 19 (183, representing 98.4% of the sequence) were assigned to individual chromosomes purified by pulse-field gel electrophoresis and hybridized to DNA microarrays. Nine supercontigs contained markers from 2 different chromosomes. For Assembly 21, using a synteny analysis with preliminary versions of the *Candida dubliniensis* genome assembly, bioinformatics, a Sequence Tagged Site (STS) map of overlapping fosmid clones, and an optical map we determined the sequence of each of the eight chromosomes. The orientation and order of the contigs on each chromosome, repeat regions too large to be covered by a sequence run, such as the ribosomal DNA cluster and the MRSs; and telomere placement were determined using the STS map. The overall assembly was compared to an optical map; this identified some misassembled contigs and gave a size estimate for each chromosome. Assembly 21 reveals a chromosome fusion, a number of small internal duplications followed by inversions, and a subtelomeric arrangement including a new gene family, the *TLO* genes. Correlations of position with relatedness of gene families imply a novel method of dispersion. The sequence of the individual chromosomes of *C. albicans* raises interesting biological questions about gene family creation and dispersion, subtelomere organization, and chromosome evolution.

2.1.2 INTRODUCTION

In the last 25 years, the opportunistic human pathogen *Candida albicans* has become a serious medical problem. This fungus is now fourth on the list of hospital-acquired infections, ahead of Gram-negative bacteria, and despite the recent introduction of a new class of antifungals, drug resistance continues to be a problem (159). In the last

15 years molecular techniques have been applied to understand the pathogenesis of this organism as well as to search for novel drug targets. However, *C. albicans* presents several difficulties for molecular biologists. It is diploid; only a part of a sexual cycle has been demonstrated; it has a very plastic genome; and it is highly heterozygous. Each of these properties is best investigated through a genomic approach. Hence, knowledge of the genome sequence has been an important goal for the last 10 years. More recently, genome structure and dynamics have become increasingly important in this organism as wide-spread aneuploidy (49, 268), the role of repeated DNA in chromosome loss (183), and chromosome rearrangement leading to drug resistance (269) have been reported.

The steps taken to create *C. albicans* genome Assembly 19 are described in section 1.4.3, but Assembly 19 was not a finished sequence, since it contained a total of 412 contigs, of which 266 were the haploid set. In order to provide a finished sequence, we used hybridization of chromosomes partially purified by pulse-field elecrophoresis as well as an STS map based on a fosmid library to identify the chromosomal location of various contigs. We then utilized bioinformatics to analyze both the emerging sequence of the sister species *Candida dubliniensis* and the primary traces used to generate Assembly 4, and coupled this with the STS map and a whole-chromosome optical map to construct Assembly 21. This assembly has eight linear DNA sequences including 9 copies of the intermediate repeat called the Major Repeat Sequence (MRS), of which 3 have been completely sequenced. The MRS is made up of three subrepeats, called RB2, RPS, and HOK (59). In addition to the intact MRS sequences, there are 14 RB2 sequences and 2 HOK sequences. The ribosomal DNA constitutes another repeat, which is not included in the assembly. In addition to its usefulness for gene mapping, Assembly

21 reveals some interesting biological features, including a putative transcription factor gene family with members proximal to 14 of the 16 telomeres, a telomere-like sequence in the middle of Chromosome 1, information on the relationships of chromosome location to similarity of gene families, and a revised ORF list.

2.1.3 MATERIALS AND METHODS

Assembly 21. The first step in Assembly 21 was to assign contigs from Assembly 19 to the appropriate chromosomes. One hundred contigs were anchored by probes on the physical map. To assign the rest, chromosomes were separated by pulse-field gel electrophoresis. Chromosomal bands were eluted from the gel, labeled with either Cy3 of Cy5 dyes, and hybridized to a microarray (224) based on ORFs identified from Assembly 4. As a control, total genomic DNA was also labelled with the reciprocal dye and co-incubated along with the partially purified chromosomes. At least two individual hybridizations (including dye swap controls) were conducted for each of the 8 chromosomes. Results and details from these experiments can be seen on our web page at http://candida.bri.nrc.ca/index.cfm?page=CaChrom. Fluorescence ratios were interpreted visually and it was very easy for us to assign chromosomal localization for 183 Assembly 19 contigs representing 98.4% of DNA sequences and 97.3% of known genes. This analysis also identified nine misassembled contigs by the fact that genes assigned to them hybridized to different chromosomes. The emerging sequence from the closely related species *Candida dubliniensis* was then aligned, using megablast, to the Assembly 19 contigs. This alignment is shown at http://candida.bri.nrc.ca/dubliniensis/. After the release of the Assembly 19 traces, Phrap was used to connect pairs of contigs which were assigned to the same chromosome, and were found to be adjacent based on the *C. dubliniensis* overlap. Phrap and the traces were also used to locate misassemblies and to correct them.

The emerging alignments were compared with the physical map and areas of disparity mapped with more probes. Attempts to assemble the contigs into whole

chromosomes with Phrap failed because of the large number of repeated elements in the *C. albicans* genome. A custom-made stitcher script (which utilized a 100 nucleotide pure text string starting from the previous alignment) was used to assemble the contig-alignments into chromosomes. The resulting assembly was checked for the presence of all the ORFs from the community annotation (33), and very few ORFs were missing. The penultimate assembly was compared with the optical genome map and the physical map and disparities resolved.

One major discrepancy was the orientation of the sequence between the MRS regions on Chromosome 4. The optical map and the physical map showed it in one conformation, while the assembly showed it in the other. The final orientation, consistent with the physical and optical maps, was confirmed by PCR of the border fragments of one MRS and by restriction digestion of genomic DNA and hybridization of a Southern blot with probes from the regions which flank the two MRSs on both sides.

The remaining gaps were filled by PCR followed by sequencing of the products. One gap proved to be the result of a misassembly in a contig. The insertion of a partially overlapping contig closed it. Two gaps were filled by using the draft sequence of Broad Institute's WO-1 assembly (http://www.broad.mit.edu/annotation/genome/candida_albicans/Info.html). One gap was filled by a sequence from Selmecki, et al (269). The result is an assembly of eight chromosomes with one unresolved (not connected) contig pair. Two gaps were introduced due to the flip of the area between two MRSs on Chromosome 4. Two gaps are located around the rDNA region. In addition, most of the MRS regions have not been resolved and can be considered gaps, with the exception of the two MRSs in

Chromosome 7 and the one MRS in Chromosome 6, which were cloned and sequenced at Chiba University.

The physical map. A fosmid library was constructed from strain 1161 (104) by M. Strathman. Sau3AI-digested genomic DNA was inserted into the fosmid vector pFOS1 (156). The library consists of 3840 clones with an average insert size of 40 kb (10x genome coverage). For probing, the library was arrayed in 10 384-well plates. Two plates each were printed on a 12x16 cm nylon membrane (Hybond-N+, Amersham Biosciences), giving a complete library set on 5 membranes. For printing, freshly thawed fosmid clones were transferred, using a 384-point replicator, to the membrane overlaid on Luria agar containing 20 μg chloramphenicol. The plates were grown overnight at 37°C and the membranes were processed according to the manufacturer's instructions for colony lifts.

Probes were variously clones of genomic DNA, T- and S-end probes from fosmids (52), and PCR products generated from SC5314 DNA using primers based on the public genomic sequences. Probes were randomly labeled with ^{32}P. The mapping was carried out as described in Chibana, et al (55). Briefly, membranes containing the library were hybridized with probes and fosmids hybridizing with the same probe were considered to overlap. Probes were also hybridized to Southern blots of pulse-field separations of chromosomes and genomic *Sfi*I digests. The overlapping fosmids were arranged in a linear fashion with chromosome markers like the MRS used to orient the array. The complete set of overlapping fosmids was trimmed to make a minimum tiling set. The physical map is available at http://albicansmap.ahc.umn.edu/index.html.

The optical map. The optical map was prepared by OpGen (Madison, WI), using a proprietary technique. The technique involves preparation of large (> 500 kb) molecules of genomic DNA, and spreading them on a microfluidic device, which causes them to elongate and adhere to the substratum. They are then treated with a restriction enzyme (in this case, *Xho*I). Since the molecules are under slight tension, restriction sites appear as breaks in the molecules. The molecules are then stained with a fluorescent dye, scanned, and the positions of the restriction sites determined, using the amount of fluorescence to calculate the mass of the fragments. The results are averaged over 500 molecules, allowing the preparation of a macro-restriction map. This can be compared with the restriction map derived from the assembly. A more complete description of the technique and more information can be found at http://www.opgen.com/. Since the optical map software searches the image database for matches and then constructs the map based on the consensus pattern, the optical map of a chromosome describes only one of the two homologues, and no data on heterozygosity are included.

Although in most cases the three methods were in agreement on the sequence assembly, in cases where disagreements could not be resolved, concurrence of two of the three approaches was deemed sufficient.

The complete Assembly 21 has been submitted to the Candida Genome Database (http://www.candidagenome.org/), which will be in charge of its maintenance and curation.

2.1.4 RESULTS

Assembly 21. The completed Assembly 21 contains 15.845 Mb of DNA, organized into the 8 *Candida albicans* chromosomes. The assembly does not include the complete telomeric sequences for every chromosome, and includes only one copy of the normally repeated rDNA on Chromosome R. All but two chromosomes end with the subtelomeric repeat CARE-2 (Rel-2) (171, 286) at each telomere. This repeat was originally shown to be telomere-associated on chromosome 7 (55). Macro-restriction maps locate Major Repeat Sequences on all chromosomes but Chromosome 3, but since the MRSs are highly repeated and more than 16 kb in size, they are represented but not included in the assembly. On Chromosomes 7 and 6, where the MRSs have been sequenced, they are inserted into the sequence. The finished sequence of Chromosome 7 has been published elsewhere (57); this assembly includes a slightly revised version. In addition to the missing MRS regions there is one gap, on Chromosome 3. Assembly 21 is a haploid assembly; in cases where Assembly 19 detected heterozygosity, allelic contigs 19-1XXXX and 19.2XXXX were assembled. In regions where there were allelic contigs in Assembly 19, only the 19-1XXXX contigs were used to construct Assembly 21, so it provides no information about heterozygosity.

Chromosome size and structure. Figure 2.1.1 and Table 2.1.1 show structures and the sizes of the individual chromosomes in Assembly 21 and compares the size of the sequence with the size of each chromosome as determined by the optical map. The Assembly 21 size in Table 2.1.1 does not include the MRS if one or more are present, and for Chromosome R, only one copy of the rDNA repeat is included. The chromosomes range in size from 3,190,598 bases for Chromosome 1 to 943,480 bases for Chromosome

7. For Chromosome R, the rDNA repeat adds 350 kb to one homologue and 800 kb to the other. Where chromosome homologues are of different sizes, as with Chromosome R, the optical map software will choose one homologue. The optical map thus gives the size, including the rDNA, of the smaller homologue of Chromosome R. The larger homologue is very close to Chromosome 1 in size, about 3.1 Mb. The fact that the subtelomeric repeat CARE-2 is missing from one telomere may indicate that the sequence does not extend to the end of the chromosome. On Chromosomes 2 and 7 both the *TLO* gene and the CARE-2 sequences are missing. We were able to map 233,091 out of 250,884 (93%) of the original sequence traces on Assembly 21. The remaining 17,793 sequences probably represent some of these missing sequences as well as allelic variants but it was impossible to map them to the current assembly.

Assembly 21 demonstrates that the sub-telomeric repeats found on Chromosome 7 (57) are characteristic of all the chromosomes. The common factor is all or part of the repeat CARE-2 (171), which includes the LTR kappa (AF041469) and shares some sequence with Rel-2 (286). There are several other LTRs at the telomeres, and on Chromosome 1R there is an intact transposon pCa1 (AF007776). Although one telomere on each of Chromosomes 2 and 7 in Assembly 21 is missing CARE-2 (and the *TLO* gene, see below), the fosmid map shows these repeats on every telomere. Sequence near the telomere is hard to clone, and the most likely interpretation of this discrepancy is that the appropriate clones for these regions were underrepresented in the Stanford library. Since the repeats tend to be telomere-proximal to the *TLO* gene, the fact that *TLO* is missing on Chromosomes 2 and 7 also suggests that some sequence is missing. The detailed organization of the sub-telomeric repeats is complex and differs at each telomere.

Centromeres. Sequences which bind the Cse4p protein (the CENP-A orthologue) from *C. albicans* have been identified on each chromosome by Sanyal and Carbon, who suggested that they are at least part of the centromere (264). This interpretation is supported by the observation that the sequence identified on Chromosome 7 contributes to its mitotic stability. Table 2.1.1 gives the address (distance in bases from the left-hand end of the chromosome) of each centromere and shows that the chromosomes are generally metacentric, with the centromere sequences located near the center of the chromosomes. However, Chromosomes 2 and 6 are acrocentric. On Chromosome 2 the centromere is about 85% of the way toward the right telomere, and on 6 it is more than 95%.

ORF analysis. The ORF analysis of Assembly 21 differs slightly from that of Assembly 19. The human-curated annotation of Assembly 19 identified 6354 genes. The Candida Genome Database contains 12015 genes, including allelic variants. Assembly 21 contains 6075 genes. Of these, the identity and sequence of 6065 has not changed significantly from Assembly 19 (>98% sequence identity). Fourteen new ORFs have been added to Assembly 21. Thus 290 genes from the annotation of Assembly 19 are not in Assembly 21. Of these, 192 have been found to be identical to other Assembly 21 genes and have >90% of their sequence incorporated in these genes, while an additional 19 have >50% of their sequence incorporated into another ORF. The 79 remaining Assembly 19 ORFs have no strong Blast hits and include 55 hypotheticals, 15 gene family members, 7 that were truncated, 1 that was overlapping, and 2 putative sequencing errors. The Candida Genome Database provides an extensive analysis of the changes in ORF classifications between the two assemblies. http://www.candidagenome.org/). Some discrepancies

between their numbers and ours arise from the fact that they started from a greater number of orf19 genes.

Repeated DNA. *Candida albicans* was shown to have 8 chromosomes by pulse-field electrophoresis. Early studies demonstrated not only that the organism was diploid but that it contained several large blocks of repeated DNA, the Major Repeat Sequence (MRS), with a complete or partial copy on each of the 8 chromosomes (53, 58, 59, 140).

Analysis of the emerging sequence demonstrated the existence of a large number of LTRs and other repeated sequences related to transposons (103). Gene families also constitute a significant source of repeated DNA. Some of these repeated sequences, especially the MRSs, are too large to be crossed in a single sequencing run, so that assembly-using bioinformatics is blocked. The sequences of the MRSs on chromosomes should be considered unreliable due to the repeated nature of MRSs, which attract traces from many different other MRSs in the genome. The smaller repeated DNA regions are subject to the same potential problem and led to erroneous assembly in some of the Assembly 19 contigs. The fosmid map and the optical map were very helpful in identifying these errors and correcting them.

The *TLO* gene family. The putative transcription factor gene *CTA2* was identified by Kaiser, *et al* in a one-hybrid screen in *S. cerevisiae* (150). This sequence has homology to members of a gene family found at 14 of the 16 telomeres. Since *CTA2* is a gene name, we renamed the family *TLO* for TeLOmere-associated genes. On this assembly, they are numbered so that the left arm of the chromosome has an odd-numbered *TLO* gene and the right *TLO* gene has an even number, ascending from Chromosome R through Chromosome 7. Thus, Chromosome R has *TLO1* on the left and *TLO2* on the right.

Although the original member of this gene family was isolated because it has transactivating activity in *S. cerevisiae*, the function of these genes in *C. albicans* has not been determined. There are 15 members of this gene family in Assembly 21. Fourteen are located within 14 kb of an end of a chromosome contig, and all are oriented in a 5'-3' direction toward the centromere and away from the telomere (Figure 2.1.1). In addition, all but one (Chromosome 1R) have the LTR kappa 5' to the ORF, usually immediately adjacent but sometimes a few kb away. One member of the *TLO* gene family (*ORF19.2661*) is found in the interior of Chromosome 1, 1.29 Mb from the left end of this 3.19 Mb chromosome. Like the other family members, this ORF points toward the centromere and away from the telomere and has a kappa sequence from the subtelomeric repeat CARE-2 in its 5' region. In this case, the kappa sequence is the only part of the CARE-2 sequence present. Thus, this particular copy resembles a telomere but is located in the middle of Chromosome 1, and we have numbered it *TLO34*, since it is located between *TLO3* and *TLO4*. In order to keep the numbering system consistent, we have saved the names *TLO6* and *TLO15* for the genes on 2R and 7L that we expect will eventually be identified. The alignment of the *TLO* genes shows that *TLO13* (orf19.6337, Chromosome 6) has an intron while *TLO3, TLO4,* and *TLO15* are truncated at the 5' end (orf19.2661 and orf19.7276, both on Chromosome 1, and 19.7127, Chromosome 7). Each of these truncations contains 5' sequences with homology to the *ORF19.6337* intron, which suggests the presence of unidentified exons. This gene family can be clearly separated into two sub-groups based on their carboxy-terminal sequences. The existence of *TLO14*, on the right arm of Chromosome 6, has been confirmed by PCR but this gene has yet to be sequenced.

Other gene families. As noted by Braun and coworkers (33), *C. albicans* has a number of gene families. These range in size from 2 members to as many as 26 (the ABC transporter superfamily) (97). Several are clustered on one or two chromosomes. For example, Chromosome 6 contains members of 6 families, with 5 represented by more than one member, and one, ALS, with 5 members. Most of the families with more than two members have representatives on more than one chromosome. In addition, some two-member families are on different chromosomes. Examination of the genes surrounding family members on different chromosomes for the most part gives no hint as to the mechanism by which homologous sequences arrived at different locations.

The arrangement of these families on the chromosomes is not random. Figure 2.1.2A shows the ORFs from Assembly 21 aligned along the eight chromosomes. ORFs with 80% or more similarity are connected by lines. There are some areas that appear to be tandem duplications. An example from Chromosome 3 is shown in Figure 2.1.2B. The sequence containing the gene *DTD2* and those for two hypothetical proteins seems to have undergone a duplication accompanied by an inversion. A similar event seems to have occurred 40 kb away involving the genes *DAL5, ECM18,* and *FUR4*. There are no sequences such as LTRs near these duplications. Although the most plausible mechanism for the generation of gene families whose members are close together on one chromosome is tandem duplication, in many cases of gene families that contain tandemly repeated genes, the most closely related members are dispersed. For the SAP gene family, only *SAP5* and *SAP6*, 84 kb apart on Chromosome 6, are both nearest neighbors and most closely related in sequence. *SAP1* and *SAP4* are adjacent and each is most closely related to another member. For the LIP gene family, *LIP5* is most similar to *LIP9*, and they are

12 kb apart on Chromosome 7. Family members *LIP1, 6,* and *10* are adjacent on Chromosome 1 and each is most homologous to a distal gene (*LIP3* for *LIP1* and *LIP2* for *LIP6* and *LIP 10*). It is interesting to note from Figure 2.1.2A that the number of highly similar ORFs outside the *TLO* and MRS sequences (represented by the blue connecting lines) is relatively small compared to the number of gene families.

2.1.5 DISCUSSION

Assembly 19, the diploid assembly of the genome of *Candida albicans* strain SC5314 was a very important achievement. It provided a great deal of insight into many aspects of genomic organization, especially the large amount of heterozygosity (148). The subsequent annotation of the assembly by the community demonstrated a number of important properties of the genome, including the number of genes (6,354), the number with introns (224), the frequency and characteristics of short tandem repeats, and the characteristics of several multigene families. Braun, *et al.* also identified putative spurious genes and genes either on overlapping contigs or truncated by the end of contigs. However, they did not address chromosome location nor try to join the 266 haploid contigs of Assembly 19 into chromosome-sized assemblies. Thus, although these two projects brought the *C. albicans* genome to a very useful state, they still left it incomplete, lacking chromosome-size contigs and with some genes in an ambiguous state. Subsequently, Chibana et al (57) completed the sequence of Chromosome 7, identified 404 genes, and compared the synteny to the *S. cerevisiae* genome. They sequenced the Major Repeat Sequences and the gaps left in Assembly 19. They then aligned the sequence on the chromosome as determined by the physical map (55). Our

ultimate objective is to increase our understanding of the evolutionary mechanisms that affect the genome of this fungal pathogen.

There are chromosome size discrepancies between Assembly 21 and the optical map; these are attributable to several causes. Where the Assembly 21 size is smaller than the optical map size, the explanation may be the missing MRS, missing telomere-associated sequences, or size heterozygosity between the homologues. For example, we know that on Chromosome 5 the MRS is 50 kb in size (183), very close to the difference between the two estimates. Where the size determined by the optical map is smaller (Chromosomes 2, 3, and 4), the difference seems most likely to be heterozygosity for insertions of retrotransposon-related sequences. In these cases, the optical map of this chromosome is probably derived from the smaller homologue. For Chromosome 2, the discrepancy is rather large, given that this chromosome in Assembly 21 lacks the MRS and probably some telomere sequences. Interestingly, the size estimates in Jones *et al*, for the various chromosomes are remarkably close to the sizes determined by the optical map in Table 2.1.1 (148).

The amount of repeated DNA in *C. albicans* is significant. The MRSs were a major problem and their placement on the chromosomes required the physical and optical maps. Chromosomes 4 and 7 each have two MRSs forming an inverted repeat, and in principle the internal DNA fragment could invert via mitotic recombination. In strain SC5314 and its derivative, CAI-4, this inversion seems to occur very rarely, at least in the laboratory. In spite of the fact that most of the known translocations in *C. albicans* occur at the MRS, suggesting that this is a hot spot for recombination, there is no evidence on either chromosome for a flip of the bracketed sequence.

The specific sequences of 6 of the 9 MRSs are unavailable. This is only a problem if sequence variation in the MRS plays a biological role, and there is no evidence that it does. In addition to the MRSs and the subtelomeric repeats, there are more than 350 LTR sequences belonging to 34 different families scattered throughout the genome (103), and several of these are found clustered at telomeres. The subtelomeric repeat CARE-2 contains an LTR called kappa (102), which is found at the 5' end of each member of the *TLO* gene family. Whether this is related to the expansion of this family to the telomeres is not clear. The repeated DNA led to misassembly of some contigs in Assembly 19, including chimeras, artifactual duplications, and omitted sequence. The two physical maps and the *C. dubliniensis* sequence were essential in sorting out these artifacts.

The numerous gene families in *C. albicans* distinguish it from *S. cerevisiae.* A very common feature of these families is a clustering of members on a particular chromosome, which might reflect an ontology wherein a single copy undergoes tandem duplication and then sequences diverge as function diverges. There are several instances where similar but oppositely oriented gene clusters suggest that an inverted duplication of a region larger than a gene has occurred (Figure 2.1.2A).

The model for gene family ontology of duplication followed by dispersion would predict that, in general, similarity should be related to proximity. The arrangements of the two families we examined in detail, the *SAP* family and the *LIP* family, raise some questions about this model. In only two cases are the most similar family members the closest neighbors (*SAP6* and *SAP5*; *LIP9* and *LIP5*). However, the members clustered on one chromosome tend to be most closely related. For the *LIP* gene family, Hube and coworkers showed that *LIP5, 8,* and *9*, on Chromosome 7 form a group and *LIP1,2,3,6,*

and *10*, on Chromosome 1 are a related but distinct group (132). *LIP7*, on Chromosome R is an outlier, only distantly related, while *LIP4*, on Chromosome 6, fits with the Chromosome 7 group. For the *SAP* gene family, *SAP4, 5,* and *6* (Chromosome 6) form a highly related cluster, while the rest of the group, on Chromosomes R, 3, 4, and 6 form a loose association, with the highest similarity being between *SAP2* on Chromosome R and *SAP1*on Chromosome 6. These relationships suggest that the families originate on one chromosome and expand there, and when one member is duplicated on another chromosome, the pattern may or may not be repeated. The large number of gene families whose members are dispersed but not randomly would suggest that *C. albicans* is efficient at gene duplication at a distance. However, there are no hints of a specific mechanism in the sequence, such as homology between flanking sequences on different chromosomes or traces of mobile genetic elements. The relatively small number of highly similar ORFs suggests that the gene family members either diverged some time ago or are under strong selection to perform specific functions.

The *TLO* gene family is unique in *C. albicans* because it is found on every chromosome, and there are no closely adjacent members. This suggests that it arose by a mechanism different from, for example, the *LIP* family. One clue is that in all cases it is flanked on its 5' side by the LTR kappa. It seems possible that it has moved via genomic rearrangements caused by the transposon for which kappa is the LTR. On Chromosome 1, there is an interior *TLO* gene, as well as one near each telomere. A plausible explanation for this arrangement is that a chromosome translocation has occurred, with DNA being added to the end of a smaller precursor of Chromosome 1, followed by reconstitution of the telomere at the new end generated. There are only three genes in the

emerging *C. dubliniensis* sequence with similarity to the *TLO* family, and they are not located at the telomeres. On chromosomes 1 and R in *C. dubliniensis*, the genes adjacent to the *TLO* family member are present and are several kilobases from the end of the assembled sequence, suggesting that their absence is not due to missing telomere-proximal sequence. Since there are significant differences in virulence between *C. albicans* and *C. dubliniensis,* there may be a role for the *TLO* gene family in some aspect of pathogenesis. The function of the *TLO* genes is unknown. Although a member of this gene family was isolated as a potential trans-activating protein (and named *CTA2*), based on a one-hybrid screen in *S. cerevisiae*, there is no evidence beyond those experiments as to function (150).

Table 2.1.1. Chromosome Size and Features

Chromosome	*Size (bp)*		*Centromere location*	*MRS*	*Features*
	Assembly 21	Optical Map*			
R	2,294,279	2,709,974	1,748,965	1	Ribosomal DNA cluster is ~800 kb on one homologue, ~350 kb on the other. Lacks CARE-2 on the right telomere.
1	3,190,598	3,218,448	1,561,879	1	
2	2,233,511	2,228,646	1,924,678	1	The left-hand telomere lacks *TLO* and CARE-2
3	1,798,342	1,794,194	816,770	0	One gap
4	1,622,838	1,619,262	1,000,800	2	
5	1,191,532	1,246,010	465,800	1	
6	1,030,364	1,057,155	975,879	1	
7	943,480	961,199	423,765	2	The right-hand telomere lacks *TLO* and CARE-2

* The MRS is included in all the optical map sizes

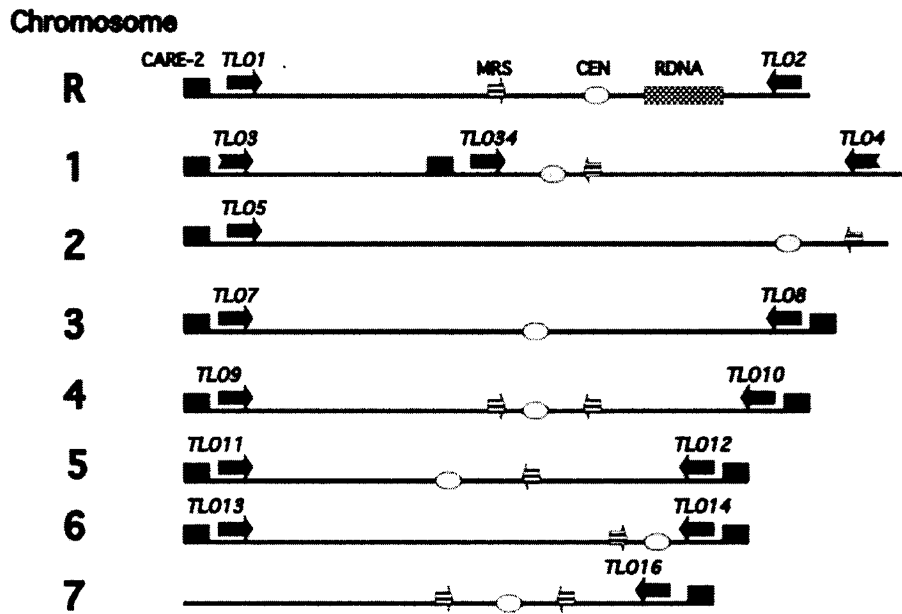


Figure 2.1.1. Schematic representation of the chromosomes in Assembly 20.

The major structural features, including the centromeres, the MRS sequences, the CARE-2 sequences, and the TLO genes are represented. The positions are approximately to scale. The smaller homologue of Chromosome R (containing a 350 kb rDNA repeat) is shown.

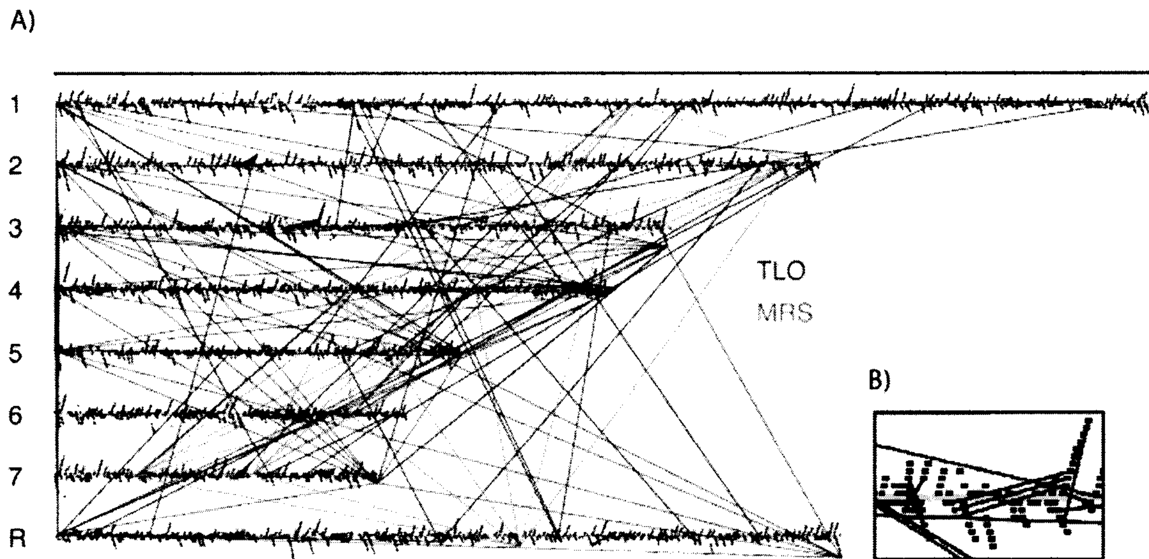


Figure 2.1.2. Highly related ORFs in the *C. albicans* genome.

A. The eight chromosomes are shown with the ORFs indicated by the small black lines. The red lines connect the TLO genes, while the green lines connect the ORFs in the MRS sequence. The blue lines connect ORFs with relatedness greater than 80%.

B. Close-up of Chromosome 3 from base 622,329 to base 785,321 showing two areas which appear to have undergone duplication followed by inversion.

Section 2.2: The Annotation of the *Candida albicans* Genome.

From Assembly 19 we identified 11,615 putative open reading frames (ORF), which encode proteins grater than 150 amino acids, as well as smaller genes with good possibility to encode functional polypeptides (33). We used a special program (Artemis annotation software (258)) and a custom database to "walk" along the individual contigs. We validated and corrected the coordinates and descriptions of each putative ORF while removing small and overlapping ORFs with no sequence similarity in other organisms to yield 6,354 *C. albicans* genes.

A Human-Curated Annotation of the *Candida albicans* Genome.

Burkhard R. Braun[1], Marco van het Hoog[2], Christophe d'Enfert[3], Mikhail Martchenko[2], Jan Dungan[4], Alan Kuo[4], Diane O. Inglis[1], M. Andrew Uhl[1], Hervé Hogues[2], Matthew Berriman[5], Michael Lorenz[6], Anastasia Levitin[2], Ursula Oberholzer[2], Catherine Bachewich[2], Doreen Harcus[2], Anne Marcil[2], Daniel Dignard[2], Tatiana Iouk[2], Rosa Zito[2], Lionel Frangeul[7], Fredj Tekaia[8], Kim Rutherford[5], Edwin Wang[2], Carol A. Munro[9], Steve Bates[9], Neil A. Gow[9], Lois L. Hoyer[10], Gerwald Köhler[4], Joachim Morschhäuser[11], George Newport[4], Sadri Znaidi[12], Martine Raymond[12], Bernard Turcotte[13], Gavin Sherlock[14], Maria Costanzo[14], Jan Ihmels[15], Judith Berman[16], Dominique Sanglard[17], Nina Agabian[4], Aaron P. Mitchell[18], Alexander D. Johnson[1], Malcolm Whiteway[2], André Nantel[2*]

1- Department of Microbiology and Immunology, University of California-San Francisco, 600 16th Street, Suite N372, San Francisco, CA 94143, USA.

2- Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount, Montreal, Quebec, H4P 2R2, Canada.

3- Unité Postulante Biologie et Pathogénicité Fongiques, INRA USC 2019, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France.

4- Department of Stomatology, University of California-San Francisco, San Francisco, CA 94143, USA.

5- The Sanger Centre, Wellcome Trust Genome Campus, Hinxton Cambridge, CB10 1SA, United Kingdom.

6- Department of Microbiology and Molecular Genetics, UT-Houston Medical School, 6431 Fannin Street, Houston, TX 77030, USA.

7- Plate-forme Intégration et Analyse Génomique, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France.

8- Unité de Génétique Moléculaire des Levures, CNRS URA 2171, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France.

9- School of Medical Sciences, University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, UK.

10- Department of Veterinary Pathobiology, University of Illinois at Urbana-Champaign, Urbana, IL 61802, USA.

11- Institut fur Molekulare Infektionsbiologie, Universitat Wurzburg, Rontgenring 11, Wurzburg, Germany.

12- Institut de Recherches Cliniques de Montreal, Montreal, Quebec, H2W 1R7, Canada.

13- Department of Medicine, Royal Victoria Hospital, McGill University, 687 Pine Ave. West Montreal, Quebec, H3A 1A1, Canada.

14- Department of Genetics, Stanford University School of Medicine, Palo Alto, CA 94305-5120, USA.

15- Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.

16- Department of Genetics, Cell Biology & Development, University of Minnesota, 6-160 Jackson Hall, 321 Church St. SE, Minneapolis, MN 55455, USA.

17- Institute of Microbiology, University Hospital Lausanne (CHUV), Rue de Bugnon 44 CH-1011, Lausanne, Switzerland.

18- Department of Microbiology and Institute of Cancer Research, Columbia University, New York, NY 10032, USA.

PLoS Genetics 2005 Jul; 1(1):36-57 © 2005 Public Library of Science.

2.2.1 ABSTRACT

Recent sequencing and assembly of the genome for the fungal pathogen *Candida albicans* used simple automated procedures for the identification of putative genes. We have reviewed the entire assembly, both by hand and with additional bioinformatic resources, to accurately map and describe 6,354 genes and to identify 246 genes whose original database entries contained sequencing errors (or possibly mutations) that affect their reading frame. Comparison with other fungal genomes permitted the identification of numerous fungus-specific genes that might be targeted for antifungal therapy. We also observed that, compared to other fungi, the protein-coding sequences in the *C. albicans* genome are especially rich in short sequence repeats. Finally, our improved annotation permitted a detailed analysis of several multigene families while comparative genomic studies show that *C. albicans* has a far greater catabolic range, encoding respiratory Complex 1, several novel oxidoreductases and ketone body degrading enzymes, malonyl-CoA and enoyl-CoA carriers, several novel amino acid degrading enzymes, a variety of secreted catabolic lipases and proteases, as well as numerous transporters to assimilate the resulting nutrients. The results of these efforts will thus ensure that the *Candida* research community has uniform and comprehensive genomic information for medical research as well as for future diagnostic and therapeutic applications.

2.2.2 INTRODUCTION

Candida albicans is a commonly encountered fungal pathogen responsible for infections generally classed as either superficial (thrush and vaginitis) or systemic (such as life-threatening blood-borne candidiasis) (91, 231). Its life cycle has fascinating aspects that have generated great excitement over the last decade, with an influx of workers and new molecular techniques brought to bear on long-standing problems (18). Topics of particular interest are the organism's capacity to shift into several different phenotypic states, each with distinct roles in infection, and its recently discovered capacity to mate, providing at least part of a sexual cycle to what was previously believed to be a strictly clonal diploid population. Other special adaptations for infection include a battery of externally displayed proteins and secreted digestive enzymes, and complex interactions with the host immune system that normally keep *C. albicans* at bay as a minor part of the mucosal flora.

Here, we report a detailed annotation of the genome sequence of this organism, bringing the previously available raw sequence to a new level of stability and usability. The genome of *C. albicans* has previously been shotgun sequenced to a level of 10.9-fold coverage (148). However the assembly of this sequence faced special difficulties because the organism is a diploid with little or no gene exchange in the wild. Thus homologous chromosomes show substantial divergence, and many genes are present as two distinctive alleles. This required that the assembly process be aware of the diploid status and be prepared to segregate reads into two alleles for any section of the genome. At the same time, the genome is rich in recently diverged gene families that are easily confused with alleles. This task was further complicated by the absence of a complete physical map of

the *C. albicans* genome. Nevertheless, this arduous assembly process resulted in a data set (assembly 19, with 266 primary contigs over 8 chromosomes) that has already yielded a number of significant advances including the production of DNA microarrays (224), libraries of systematic gene knockouts (254), large scale transposon mutagenesis (300) and the ability of many individual researchers to identify novel genes using bioinformatic tools (299). Unfortunately, due to the mostly computational methods used in its development, the current genome assembly still contains a significant number of predicted genes that are fragmented, overlapping, or otherwise erroneous. As a consequence, different groups have been using different methods for the identification and classification of *C. albicans* genes, which has hindered communication and complicated comparisons between large-scale data sets.

Following the publication of these early functional genomics studies, it was realized that the needs of the *C. albicans* research community would be better served by a unified gene nomenclature. The results of this community-based effort were initially based on the version 19 computational assembly and preliminary annotation produced independently by various research groups. We used visual inspection of 11,615 putative coding sequences and various bioinformatic tools to refine the quality and description of each open reading frame (ORF).

In the end, we provide unique identifiers, coordinates, names and descriptions for 6,354 genes. With the exception of certain large gene families, we have not annotated the portion of the Assembly 19 DNA that was set-aside as secondary alleles, instead concentrating on the primary sequence that forms one haploid genome equivalent. Investigation of the identity and relative divergence of all alleles will be an important

further project for the *C. albicans* genome, as will finishing and linking the small number of gaps that remain in the primary sequence. In addition, we describe a variety of gene families and we discuss insights into virulence. Finally, we used comparative genomics to point out a variety of additional insights that are illuminated by the high-quality annotation provided here. This project serves as a model for community-based annotation that could be applied by other research communities that wish to improve on automated sequencing pipeline output that may be available for their organisms of interest.

2.2.3 MATERIALS AND METHODS

Identification of *C. albicans* ORFs and merging of preliminary annotations

Nucleotide sequence data for Assembly 19 were retrieved from the Stanford Genome Technology Center (SGTC) web site (http://www-sequence.stanford.edu//group/candida/). Assembly 19 is composed of a haploid supercontig set (contigs 19-831 to 19-10262), here referred to as the haploid set, and a allelic supercontig set (contigs 19-20001 to 19-20161), here referred to as the allelic set (148).

The 11615 orf19 ORFs identified in the haploid set were compared by reciprocal BLASTP to the 9168 ORFs identified by the SGTC using Assembly 6 of the *C. albicans* genome sequence (designated orf6.n). A similar reciprocal comparison was run using the set of 6165 *C. albicans* proteins available in the CandidaDB database that have been defined by applying a procedure similar to that outlined above on Assembly 6 and through a manual curation aiming to reach a non-redundant protein set (http://genolist.pasteur.fr/CandidaDB; (66)). Furthermore, orf19 ORFs were reciprocally

compared to the *S. cerevisiae* proteome using data available at the Saccharomyces Genome Database (60). All data were parsed using *readblast* (284) and a matrix was generated that correlated ORFs from each data set.

We used the genome annotation tool Artemis (258), which provides very detailed annotation capability, visually mapping desired features onto the target sequence. All ORFs were assigned a color code in order to facilitate annotation using Artemis. All orf19 corresponding to an Integrated Processing Facility (IPF) classified as FALSORF and orf19 identified at the SGTC and not found among IPFs were color-coded in grey. orf19 with an unambiguous allele (90% identical amino acids over the whole length of the longest ORF) were color-coded in red (> 150 codons) or pink (< 150 codons). orf19 with a questionable allele (90% identical amino acids over the whole length of the shortest ORF) were color-coded in green (> 150 codons) or pale green (< 150 codons). orf19 without a clear allele (less than 90% identical amino acids over the whole length of the shortest ORF or no reciprocal match) were color-coded in blue (> 150 codons) or light blue (< 150 codons).

From this "master" file of sequences and their corresponding preliminary annotation, groups of contigs were selected and saved as partially annotated subsequences that were reserved and retrieved by members of the annotation consortium, and once fully annotated, were returned to a central web site. A version of Artemis was distributed to the consortium that included a modified "options" file (20) allowing project-specific qualifiers to be used, and also featuring the *C. albicans*-specific translation table.

Whole genome BLAST searches and visualization of sequence homologies

ORF sequences were translated to proteins using the translation table for *C. albicans* (232), and compared, using the BLASTP algorithm (5) with the NR database, the *C. albicans* proteome itself, the putative proteomes of five fungi; *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Neurospora* crassa, *Aspergillus nidulans, Magnaporthe grisea* and the proteomes of five other eukaryotes; *Arabidopsis thaliana, Drosophila melanogaster, Caenorhadbitis elegans, Mus musculus* and *Homo sapiens*. Sequence data was obtained from the EMBL-EBI Integr8 Browser [http://www.ebi.ac.uk/integr8/EBI-Integr8-HomePage.do] and the Broad Institute for Genome Research [http://www.broad.mit.edu/annotation/]. For the most similar proteins by BLAST, negative exponents of the expect (E) values were parsed from the output files, collected in a relational database and visualized as a color range associated with each reference protein (Figure 2.2.1). For this purpose, a web-accessible visualization tool was constructed using Macromedia Cold Fusion and an Apache2 web server. These results can be consulted at [http://candida.bri.nrc.ca/candida/index.cfm?page=blast].

2.2.4 RESULTS AND DISCUSSION

The annotation process

Compilation of *Candida* annotation data

As detailed in the Methods section, we used assembly version 19 of the *C. albicans* genome (148) to identify 11,615 putative ORFs. These include genes encoding proteins greater than 150 amino acids (a.a.) as well as smaller genes of between 150-447 b.p. that have a coding function > 0.5 as determined with a GeneMark matrix (197).

These ORFs were then compared to the set of 7680 *C. albicans* ORFs defined by the Stanford Genome Technology Center (SGTC), thus permitting their classification using the same systematic identifiers of the format orf19.n (148). The 3936 novel ORFs without an orf19.n counterpart were assigned a new reference number of the format orf19.n.i where orf19.n is the 5-prime closest (contig-wise) ORF defined by the SGTC and i is an integer that varies between 1 and the number of novel ORFs found in the orf19.n - orf19.n+1 interval. To simplify correlation with previously published data that use the orf6.# or earlier nomenclatures, we have produced a web-accessible translation tool [http://candida.bri.nrc.ca].

Positional information for each ORF was merged with data from a variety of different sources, including the SGTC (http://www-sequence.stanford.edu/group/candida/index.html), CandidaDB (http://genolist.pasteur.fr/CandidaDB; (66)), the Agabian lab annotation (http://agabian.ucsf.edu/canoDB/anno.php) and annotation data from the Johnson/Fink laboratories (15) that had been updated with a Magpie annotation (224). This large dataset was then reformatted into EMBL-style files thus allowing for input in the Artemis annotation software (258). Volunteer annotators accessed a custom-made database to reserve and download EMBL files containing sequence and annotation data for each of the 266 DNA sequence contigs. To help in validating various and sometimes conflicting sources of information, translated protein sequences from putative *C. albicans* ORFs were compared to putative protein sequences extracted from five fungal genomes; *Saccharomyces cerevisiae* (99, 153), *Schizosaccharomyces pombe* (309), *Neurospora crassa* (94), *Aspergillus nidulans* (The Aspergillus database, http://www-

genome.wi.mit.edu/annotation/fungi/aspergillus/) and *Magnaporthe grisea* (The Magnaporthe database, http://www-genome.wi.mit.edu/annotation/fungi/magnaporthe/) as well as to the genomes of five other eukaryotes; *Arabidopsis thaliana* (The *Arabidopsis* Sequencing Consortium, 2000), *Drosophila melanogaster* (2), *Caenorhadbitis elegans* (The *C. elegans* Sequencing Consortium, 1998), *Mus musculus* (302), *Homo sapiens* (167), and the GenBank non-redundant (NR) protein database. Comparisons against the translated *C. albicans* genome were also performed to help identify overlapping genes and putative gene families. To help interpret such a large number of sequence comparisons, we organized sequence similarity data in a web-accessible database using a novel visualization concept whereby we used a colorimetric display to indicate BLAST similarity, which was easy and rapid to scan visually (Figure 2.2.1). The annotators could thus rapidly determine which genes are potentially unique to *C. albicans* (e.g., orf19.4741 and orf19.4786), those that are members of gene families (e.g., orf19.4736 and orf19.4779), genes that only have homologs in fungal genomes (e.g., orf19.4756 and orf19.4778) or those with homologs in all eukaryotic genomes (e.g., orf19.4732 and orf19.4784). Finally, a strong hit against the complete NR database, but not in the other genomes (orf19.4772 and orf19.4800), allowed us to identify *C. albicans* genes that had already been described and submitted to the sequence databases prior to the publication of assembly orf19. Clicking on the relevant boxes opened an additional window containing the precompiled sequence alignments, thus permitting the validation of interesting observations. These visualization tools and the results of sequence comparisons are available at [http://candida.bri.nrc.ca/candida/index.cfm?page=blast].

The coordinates and annotations for all 11,615 putative ORFs were thus verified, corrected and (if necessary) rewritten by the annotators. We removed ORFs smaller than 300 b.p. with no significant sequence similarity to other genes, either within the *C. albicans* genome or in the sequence databases. In cases where two ORFs overlapped by more than 50%, the smallest gene was removed unless it showed even a slight sequence similarity to another gene in the sequence databases. In other cases, we encountered two, or more, contiguous ORFs that obviously were part of the same gene. These interruptions were usually due to unidentified introns or presumed sequencing errors. In these cases, we decided to merge the relevant gene fragments into a single entry. A total of 5262 ORF entries were thus removed from the database, or merged with neighboring ORFs, leaving 6354 confirmed genes.

A nomenclature for *C. albicans* genes

Following consultations with the *C. albicans* research community during the 5th and 6th ASM Conferences on Candida and Candidasis, it was agreed that *C. albicans* gene names should follow the format established for *S. cerevisiae* (51). Gene names consist of three letters (the gene symbol) followed by an integer (e.g., *ADE12*); the gene symbol should be an acronym for, or relate to, the gene function, gene product, or mutant phenotype. It is preferable that a given gene symbol have only one meaning, so that all genes using that symbol are related in some way, for instance by sharing a function, participating in a pathway, or belonging to a gene family. In addition, gene symbols that are used in *S. cerevisiae* gene names should retain the same meaning when used for *C. albicans* genes. The prefix 'Ca' has sometimes been used on gene names to denote that a

gene is derived from *C. albicans*; however, while the use of prefixes adds clarity to discussions of genes from different species that share a name (e.g., comparing Ca*URA3* to Sc*URA3*), the prefix is not considered part of the gene name proper. For more details on genetic nomenclature, see the Candida Genome Database (CGD; (6)) website [http://www.candidagenome.org/Nomenclature.html].

Wherever possible, genes that are orthologous between *C. albicans* and *S. cerevisiae* should share the same name. We have provided 3409 suggested names (in the /SuggGene field of the EMBL files) for many *C. albicans* open reading frames based on their orthology to *S. cerevisiae* genes; these are not yet considered the standard *C. albicans* gene names, but rather provide guidance for investigators wishing to name these genes. CGD assigns standard names to *C. albicans* genes for which there are published data (the /PubGene field). The annotation contains 355 such entries. Generally CGD considers the first published name in the correct format to be the standard name; common usage and uniqueness are also considered. All names that have been used for a gene are collected in CGD, regardless of their format, so that information from the literature can be traced to the correct gene. In the current annotation, additional published gene names have been placed in the /Synonym field. The complete annotation dataset, results of BLAST sequence similarity searches and the identification of conserved protein domains can be obtained from our web page at [http://candida.bri.nrc.ca/candida/index.cfm?page=CaGeneSearch]. Furthermore, the Candida Genome Database (CGD; www.candidagenome.org), funded by the National Institute for Dental and Craniofacial Research (NIDCR) at the NIH, will curate the

scientific literature and provide tools for accessing and analyzing the *C. albicans* genome sequence.

Content and general statistics

We identified 6,354 genes in version 19 of the *C. albicans* genome assembly. This number is certain to change slightly with time as more data come to light. For instance, 80 of these genes are probably duplicates, having almost identical counterparts near the extremities of sequence contigs. Novel genes may also lie in unsequenced/unassembled gaps between the DNA sequence contigs. We identified 246 genes containing mutations or sequencing errors that result in a frameshift, or the insertion of a stop codon, that will have to be confirmed through resequencing. In the meantime, these elements have been joined as a single ORF entry and tagged with the entry "sequencing error?" inside their /Note field. We have also identified 190 genes truncated at the ends of contigs, only 35 of which have an identical counterpart on a potentially overlapping contig. New information will be continuously integrated into the community data as it is submitted.

The mean protein coding length of 1439 b.p. (480 a.a.) is almost identical to what has been observed in *S. cerevisiae* and *S. pombe* while the gene density stands at one gene per 2342 b.p. Short descriptions for all gene products were provided by annotators, usually based on sequence similarity. 1218 (19.2%) genes encode unique proteins with no significant homologs in the sequence databases, an almost identical percentage with that observed in the current version of the *S. cerevisiae* annotation (153). An additional 819 (12.9%) gene products exhibited significant similarities to other proteins of unknown

function. Furthermore, we have provided Enzyme Commission numbers and GO terms to 1334 and 3586 gene products respectively.

Analysis of protein domains

Compared to the *S. pombe* and *S. cerevisiae* proteomes, the *C. albicans* proteome shows a slight increase in the abundance of leucine-rich repeats (IPR001611), some zinc finger transcription factors (IPR001138), esterases/lipases (IPR000379) and trans-membrane transporters for polyamines (IPR002293) and amino acids (IPR004841). If the analysis is expanded to the other fungal proteomes, only the increased abundance in leucine-rich repeats appears to be unique to *C. albicans.*

Genome-based identification of antifungal targets

One of the main arguments supporting large-scale sequencing projects for fungal pathogens is the hope of finding novel antifungal targets, particularly those that are absent from the genome of their host. 228 *C. albicans* genes have a very strong sequence homolog (Based of a top hit BLASTP E-value < 1e-45) in all of 5 fungal genomes but no significant sequence similarity (best BLASTP E-value > 1e-10) to genes in the genomes of either humans or mice. For example, these include *FKS1* which encodes a 1,3-beta-glucan synthase that is the target for the cell wall-agents echinocandins (28). The list includes 46 gene products that are assumed to be located on the plasma membrane, 71 that are predicted to be involved in the transport of small molecules and 21 that appear to be involved, directly or indirectly, with cell wall synthesis. Furthermore, 41 gene products have been associated with an E.C. number, indicating an enzymatic activity,

with phospholipases being the most abundant. The roles and sites of action of these gene products suggest that they would be both more accessible, and theoretically amenable to inhibition by small molecules.

Multigene families

Many putative and demonstrated virulence factors of *C. albicans* are members of large multigene families. Well-known examples of such families encode secreted aspartyl proteinases (SAPs; (203), (304)), agglutinins (ALSs; (129)), secreted lipases (LIPs; (132)), high-affinity iron transporters (FETs; (243)), and ferric reductases (FREs/CFLs; (166)). Members of each of these families are differentially expressed as a function of the yeast-hyphae transition, phenotypic switching, or timing during experimental infection. Also, each of these families is large relative to the corresponding homolog or family of homologs in *S. cerevisiae*, leading to the concept that expansion of many *C. albicans* gene families may be an adaptation to a commensal lifestyle and may be, in part, responsible for *C. albicans*' unusual ability to occupy a variety of host niches.

A striking difference between *C. albicans* and *S. cerevisiae* is the manner in which they acquire nutrients from the environment. In addition to the well-described SAPs, LIPs, and FETs, *C. albicans* possesses expanded families of acid sphingomyelinases (with 4 genes per haploid genome), phospholipases B (6 genes), oligopeptide transporters (7 genes), and amino-acid permeases (23-24 genes). Another striking difference is the emphasis by *C. albicans* on respiratory catabolism, as reflected in expanded families of peroxisomal enzymes. These include families of acyl-CoA

oxidases (3 genes), 3-ketoacyl-CoA thiolases (4 genes), acyl-CoA thioesterases (3-4 genes), fatty acid-CoA synthases (5 genes), and glutathione peroxidases (4 genes).

Additional families that may pertain to colonization or pathogenesis include those encoding the estrogen-binding protein *OYE1* (7 genes), the fluconazole-resistance transporter *FLU1* (13 genes), and the vacuolar protein *PEP3/VPS16* (4 genes), whose *Aspergillus* homolog is required for nuclear migration and polarized growth.

Zinc cluster transcription factors

Proteins of the zinc finger superfamily represent one of the largest classes of DNA-binding proteins in eukaryotes. Several different classes of zinc finger domains exist which differ in the arrangement of their zinc-binding residues (213). One of these domains, which appears to be restricted to fungi, consists of the $Zn(II)_2Cys_6$ binuclear cluster motif in which six cysteines coordinate two zinc atoms (265, 288). *S. cerevisiae* possesses fifty-four zinc cluster factors defined by the presence of the zinc cluster signature motif $CX_2CX_6CX_{5\text{-}16}CX_2CX_{6\text{-}8}C$ which is generally located at the N-terminus of the protein. These proteins function as transcriptional regulators involved in various cellular processes including primary and secondary metabolism (e.g. Gal4p, Ppr1p, Hap1p, Cha4p, Leu3p, Lys14p, Cat8p), pleiotropic drug resistance (e.g. Pdr1p, Pdr3p, Yrr1p) and meiosis (Ume6p) (3, 288). Quite often, they bind as homo- or heterodimers to two CGG triplets organized as direct, indirect or inverted repeats and separated by sequences of variable length (118, 288). A large proportion of these factors (50%) also contain a middle homology region (MHR; Fungal_trans in PFAM) located in the central

portion of the protein that has been proposed to participate in DNA-binding and to assist in DNA target discrimination (265).

Analysis of the *C. albicans* proteome using a combination of sequence analyses tools (SMART, PFAM, PHI-BLAST) allowed us to identify 77 binuclear cluster proteins. These factors are characterized by the presence of the zinc cluster signature motif $CX_2CX_6CX_{5-24}CX_2CX_{6-9}C$ generally located at the N-terminus of the protein (72 out of 77) and with spacing between cysteines 3 to 4 and 5 to 6 slightly different from the *S. cerevisiae* motif. To our knowledge, only six of the *C. albicans* zinc cluster genes have been characterized in detail, including *SUC1* involved in sucrose utilization (154), *FCR1* implicated in pleiotropic drug resistance (281), *CWT1* required for cell wall integrity (218) and *CZF1*, *FGR17* and *FGR27* involved in filamentous growth (35, 300). The functions of many uncharacterized *C. albicans* zinc cluster factors (approximately 20%) can be inferred from the fact that they display high levels of sequence similarity (Top BLASTP E value $\leq$ 1e-20) with the products of *S. cerevisiae* genes with a known function. In the case of *GAL4*, however, the *C. albicans* homologous ORF identified (orf19.5338) encodes a significantly smaller protein (261 a. a.) than *S. cerevisiae* Gal4p (881 a. a.), lacking the C-terminal two thirds of the protein that contains one of two transcriptional activating domains, and must therefore have somewhat different functions. Finally, it is noteworthy that many of the zinc cluster factors known to be involved in pleiotropic drug resistance in *S. cerevisiae*, such as Pdr1p, Pdr3p, Yrr1p, Yrm1p, Rds1p, and Rdr1p, do not appear to possess close structural homologs in *C. albicans*. Since pleiotropic drug resistance is frequently observed in *C. albicans*, it is likely that this

organism possesses functional homologs of these genes or other novel processes which remain to be identified.

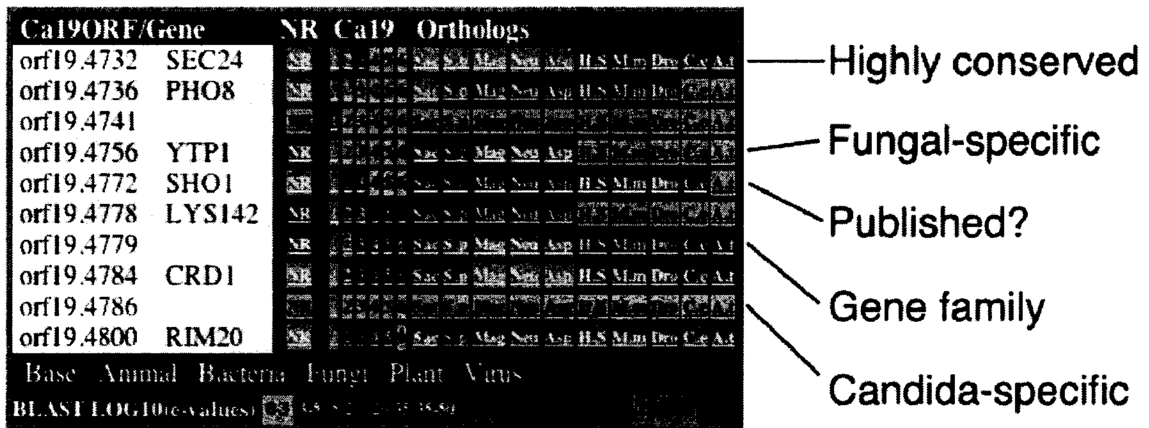


Figure 2.2.1. Visualization of proteins sequence similarities

Sample from a web page used by annotators of the *C. albicans* genome to visualize the significance of the best hit from whole proteome BLASTP searches. Each putative ORF was compared to the non redundant database (NR), the *Candida* ORF list itself (Ca19: showing results from the 4 top hits) and amino acid sequences from the proteomes of *Saccharomyces cerevisiae* (Sac), *Schizosaccharomyces pombe* (S.p), *Magnaporthe grisea* (Mag), *Neurospora crassa* (Neu), *Homo sapiens* (H.S), *Mus musculus* (M.m), *Drosophila melanogaster* (Dro), *Caenorhadbitis elegans* (C.e) and *Arabidopsis thaliana* (A.t). The BLASTP E-value from the top hit was converted to a color scale as indicated. Examples of *C albicans* genes with interesting similarity patterns are indicated.

CHAPTER 3. POSTGENOMIC STUDIES OF *CANDIDA ALBICANS*: INTRAGENIC COMPARATIVE GENOMIC STUDIES OF *C. ALBICANS*.

We compared the translated amino acid sequences to the proteomes of five fungi (*S. cerevisiae, Schizosaccharomyces pombe, Neurospora crassa, Aspegilus nidulans*, and *Magnaporthe grisea*) and five higher eukaryotes (*Caenorhabditis elegans, Drosophila melanogaster, Arabidopsis thaliana, Mus musculus,* and *Homo sapiens*) (33). Comparative genomic studies have provided a better understanding of the *C. albicans* – specific biological processes: we observed that close to 20% of *C. albicans* genes are specific to that organism. In addition, 228 genes showed very high sequence similarity among six fungal genomes but are completely absent from mammalian genomes. The products of those genes present attractive antifungal drug targets whose inhibition would unlikely provoke secondary effects in the host. These include trans-membrane transporters and enzymes. The *C. albicans* genome is also rich in gene families that encode products with a role in pathogenesis. These include secreted lipases (132) and aspartyl proteases (222), the ALS family of cell surface agglutinins (129), high affinity iron transporters and ferric reductases (243), ABC transporters that can pump out drugs (261, 262), and superoxide dismutases (134, 162, 211, 249).

We have characterized the function and the transcriptional regulation of one member of the copper-zinc superoxide dismutase, *SOD5*, which is needed for the pathogenicity of this fungus. The expression of this gene was induced during the yeast to hyphal transition of *C. albicans*, and when the fungus was challenged with osmotic and oxidative stresses. *SOD5* transcription was also induced when *C. albicans* cells were

grown on nonfermentable carbon sources. Transcription factors Rim101p and Efg1p were found to regulate the transcription of the *SOD5* gene. To characterize the function of *SOD5* we created a *SOD5* deletion strain, which showed sensitivity to hydrogen peroxide, and established that the *SOD5* product was necessary for the virulence of *C. albicans* in a mouse model of infection. While we observed that a *sod5* mutant strain is resistant to macrophage attack, Fradin *et al.* showed a reduced survival of a *sod5* strain when mixed with neutrophils (90). Finally, the Sod5 protein was shown to localize on the cell surface of *C. albicans* through glycosylphosphatidylinositol (GPI) anchorage (73), which suggests that its role is to neutralize neutrophil-generated superoxide radicals at the surface of the fungal cells, while the roles of CuZnSod1p and MnSod2p are to inactivate cytosolic and mitochondrial superoxide radicals (134, 249). We propose that *C. albicans* has its duplicated *CuZnSOD*s downstream of different promoters to allow response to different conditions that demand antioxidant enzymes. Our results show that under certain hyphal-inducing conditions *SOD5* is the major player among *CuZnSOD*s; a recent study shows *SOD4* to be regulated in White-Opaque phenotype switching, which is involved in the mating processes of *C. albicans* (165).

Section 3.1 is an article that describes the characterization of *SOD5*, and section 3.2 is an article that describes activation domains of *C. albicans* transcription factors.

Section 3.1: Superoxide dismutases in *C. albicans*.

CuZn superoxide dismutases in *Candida albicans*: Transcriptional regulation and functional characterization of the hyphal induced SOD5 gene.

Mikhail Martchenko[1], Anne-Marie Alarco[2], Doreen Harcus[2], and Malcolm Whiteway[1,2*].

[1]Department of Biology, McGill University, Stewart Biology Building, 1205 Avenue Docteur Penfield, Montreal, Quebec, Canada H3A 1B1;

[2]Genetics Group, Biotechnology Research Institute, National Research Council, Montreal, Quebec, Canada H4P 2R2

Molecular Biology of the Cell 2004 Feb;15(2):456-67 © 2004 The American Society for Cell Biology.

3.1.1 ABSTRACT

Superoxide dismutases (SOD) convert superoxide radicals into less damaging hydrogen peroxide. The opportunistic human pathogen, *Candida albicans,* is known to express CuZnSOD (*SOD1*) and MnSOD (*SOD3*) in the cytosol and MnSOD (*SOD2*) in the mitochondria. We identified three additional CuZn containing superoxide dismutases, *SOD4, SOD5* and *SOD6* within the sequence of the *C. albicans* genome. The transcription of *SOD5* was upregulated during the yeast to hyphal transition of *C.*

albicans, and *SOD5* was induced when *C. albicans* cells were challenged with osmotic or with oxidative stresses. *SOD5* transcription was also increased when cells were grown on non-fermentable substrates as the only carbon source. The Rim101p transcription factor was required for all inductions observed, while the Efg1p transcription factor was specifically needed for serum-modulated expression. Deletion of *SOD5* produced a viable mutant strain that showed sensitivity to hydrogen peroxide when cells were grown in nutrient limited conditions. Sod5p was found to be necessary for the virulence of *C. albicans* in a mouse model of infection. However, the *sod5* mutant strain showed the same resistance to macrophage attack as its parental strain, suggesting that the loss of virulence in not due to an increased sensitivity to macrophage attack.

3.1.2 INTRODUCTION

Aerobic eukaryotic pathogens can encounter superoxide radicals (O_2-) generated from several sources. These sources can be internal or external. An important internal source is the mitochondrial respiratory chain (29, 46, 181), and thus the rate of respiration can have a significant impact on reactive oxygen species (ROS) production. A key external source of ROS encountered by pathogens is from phagocytes. The superoxide radical is the first intermediate in the oxidative burst generated in the phagosome, and this burst is thought to be involved in pathogen killing (247). The superoxide radicals are known to inactivate [4Fe-4S] cluster-containing enzymes by oxidizing one iron and releasing it from the cluster (93, 186). Free iron can react with hydrogen peroxide to generate toxic hydroxyl radicals (OH-) by Fenton chemistry (92, 215). The hydroxyl and superoxide radicals react with cellular components resulting in oxidation of proteins and

nucleic acids as well as lipid peroxidation. These effects can lead to inactivation of enzymes, to disruption of membranes, to mutations, and to cell death (111, 112).

In order to reduce the harmful effects of superoxide radicals, cells express detoxifying enzymes. Superoxide dismutase (SOD) is an antioxidant enzyme involved in elimination of superoxide anions; it catalyzes the reaction: O_2- + O_2- + $2H^+ \rightarrow H_2O_2$ + O_2. Normally, hydrogen peroxide (H_2O_2) is still toxic to the cell, therefore another enzyme, catalase, converts it to water. Superoxide dismutases can be classified according to metal cofactor(s) bound to them: there are iron (FeSOD), manganese (MnSOD), copper and zinc (CuZnSOD) and nickel (NiSOD) versions of the enzymes. The CuZnSOD family is thought to have evolved separately, and thus these enzymes share no sequence similarity with other SOD families (151, 275).

An important aerobic eukaryotic pathogen, *Candida albicans*, causes the majority of human fungal infections. These infections range from thrush in immunocompetent colonized hosts, to life-threatening systemic infections in immunocompromised individuals such as patients with cancer (64). *C. albicans* can grow in several morphological forms: it can proliferate as a budding yeast and it can also form filaments, such as true hyphae and pseudohyphae. Hyphal formation of this fungus can be induced by high temperature, basic pH, and by the presence of serum, and this morphological transition is implicated in pathogenesis of *C. albicans* (35).

Typically, eukaryotes express cytosolic CuZnSOD and mitochondrial MnSOD (93). *C. albicans* possesses both cytosolic CuZnSOD and mitochondrial MnSOD encoded by the *SOD1* and *SOD2* genes respectively (134, 249). In addition to these proteins, a *C. albicans* cytoplasmic MnSOD (the *SOD3* gene product) was recently identified (162).

SOD1 was shown to form homodimers (162); while *SOD2* and *SOD3* were shown to form homotetramers (162, 249). *C. albicans* Sod1p was shown to protect cells against superoxide radicals produced by macrophages, and it was shown to be important for the virulence of *C. albicans* in a mouse model (135). *C. albicans* catalase was also shown to be involved in virulence of this pathogen. *C. albicans* Sod2p was shown to have no effect on the virulence aspect of this fungus, but rather protects the cell against intracellularly-produced superoxides (133). *C. albicans* appears to coordinate the production of CuZnSOD and MnSOD in a reciprocal manner: CuZnSOD shows the highest activity in the mid-exponential phase, while MnSOD has the highest activity upon the entry to stationary phase (162, 249). The morphology of *C. albicans* is coupled to CuZnSOD production since cells generate increased amounts of CuZnSOD during the switch from yeast to hyphal forms (109). In particular, conditions of elevated production of CuZnSOD coincide with increased levels of ROS produced by *C. albicans* as cells produce the highest levels of ROS, including superoxide, during hyphal growth and in exponential phase (266). In addition, *C. albicans* generates higher amounts of ROS when grown in the presence of salt (260, 266).

We have identified three new potential *C. albicans* superoxide dismutase genes. We studied the expression, regulation and function of one of these genes, *SOD5*, and show that this gene is under a complex regulation of expression. Functional tests show no involvement of *SOD5* in defense against macrophage-induced killing. However, the effects of the deletion of *SOD5* on virulence may suggest the involvement of Sod5p in removal of superoxides produced by the host.

3.1.3 MATERIALS AND METHODS

***C. albicans* and Bacterial Strains.** The *C. albicans* strains are listed in Table 1; plasmids and oligonucleotides used in this study are listed in Tables 2 and 3. The *C. albicans* strains SC5314 (88), HLC52 (190), JKC19 (188), DAY25 and DAY185 (69), BWP17 (305), and GKO6 (71) were used for transcription analysis; RM1000 (226) was used as a parental strain to generate the RM1000-HU, H10, UH16, UH16-1, UH16-2, and UH16-3 strains. The *Escherichia coli* strain DH5α was used for all plasmid constructions and maintenance. The plasmid pVEC (205) was used for the generation of plasmids p37, p37H, p37U, and p*SOD5*-110; plasmids p5921 (88) and pRM100 (239) were used for the amplification of *hisG-CaURA3-hisG* and *CaHIS1* markers respectively.

Media and Culture Growth Conditions. *C. albicans* strains were routinely cultured on YPD (1% yeast extract, 2% peptone, 2% dextrose, pH6) medium at 30°C. Cells with integrated selection markers or disrupted genes were cultured in synthetic glucose medium containing 0.67% yeast nitrogen base, 0.15% amino acids, 2% dextrose and appropriate supplements. All media were supplemented with uridine (25 μg/ml) for the growth of *C. albicans* Ura auxotrophic strains.

Since under certain conditions *C. albicans* cells form hyphae at 37°C, it is difficult to accurately measure the growth of cells in the culture media. We therefore prepared standard growth-curves in YPD media at 37°C. We observed that it takes 300 minutes for SC5314 strain to reach an OD_{600} of 1.0 at 30°C, and it takes 225 minutes for SC5314 strain to reach an OD_{600} of 1.0 at 37°C. It takes 325 minutes for RM1000-HU and UH16-2 strains to reach an OD_{600} of 1.0 at 30°C, and it takes 250 minutes for

RM1000-HU and UH16-2 strains to reach an OD_{600} of 1.0 at 37°C. These standard curves were used to approximate the time when the conditions should have changed.

For the Northern analyses performed with the yeast and mycelial forms of *C. albicans*, blastospores were first grown overnight at 30°C in the following media: mammalian cell culture media IMDM (pH6.5) (Gibco-Invitrogen), YPD, YP, or YP supplemented with either 2% galactose (YPGal), 3% glycerol (YPGlyc) or 1.5% methanol (YPMeth). The precultures were then diluted to an OD_{600} of 0.1 and incubated in the same media at either 30°C until the OD_{600} reached 1.0, or at 37°C until the OD_{600} reached 2.5. In other cases, the preculture grown in YPD medium was diluted to an OD_{600} of 0.1 and then incubated until the OD_{600} reached 1.0 at either 30°C or 37°C in the following media: YPD, YPD supplemented with 10% serum, YPD adjusted to pH8, YPD with 0.3M NaCl, and YPD with 0.3M NaCl at pH8.

For some samples, the preculture grown in YPD medium was diluted to an OD_{600} of 0.1 and then incubated at either 30 or 37°C in YPD. In those cultures the pH, the NaCl concentration or both of those factors were increased 10, 30, or 60 minutes before the OD_{600} reached 1.0. In those conditions the pH of the media was increased to 8.0 with 10N NaOH; the salt concentration was increased to 0.3M by adding an appropriate amount of 5M NaCl solution into the media.

In other cases, the preculture grown in YPD medium was diluted to an OD_{600} of 0.1. The culture was then supplemented with reactive oxygen species-generators one hour before the OD_{600} reached 1.0 at 30°C, or one hour before the OD_{600} reached 2.5 at 37°C. In this case cells were grown in YPD with 0.4mM H_2O_2, YPD with 0.5mM Menadione (Sigma), or YPD with 0.5mM Riboflavin (Sigma) (illuminated with 100W table-lamp).

All liquid cultures were rotated at 250 rotations per minute. Spot-assays were done on media solidified with 2% agar. 5'-Fluoroorotic acid (5-FOA) (Diagnostic Chemicals Limited) containing medium was prepared as described before (25), except that uridine was substituted for uracil. *C. albicans* cells were transformed by the previously described method (48).

RNA Preparation and Northern Blot Analyses. Total RNA was extracted by the hot phenol extraction method (44). The Northern blot analyses were performed by the previously described method (4) with the following modifications. RNA samples (80μg) were electrophoresed on a 7.5% formaldehyde, 1% agarose gel and transferred by capillarity to a Zeta-Probe nylon membrane (Bio-Rad, Mississauga, Ontario). Detection of specific RNAs was performed by hybridization at 65°C in 0.5 M NaPO4, pH7.2, 1 mM EDTA, 7% SDS, 1% bovine serum albumin with 50ng of 32P-labeled DNA probes (ssDNA was omitted from this solution). Probes were generated by using the random prime labeling system (Amersham Pharmacia Biotech) with the following modifications: 50ng (rather than 25ng) of DNA was added to the reaction, the reaction was incubated at 37°C for 45 minutes (rather than for 10 minutes), and all of the labeled probe (rather than 1.25ng) was used for hybridization. The unincorporated labeled nucleotides were removed by G-50 micro columns (Amersham Pharrmacia Biotech). Membranes were washed once at 65°C with a solution containing 40 mM NaPO4, pH 7.2, 5% SDS, 1 mM EDTA, 0.5% bovine serum albumin for 10 minutes (rather than twice for 20 minutes), and once with a solution containing 40 mM NaPO4 pH 7.2, 1% SDS, and 1 mM EDTA for 20 minutes (rather than four times for 20 minutes) and were visualized using a PhosphorImager (Molecular Dynamics, Model # 4255).

All Northern probes were made by PCR: the 0.24kb long *SOD5* probe used primers OMM8 and OMM9, the 0.24kb long *SOD4* probe used primers OMM10 and OMM11, the 0.21kb long *SOD1* probe used primers OMM14 and OMM15, and the 0.5kb long *SOD6* probe used primers OMM16 and OMM17. A *CaACT1* probe was generated with UO106 and UO107 oligos and was used as an internal control to monitor RNA loading and transfer (a gift from Dr. Ursula Oberholzer). The Dot-Blot was performed the same way as the Northern, with the difference that 80μg of *SOD1*, *SOD4*, *SOD5*, and *SOD6* unlabeled probes (rather than 80μg of RNA) were spotted on the membranes.

Deletion and Reversion of *SOD5*. In order to delete the *SOD5* gene, the *SOD5* ORF and its 1kb flanking regions were PCR-amplified from genomic DNA of the CAI4 strain with oligos ODH157 and ODH158 that contained *Eco*RI sites. The PCR product (2.6kb) and plasmid pVEC were digested with *Eco*RI, and a 4.7kb *Eco*RI fragment of pVEC was gel-purified and ligated with the *Eco*RI digested 2.6kb *SOD5* PCR product. Clone p37 contained an insert with the *SOD5* ORF running in the opposite orientation with respect to that of the Amp marker. The *SOD5* ORF was deleted from p37 by the use of inverse PCR using either oligonucleotides OMM1 and OMM2 or OMM1 and OMM3. The *hisG-URA3-hisG* marker was cut from p5921 with *Bgl*II and *Bam*HI; OMM1 and OMM2 create *Bam*HI restriction sites in p37 into which the *hisG-URA3-hisG* marker was inserted to yield p37U. The *HIS1* marker was PCR amplified from pRM100 with OMM4 (containing an *Xba*I site) and CB20R (containing a *Bam*HI site, a gift from Dr. Catherine Bachewich); OMM1 and OMM3 created *Bam*HI and *Xba*I restriction sites respectively in p37 into which the *HIS1* marker was ligated to yield p37H. The *HIS1* and *SOD5*-1kb-flanking regions were PCR amplified from p37H with oligos ODH157 and ODH158, and

this PCR product was transformed into RM1000 to yield the H10 strain, in which the *sod5::HIS1* fragment replaces the chromosomal *SOD5* ORF. The *hisG-URA3-hisG* and *SOD5*-1kb-flanking regions were cut out of p37U with *Xmn*I and *Spe*I, and this fragment was transformed into H10 to yield the UH16 strain, in which the *sod5::hisG-URA3-hisG* fragment replaces the second chromosomal *SOD5* ORF. Strain UH16 was plated on 5-FOA-containing plates to select for Ura-auxotrophs. As a result, the UH16-1 strain was created in which one *SOD5* ORF was replaced with *HIS1* marker, and the second *SOD5* ORF was replaced with *hisG*. In order to create the prototrophic strain UH16-2, UH16-1 was transformed with the p37U construct, which was targeted to the chromosomal *SOD5*-promoter region by cutting the construct in the *SOD5*-promoter region with *NsiI*. The *HIS1 URA3* – auxotrophic wild type stain was transformed with the p37U and p37H constructs in a sequential manner; these were targeted to the chromosomal *SOD5*-promoter region by cutting the constructs in the *SOD5*-promoter region with *NsiI*, creating strain RM1000-HU. For *SOD5* reversion studies, the *SOD5* ORF and its 1kb flanking regions were PCR-amplified from CAI4 genomic DNA with the OMM18 and OMM19 oligos that contained *Bam*HI and *Xba*I sites respectively. The resulting PCR product (2.6kb) and plasmid pVEC were double-digested with *Bam*HI and *Xba*I restriction enzymes and ligated together to yield p*SOD5*-110. p*SOD5*-110 was transformed into UH16-1 to yield revertant strain UH16-3.

Genomic DNA was prepared from all of the obtained strains, digested with *Nhe*I and *Xho*I, and analyzed for the correct integration events by Southern blotting with a digoxigenin (DIG) system (Boehringer Mannheim). The DIG probe was generated by PCR with ODH157 and OMM1 oligos that amplified the 1kb *SOD5* promoter sequence

from p37. To confirm the absence of the *SOD5* ORF, genomic DNA from UH16, UH16-1, and UH16-2 strains were cut with *Dra*I, and analyzed by Southern blotting with a probe that binds inside the *SOD5* ORF. This probe was also used for Northern analysis to confirm the absence of *SOD5* mRNA in the *SOD5* inducing conditions (see "experimental" section for results).

Protein preparation and In-gel SOD activity assay. Protein preparation was done as previously described (139). The protein concentration was determined by the Bradford method using an assay kit (BIO-RAD), as described by the provider. SOD activity in polyacrylamide gels was detected using nitroblue tetrazolium (NBT) (13), with the modification that 5mg of cell extracts were electrophoresed on 10% non-denaturing polyacrylamide gels in triplicates. After electrophoresis, one gel was used for Coomassie blue staining (207). The other two gels were used for determining SOD activity with or without 4mM KCN (139).

Spot Assay and Growth Assay. Overnight cultures of *C. albicans* cells were concentrated to OD_{600} of 20.0. Serial twofold dilutions were made and 5ul of cells were spotted onto a plate and incubated for 48 hours at either 30°C or 37°C. Spot assays were prepared on the types of media described in the results section. The growth assay of cells in liquid media was made by diluting overnight cells in those media to OD_{600} of 0.1. The growth rate was monitored at OD_{600} for 24 hours at 30 and 37°C.

Macrophage and Virulence assays and fungal burden. The end point dilution survival assay was performed as described previously (208). Dilutions of *C. albicans* without macrophages where colonies could be distinctively visualized were counted toward the

lower limit and compared to the same dilutions with macrophages. Colonies from a total of at least 72 wells per condition were used to provide data.

For virulence assays, *C. albicans* strains were passaged overnight at 30°C in YPD prior to infection. Cells were harvested, washed in cold phosphate-buffered saline (PBS) (Wisent inc., St.-Bruno, Canada), and counted using a hemacytometer. Cells were then resuspended at a density of 5.75×10^6 cells / ml. B6 mice (8 weeks old, females) (Charles River Laboratories, Wilmington, MA, USA) were infected through the lateral tail vein with 200 μl (1.15×10^6 organisms). Mouse survival was monitored at least three times daily, and moribund mice were euthanized by CO_2 gas. We used *C. albicans* strain RM1000-HU to infect 13 mice, UH16-2 and UH16-3 to infect 12 mice. *C. albicans* colony forming units (cfu) in kidney were counted from live animals sacrificed at day five after the infection. The kidneys were removed aseptically from four mice per strain; the organs were weighed, homogenized, diluted in sterile saline and plated on YPD plates. Colonies were counted after incubation of the plates at 30°C for 48 hours and the results were expressed as cfu per gram of infected organ.

3.1.4 RESULTS

Identification of *SOD4*, *SOD5*, and *SOD6*

The mammalian pathogen, *Candida albicans*, is known to possess both a copper-zinc dependent superoxide dismutase (CuZnSOD encoded by *SOD1*) (134) and a manganese dependent superoxide dismutase (MnSOD, encoded by *SOD3*) (162) in the cytosol, and to encode a second MnSOD (*SOD2*) that is localized to mitochondria (249). The *C. albicans* genome sequence revealed three additional open reading frames with

high sequence homology to the copper-zinc family of superoxide dismutases: orf 19.2062, orf 19.2060 and orf 19.2108 in version 19 of Stanford's *Candida albicans* genome sequence assembly (Stanford Genome Technology Center, http://www-sequence.stanford.edu/group/candida/). We have designated these new genes as *SOD4*, *SOD5* and *SOD6* respectively. All three of these genes share sequence homology to the previously described *SOD1* and to each other, but share no homology to the manganese-containing superoxide dismutases (Figure 3.1.1A). All three predicted proteins (Sod4p, Sod5p and Sod6p) contain characteristic residues involved in copper and zinc binding, and amino acids involved in dimer contact (86) (Figure 3.1.1B). Northern analysis (described below) showed that the mRNAs for each of the three genes were approximately 700 nucleotides in length. However, the predicted length of Sod6p of 316 amino acids was too large to be easily accommodated within this size of message. This discrepancy could be resolved if *SOD6* contained an intron. Nucleotide sequence analysis shows that *SOD6* contains a well-characterized 3' splicing site "CTAGG" motif starting at nucleotide 752, a signal for lariat formation "AACTGAC" starting at nucleotide 723, and a possible 5' splicing site "GTCGGT" starting at nucleotide 462, which closely resembles the previously reported "GTANGT" motif (9). Splicing of this potential intron would generate a predicted ORF of 218 amino acids, close in size to the predicted 232 amino acids and 228 amino acids of Sod4p and Sod5p. All these proteins are likely to be cytosolic, since they lack mitochondria-targeting sequences. Intriguingly, *SOD4* and *SOD5* are close together in the genome, with *SOD4* located 2kb upstream of *SOD5* (224).

SOD5 is transcriptionally induced during yeast to hyphal transitions

Hyphal formation in *C. albicans* can be induced by incubation of the cells in the presence of serum and high temperature. Recent transcriptional profiling experiments established that the *SOD5* gene was induced during the yeast to hyphal switch (224). In order to study the expression pattern of the copper-zinc SOD genes, we generated probes to unique sequences within each gene (Figure 3.1.1B). Dot-blot analysis confirmed the specificity of each probe; there was an insignificant cross-hybridization between the *SOD4* and *SOD5* probes, and no cross-hybridization for any of the other combinations (our unpublished results). These probes were used to investigate transcriptional regulation of the *SOD5* gene under a variety of conditions. Northern blot analysis of the Actin (*ACT1*) mRNA in *C. albicans* strains was used as an internal loading control, since the transcription of this gene was shown to be unaffected under most of the growth conditions tested in our study (224). In agreement with microarray data, Northern analysis demonstrated that the *SOD5* transcription increases at 37°C in the presence of serum, compared to the control condition of no serum and 30°C (Figure 3.1.2A, lanes 1 and 4). We investigated whether serum, temperature or both of these factors regulated the expression of *SOD5*. In the absence of serum, a temperature increase from 30°C to 37°C reduced the expression of *SOD5* from a relative intensity of 1.0 to 0.3 (Figure 3.1.2A, lanes 1 and 2). In contrast, the presence of serum caused high levels of *SOD5* expression (Figure 3.1.2A, compare lane 1 with lanes 3 and 4). This suggests that serum, rather than temperature, controls the induction of *SOD5* expression.

The mammalian cell culture medium IMDM lacks serum, but allows *C. albicans* to grow exclusively in a hyphal form at both 30°C and 37°C. *SOD5* was highly expressed

in this medium at both temperatures (Figure 3.1.2B). As in (162), we observed the accumulation of a new, 3.4kb mRNA species hybridizing to the *SOD5* probe (lane 3), which does not correspond to the specific *SOD1*, *SOD4* or *SOD6* mRNAs (our unpublished results). The nature of this mRNA is unclear, but since the distance between the start codon of *SOD5* and the stop codon of *SOD4* on the chromosome is 3.3kb it is possible that the *SOD5* probe hybridized to a mRNA species that contains both *SOD4* and *SOD5*. This idea is supported by the detection of the same message by a *SOD4* probe (our unpublished results).

Another way to induce the yeast-to-hyphal transition is to increase the pH and the temperature of the medium. At both temperatures, 30°C and 37°C, an increase of pH from 6.0 to 8.0 increases the expression of *SOD5* (Figure 3.1.3a, lanes 6-9 and 18-21). At both temperatures the rise in *SOD5* mRNA expression persists up to 300 minutes for 30°C or 225 minutes for 37°C of growth in the media with an increased pH. The induction peaks at 30 minutes for cells grown at 30°C, and at 10 minutes for cells grown at 37°C.

Osmotic and oxidative stresses affect *SOD5* transcription

C. albicans cells grown in the presence of salt were shown to generate higher amounts of ROS (260, 266). Intriguingly, previous transcription profiling established that osmotic stress also induced the expression of *SOD5* (82). We confirmed these results by Northern analysis, as cells grown in the presence of high salt concentrations showed an increased transcription of *SOD5* at 30°C (Figure 3.1.3a). The levels of *SOD5* mRNA increase as early as 10 and 30 minutes after the addition of salt to the media at both 30°C

and 37°C, but stay up after 60 and 300 minutes only at 30°C (Figure 3.1.3a, lanes 1-5 and 13-17).

When pathogens are engulfed by neutrophils there is an increase in both pH and salt concentration inside the phagosome (247). Intriguingly, high salt and high pH have a cumulative effect on *SOD5* transcription at 30°C; at 37°C high salt and high pH yield to a higher *SOD5* response compared to the levels of *SOD5* transcripts induced separately by high pH and high salt (Figure 3.1.3a, lanes 10-13 and 22-23). The fact that high salt and high pH together lead to increased *SOD5* expression was expected for a model that the expression increase permits the mobilization of superoxide dismutase activity to defend against phagosome mediated killing.

In some organisms such as *S. cerevisiae*, the transcription of the gene encoding the cytoplasmic superoxide dismutase is induced by treatment with various oxidative agents like hydrogen peroxide and menadione (162, 174, 311). We investigated the transcription profile of the *C. albicans* CuZn superoxide dismutases in cells treated either with hydrogen peroxide or superoxide radicals. We used two types of superoxide generators: riboflavin, which generates superoxides outside of the cell (139), and menadione, which creates superoxides inside the cell (174). The presence of hydrogen peroxide caused increased levels of *SOD1* expression (Figure 3.1.3b, compare lane 1 with lanes 3 and 4) and moderate increases in transcription of *SOD5*. Menadione leads to moderate increases in transcription of *SOD1* at both temperatures and that of *SOD5* (Figure 3.1.3b, lanes 5 and 6). Riboflavin treatment results in elevated levels of *SOD1* at 37°C and that of *SOD5* at both temperatures (Figure 3.1.3b, lanes 7 and 8). The *SOD1* induction was weak, as previously described (162). The fact that the transcription of

SOD5 was increased upon the induction of oxidative stress is also expected if the gene is involved in defense against a phagosome-generated oxidative burst. Since *SOD5* transcription was induced by both intracellular and extracellular superoxide generators, it provides us with the first evidence that Sod5p is needed for elimination of superoxides whose source is from within the cell.

Non-fermentable carbon sources affect *SOD5* expression

Since we saw that the intracellular superoxide generator, menadione, induced *SOD5* levels, we further investigated the effect of intracellularly-generated superoxide on *SOD5* transcription. Studies of isolated mitochondria showed that up to 80% of the superoxide in eukaryotic cells was produced in the mitochondria by electron leakage at the QH2:cytochrome *c* segment (complex III) during respiration (47). Therefore, the effect of respiration on *SOD5* transcription was investigated. One of the ways to increase respiration is to grow the cells in media with a non-fermentable carbon source, such as glycerol or methanol, while providing glucose or galactose as the sole carbon source allows cells to grow by fermentation (72, 193).

At 30°C we grew cells until they reached an OD_{600} of 1.0. We observed that the growth on glucose or galactose as the only carbon source leads to relatively low levels of *SOD5* transcription (Figure 3.1.4, lanes 1 and 3). When cells were grown in the presence of glycerol or methanol as the only carbon source at 30°C *SOD5* expression was moderately increased (Figure 3.1.4, lanes 5 and 7), compared to fermentable carbon sources at the same temperature (compare with lanes 1 and 3). At 30°C, the highest

SOD5 transcript levels were detected in cells grown in the absence of any carbon source (YP) (Figure 3.1.4, lane 9).

At 37°C we grew cells until they reached an OD_{600} of 2.5. We observed that the growth on glucose or galactose as the only carbon source leads to very low levels of *SOD5* transcription (Figure 3.1.4, lanes 2 and 4). When cells were grown in the presence of glycerol or methanol as the only carbon source at 37°C *SOD5* expression was highly increased (Figure 3.1.4, lanes 6 and 8), compared to fermentable carbon sources at the same temperature (compare with lanes 2 and 4). At 37°C, the highest *SOD5* transcript levels were detected in cells grown in the absence of any carbon source (YP) (Figure 3.1.4, lane 10). These results show that growth on non-fermentable carbon sources increase *SOD5* transcription, which might suggest that one of the roles of Sod5p is to eliminate intracellularly generated superoxide radicals. *SOD1* and *SOD4* were also up-regulated by increased temperature when grown on non-fermentable carbon sources, and *SOD1*, *SOD4*, and *SOD6* were also up-regulated in the absence of carbon sources (YP) (our unpublished results).

Transcription Factors involved in the regulation of *SOD5* expression

Previous transcription profiling established that the non-hyphal *efg1 cph1* double mutant failed to induce *SOD5* expression under hyphal inducing conditions (224). We used Northern analysis to determine the relative impact of each transcription factor on the expression of the *SOD5* gene. For each of the individual null mutants we attempted to induce *SOD5* transcription by one of the following conditions: serum and high temperature, osmotic stress (0.3 mM NaCl) and basic pH (pH8). In all of these conditions

the *cph1* mutant showed the same pattern of *SOD5* expression as the wild type strain: *SOD5* was induced in all three conditions as compared to the non-induced condition (YPD 30°C) (Figure 3.1.5, lanes 1-8). On the other hand, the *efg1* mutant failed to increase the transcription of *SOD5* specifically under serum inducing conditions (Figure 3.1.5, lanes 9-12, compare lanes 2 and 10). In addition, under non-inducing conditions the *efg1* mutant showed generally elevated levels of *SOD5* message (Figure 3.1.5, compare lanes 1 and 9). Interestingly, the addition of salt to the media does not increase the *SOD5* mRNA levels, compared to *SOD5* levels during the growth of *efg1* mutant strain under non-inducing conditions (Figure 3.1.5, lanes 9 and 11). The *SOD5* mRNA levels were increased in the *efg1* mutant in alkaline-inducing conditions, although not to the same extent as the corresponding increase in the wild type strain (Figure 3.1.5, lane 12). To investigate the factors responsible for the pH-mediated induction of *SOD5* expression we looked at cells defective in either the Rim101p transcription factor or in Mds3p; these factors represent independent branches of the pH response (71). In contrast to both the *efg1* and *cph1* mutants, the *rim101* mutant blocked *SOD5* induction under all of the conditions tested compared to *SOD5* expression in its parental strain (Figure 3.1.5, lanes 13-20). Just as in the wild type and the *cph1* mutant, the *mds3* mutant showed that *SOD5* was induced in all three conditions (YPD 30°C) (Figure 3.1.5, lanes 21-28). Therefore, the Cph1p transcription factor and Mds3p have essentially no effect on *SOD5* transcription under the conditions tested, Efg1p is required for *SOD5* transcription in the presence of serum and perhaps for suppressing the *SOD5* expression under the non-inducing condition, and Rim101p is important for the normal *SOD5* transcription under all of the conditions tested.

SOD5 is not essential for viability in _C. albicans_

To study the function of Sod5p in *C. albicans*, we created a strain in which both copies of *SOD5* were deleted. We generated a strain UH16 with the *sod5::HIS1 sod5::hisG-URA3-hisG* constructs replacing both copies of the *SOD5* gene (see "Materials and Methods" for details). The correct position of insertions during these steps was analyzed by Southern blotting (Figure 3.1.6a). The absence of the *SOD5* gene was also confirmed with a Southern blot probed with a *SOD5* internal fragment probe. This probe binds inside the *SOD5* open reading frame exclusively, and it fails to generate a *SOD5*-corresponding band in the *SOD5* double-mutant (our unpublished results). When the DNA was restricted with *Dra*I, the *SOD5* probe hybridized to an unexpected 7kb-DNA fragment, which could be derived from the accumulation of a 7kb DNA fragment seen with ethidium bromide staining. The nature of this piece of DNA is unclear, but it might be derived from the multiple repeat sequence whose *DraI* restriction digest yields a 7kb DNA fragment that has an affinity to the *SOD5* probe (our unpublished results). In addition, the absence of *SOD5* mRNA under either *SOD5* inducing conditions (the presence of either serum or alkaline conditions or high pH) or non-inducing conditions (YPD) was confirmed by Northern blotting using the same internal probe (Figure 3.1.6b). No increase in *SOD1*, *SOD4*, and *SOD6* transcript levels was observed in the *sod5* mutant strain (UH16-2) (our unpublished results). Finally, we performed an in-gel activity assay that detects the activity of both types of superoxide dismutases, CuZnSOD and MnSOD. RM1000-HU cells grown in IMDM media at 37°C express both MnSOD (Figure 3.1.6c, lane 1, upper band) and high amounts of CuZnSOD (Figure 3.1.6c, lane 1, lower band).

On the other hand, UH16-2 cells showed reduced levels of the CuZnSOD activity (Figure 3.1.6D, lane 2, lower band is missing) under the conditions tested. The CuZnSOD activity is restored in the UH16-3 revertant strain (Figure 3.1.6c, lane 3, lower band). The *rim101* mutant strain also showed reduced levels of the CuZn superoxide dismutase activity band (Figure 3.1.6c, lane 4). This observation is consistent with the results in Figure 3.1.5, and shows that the *RIM101* mutant cells possess low levels of both *SOD5* transcripts and Sod5p enzyme activity. The activity of the CuZnSOD can be inhibited by cyanide (KCN), which blocked the CuZnSOD activity of RM1000-HU and UH16-3 cells grown in IMDM medium at 37°C (Figure 3.1.6c, lanes 5 and 6, lower bands are missing). This shows that in IMDM media *SOD5* seems to be a major contributor of overall CuZnSOD activity in *C. albicans*. The ability to generate the viable double-mutant strain shows that *SOD5* is not an essential gene in *C. albicans*.

Hydrogen Peroxide together with limited nutrients impairs growth of the *SOD5* mutant strain

We examined the growth properties of the *sod5* mutant strain under conditions that were shown to increase the transcription of the *SOD5* gene. The spot assay showed that there is no growth difference between the UH16-2 strain and the wild type strain in medium that contained serum, NaCl, pH8, NaCl and pH8. The *sod5* mutant strain showed no heat sensitivity at 42°C, and no growth defect on Lee and Spider medium (our unpublished results). In addition, the *sod5* mutant strain showed no defect with respect to hyphal development triggered by serum, IMDM media, or high pH (pH 8.0) at 37°C as shown by spot assays and growth in liquid media (our unpublished results).

In *S. cerevisiae* both of the single genes encoding CuZnSOD and MnSOD were found to be upregulated during the metabolic transition from fermentation to respiration (209), and the *S. cerevisiae* CuZnSOD mutant possesses a reduced ability to grow by respiration in glycerol rich medium (72, 193). We therefore investigated the ability of the *sod5* mutant strain UH16-2 to grow on media rich with non-fermentable carbons. We found that there was no growth difference in the liquid or solid medium that contained either glucose, galactose, glycerol or methanol as a carbon source; in the absence of any of above-mentioned carbon source molecules the *SOD5* mutant grew as fast as its parental strain (our unpublished results). CuZnSOD deficient *S. cerevisiae* require lysine and methionine for aerobic growth (22), whereas superoxide dismutase defective *E. coli* require branched chain amino acids (138). However, the *C. albicans SOD5* mutant showed neither lysine, methionine, nor branched amino acids auxotrophies (our unpublished results).

The *E. coli* and *Helicobacter pylori* SOD mutants were found to be more sensitive to the effect of hydrogen peroxide than its parental isogenic strain (42, 155, 270). We compared the growth rate of the *SOD5* mutant with that of its parental isogenic strain. We found no growth difference between the *SOD5* mutant and the wild type strains grown on nutrient-rich media (YPD) supplemented with 1.0mM hydrogen peroxide (Figure 3.1.7). However, the *SOD5* mutant (UH16-2) grew poorly on SC (synthetic complete) media supplemented with 1.0mM hydrogen peroxide, compared to its parental strain (Figure 3.1.7). The growth of UH16-3 strain showed an intermediate growth on SC media supplemented with hydrogen peroxide, consistent with the reintroduction of *SOD5* in one copy into the UH16-1 background. The *SOD5*-phenotype suppression is observed in

UH16-2 cells after incubation in SC media, which contains hydrogen peroxide and slight excess of fermentable glucose (2.5% glucose, rather than 2.0% of glucose) or in YPD media, which contains hydrogen peroxide (Figure 3.1.7).

SOD5 **is involved in virulence of** ***C. albicans*** **but is not involved in defense of** ***C. albicans*** **against the fungicidal activity of macrophages**

The sensitivity to hydrogen peroxide is often linked to a decrease in virulence. To test directly the *SOD5* involvement in macrophage-defense the UH216-2 strain was incubated with RAW 264.7 mouse macrophage cells and the percent survival (28.9±5.7%) was not different from the survival of its parental RM1000-HU strain (27.6±7.0%) and from the revertant UH16-3 strain (28.9±7.0%). These results suggest that Sod5p is not required for *C. albicans* survival when cells are subjected to the fungicidal activity of macrophages.

To test directly the *SOD5* involvement in an organism-defense, we tested the virulence of the *sod5* mutant strain in a mouse systemic infection model by intravenous tail infection. The results showed that 90% of the mice infected with the wild type (RM1000-HU) died within 8 days of being inoculated. However, the homozygous *sod5* mutant strain (UH16-2) had markedly diminished virulence compared to the wild type parental strain (RM1000-HU). Mice infected with UH16-3 strain showed an intermediate survival rate consistent with the reintroduction of one copy of *SOD5* gene into the UH16-1 background.

The colony forming units (cfu) in mice kidneys were determined on the fifth day after infection. One hundred percent of the kidneys recovered from mice infected with the

wild type parental strain (RM1000-HU) revealed an average of 1.0 x 10^6 cfu g^{-1} kidney at day of sacrifice. Seventy five percent of the kidneys recovered from mice infected with the *sod5* revertant strain (UH16-3) revealed an average of 7.0 x 10^5 cfu g^{-1} kidney at the day of sacrifice, and 2.0 x 10^4 cfu g^{-1} kidney in the remaining twenty five percent. Seventy five percent of the kidneys recovered from mice infected with the *sod5* strain (UH16-2) revealed an average of 5.0 x 10^4 cfu g^{-1} kidney on the day of sacrifice, and 5.0 x 10^5 cfu g^{-1} kidney of the remaining twenty five percent, indicating an effective clearance of the *sod5* knock out strain. These results demonstrate that a functional Sod5p plays an important role in the virulence of *C. albicans*.

3.1.5 DISCUSSION

C. albicans is known to express a CuZnSOD (*SOD1*) and a MnSOD (*SOD3*) in the cytosol and a second MnSOD (*SOD2*) in the mitochondria (134, 162, 249). In this study we report the existence of three additional candidate CuZn containing superoxide dismutases, the products of the *SOD4*, *SOD5* and *SOD6* genes. We have characterized the *SOD5* gene that appears to be important factor in the virulence of this fungus. Virulence of *C. albicans* is often attributed to its ability to switch from yeast to hyphal forms. Consistent with this, *SOD5* expression increases with several hyphal-inducing conditions, *SOD5* expression is dependent on known components of hyphal inducing signaling pathways, and the absence of this gene leads to a defect in the virulence of this organism. As expected for superoxide dismutases, *SOD5* has some roles in responding to certain stresses.

Superoxide dismutases serve to convert damaging superoxide radicals, a key form of reactive oxygen species (ROS), to less damaging hydrogen peroxide that can be converted into water by catalase action. *SOD5* clearly encodes a CuZn-containing superoxide dismutase: it has the highest sequence similarity to the CuZn superoxide dismutase family, and deletion of the *SOD5* gene eliminates the major source of CuZn dependent (cyanide sensitive) superoxide dismutase activity detected in IMDM grown cells.

What is the significance of *C. albicans* having multiple CuZn superoxide dismutase genes, most of which are potentially expressed in the cytosol? For pathogens like *C. albicans*, there are both internal and external sources of superoxides. An attractive model would be to have high levels of SOD activity to protect the fungal cell against the oxidative burst inside the phagosome. The fact that *SOD5* is highly expressed in the hyphal state of *C. albicans* (this study and (224)), and hyphal cells are typically the form that interacts with the host phagocytic cells (reviewed in (105)), would support such a model. In addition, increases in pH, salt concentration, or presence of oxidative species lead to increases in *SOD5* transcription, the same events happen inside the phagosome (247), further strengthening a model that *SOD5* could have a defense-related function. Most significantly, the strongest evidence that supports this model is the fact that the *sod5* knock out strain is defective in mouse killing. It was previously shown that the *C. albicans sod1* mutant strain is defective in virulence and in macrophage killing. Our results have shown that the *sod5* mutant behaves similarly to *sod1* with respect to the virulence in mice, but surprisingly Sod5p does not affect macrophage killing. This would

suggest that there are other phagocytic mammalian cells, like neutrophils or dendritic cells, to which Sod5p is critical for a wild type response.

We investigated the possibility that a further function of *SOD5* is to protect *C. albicans* cells against intracellularly-generated superoxide radicals. *C. albicans* produces the highest levels of CuZnSOD during the mid-exponential phase (162, 249) and it generates increased amounts of CuZnSOD during the switch from yeast to hyphal forms (109). During the same conditions, hyphal growth and in exponential phase of the growth, *C. albicans* is known to produce more ROS, including superoxide radicals (266). These two parameters, superoxide generation and superoxide detoxification, were correlated to respiration of *C. albicans* (139). Mitochondrial respiration was shown to be a powerful source of superoxide radicals in yeast (47), and factors like non-fermentable carbon sources were linked to increased superoxide production by electron transport chain in *C. albicans* (252), *S. cerevisiae* (67, 193, 209), and in *Schizosaccharomyces pombe* (143). In our study, these conditions that increase ROS production also increase *SOD5* transcription. *SOD5* expression levels were increased when cells grew in the presence of salt or on non-fermentable carbon as the only carbon source. In addition, an increased amount of fermentable glucose restored the growth ability of the *SOD5* mutant cells grown in nutrient limiting media in the presence of hydrogen peroxide. We also observed that *SOD5* transcription was induced by both menadione and riboflavin, which are intracellular and extracellular superoxide generators respectively. This also provides us with the evidence that Sod5p is needed for elimination of superoxides found on both sides of the cell membrane. This might suggest that Sod5p shares the redundancy with

other superoxide dismutases in eliminating extracellularly and intracellularly-generated superoxide radicals in *C. albicans*.

All the increases in *SOD5* transcript levels were observed in *C. albicans* grown in exponential phase, the same phase in which *C. albicans* produces more ROS. All hyphal-inducing conditions induced *SOD5* transcription, and *SOD5* transcription was found to be regulated by transcription factors regulating the yeast to hyphal switch of *C. albicans*. All these lines of evidence might suggest that hyphal growth and log-phase growth processes increase mitochondrial respiration in aerobically growing *C. albicans*, which subsequently increase mitochondrially generated superoxide radicals, leading to increased levels of detoxification enzymes. It is possible that these detoxification enzymes include proteins like *SOD5*.

Our results also show that *SOD5* possesses a complex regulation of expression. Intriguingly, Efg1p was found to positively regulate *SOD5* transcription in the presence of serum as seen in Nantel *et al.*, 2002, and negatively regulate *SOD5* transcription under non-inducing conditions. In addition, *SOD5* transcription was found to be dependent on the Rim101 transcription factor, which is required for alkaline-induced filamentation (70). This implicates the Rim101p pathway in the response of the *C. albicans* cell to oxidative stress. Rim101p was shown to be required for *SOD5* induction in response to a variety of external signals, not just pH. Interestingly enough, the growth of the *S. cerevisiae rim101* mutant was found to be inhibited at basic pH and in the presence of NaCl (163), the same conditions in which *SOD5* transcription was found to be up-regulated. The 1kb sequence upstream of the *SOD5* start codon revealed four potential

Rim101p binding sites, which closely resemble the previously described CCAAGAAA – Rim101p binding site (244), and four potential Efg1p binding sites (CAWWTG) (182).

All recent evidence suggests that *C. albicans* possesses a complicated cytosolic detoxification system. During the log-phase growth of *C. albicans*, *SOD1* and *SOD5* (CuZn-containing superoxide dismutases) are expressed (this study and (162)). *SOD1* appears to be primarily responsible for removal of extracellular, macrophage-generated superoxide radicals (135). *SOD5* appears to be responsible for detoxification of extracellularly and intracellularly, possibly mitochondrially-generated superoxide radicals during log-phase and hyphal growth. As *C. albicans* cells enter the stationary phase, the expression of CuZn-containing superoxide dismutases is repressed and the expression of Mn-containing cytosolic superoxide dismutase (*SOD3*) is increased (162).

The search for sequences similar to CuZnSOD in genomes of *Neurospora crassa, Aspergillus fumigatus, S. cerevisiae* and *Schizosaccharomyces pombe* revealed that all of these organisms have *SOD1* homologues, but only filamentous fungi (*N. crassa* and *A. fumigatus*) have additional sequences homologous to *SOD5* (NCU03013.1(161682-160794) and contig:69:a_fumigatus (74890-75198), respectively). Perhaps filamentous fungi have more sources of production of harmful superoxide radicals, which applies more pressure on these fungi to possess extra isoforms of superoxide detoxificators, like *SOD5*. The role of *SOD4* and *SOD6* remain to be established, although *SOD4* was recently shown to be regulated in White-Opaque phenotype switching (164).

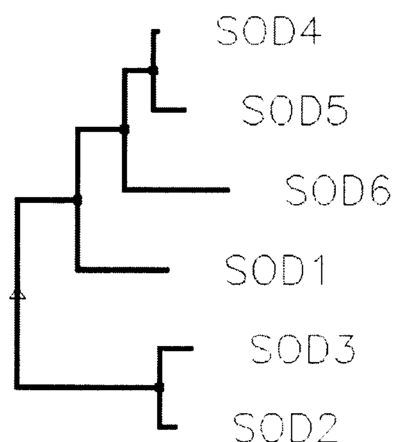
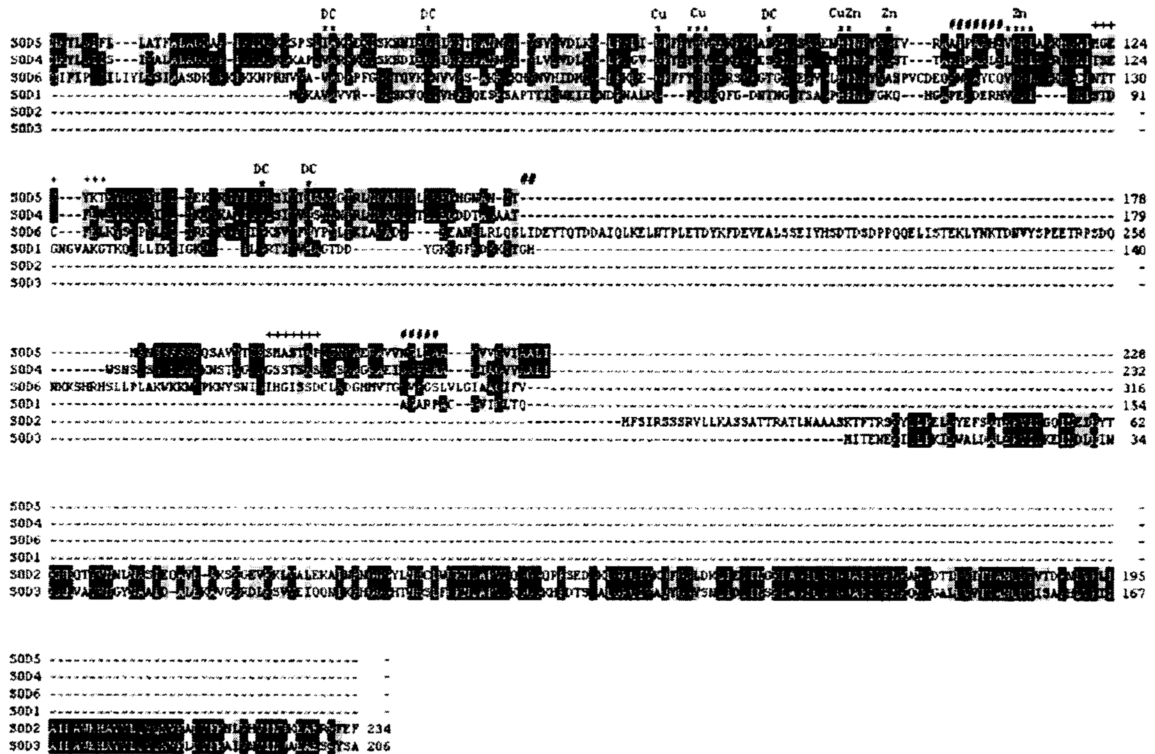


Figure 3.1.1. Similarity of the CuZn-containing and Mn-containing superoxide dismutases in *C. albicans*. **(a)** Phylogenetic tree shows the similarity between all *C. albicans*' superoxide dismutases: Sod4p, SOD5p, and Sod6p are homologous to Sod1p, and the CuZnSODs share no sequence similarity with the MnSODs. The analysis was produced by the Multiple Sequence Alignment program of Florence Corpet (http://prodes.toulouse.inra.fr/multalin/multalin.html).



(b) Amino acid sequence alignment of CuZn Superoxide dismutases (*SOD1, SOD4, SOD5, SOD6*) and Mn Superoxide dismutases (*SOD2, SOD3*). Deduced amino acid sequences were aligned with the CLUSTALW program. Identical amino acids are boxed in black and conservative changes are boxed in grey. Hyphens indicate gaps that were introduced to maximize the alignment. "+" indicate oligos used for Northern and Southern analysis for Sod4p, Sod5p and Sod6p. "#" indicate oligos used for Northern and Southern analysis for Sod1p. The Sod4p, Sod5p, and Sod6p contain CuZn-SOD characteristic residues. Asterisks below Cu and Zn stand for copper and zinc metals binding respectively; asterisks below DC stands for amino acids involved in dimer contact (86). Numbers at the ends of the sequences represent the amino acid lengths of the proteins.

Figure 2A.

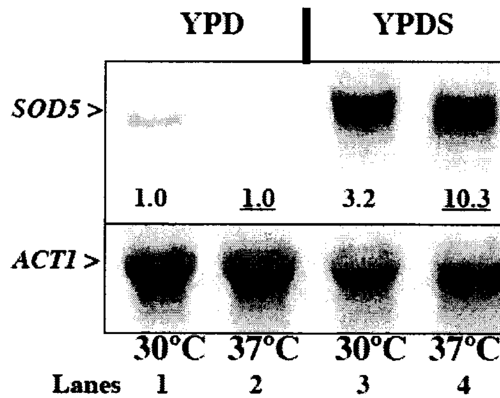


Figure 2B.

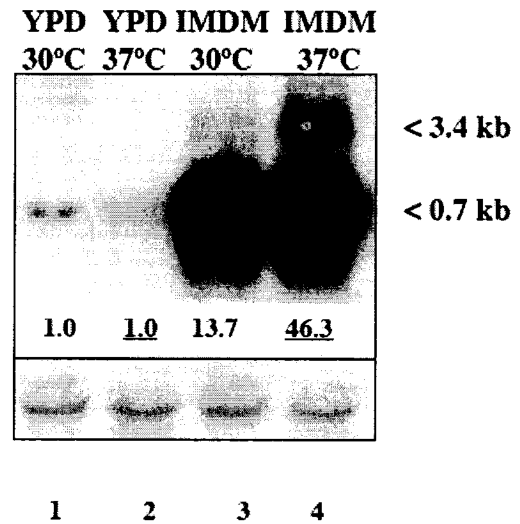


Figure 3.1.2. Hyphal-inducing conditions induce *SOD5* expression. **(a)** Serum, rather than temperature, induces *SOD5* transcription. Northern blot analysis of the *SOD5* transcripts in *C. albicans* strain SC5314 grown in the log-phase in indicated media (YPD and YPD containing 10% serum) at indicated temperatures. **(b)** *SOD5* transcription is up-regulated in IMDM medium. Northern blot analysis of the *SOD5* transcripts in *C. albicans* strain SC5314 grown in the log-phase in IMDM medium at 30 and 37°C. Northern blot analysis of the Actin (*ACT1*) transcripts in *C. albicans* strain SC5314 was used as an internal loading control. Relative intensities are shown below each of the genes analyzed. Relative intensities of cells grown at 37°C are underlined and quantified independently from *SOD5* intensities obtained at 30°C. Quantifications were performed by NIH Image software (RSB-NIMH-NIH http://rsb.info.nih.gov/nih-image/).

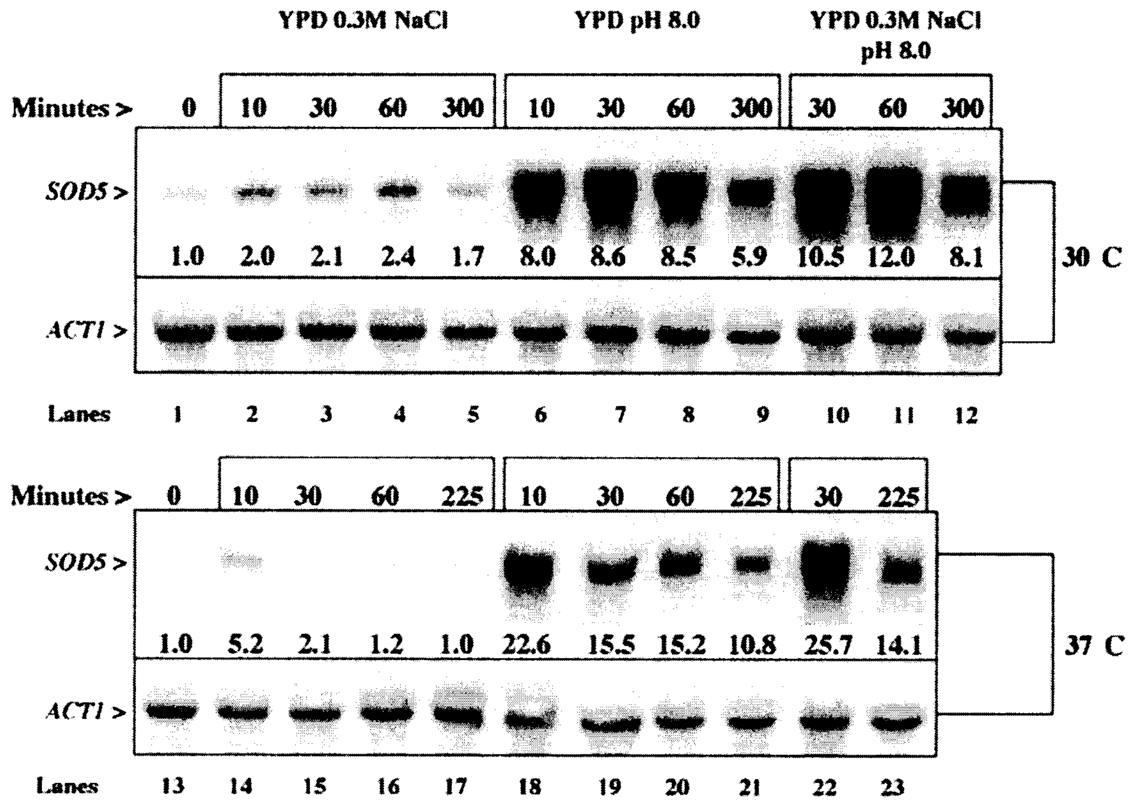
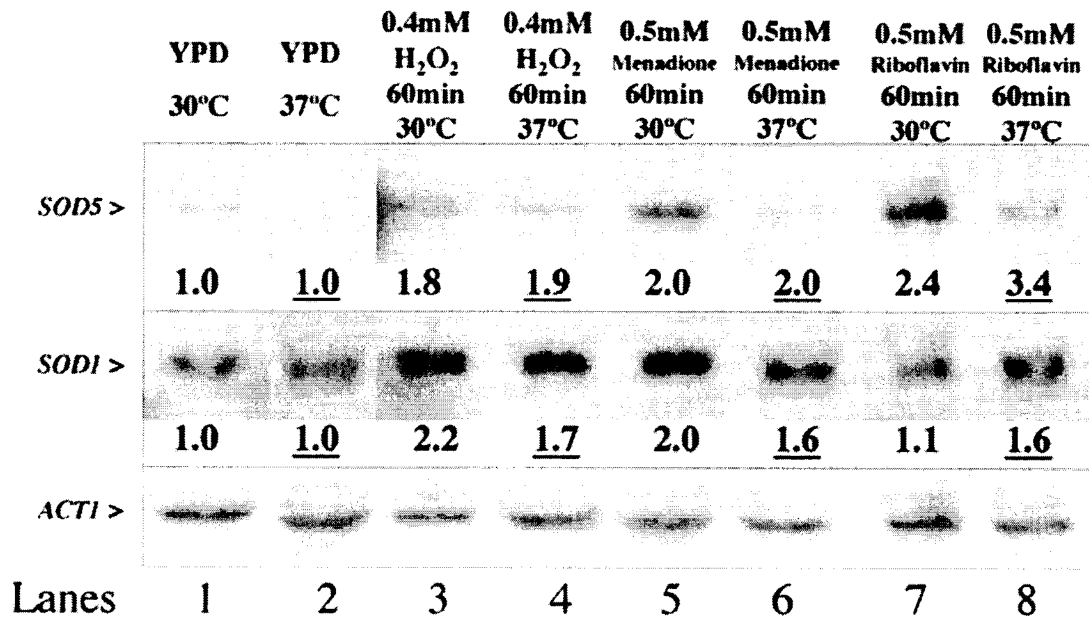


Figure 3.1.3. Alkaline conditions, Osmotic and Oxidative stresses affect SOD5 transcription. **(a)** *SOD5* transcription is upregulated in alkaline conditions and in the presence of increased amounts of salt. Northern blot analysis of the *SOD5* transcripts in *C. albicans* strain SC5314 diluted to an OD_{600} of 0.1 and then incubated at either 30 or 37°C in YPD. In those cultures the pH, NaCl concentration or both of those factors were increased 10, 30, 60 or 300 / 225 minutes before the OD_{600} reached 1.0.



(b) Effects of oxidative shock on *SOD5* and *SOD1* of *C. albicans*. Northern blot analysis of the *SOD5* transcripts in *C. albicans* strain SC5314 grown in log-phase, from OD_{600} of 0.1 to OD_{600} of 1.0 at 30°C or from OD_{600} of 0.1 to OD_{600} of 2.5 at 37°C; the hydrogen peroxide / menadione / riboflavin were supplemented one hour before reaching the final OD_{600} of 1.0 or 2.5. Lanes 1 and 2: no oxidative shock was applied. Lanes 3 and 4: hydrogen peroxide was added 1 hour before OD_{600} reached 1.0. Lanes 5 and 6: menadione was added 1 hour before OD_{600} reached 1.0. Lanes 7 and 8: riboflavin and light were added 1 hour before OD_{600} reached 1.0. Northern blot analysis of the Actin (*ACT1*) transcripts in *C. albicans* strain SC5314 was used as an internal loading control. Analysis as described in Figure 3.1.2.

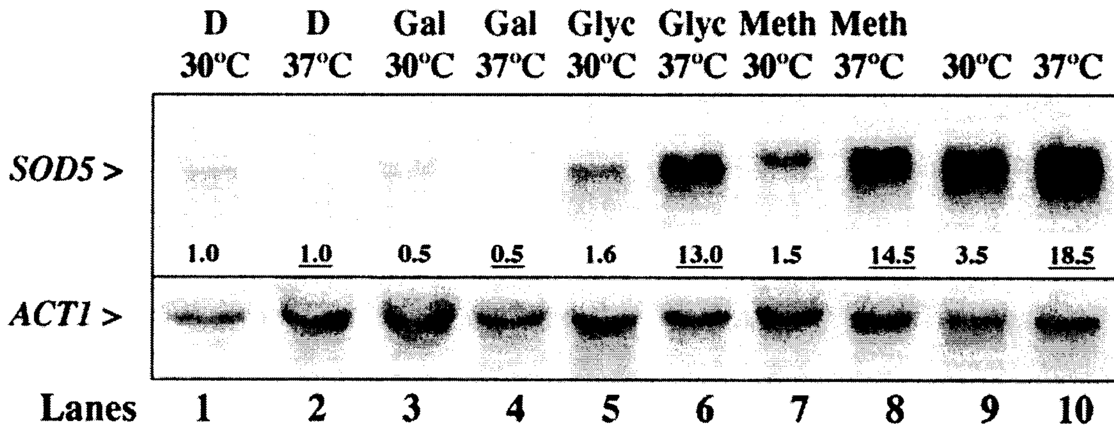


Figure 3.1.4. *SOD5* transcription is affected by non-fermentable carbon source. Northern blot analysis of the *SOD5* transcripts in *C. albicans* strain SC5314 grown in log-phase, from OD_{600} of 0.1 to OD_{600} of 1.0 at 30°C or from OD_{600} of 0.1 to OD_{600} of 2.5 at 37°C in YPD, YPGal, YPGlyc, YPMeth, and YP. Northern blot analysis of the Actin (*ACT1*) transcripts in *C. albicans* strain SC5314 was used as an internal loading control. Analysis as described in Figure 3.1.2.

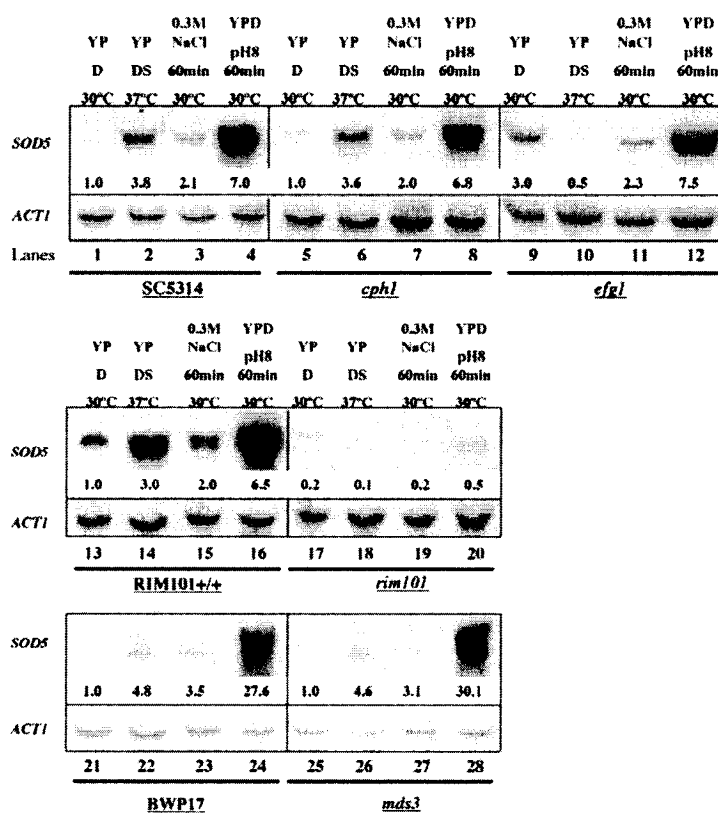


Figure 3.1.5. Transcription factors involved in the transcription of *SOD5*. Northern blot analysis of the *SOD5* transcripts in *C. albicans* strains lacking indicated transcription factors (strains SC5314, JKC19, HLC52, DAY25, DAY185, BWP17 and GKO6 (*mds3*) (see Table 1)). The potential transcription of *SOD5* was induced by one of the three inducing conditions: presence of serum (10%) at 37°C, presence of salt (0.3M NaCl) for 1 hour at 30°C, or switch to alkaline conditions (pH8) for 1 hour at 30°C. Analysis as described in Figure 3.1.2. The intensities of *SOD5* in HLC52 were compared to SOD5 intensities of JKC19. The intensities of *SOD5* in DAY25 were compared to *SOD5* intensities of DAY185. The intensities of *SOD5* in *mds3* were compared to *SOD5* intensities of BWP17. These Northern blot analyses were reproduced twice. The general drop in *SOD5* levels in *rim101* was reproduced, but the difference in relative intensities of *SOD5* in the *rim101* strain was not significant.

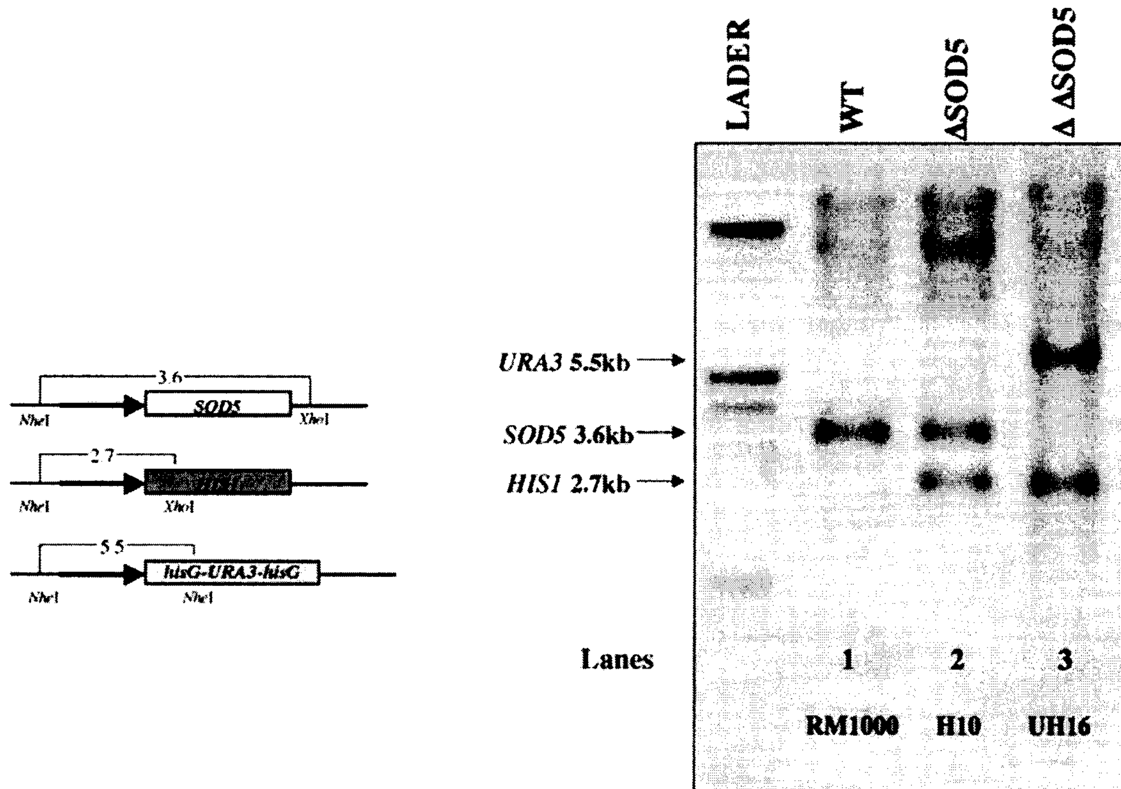
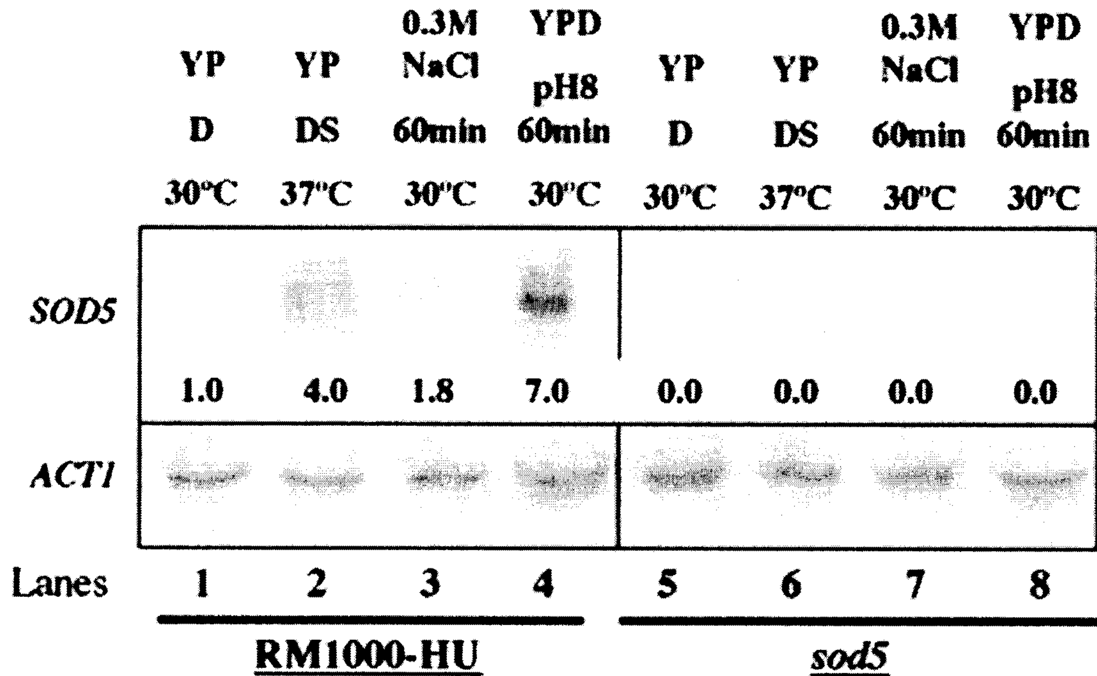
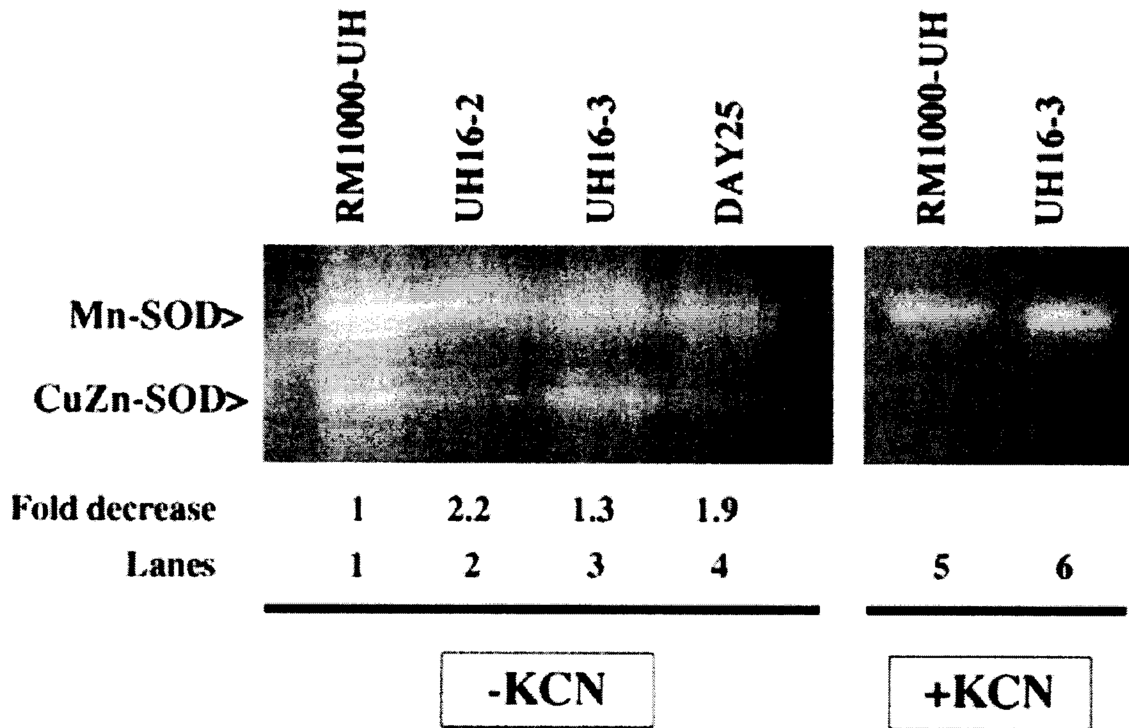


Figure 3.1.6. *SOD5* gene is deleted from UH16 strain. **(a)** Restriction endo-nuclease map of *SOD5*, of *HIS1* inserted in place of *SOD5* ORF, and of *hisG-URA3-hisG* inserted in place of *SOD5* ORF digested with *Nhe*I and *Xho*I. The arrows indicate 1kb of *SOD5* promoter region. Southern blot analysis with the use of a 1kb probe generated with ODH157 and OMM1 oligos recognizes 1kb of the *SOD5* promoter region. The genomic DNA samples, digested with *Nhe*I and *Xho*I were prepared from strains RM1000 (*SOD5/SOD5*; lane 1), H10 *(Δsod5::HIS1/SOD5*, lane 2), UH16 (*Δsod5::hisG-URA3-hisG/Δsod5::HIS1*, lane 3). The wild-type *SOD5* band is 3.6kb, the band corresponding to the *URA3* cassette is 4.1kb, and the band corresponding to the *HIS1* cassette is 2.7 kb.



(b) Northern blot analysis of the *SOD5* transcripts in the UH16-2 mutant and its parental strain, RM1000-HU. The transcription of *SOD5* was induced by one of the inducing conditions: presence of serum (10%) at 37°C, or switch to alkaline conditions (pH8) for 1 hour at 30°C, or presence of 0.3M NaCl for 1 hour at 30°C for UH16-2 and RM1000-HU strains. Northern blot analysis of the Actin (*ACT1*) transcripts in indicated *C. albicans* strains was used as an internal loading control. Analysis as described in Figure 3.1.2.



(c) Superoxide dismutase activity staining revealed SOD activity by staining the 10% native polyacrylamide gel with nitro blue tetrazolium and riboflavin. A sample (5mg) of the crude extract from exponentially growing cells in IMDM 37°C of RM1000-HU (lanes 1and 5), UH16-2 (lane 2), UH16-3 (lane 3 and 6) and DAY25 (lanes 4) was loaded into three gels. One gel (lanes 1-4) was stained for SOD activity in the absence of 4mM KCN, the second gel (lanes 5-6) was stained for SOD activity in the presence of KCN, and the third gel was stained with Coomassie blue as an internal loading control (our unpublished results). Relative intensities are shown below each of the genes analyzed. Quantifications were performed by NIH Image software (RSB-NIMH-NIH http://rsb.info.nih.gov/nih-image/).

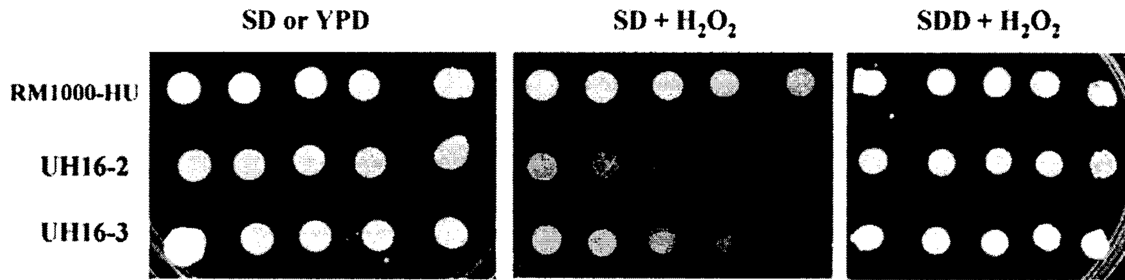


Figure 3.1.7. Hydrogen peroxide and limited nutrients inhibit growth of UH16-2 strain. Spot assay of RM1000-HU, UH16-2, UH16-3 strains on SC or YPD, SC supplemented with hydrogen peroxide, and SC supplemented with 2,5% dextrose and hydrogen peroxide (SDD) media at 30°C for 48 hours.

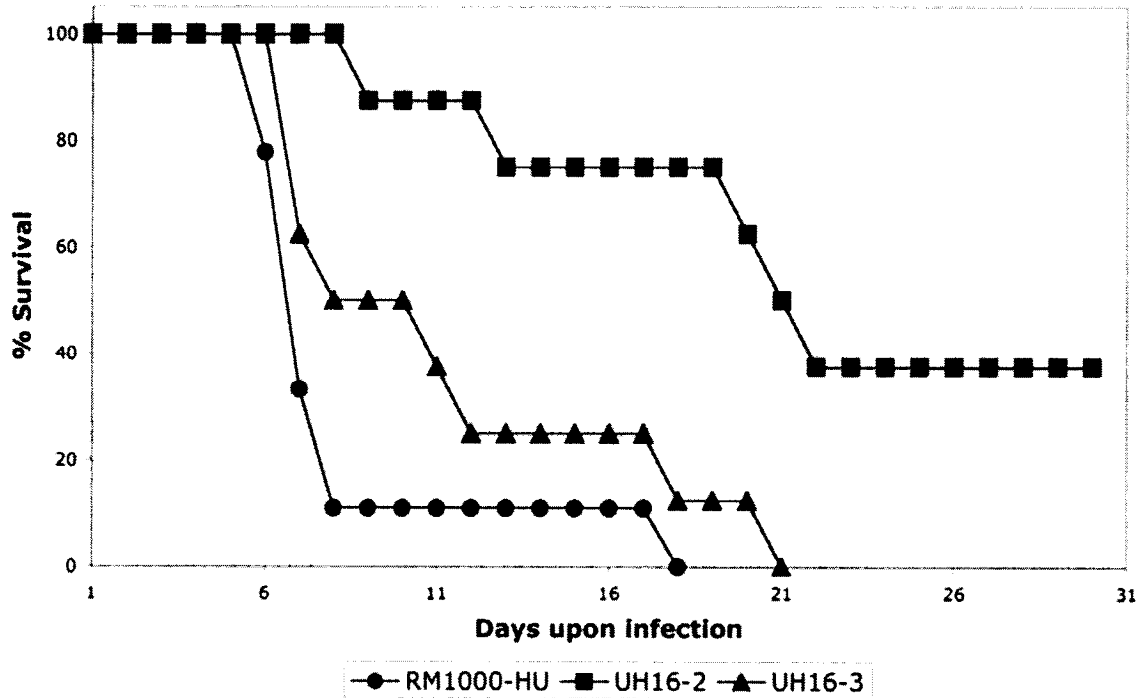


Figure 3.1.8. Sod5p activity is required for virulence in animal model. Survival of mice infected with *C. albicans sod5* mutant UH16-2 (squares), *sod5* revertant UH16-3 (triangles), and wild type parental RM1000-HU (round) strains. Mice were inoculated by tail vein injection and survival was measured over a 30-day period.

Section 3.2: Transcription factors in *C. albicans*.

As shown by Davis *et al*, a *C. albicans* transcription factor Rim101p, which is under the control of pH response pathway, is required for host-pathogen interaction (69). We showed that Rim101p regulates the expression of *C. albicans* virulence factor *SOD5*, which therefore makes it a virulence factor as well. This example shows that transcription factors, which regulate the expression of genes whose products participate in host-pathogen interaction, can be just as important as the virulence factors themselves. Thus, comparative genomic studies have also provided an opportunity to understand the mechanisms of gene expression in *C. albicans*.

Proteins of the zinc finger superfamily, for example, represent one of the largest classes of DNA-binding proteins in eukaryotes. The functions of many uncharacterized *C. albicans* zinc cluster factors could be inferred from the fact that they display high levels of sequence similarity with the products of *S. cerevisiae* genes with a known function. However, approximately half of the *C. albicans* zinc cluster genes do not appear to have homologues in *S. cerevisiae*, and are therefore likely to participate in processes specific to *C. albicans*. In addition, the majority of *S. cerevisiae* transcriptional activators that were found to have *C. albicans* homologues, one half shared homology only within a DNA binding domain.

The lack of homology in the activation domains of transactivators between *S. cerevisiae* and *C. albicans* might suggest a concomitant reduced structural similarity in the activation-domain-interacting complexes between the two species. A pair-wise sequence comparison of the transcriptional machinery between *C. albicans* and *S. cerevisiae* shows a high level of conservation in the RNA polymerase II complex. The

exceptions to this are TFIIA and the Mediator complex: *S. cerevisiae* and *C. albicans* show low levels of homology with respect to the proteins of those two complexes. Interestingly, of all RNA polymerase machinery, both of those complexes were shown to interact with transcriptional regulators. Since *C. albicans'* Med-type mediator complex shares very little homology to that of its host, it may be an attractive target for antifungal drugs.

Transcriptional activation domains of the *Candida albicans* Gcn4p and Gal4p homologs.

Mikhail Martchenko[1], Anastasia Levitin[2], and Malcolm Whiteway[1,2*].

[1]Department of Biology, McGill University, Stewart Biology Building, 1205 Avenue Docteur Penfield, Montreal, Quebec, Canada H3A 1B1;

[2]Genetics Group, Biotechnology Research Institute, National Research Council, Montreal, Quebec, Canada H4P 2R2

Eukaryotic Cell 2007 Feb;6(2):000 © 2007 American Society for Microbiology.

3.2.1 ABSTRACT

Many putative transcription factors in the pathogenic fungus *Candida albicans* contain sequence similarity to well-defined transcriptional regulators in the budding yeast *Saccharomyces cerevisiae*, but this sequence similarity is often limited to the DNA binding domain of the molecules. The Gcn4p and Gal4p proteins of *Saccharomyces cerevisiae* are highly studied and well-understood eukaryotic transcription factors of the bZip (Gcn4p) and C_6 zinc cluster (Gal4p) families; *C. albicans* has CaGcn4p and CaGal4p with DNA binding domains highly similar to their yeast counterparts. Deletion analysis of the CaGcn4p protein shows that the N' terminus is needed for transcriptional activation; an 81 amino acid region is critical for this function, and this domain can be coupled to a *lexA* DNA-binding module to provide transcription activating function in a

heterologous reporter system. Deletion analysis of the *C. albicans* Gal4p identifies a C-terminal 73 amino acid long transcription activating domain that also can be transferred to a heterologous reporter construct to direct transcriptional activation. These two transcriptional activation regions show no sequence similarity to the respective domains in their *S. cerevisiae* homologs, and the two *C. albicans* transcription activating domains themselves show little similarity.

3.2.2 INTRODUCTION

Transcriptional regulators control the expression of genes to co-ordinate the availability of cellular function with the physiological needs of the cell. Gene-specific transcriptional activation is often regulated by the binding of positively-acting proteins to upstream activating sequences (UAS) in the DNA where they recruit and control the activities of chromatin-modifying and remodeling complexes and the transcription apparatus (178). A typical transcriptional activator then interacts with the RNA Polymerase II complex through binding to an adaptor complex termed Mediator; this Mediator complex consists of about 20 proteins and is conserved from yeast to humans (27). Eukaryotic transcriptional activator proteins are generally bipartite in nature, with separate domains for DNA binding and transcriptional activation (217, 292). The transcriptional activation domains are classified according to their amino acid composition: rich in acidic (e.g. *Saccharomyces cerevisiae* Gal4p and Gcn4p) or basic residues (tobacco BBC1), or rich in glutamine (*S. cerevisiae* Mcm1p), threonine/serine (human OCT2), or isoleucine (NTF1) (8, 38, 61, 83, 125, 146, 201, 202). The DNA binding modules also fall into many classes, such as zinc-finger, leucine-zipper, and

helix-loop-helix motifs (81, 84, 306). Although the activation domain is critical for function, and can provide a level of regulation, the functional targets of such transcriptional activators are determined by the DNA-binding address of the protein.

In both *S. cerevisiae* and *C. albicans GCN4* encodes a transcriptional activator of amino acid biosynthetic genes that responds to amino acid starvation (123, 293). ScGcn4p is tightly regulated at both transcriptional and translational levels. The 5' leader region of *GCN4*, which codes for a transcriptional activator of amino acid biosynthetic genes in response to amino acid starvation, contains four small upstream ORFs (uORF1-4). These uORFs act as negative regulators of translation: the ribosome initiates translation at uORF1 and becomes re-activated for translation at subsequent uORFs. Under environmental stresses, such as amino acid starvation, the translation of *ScGCN4* is induced: ScGcn2 kinase phosphorylates eIF2, and the scanning ribosome is not re-activated until it bypasses the uORFs and initiates translation at the *GCN4* open reading frame (122). The unusually long 5'-leader sequence on the *GCN4* mRNA, which carries four upstream open reading frames is conserved in *C. albicans* (293). It was recently shown that the protein kinase, Gcn2, which is involved in transcriptional and translational regulation of Gcn4p in *S. cerevisiae*, is not involved in the regulation of *C. albicans* Gcn4p (290).

Gcn4p binds as a homodimer with its basic leucine-zipper (bZIP) found in its carboxy terminus to a TGACTC sequence located upstream of many genes induced during amino acid starvation (7, 126). In *S. cerevisiae* Gcn4p there are two transcription activation domains: one resides in an acidic segment in the center of the protein between residues 107 and 144 (78, 125, 141), the second Gcn4p activation domain is located in

the N-terminal 100 amino acids. The two activation domains are functionally redundant and can work independently to produce high-level activation (78, 141).

S. cerevisiae Gal4p represents a second well-studied transcription factor, and functions as the transcriptional activator of galactose catabolism (192, 237, 291). ScGal4p contains a DNA binding domain (amino acids 1-65), and two transcriptional activation domains, domain I (amino acids 149-196), and domain II (amino acids 768-881) (76, 202). The transcriptional activation domain II of Gal4p interacts with Gal80p in the absence of galactose, and through this contact, Gal80p inhibits Gal4p (230), preventing it from activating the expression of galactose-dependent genes. During growth on galactose, Gal3p binds Gal80p and removes it from Gal4p at the *GAL* gene's promoter and prevents Gal80p from inhibiting Gal4p function (237). Therefore, in the presence of galactose, Gal4p is freed from Gal80p inhibition, and subsequently activates the expression of the galactose regulon (192, 237, 291). The Gal4p DNA binding domain interacts with a specific upstream activating sequence UAS_G (CGG(N_{11})CCG), located in the promoter regions of the genes whose products participate in the galactose metabolism circuit, such as *GAL1, 2, 3, 7, 10, 80* (147). NMR analyses of various amino terminal Gal4p fragments and X-ray crystal structure determination of Gal4p-UAS_G complex (210) show that the C_6 zinc cluster is the DNA binding module of Gal4p.

The primary role of a transcription activator is to recruit the RNA Polymerase II machinery to the promoter to which it is bound. To achieve this, the transcriptional activation domains of both Gcn4p and Gal4p interact with Gal11p, which is a component of the mediator complex that binds the RNA Polymerase II machinery (10, 114, 142, 235, 248).

Three general models have been proposed to characterize the structure of the transcriptional activation domain. In the first hypothesis, it was proposed that activation domains are unstructured "acidic blobs" that interact with their targets via ionic interactions. This model is supported by the observation that the removal of the residues of activation domains decreases the activity gradually, rather than abruptly (271). A second model proposes that acidic activation domains form amphipathic α helices, in which acidic residues are aligned on one face of the helix. This model is supported by the observation that an artificial 15 residue peptide, designed to fold into amphipathic α helix, shows transcription activating abilities when fused to the *GAL4* DNA binding domain (202). The authors of the third model argue that the most likely secondary structure is an antiparallel β sheet (301).

We have recently established a detailed annotation of the *C. albicans* genes (33). Intriguingly, although many of the *C. albicans* transcription factors have sequence similarity to transcription factors in *S. cerevisiae*, the similarities occur primarily in the DNA binding motifs of those proteins. In addition, it has been previously shown that although *C. albicans* Rfg1p, Rap1p, Gat1p, Msn2p and Msn4p have *S. cerevisiae* DNA-binding-domain homologs, these transcription factors control the regulation of different processes in the two organisms (23, 149, 185, 228). Here we have investigated the transcription activation domains of *C. albicans* homologs of the Gcn4p and Gal4p transcription factors; these domains are serine-threonine rich and lack sequence similarity to the yeast homologs.

3.2.3 MATERIALS and METHODS

***C. albicans* and bacterial strains.** CAI8 (88) was used to generate CRC103 and CRC106 (256). CRC106 carries the *S. aureus lexA* operator; CRC103 does not and serves as a negative control for *lexA* binding. CRC106 and BWP17 (305) were used to define the transcription activation domains of *GCN4* and *GAL4*.

Plasmid construction. Plasmids and oligonucleotides are shown in Tables 2 and 3. To create p*lexA-HIS1* we PCR-amplified the *CaHIS1* open reading frame and its termination sequence from pFA-*HIS1* with OMM46 and OMM47, which contain the *lexA* binding site and the *CaADH2* TATA box. The PCR product, which contains the *lexA* binding site, TATA box, *HIS1* open reading frame and termination sequence, was cloned into pFA-*ARG4* using the *SalI* and *SunI* restriction sites. Two more *lexA* binding sites were added by annealing oligonucleotides OMM50 and OMM51 and cloning them into the *SunI* site close to the third *lexA* binding site to yield p*lexA-HIS1*. To create strain CMM25, the p*lexA-HIS1* construct was integrated into the *ARG4* locus of strain BWP17 by treating p*lexA-HIS1* with *AgeI*, which cuts once in the *ARG4* sequence. CIp-*lexA-GCN4* deletion constructs were created by PCR using divergent primers: oligo OMM56 annealed to region upstream of *GCN4* while oligos OMM60 and OMM62 align inside the *GCN4* open reading frame in the CIp-*lexA-GCN4* template. Oligos OMM66-67 were used for CIp-*lexA-GAL4*. To fuse *GAL4* in-frame with *lexA*, CIp-*lexA-GAL4* was cut with *MluI*. The resulting 5' overhangs were filled with T4 Polymerase, and the constructs were self-ligated with T4 DNA ligase. CIp-*lexA-GCN4* and CIp-*lexA-GAL4Δ1-81* were cut with *BstBI* and self-ligated to yield CIp-*lexA-GCN4Δ247-323* and CIp-*lexA-GAL4Δ1-81Δ247-323* respectfully. CIp-*lexA-GAL4* was cut with *ZraI* and *PstI* and self-ligated to create

CIp-*lexA-GAL4Δ188-261*. OMM125 and OMM57 were used to PCR amplify *GCN4AD*, which was ligated into *AatII* and *PstI* cut CIp-*lexA-GAL4* to create CIp-*lexA-GAL4-GCN4AD*. *CaHIS1* open reading frame, its promoter, and termination sequence was cut out from pFA-*HIS1* using *NotI* and ligated into *NotI* - cut pOP*lacZ* to create pOP*lacZ-HIS1*. CMM85 strain (*gcn4* with pOP*lacZ-HIS1*) was created by first converting *gcn4* (229) into *URA3* auxotroph (*gcn4-ura3*) using 5-FOA containing media, followed by an integration of *XcmI* – cut pOP*lacZ-HIS1* at *HIS1* locus. CMM86, CMM87, CMM88 and CMM89 were created by transforming CMM85 (*gcn4* with pOP*lacZ-HIS1*) with CIp-*lexA*, CIp-*lexA-GCN4*, CIp-*lexA-GCN4Δ1-81*, and CIp-*lexA-GCN4Δ1-81Δ247-323* respectively. All of the constructs created in this study were integrated into the genome of *C. albicans*: all of the constructs, except for pOP*lacZ* and p*lexA-HIS1* were digested with *StuI* to integrate them at *RPS1* locus. pOP*lacZ* was digested with *BamHI* and p*lexA-HIS1* was digested with *AgeI* to integrate them at *ADE2 and ARG4* loci respectively.

β-galactosidase assays. The expression level of the *lacZ* was assayed in two ways: by β-galactosidase overlay assay using independently isolated transformants grown on solid YPD, YPGal, or SC-aa media, or by β-galactosidase assays performed on mid-exponential shake-flask YPD, YPGal, or SC-aa – grown liquid cultures (255). The β-galactosidase assays were expressed in Miller units from three independent transformants using means and standard deviations.

3.2.4 RESULTS

Definition of the CaGcn4p activation domain

We have recently performed a detailed genome annotation of *C. albicans* genome (33), which showed that frequently the transcription factors of this organism share homology to transcription factors of other organisms only within the DNA binding domain. We have defined 198 *S. cerevisiae* genes whose products contain a DNA binding domain and are classified as transcription factors by combining the list of transcriptional regulators from Harbison *et al.* (115) with the list from http://www.yeastract.com/tflist.php. Of these, 32 were experimentally shown to be transcriptional repressors in *S. cerevisiae*. Ninety-nine of the remaining 166 *S. cerevisiae* transcriptional activators were found to have *C. albicans* homologs, one half of which share homology only within a DNA binding domain (Figure 3.2.1). A detailed assessment of global and transcriptional activation domain similarities are provided (Table 3.2.1) Since there is no primary sequence that defines the activation domain as a module, the nature of the activation domain is based on the experimentally defined part of transcription factor. A majority of *S. cerevisiae* transcription factors, such as Gal4p, Gcn4p, Upc2p, Leu3p, and Arg81p (68, 78, 125, 141, 202, 242, 313) were experimentally shown to have acidic activation domains. When we compared the transcription factors of *S. cerevisiae* with the transcription factors of *C. albicans* we observed that in some cases the sequence of the experimentally defined activation domains of *S. cerevisiae* is very well conserved in the *C. albicans* transcription factors homologs, such as in Upc2, Leu3, and Arg81 (68, 242, 313). In other cases the sequence of experimentally defined activation domain of *S. cerevisiae* is not detectable in the *C. albicans* transcription factor

homologs, such as in Gal4p and Gcn4p. The presence of homology in the DNA binding domain of the *C. albicans* transcription factors like Gcn4p and Gal4p tells us that these might be transcriptional regulators, but the absence of homology in the activation domain makes it difficult for us to predict whether these could work as activators or repressors. We therefore directly investigated the function of the regions outside the DNA binding motif.

We have investigated the role of the non-homologous regions of candidate transcription factors to establish if they play a role in transcriptional activation. Within the DNA binding domain, CaGcn4p shares strong sequence similarity (88%) with *S. cerevisiae* Gcn4p. To establish if this protein possesses a functional activation domain we made use of the *S. aureus lexA* one-hybrid system (256). This system contains the *C. albicans* actin promoter (p*ACT1*) placed upstream of the *S. aureus lexA* open reading frame to create CIp-*lexA*, and the *S. aureus lexA* operator upstream of both an *ADH1* basal promoter and a *lacZ* open reading frame, creating pOP*lacZ* (256); as well, we placed the *S. aureus lexA* operator upstream of the *HIS1* open reading frame to create p*lexA-HIS1*. We integrated pOP*lacZ* into strain CAI8 to yield reporter strain CRC106 (256) and integrated p*lexA-HIS1*into strain BWP17 to yield reporter strain CMM25. Fusions were constructed in CIp-*lexA* and introduced into these two reporter strains. In the absence of any transcriptional activator fused to *lexA*, the reporter CRC106 derivative yielded basal levels of β-galactosidase and the reporter CMM25 derivative produced no growth in the absence of histidine; when a trans-activator is fused to *lexA*, the system yielded higher levels of β-galactosidase and permitted growth in the absence of histidine. A full length *CaGCN4* cloned downstream of the *lexA* open reading frame, creating CIp-

lexA-GCN4, was fully capable of trans-activation in the *C. albicans* assays. This construct generated five times higher β-galactosidase activity when transformed into CRC106 to generate CMM14 and permitted growth in one day in the absence of histidine when transformed into CMM25 to create CMM30, compared to the appropriate controls CMM10 and CMM26 which contain CIp-*lexA* (Figure 3.2.2). This result suggests that *C. albicans GCN4* contains a transcription activation domain (256).

The C-terminal systematic deletions of *lexA-GCN4* identified the N-terminal 81 amino acid region serving as an activation domain; CIp-*lexA-GCN4Δ247-323,* CIp-*lexA-GCN4Δ161-323,* CIp-*lexA-GCN4Δ123-323,* and CIp-*lexA-GCN4Δ82-323* were as active as full-length CIp-*lexA-GCN4* in both the *lacZ* background and *HIS1* background. Further C-terminal deletions gradually reduced both *lacZ* and *HIS1* activities, suggesting that the Gcn4p activation domain is at least 81 amino acids long (see CIp-*lexA-GCN4Δ69-323*, CIp-*lexA-GCN4Δ56-323*, and CIp-*lexA-GCN4Δ42-323*) (Figure 3.2.3). The deletion of the proposed N-terminal activation domain in the context of the full length *GCN4* (CIp-*lexA-GCN4Δ1-81*) showed the same activity of β-galactosidase in CMM21 compared to the activity of full length *GCN4* in CMM14, while CIp-*lexA-GCN4Δ1-81* in CMM37 showed a slightly reduced *HIS1* activity compared to CMM30. This observation could be explained either by a second transcription activation domain (as in *S. cerevisiae* Gcn4p) located between the N-terminal activation domain and C-terminal DNA binding domain, or due to the CIp-*lexA-GCN4Δ1-81* interaction with the endogenous wild type Gcn4p through a dimerization domain located at the C' terminus of the protein (126), generating an activating heterodimer. We directly tested the capacity of the region between the DNA binding domain and the N-terminal activating domain to

allow transcriptional activation by creating CIp-*lexA-GCN4Δ1-81Δ247-323*, in which both the activation and the DNA binding domains were deleted; this construct generated background β-galactosidase activity and no *HIS1* activity (see CMM64 and CMM65), suggesting that either the activation domain at the N' terminus is the only CaGcn4p activation domain, and that CIp-*lexA-GCN4Δ1-81Δ247-323* yields an unstable protein, or that the Gcn4p DNA binding domain directs transcriptional activation. To distinguish these hypotheses we tested *lexA-GCN4Δ1-81*, which lacks N-terminal activation domain, for its ability to activate the expression of *lacZ* in the absence of endogenous Gcn4p (*gcn4* strain). Although *lexA-GCN4Δ1-81* resulted in high expression of *lacZ* in the wild type strain (CMM21), the *lacZ* expression was dropped down to background levels in the absence of the endogenous *GCN4* (CMM88). At the same time, *lexA-GCN4* showed high *lacZ* expression in both wild type and *gcn4* backgrounds (CMM14 and CMM87), while *lexA* and *lexA-GCN4Δ1-81Δ247-323* showed low *lacZ* levels of expression in both *GCN4* and *gcn4* backgrounds (CMM10, CMM86, CMM64, and CMM89). These results suggest that the activation domain at the N' terminus of *C. albicans* Gcn4p is the only CaGcn4p activation domain. At the amino acid level, this activation domain is nucleophilic (has a composition of 20% serine-threonine) and shares no similarity with the activation domain of ScGcn4p.

Definition of the CaGal4p-homologue activation domain

We examined the activation domain of a candidate *C. albicans* version of the yeast Gal4p protein. In *S. cerevisiae* Gal4p is a highly studied transcription factor, and the structure of its DNA-binding and transcriptional activation modules as well as the target

promoters have been extensively investigated (192, 237, 291). Within the DNA binding domain, the putative Gal4p protein, encoded by *C. albicans ORF19.5338*, shares strong sequence similarity (86%) with *S. cerevisiae* Gal4p; the DNA binding domain has the six cysteine residues, the linker region, and the dimerization region all well conserved. A blast search of the ScGal4p sequence in the *C. albicans* genome yields Orf19.5338 as its closest homologue; at the same time, searching the Orf19.5338 sequence in the *S. cerevisiae* genome yields ScGal4p as its closest homologue. Since ScGal4p and Orf19.5338 form a "reciprocal best hit" relationship, we named Orf19.5338 CaGal4p. Although this *C. albicans* Gal4p homologue binds 5'-CGGN_{11}CCG-3', the upstream activating sequence (UAS_G) to which Gal4p binds in *S. cerevisiae* (data not shown), the promoters of *C. albicans GAL* genes lack UAS_G. Rather, UAS_G are found upstream of *C. albicans* subtelomeric and glycolysis genes (data not shown). In addition, this *C. albicans* gene encodes a much smaller 261 amino acids long protein, compared to *S. cerevisiae* Gal4p of 881 amino acids. The regions outside the DNA binding domain of those two proteins share no similarity, and the negatively charged region that serves as the interaction domain for ScGal80p is missing in CaGal4p. Interestingly, although *C. albicans* can grow on galactose (89), its genome also lacks a Gal80p homologue.

To establish if CaGal4p contains a transcriptional activation domain, full-length *GAL4* was cloned downstream of the *lexA* open reading frame to create CIp-*lexA-GAL4*. We transformed this construct into strain CRC106 to create CMM11; CIp-*lexA-GAL4* showed five times more trans-activating ability compared to the activity of the control vector CIp-*lexA*, which was transformed into CRC106 to create strain CMM10 (Figure 3.2.2). Similarly, strain CMM25, which contains the *lexA* operator in front of *HIS1*, was

transformed with CIp-*lexA* and CIp-*lexA-GAL4* to create CMM26 and CMM27 respectively. As was found for the *lacZ* system, *lexA-GAL4* showed high transcription activating ability compared to the activity of the vector alone, since only CMM27 grew in the absence of histidine. These results suggest that *C. albicans* Gal4p can act as a transcriptional activator.

We examined which part of CaGal4p was essential for the trans-activating capacity. Deletion of the C-terminal 71 amino acids, creating (CIp-*lexA-GAL4Δ188-261*), abolished the trans-activating ability of *lexA-GAL4* when introduced into both CRC106 to create CMM13 and into CMM25 to create CMM29. As well, fusion of *lexA* to the C-terminal 71 amino acids of CaGal4p showed transcription activating abilities similar to the full-length *GAL4* (see CMM60 and CMM61) (Figure 3.2.4). Similarly to the Gcn4p activation domain, the CaGal4p activation domain showed a nucleophilic nature; it has a 30% serine-threonine composition, but shares no other similarity with the activation domain of ScGcn4p.

It is currently believed that the gene specificity of the transcription factor comes from its DNA binding domain: this domain binds to a nucleotide motif on the promoters and recruits the RNA Polymerase II machinery (178). To see whether the transcriptional activation domain plays a role in the transcriptional selectivity of CaGal4p, we replaced the 71 amino acid-long CaGal4p activation domain (Gal4AD) with the 81 amino acid-long Gcn4p activation domain (Gcn4AD). We observed that *in vivo* the *lexA-GAL4-GCN4AD* construct in CMM12 and CMM28 showed the same transcription activating ability as the *lexA-GAL4-GAL4AD* in CMM11 and CMM27 respectively (Figure 3.2.4).

3.2.5 DISCUSSION

Eukaryotic transcription factors are typically bipartite in nature, with a region (the DNA binding or DB domain) specifically designed to interact with a defined DNA sequence, and a region (the transcriptional activation or TA domain) required to interface the factor with the transcriptional machinery. There are several classes of each of these modules, and they are connected together in a variety of ways. Within the transcriptional activation modules there are domains rich in acidic or basic residues, or rich in glutamine, threonine/serine, or isoleucine residues (8, 38, 61, 83, 125, 146, 201, 202). In this study we defined the transcription activation domains in a pair of *C. albicans* transcription factors that share sequence similarity with their *S. cerevisiae* homologs only within their DNA binding domains. The Zn(II)2Cys6 (or C6 zinc) binuclear cluster DNA binding domain is one of the largest classes of fungal DNA binding proteins, the best characterized of which are Gal4p, Ppr1p, Leu3p, Hap1p, and Put3p. Although the DNA binding sequence of ScPut3p (CGGN_{10}CCG) is very similar to that of ScGal4p (CGGN_{11}CCG), the distinction in recognition sequences in conserved; *C. albicans* possesses homologs of all of these *S. cerevisiae* Gal4-like Zn(II)2Cys6 proteins, including Put3p.

In *S. cerevisiae* it was shown that Gal4AD and Gcn4AD have an acidic amino acid – rich nature and are located in the C' and N' termini respectively (78, 125, 141, 201, 202). We analyzed the transcription activation domains of the *C. albicans* Gcn4p (*CaGCN4*) and Gal4p (*CaGAL4*) homologs, and found that just as in *S. cerevisiae* they are positioned at the N' and C' termini of the respective proteins. However, the *C. albicans* Gcn4p and Gal4p activation domains do not share sequence similarity either to each other or to the activation domains of their *S. cerevisiae* homologs, and *C. albicans*

Gal4p and Gcn4p have nucleophilic activation domains. Nucleophilic transcriptional activation regions have been previously seen almost exclusively in higher eukaryotic transcription factors (50, 61, 110). A screen for *C. albicans* transcriptional activation domains using a genomic library fused downstream of *lexA* yielded an active fragment containing a normally non-coding region that expressed 33% serines and threonines in the fusion construct (data not shown), which also suggests that nucleophilicity can be an important feature of *C. albicans* activation domains. The serine and threonine amino acids could potentially be converted into an acidic form by phosphorylation.

The *S. cerevisiae* Gal4p and Gcn4p proteins each contain two transcriptional activation domains (78, 126, 141, 201, 202). In contrast, the *C. albicans* Gcn4p and Gal4p homologs appear to each contain only one transcriptional activation domain (Figures 3.2.3 and 3.2.4). Each of the two ScGcn4p activation domains seems to be composed of two or more small sub-domains that have additive effects on transcription and that can cooperate in different combinations to promote high-level expression of the Gcn4p-dependent genes (78, 141). These results are consistent with our observation that the C-to-N-terminal deletions within the CaGcn4p activation domain lead to a gradual, rather than to an abrupt, reduction of the transcription activating abilities of the fusion protein (Figure 3.2.3).

To determine when the changes in the activation domains of Gcn4p and Gal4p occurred during the evolution of the yeast species, we used available genomic data of the ascomycota (*Schizosaccharomyces pombe*, *Neurospora crassa, Aspergillus niger, S. cerevisiae, S. paradoxus, S. mikatae, S. kudriavzevii, S. bayanus, C. glabrata, S. castellii, Kluyveromyces lactis, Ashbya gossypii, Debaryomyces hansenii, C. tropicalis, C.*

dubliniensis, and *C. albicans*) (Figure 3.2.5A and 3.2.B). Archiascomycetes were observed to lack the activation domains of either ScGcn4p or CaGcn4p. Euascomycetes possessed the ScGcn4p activation domain II (ADII). We noted that the common ancestor of *D. hansenii, C. tropicalis, C. dubliniensis,* and *C. albicans* lost the activation domain I (ADI) of ScGcn4p and acquired the activation domain of CaGcn4p. We also observed that the ancestor of *C. tropicalis, C. dubliniensis,* and *C. albicans* lacked ScGcn4p ADII (Figure 3.2.5A). In addition, the ancestor of *C. tropicalis, C. dubliniensis,* and *C. albicans* lacked ScGal4p ADI, while the ancestor of *D. hansenii, C. tropicalis, C. dubliniensis,* and *C. albicans* missed ScGal4p ADII (Figure 3.2.5B). In both cases, *D. hansenii* represents an intermediate with both *S. cerevisiae* and *C. albicans* activation domains. These observations show that the changes in the activation domain of Gal4p and Gcn4p of *C. albicans* occurred relatively recently on the evolutionary scale.

The lack of homology in the activation domains of transcriptional activators between *S. cerevisiae* and *C. albicans* might suggest a concomitant reduced structural similarity in the activation-domain-interacting complexes between the two species. A pair-wise sequence comparison of the transcriptional machinery between *C. albicans* and *S. cerevisiae* shows a high level of conservation in the RNA polymerase II complex. The exceptions for this are TFIIA and the Mediator complex: *S. cerevisiae* and *C. albicans* show low levels of homology with respect to the proteins of those two complexes, that also interact with transcriptional regulators (Figure 3.2.6, Table 1.3.1).

The characterization of the bipartite structure of eukaryotic transcription factors like yeast Gal4p was a fundamental conceptual advance (152), and has led to important technical developments like the yeast two-hybrid system (85). In general *C. albicans*

transcription factors follow the pattern of distinct DNA binding and transcriptional activation domains, and many show strong sequence similarity, extending to both domains, to specific yeast transcription factors. However, a large number of *C. albicans* proteins have strong sequence similarity that is limited only to the DNA binding module of a yeast transcription regulator. We have shown that although the well studied Gal4p and Gcn4p proteins of yeast share similarity only to the DNA binding regions of the Gcn4p and Gal4p proteins of *C. albicans*, the *Candida* proteins still contain transcriptional activation capacity. Further work will be necessary to establish the molecular logic of linking common DNA binding modules to distinct activation domains in these two fungi, in particular in cases such as Gcn4p where similar cellular processes are regulated by the two proteins.

Table 3.2.1: Genome-wide comparison of transcriptional activators between *C. albicans* and *S. cerevisiae.*

One hundred and sixty six transcriptional activators of *S. cerevisiae* were tested by Blast against the *C. albicans* genome to predict 99 *C. albicans* homologues out of 216 transcription factors. Fifty of those 99 transcription factors share homology in both their DNA binding domains and potential transcriptional activation domains, while 49 share sequence similarity only within the DNA binding domain. A pair-wise sequence comparison was done on the full length (FL), DNA binding domain (DBD), and on the sequence outside of the DBD (Non-DBD). Analysis was done using a standard pair-wise blastp setting on http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi. The values are

shown in "-LOG10(e-values)", which is a statistical measure of the similarity between the two proteins: the higher the value of P – the more homologous the two proteins are.

Sc TF Name	FL	DBD	Non-DBD	orf19.####	Ca TF Name
Sefl	2.00E-169	5.00E-25	4.00E-147	orf19.3753	Sefl
Put3	3.00E-163	9.00E-14	1.00E-143	orf19.6203	Put3
Asg1	7.00E-132	1.00E-16	1.00E-126	orf19.166	Asg1
Dal81	2.00E-130	0.004	4.00E-122	orf19.3252	Dal81
Aro80	2.00E-130	3.00E-14	3.00E-116	orf19.3012	Aro80
Upc2	6.00E-116	1.00E-13	6.00E-103	orf19.391	Upc2
Rds2	3.00E-108	1.00E-07	5.00E-102	orf19.5849	Cwt1
Stb5	6.00E-98		4.00E-98	orf19.3308	Stb5
Snt2	1.00E-99	3.00E-13	1.00E-86	orf19.1259	orf19.1259
Arg81	5.00E-83	1.00E-13	4.00E-83	orf19.4766	Arg81
Ppr1	1.00E-92	4.00E-11	8.00E-80	orf19.3986	Ppr1
YER051W	2.00E-69		2.00E-69	orf19.3281	orf19.3281
Cat8	6.00E-94	3.00E-22	5.00E-68	orf19.5097	Cat8
YBR239C	6.00E-76	2.00E-11	1.00E-62	orf19.2423	Zcf11
YML081W	8.00E-79	8.00E-23	2.00E-61	orf19.5026	orf19.5026
Sip3	3.00E-94	7.00E-24	6.00E-58	orf19.3047	orf19.3047
Stb4	1.00E-57		3.00E-57	orf19.4288	Cta7
YPR022C	1.00E-61	8.00E-13	2.00E-48	orf19.7397	orf19.7397
Tea1	8.00E-47	1.20E-02	9.00E-45	orf19.6985	Tea1
War1	4.00E-54	2.00E-08	2.00E-44	orf19.1035	War1
Pdc2	2.00E-73	6.00E-18	2.40E-44	orf19.4863	Pdc2
Skn7	1.00E-76	5.00E-32	9.00E-42	orf19.971	Skn7
Uga3	3.00E-50	7.00E-12	3.00E-38	orf19.7570	Uga3
Pho2	4.00E-48	2.00E-10	2.00E-34	orf19.4000	orf19.4000
Lys14	4.00E-46	4.00E-08	1.00E-33	orf19.4776	Lys143
Mbp1	3.00E-72	1.00E-38	2.00E-29	orf19.5855	orf19.5855
Spt23	2.00E-68	1.00E-39	3.40E-28	orf19.1751	Spt23
Stp2	3.00E-37	6.00E-05	1.00E-23	orf19.4961	Stp2
Sfp1	1.00E-70	5.00E-46	5.00E-23	orf19.5953	Sfp1
Yap1	1.00E-23	1.00E-03	1.00E-20	orf19.1623	Cap1
Stp4	3.00E-33	1.00E-14	3.00E-19	orf19.909	Stp4
Mal13	1.00E-33	5.00E-05	5.00E-16	orf19.7319	Suc1
Yap3	8.00E-15		8.00E-15	orf19.3193	Fcr3
Msn1	2.00E-13		2.00E-13	orf19.3328	orf19.3328
Rts2	2.00E-31	5.00E-12	2.00E-13	orf19.3817	orf19.3817
Bas1	5.00E-60	7.00E-40	1.00E-08	orf19.3809	Bas1
Hal9	4.00E-11	1.40E-01	2.00E-07	orf19.3188	Tac1
Ste12	2.00E-72	4.00E-47	6.00E-07	orf19.4433	Cph1
YDR266C	2.00E-74	3.00E-67	7.00E-07	orf19.3449	orf19.3449
Met4	7.00E-05		2.00E-06	orf19.5312	orf19.5312
Azf1	3.00E-67	5.00E-58	9.00E-06	orf19.173	orf19.173
Gts1	9.00E-44	5.00E-39	1.00E-05	orf19.6393	orf19.6393
Adr1	4.00E-23	9.00E-14	2.00E-05	orf19.2752	Adr1
Mcm1	2.00E-26	1.00E-14	3.00E-04	orf19.7025	Mcm1

Dot6	3.00E-18	8.00E-18	0.003	orf19.2545	Dot6
Aca1	2.00E-09	3.00E-07	0.011	orf19.6102	orf19.6102
Ace2	1.00E-29	3.00E-21	0.03	orf19.6124	Ace2
Imp2'			0.35	orf19.3136	orf19.3136
Hap1	5.00E-08	1.00E-09	0.41	orf19.7372	Zcf36
Hac1	5.00E-17	5.00E-17	0.999	orf19.2432	Hac1
Gzf3	9.00E-19	4.00E-20	2.1	orf19.2842	Gzf3
Ino4	3.00E-03	3.00E-03		NA	NA
Sko1	4.00E-07	3.00E-08		orf19.1032	Sko1
Rpn4	5.00E-15	3.00E-16		orf19.1069	Rpn4
Flo8	0.12	0.12		orf19.1093	Flo8
Hap2	5.00E-23	6.00E-24		orf19.1228	orf19.1228
Pho4	5.00E-05	4.00E-06		orf19.1253	orf19.1253
Gat1	3.00E-19	1.00E-20		orf19.1275	Gat1
Gcn4	4.00E-18	7.00E-11		orf19.1358	Gcn4
Gat3	1.00E-04	9.00E-06		orf19.1577	orf19.1577
Met31	1.00E-15	7.00E-17		orf19.1757	Met31
Rap1	7.00E-17	9.00E-18		orf19.1773	Rap1
Hap5	2.00E-42	1.00E-42		orf19.1973	Hap5
Ndt80	9.00E-49	4.00E-49		orf19.2119	Ndt80
Fhl1	6.00E-51	1.00E-52		orf19.2236	orf19.2236
Rtg3	1.00E-15	9.00E-17		orf19.2315	orf19.2315
Ecm22	0.21	0.007		orf19.2623	Ecm22
Rgt1	1.00E-08	4.00E-10		orf19.2747	Rgt1
Cbf1	2.00E-28	3.00E-22		orf19.2876	Cbf1
Zap1	7.00E-51	5.00E-51		orf19.3794	Csr1
Gln3	4.00E-15	3.00E-16		orf19.3912	orf19.3912
Mga1	9.00E-11	5.00E-12		orf19.3969	orf19.3969
Leu3	6.00E-130	6.00E-19		orf19.4225	Leu3
Ifh1	2.00E-07			orf19.4281	orf19.4281
Sut1	7.00E-04	3.00E-04		orf19.4342	orf19.4342
Rme1	6.00E-07	4.00E-08		orf19.4438	Rme1
Hap3	9.00E-25	1.00E-25		orf19.4647	Hap3
Rlm1	5.00E-25	9.00E-20		orf19.4662	Rlm1
Rtg1	7.00E-24	1.00E-24		orf19.4722	orf19.4722
Msn4	1.00E-17	4.00E-19		orf19.4752	Msn4
Hcm1	1.00E-26	7.00E-28		orf19.4853	Hcm1
Tye7	9.00E-15	2.00E-15		orf19.4941	Tye7
Cup2	4.00E-16	3.00E-14		orf19.5001	Cup2
Gal4	1.00E-18	7.00E-19		orf19.5338	Gal4
Fkh2	3.00E-89	2.00E-89		orf19.5389	Fkh2
Efh1/Sok2	1.00E-31	5.00E-33		orf19.5498	Efh1
YDR026C	1.00E-44	1.00E-45		orf19.5722	orf19.5722
Tec1	5.00E-15	3.00E-14		orf19.5908	Tec1
Stp1	2.00E-30	2.00E-31		orf19.5917	Stp3
Efg1/Sok1	7.00E-50	3.00E-48		orf19.610	Efg1
Msn2	1.00E-13	4.00E-15		orf19.6121	Mnl1
Cup9	8.00E-13	9.00E-14		orf19.6514	Cup9
Met28	2.3	2.3		orf19.7046	Met28

Mac1	2.00E-06	3.00E-07	orf19.7068	Mac1
Usv1/Bcr1	1.00E-21	5.00E-23	orf19.723	Bcr1
Rim101	8.00E-32	4.00E-33	orf19.7247	Rim101
Crz1	1.00E-37	1.00E-35	orf19.7359	Crz1
Hap4	9.00E-05		orf19.740	Hap4
Oaf1	1.00E+00	3.40E+00	orf19.7583	ZCF39

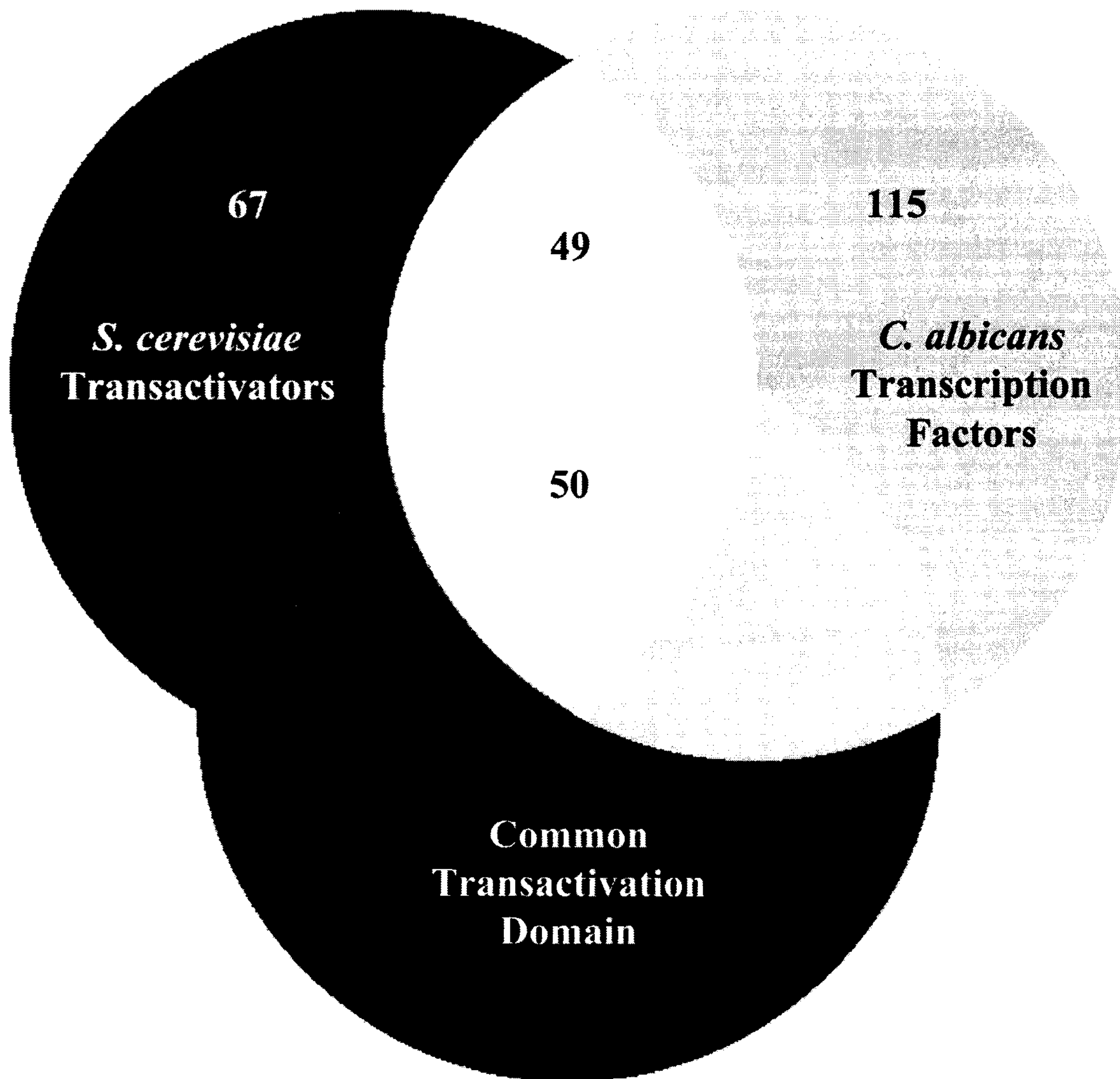


Figure 3.2.1. Genome-wide comparison of transactivators between *C. albicans* and *S. cerevisiae*.

Ninety-nine of the transcriptional activators of *S. cerevisiae* share sequence similarity to the *C. albicans* transcription factors. Fifty of those 99 transcription factors share homology in both their DNA binding domains and potential transactivation domains, while 49 share sequence similarity only within the DNA binding domain.

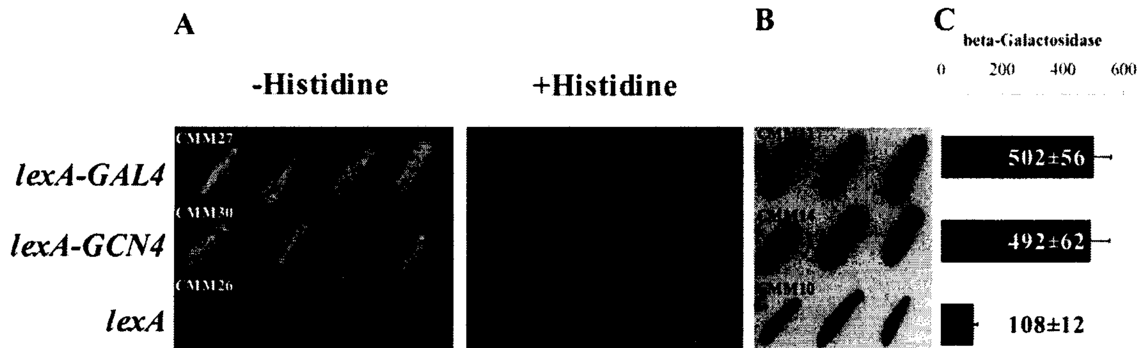


Figure 3.2.2. Gal4p and Gcn4p act as transcriptional activators in *C. albicans.*

C. albicans GCN4 and *GAL4* were fused to *lexA*, which binds to *lexA* binding DNA sequence upstream of *HIS1* and *lacZ.* The activities of the *HIS1* and *lacZ* reporters were analyzed in three independent transformants.

A. Cells were grown for 1 day in the absence of histidine.

B. *Candida* cells grew overnight on selective media; cells were permeabilized by chloroform, and overlaid with X-gal. The color was allowed to develop for 4 hours. Liquid beta-galactosidase assays are expressed in Miller units.

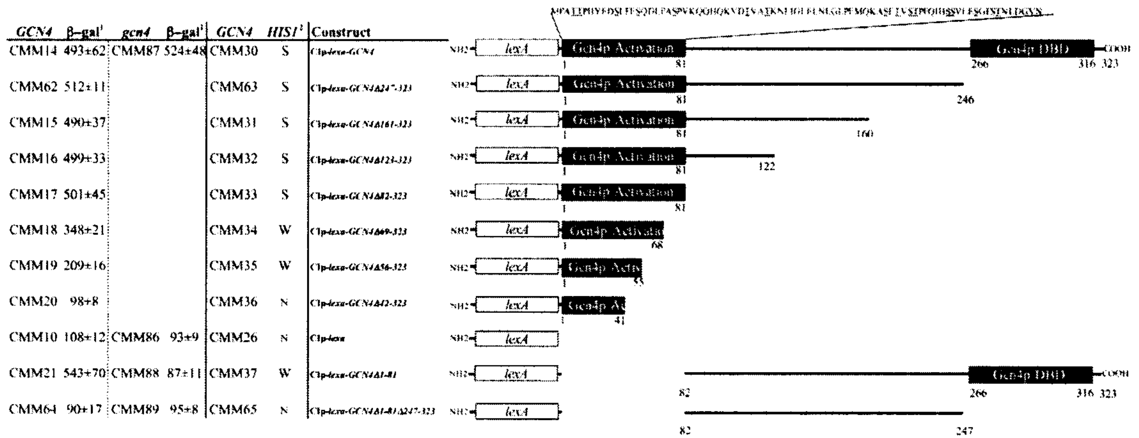


Figure 3.2.3. Identification of CaGcn4p Transcriptional Activation Domain.

Structural and functional analysis of *lexA-GCN4* derivatives. The structures of various *lexA-GCN4* derived proteins are indicated by grey and black bars respectively, with the N- and C- terminal residues defined as in the wild type proteins. 1: Liquid beta-galactosidase assays are expressed in Miller units; 2: the expression of the *HIS1* reporter is either strong, S (1 day it took for a strain to grow in the absence of histidine), weak, W (2 days it took for a strain to grow in the absence of histidine), or no growth, N.

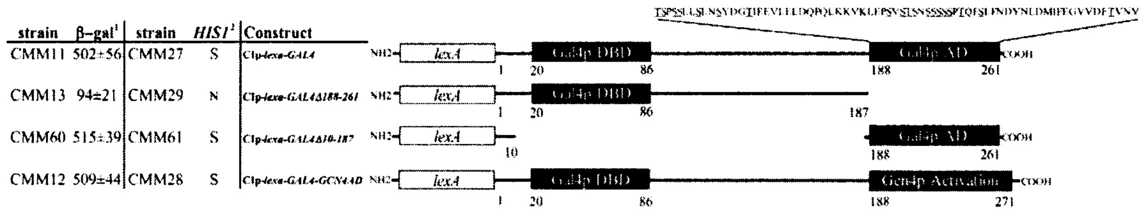


Figure 3.2.4. Identification of CaGal4p Transcriptional Activation Domain.

Structural and functional analysis of *lexA-GAL4* derivatives. The structures of various *lexA-GAL4* derived proteins are indicated by grey and black bars respectively, with the N- and C- terminal residues defined as in the wild type proteins. 1: Liquid beta-galactosidase assays are expressed in Miller units; 2: the expression of the *HIS1* reporter is either strong, S (1 day it took for a strain to grow in the absence of histidine), weak, W (2 days it took for a strain to grow in the absence of histidine), or no growth N.

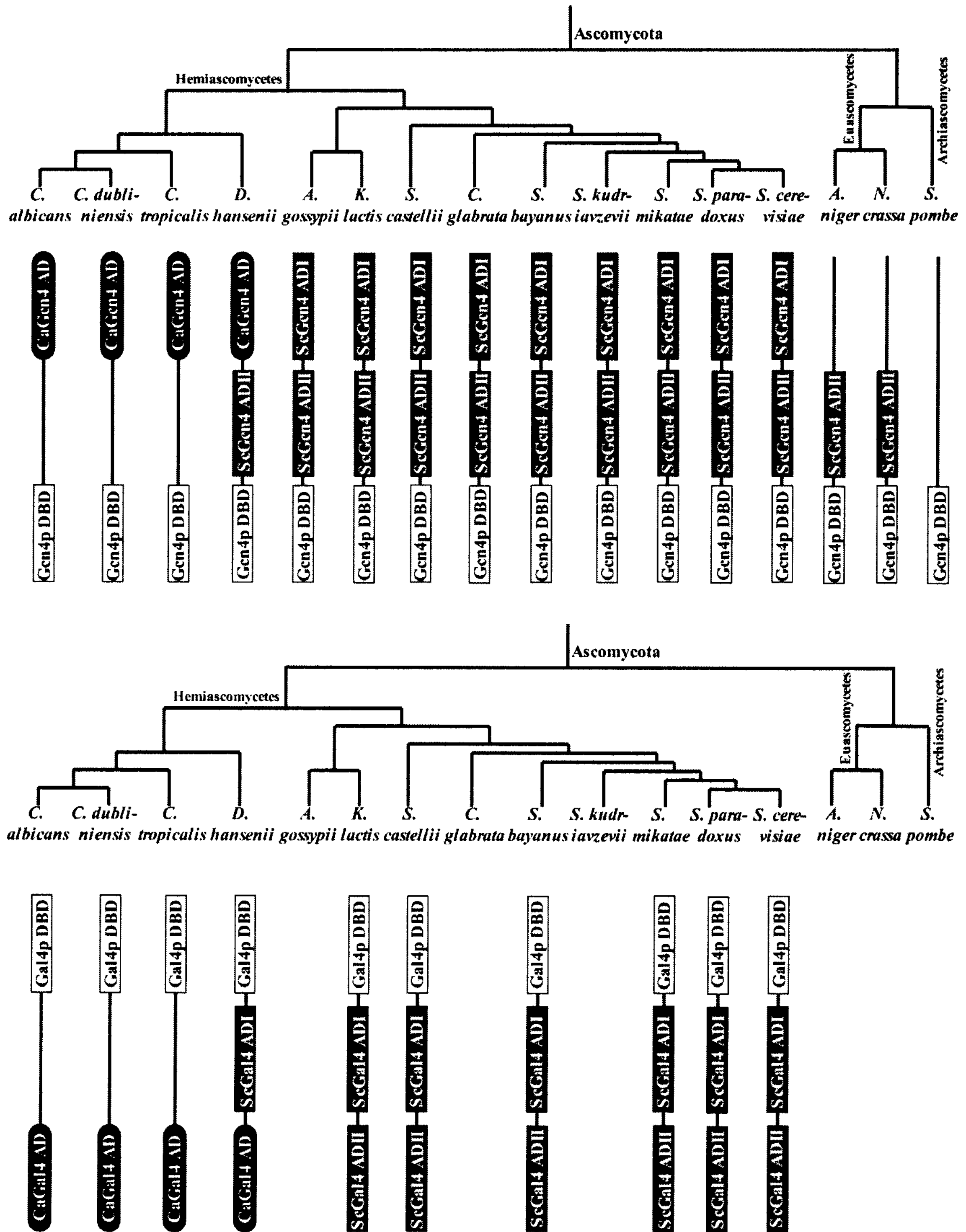


Figure 3.2.5. The phylogenetic comparison of Gcn4p and Gal4p orthologs across the Ascomycota. The changes in the Gcn4p activation domain of *C. albicans* occurred relatively recently on the evolutionary scale. We used available genomic data of the

Ascomycota *Schizosaccharomyces pombe*, *Neurospora crassa, Aspergillus niger, S. cerevisiae, S. paradoxus, S. mikatae, S. kudriavzevii*, *S. bayanus, C. glabrata, S. castellii, Kluyveromyces lactis, Ashbya gossypii, Debaryomyces hansenii, C. tropicalis, C. dubliniensis,* and *C. albicans* to determine at what evolutionary stage the Gcn4p (A) and the Gal4p (B) have changed the sequence of their activation domains. The sequences of CaGcn4p, ScGcn4, CaGal4p and ScGal4p were blasted against the genomes of ascomycota using the standard blastp setting on http://seq.yeastgenome.org/cgi-bin/blast-fungal.pl, and confirmed using a standard pair-wise blastp setting on http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi. ScGcn4 activation domain II (ScGcn4 ADII) lies between amino acids 107-144, and ScGcn4 ADI is between amino acids 12-92. ScGal4 ADI is between amino acids 149-196, and ScGcn4 ADII is between amino acids 768-881. *A. gossypii, C. glabrata,* and *S. kudriavzevii*, are unable to metabolize galactose and lack Gal4p. *S. pombe*, *N. crassa,* and *A. niger* do not posses close Gal4p homologue. The phylogenetic relationship between the fungi is shown as in (74).

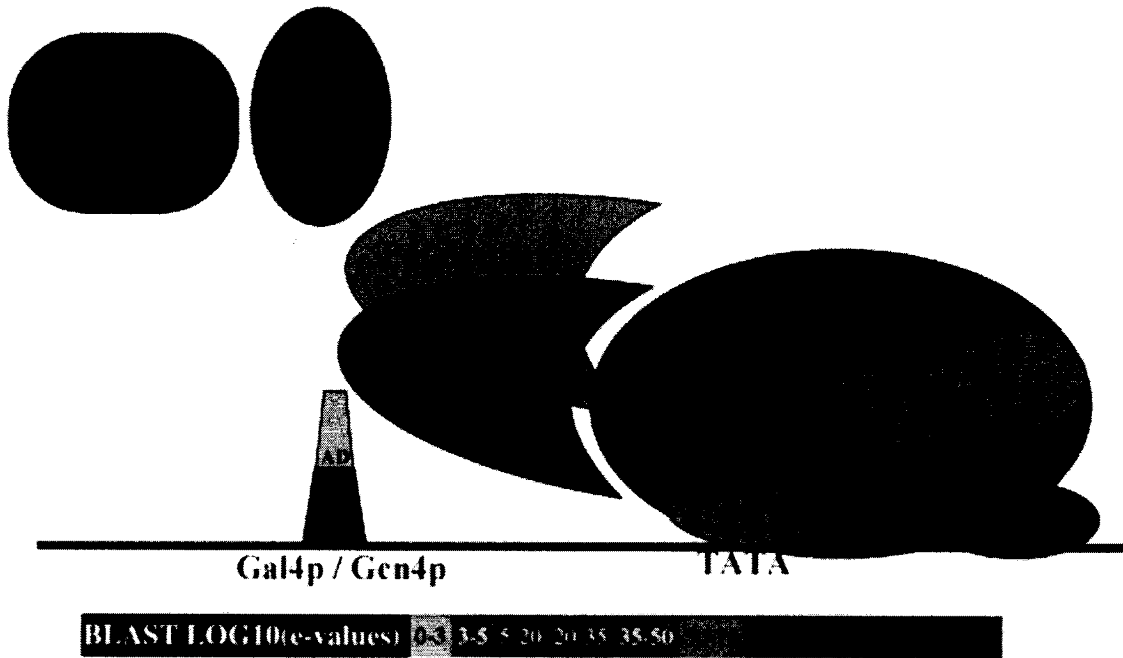


Figure 3.2.6. The comparison of the transcriptional machinery between *C. albicans* and *S. cerevisiae*.

The constituents of the transcriptional complexes were compared using a standard pair-wise blastp setting on http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi. The log values were averaged and shown by the associated color.

CHAPTER 4: POSTGENOMIC STUDIES OF *CANDIDA ALBICANS*: RESULTS OF THE COMPARATIVE GENOMICS STUDIES OF *C. ALBICANS* INTERGENIC REGIONS.

Section 4.1: Transcriptional rewiring in *C. albicans.*

With the recent sequencing of many yeast genomes, the hemiascomycete lineage emerged as a central paradigm for studying the genetic basis of phenotypic diversity (238). Despite a strong conservation of gene content, yeast species exhibit major phenotypic differences. For example, almost two-thirds of *C. albicans*' ~6,000 open reading frames are orthologous to *S. cerevisiae* genes, yet their life styles are very different, probably due to the difference in gene expression. Differences in gene expression underlie many of the phenotypic variations between related organisms, yet approaches to characterize such differences on a genome-wide scale are not well developed. The availability of the genome sequences of both *S. cerevisiae* and *C. albicans*, which are related organisms that span a significant evolutionary distance, provides a useful framework to develop computational tools for comparative genomics.

Phenotypic diversity can often be traced to the differential expression of specific regulatory genes (1, 45, 101, 296, 310). Recently, microarray experiments revealed large-scale differences in the genome-wide transcription response of related organisms to equivalent environmental conditions. For example, the transcription program underlying insect metamorphosis differs considerably between related species of the *Drosophila melanogaster* subgroup (250). Similarly, both the meiotic and the mitotic cell cycle transcription program have diverged significantly between the budding and the fission

yeasts (257). The impact of such large-scale variations in gene expression on the phenotypes of the organisms is not yet understood.

Transcriptional modules, i.e., groups of coregulated genes, play a central role in the organization and function of regulatory networks (137, 267, 283). Comparative studies have demonstrated that various transcriptional modules are highly conserved across a wide variety of organisms from *Escherichia coli* to humans (17, 214, 278). A common assumption is that conserved regulatory mechanisms underlie module conservation, because coregulation imposes tight constraints on the evolution of a module's promoters. Indeed, recent studies showed that orthologous transcriptional modules are often associated with conserved *cis*-elements (96). Divergence of *cis*-elements in specific promoters has been documented, but it is difficult to explain the divergence in a regulatory program associated with coexpressed genes (194, 195).

The use of comparative genomics to identify putative regulatory binding sites has attracted a lot of attention since the sequence of three non-*cerevisiae Saccharomyces* species was made public (153). It appears that for most known transcription factors the conservation of the DNA binding site amongst related species is a reliable detection strategy. *C. albicans* and its two closely related species *C. dubliniensis* and *C. tropicalis* (sequenced at the Sanger Centre and the Broad Institute respectively) form a comparative genomic system similar to the four *Saccharomyces* species. Although *C. tropicalis* is more distant to *C. albicans* than any of the *Saccharomyces* species, the gene conservation and the high level of synteny with *C. albicans* makes *C. tropicalis* an excellent organism for a comparative genomics study.

The previously discussed genome assembly and the release of a genome

annotation of *C. albicans* (33) also provide functional comparison of the two groups of yeasts. In this study, we show that the identification of *cis*-regulatory elements from three *Candida* species (*C. albicans, C. dubliniensis, C. tropicalis*) and four *Saccharomyces* species (*S. cerevisiae, S. paradoxus, S. mikatus, S. bayanus*) shows reorganization of the transcriptional regulatory network between these two groups of yeasts.

For each group of yeast the contigs were first aligned and the intergenic sequences from syntenic regions extracted. A constrained sampling method was then used to find all the conserved mini-alignments from each intergenic region. To identify regulatory elements, the intersection between the set of genes that have a conserved mini-motif and every Gene Ontology (GO) set was systematically tested and scored. Randomized gene sets were used to establish the false discovery threshold. Every motif-GO pair found in one genus was further tested and scored in the other genus. Our analysis rediscovered most known *Saccharomyces* regulatory motifs as well as new motifs specific or common to *Saccharomyces* and *Candida*. The comparison of the detected motifs and their target genes reveals important rewiring events.

Many conserved transcription factors in *Saccharomyces* and *Candida* have kept their preferred DNA binding motif but the sets of genes that they regulate show significant variability. For example Gal4p, Ino4p and Cbf1p are involved in such rewiring events. The Gal4p binding site (CGG(N_{11})CC) is no longer found conserved upstream of galactolysis genes in *Candida*. Instead, it is enriched in glycolysis genes and a different motif is found to regulate galactolysis in *Candida*. The Ino4p motif (CATGTGAA), upstream of lipid biosynthesis genes in *Saccharomyces* is conserved upstream of multiple peroxisomal genes in *Candida*. The centromere binding factor

Cbf1p (CACGTG), while maintaining its role in sulfur and methionine metabolism (24), is also found upstream of more than 25 ribosomal genes in *Candida*. These examples illustrate the adaptation of transcriptional regulation over the 100 million years of evolution that separate the *Saccharomyces* and *Candida* species and provide insight on fundamental aspects of their respective biology.

The Gal4p mediated activation of the Leloir pathway (*GAL1*, *GAL7*, *GAL10*) genes of *S. cerevisiae* in response to galactose is among the most intensively studied and best understood of all metabolic regulatory pathways. Microarray analysis of *C. albicans*, grown on galactose as the sole carbon source shows transcriptional activation of the *Candida* homologs of the *GAL1*, *2*, *10*, and *7* genes, while loss of CaGal4p does not modulate growth on galactose or the transcriptional activation of the Leloir pathway genes. CaGal4p instead acts as a transcriptional activator of the telomeric *TLO* gene family as well as genes whose products are involved in glycolysis and the TCA cycle. We identified *cis*-acting regions in the promoters of the *C. albicans GAL* genes, including a unique *GAL*-specific palindrome. Rim101p and Rmd5p transactivate the expression of *GAL10* in the presence of dextrose and galactose respectively, while Cph1p acts as an activator in galactose through the *GAL*-specific palindrome. Our studies show that *C. albicans* and *S. cerevisiae* can regulate the same process by different regulatory circuits.

The full spectrum of evolutionary events we discovered, encompassing both conservation and divergence, provides a general framework for the study of the evolution of transcription regulation and highlights the flexibility and evolvability of *cis*-regulatory programs. The evolution of transcriptional modules is an important aspect in understanding regulatory networks.

Transcription rewiring within the galactose metabolism circuitry of *Candida albicans.*

Mikhail Martchenko[1], Anastasia Levitin[2], Wiriya Chiranand[3], Herve Hogues[2], Andre Nantel[2], Michael Gustin[3], and Malcolm Whiteway[1,2*].

[1]Department of Biology, McGill University, Montreal, Quebec, Canada H3A 1B1;
[2]Genetics Group, Biotechnology Research Institute, National Research Council, Montreal, Quebec, Canada H4P 2R2
[3]Dept Biochemistry, Rice University, Houston, TX, USA 77251-1892

4.1.1 Abstract

The Leloir pathway genes encode the enzymatic machinery involved in the metabolism of galactose. In the distantly related fungi *Saccharomyces cerevisiae* and *Candida albicans* the genes encoding these enzymes are syntenically arranged but the upstream regulatory regions are highly divergent. In *S. cerevisiae* the Leloir pathway genes are positively regulated by Gal4p acting through the UAS_G sequence $CGG(N_{11})CCG$. However, in *C. albicans* the Gal4p and UAS_G combination is found to regulate genes unrelated to galactose metabolism. We identified a palindromic sequence that acts to control *GAL10* expression in *C. albicans* in the presence of galactose. This palindrome is found upstream of other Leloir pathway genes in *C. albicans*, and in the absence of other regulatory sequences, activation of expression through this sequence in the presence of galactose requires Cph1p, the homolog of the Ste12p transcription factor

of *S. cerevisiae*. Thus, although the cellular process of galactose induction of the Leloir pathway is conserved between the two organisms, the regulatory circuits achieving the cellular process are completely distinct.

4.1.2 Introduction:

The diploid fungal pathogen, *Candida albicans*, which causes the majority of human fungal infections, diverged from the bakers or brewers yeast *S. cerevisiae* approximately 250 million years ago (116). With the recent sequencing of many yeast genomes, the hemiascomycete lineage has emerged as an important model for studying the genetic basis of functional diversity (238), and despite a strong conservation of gene content, yeast species exhibit major phenotypic differences. For example, almost two-thirds of *C. albicans*' ~6,000 open reading frames are orthologous to *S. cerevisiae* genes (33), yet the organism's life-styles are very different. This suggests that differential regulation of gene expression, and not simply differential gene content, plays an important role in phenotypic variation among related organisms.

The evolution of gene regulation could occur either through modifications to the *trans*-acting regulatory protein or by changes to the *cis*-acting DNA elements in promoters. Recent studies show that both processes function to direct transcriptional rewiring, which can drive regulatory divergence (121, 298). For example, a group of mating genes that is transcriptionally regulated by an activator in *C. albicans* is transcriptionally regulated by a repressor in *S. cerevisiae*, a modification that was achieved by changing both *cis* and *trans*-acting elements (298).

Because many of the fundamental biological processes of *C. albicans* are similar to those of *S. cerevisiae*, budding yeast is an important guide for studying the pathogen. Yet, as these related organisms span a significant evolutionary distance there are also significant differences in cellular function. Unlike *S. cerevisiae*, *C. albicans* can form true hyphae, can escape mammalian immune cells, and can proliferate in mammalian hosts, and thus has the tools to be a pathogen. Despite a strong conservation of gene content, yeast species exhibit major phenotypic differences due to the different environmental selections imposed on them; these pressures may in turn drive changes in gene regulation. For example, unlike *S. cerevisiae* Gcn2p, *C. albicans* Gcn2p does not control the translation of Gcn4p, the regulator of amino acid biosynthetic genes (290). The regulation of the ribosomal protein genes in *C. albicans* is different from that in *S. cerevisiae* (282), and perhaps because of the different stress conditions imposed by the environment, the *C. albicans* homologues of the *S. cerevisiae* general stress response transcription factors Msn2p and Msn4p play no obvious role in stress response (228).

We are interested in whether transcriptional rewiring occurs for conserved processes such as sugar and amino acid metabolism. Galactose is utilized by almost all organisms through its conversion to glucose-6-phosphate in a reaction catalyzed by the enzymes of the Leloir pathway (179). In *S. cerevisiae* these enzymes are encoded by *GAL1* (galactokinase), *GAL7* (galactose-1-phosphate uridyl transferase), *GAL10* (UDP-glucose-4-epimerase), and *GAL5* (phosphoglucomutase) (75, 77). Galactose enters *S. cerevisiae* cells through a specific permease, encoded by *GAL2* (297). Because eukaryotic genes are not organized in operons, it is intriguing that the *GAL1*, *GAL7*, and *GAL10* genes are clustered near the centromere of chromosome II (12).

The expression of these *GAL* genes is controlled by a network of activating and repressing activities, encoded by the specific regulatory genes *GAL3*, *GAL4*, and *GAL80*. In *S. cerevisiae GAL4* serves as the transcriptional activator of galactose catabolism (172, 192), while Gal80p is involved in the repression of *GAL* genes by binding the transcriptional activation domain of Gal4p (287). Gal3p functions by forming a complex with Gal80p to relieve inhibition of Gal4p; in the presence of galactose and ATP, Gal3p sequesters Gal80p in the cytoplasm, preventing inhibition of Gal4p and thus activating *GAL* gene expression (237, 240). ScGal4p contains a DNA binding domain that interacts with a specific upstream activating sequence (UAS_G; CGG(N_{11})CCG), located in the promoter regions of *GAL1, 2, 3, 7, 10, 80* (303); the *GAL5* gene is unregulated, being expressed under all conditions, and lacks the UAS_G (21).

This genetic regulatory circuit acting on the *S. cerevisiae GAL* genes is among the most highly studied and best understood of eukaryotic metabolic control pathways. As such it has acted as a paradigm for general principles of metabolic control (147, 192), and has served as well as a template for the application of a systems biology approach to metabolic circuitry (136). In this investigation we studied the regulation of galactose metabolism in the pathogenic fungus *Candida albicans*. This work helps define the network in *C. albicans* that regulates *GAL* genes and shows that *C. albicans* and *S. cerevisiae* control the same process through different regulatory circuits.

4.1.3 Materials and Methods

***C. albicans* strains and plasmid constructions:** The *C. albicans* strains used in this study are listed in Table 1. BWP17 (305) was used to generate *GAL4/gal4*, *gal4* and its wild type phototrophic equivalent BWP17-HA. Strains CAI4 (88) and CAI8 (88) were used to map p*GAL10* regulatory motifs. Strains *rmd5, rim101, rim13*, DAY286 (71), and JKC19 (188) were used to identify the transcriptional regulators of the *GAL10* promoter. Strains *gal4* and its wild type equivalent, CMM1, were used to show that CaGal4p regulates the expression from the *TLO* and *LAT1* promoters through the Gal4p-DNA binding site. JKC19 (*cph1*) was used to show that Cph1p regulates expression from the *GAL10* promoter through a *GAL*-specific palindrome.

Plasmids and oligonucleotides are shown in Tables 2 and 3. The *GAL4* deletion cassettes, p*HIS1-GAL4* and p*ARG4-GAL4*, were created by cloning the 500 bp *GAL4* flanking regions into pFA-*HIS1* and pFA-*ARG4* (100). The *TLO*, *LAT1*, and *GAL10* promoters as well as their shorter versions were cloned into the plac-poly backbone (Alistair Brown, unpublished), and the *GAL* palindromic sequence was cloned into pCR*lacZ* (256). All of the constructs created in this study were integrated into the genome of *C. albicans* using digestion with *StuI* to target them to the *RPS1* locus (219).

Microarray analysis. Transcription profiling and analysis were performed using long oligonucleotides microarrays as described (225). Total RNA was extracted by the hot phenol extraction method (44). GeneSpring software (Agilent Technologies) was used for normalization and for the identification of significantly regulated transcripts. We used a T-test *P* value cutoff of 0.05 to be statistically significant. To analyze the transcriptional response under each experimental condition at least 8 individual biological replicates

were used.

C. albicans strain *gal4* and its wild type equivalent (CMM3 and CMM1) were inoculated overnight in media containing 1% yeast extract, 2% peptone, with either 2% dextrose (YPD) or 2% galactose (YPGal) supplemented with uridine (25 μg/ml) for the growth of *C. albicans* Ura auxotrophic strains. These overnight cultures were diluted to an OD_{600} of 0.1 in their respective media and grown to an OD_{600} of 1.0 (~3 generations). The wild – type strain SC5314 (88) was either grown in YPD or in YPGal media. The overnight cultures were inoculated from a fresh colony and were grown in either YPD or in YPGal at 30°C. These overnight cultures were diluted to OD_{600} of 0.1 in their respective media: overnight YPD and YPGal – grown cells were diluted into a fresh YPD and YPGal media and grown to an OD_{600} of 1.0.

β-galactosidase and growth assays. β-galactosidase assays were performed as described in (255). The growth of the *gal4* strain was assayed as described (211).

4.1.4 Results

Elements of the *C. albicans GAL* regulon:

The sequencing of the *C. albicans* genome (148) provides a framework for the post-genomic studies of this important fungal pathogen. We have recently completed the assembly of the *Candida albicans* genome (Van het Hoog *et al*, submitted) and established a detailed annotation of its genes (33). Intriguingly, although overall the synteny between *C. albicans* and *S. cerevisiae* is very low, the genomic organization of the region involving the galactose metabolism cluster in *S. cerevisiae* is remarkably similar to that found in *C. albicans*. The galactokinase, galactose-1-phosphate uridyl

transferase and UDP-glucose-4-epimerase genes show conservation of sequence (63%, 79%, and 71% similarity respectively), and these genes are clustered together on chromosome I of *C. albicans* in the same relative arrangement as was found in *S. cerevisiae*. In the fungal pathogen this cluster also includes *GAL2* and two open reading frames between *C. albicans GAL10* and *GAL7* that are conserved in *C. dubliniensis*, *C. parapsilosis*, *Debaryomyces hansenii*, and *Schizosaccharomyces pombe* (Figure 1A).

Although the galactose metabolism structural genes are well conserved between *S. cerevisiae* and *C. albicans*, the regulatory components show less similarity. We were not able to detect even weak candidates for Gal3p in the *C. albicans* genome. A candidate for Gal80p, encoded by *ORF19.6899*, shares only 40 percent sequence similarity with ScGal80p. *ORF19.5338*, annotated as CaGal4p, shares strong sequence similarity (86%) with its *S. cerevisiae* homologue only in the DNA binding domain (Figure 1B), with the six cysteine residues, the linker region, and the dimerization region all well conserved. This *C. albicans* gene encodes a much smaller protein of 261 amino acids compared to its *S. cerevisiae* homolog of 881 amino acids. The activation domains of those two proteins share no similarity, and the negatively charged region that serves as the ScGal80p interaction domain is missing in CaGal4p (212).

Galactose regulates expression of *GAL1, 2, 7,* and *10*:

To investigate whether galactose regulates expression of the *GAL1*, *2*, *7* and *10* genes in *C. albicans*, we performed a transcriptional microarray analysis of cells grown on galactose as the sole carbon source, as compared to cells grown on glucose. Just as in *S. cerevisiae*, galactose strongly induced the transcription of genes for the Leloir pathway

enzymes; the *GAL1*, *2*, *7* and *10* genes were among the most highly expressed of all the induced genes. In addition, the expression of 78 genes involved in glycolysis, glycogen synthesis, formation of the cell wall, transport, fatty acid metabolism, and a variety of unknown functions was higher in galactose grown cells compared to glucose grown cells (Figure 2A and Table 4.1.1), while the expression of a further 54 *C. albicans* genes, including genes for cell wall components, transporters, transcription factors, ribosome synthesis elements, as well as for unknown functions, was higher in glucose than in galactose (Figure 2B and Table 4.1.2).

The *C. albicans GAL10* promoter has a conserved and functionally important element:

In order to study the regulation of the galactose metabolizing genes, we cloned the promoter of *GAL10* upstream of a *lacZ* reporter (Brown AJ, unpublished), and transformed this reporter construct into a wild type strain. Cells were grown in liquid medium with raffinose as the carbon source, and shifted to medium containing either dextrose or galactose. After 90 minutes growth the galactose treated cells expressed 5-fold greater β-galactosidase activity than the dextrose treated cells. The raffinose-grown cells showed intermediate levels of *lacZ* expression (not shown); therefore, as in *S. cerevisiae, GAL10* is apparently regulated by both activation and repression mechanisms. Since the presence of galactose coordinately upregulates the expression of the *GAL* genes, we searched for an upstream activating sequence (UAS_G). We compared the DNA sequences of the *GAL1*, *GAL7*, and *GAL10* promoters, and found a common palindromic stretch 5'-TGTAACGTTACA-3' that occurs only four times in the intergenic regions of

the *C. albicans* genome, with all occurrences in the promoters of the *GAL1*, *GAL7*, and *GAL10* genes. As well, the promoter of *GAL2* has the sequence GTTACGTTAC, which is a GTTAC repeat rather than GTTAC palindrome. The palindromic sequence is also conserved in the promoters of the Leloir pathway genes of *C. dubliniensis*, *C. parapsilopsis,* and *Debaryomyces hansenii.*

To test the importance of this *GAL*–palindrome sequence in the regulation of *GAL10* expression in the presence of galactose, we deleted it from the *GAL10* promoter. When cells were grown overnight in either galactose or dextrose medium, the *GAL*-palindrome appears critical for the galactose-mediated induction of *GAL10* expression. The palindrome-less full-length *GAL10* promoter showed a slight 1.3-fold reduction of expression in the presence of dextrose, and a strong 5-fold drop of expression in the presence of galactose, compared to the intact *GAL10* promoter (Table 4.1.5). A truncated version of the promoter was also assayed. In these constructs the *lacZ* reporter was linked to a minimal promoter containing or lacking the palindromic sequence. After overnight growth the palindrome directed 10-fold higher expression in the presence of galactose compared to dextrose, which suggests that the *GAL* palindrome acts as a galactose-dextrose dependent regulator of the *GAL10* promoter under these growth conditions (Figure 5, Table 4.1.5).

Definition of the *C. albicans* Gal4p regulon:

In *S. cerevisiae* Gal4p activates the expression of *GAL* genes to metabolize the galactose present in the environment, and the deletion of ScGal4 leads to the inability of cells to grow on galactose (192). To investigate which *C. albicans* genes are regulated by

the Gal4p homologue, we created a *C. albicans gal4* knock out strain. The *C. albicans gal4* strain grew identically to the wild type strain when utilizing galactose, dextrose, ethanol, or glycerol as a sole carbon source (data not shown). Using microarrays we then compared the transcriptional behavior of the wild type and *gal4* strains when both were grown on galactose-containing media (YPGal). The expression of *GAL* genes remained unchanged (Tables 3 and 4). In addition, deletion of the *GAL4* gene had no effect on the regulation of *GAL10*-driven *lacZ* expression, suggesting that Gal4p does not transactivate the expression of *GAL10* (Table 4.1.5).

Further transcriptional analysis of the *gal4* strain compared to the control grown on galactose showed that loss of *gal4* reduces the expression of 87 genes and increases the expression of 37 genes: the Gal4p-influenced gene set represents 2% of the genome and includes genes whose products are involved in a diverse range of cellular processes. The 87 genes that are reduced in the absence of Gal4p encode transporters, cell wall proteins, nucleic acid binding proteins, and hypothetical proteins (Figure 3A, Table 4.1.3). Interestingly, one third of these downregulated genes are members of a newly discovered subtelomeric gene family *TLO* (Van het Hoog *et al.*, submitted), and 14 genes encode products involved in glycolysis (e.g. *PGI12, LSC2, LAT1*). The expression of a further 37 genes, which encode transporters, cell wall proteins, nucleic acid binding proteins, iron and copper uptake regulators, and hypothetical proteins, is elevated in the absence of CaGal4p (Figure 3B, Table 4.1.4).

The strongest Gal4p-transactivated genes are members of the subtelomeric *TLO* gene family not found in *S. cerevisiae*. Almost all of the *TLO* genes share an identical promoter region, which contains the sequence $CGG(N_{11})CCG$ that represents the binding

site for ScGal4p. When cloned upstream of *lacZ*, a *TLO* promoter that includes the Gal4 binding site was 5.5 times less active in a *gal4* mutant strain, confirming that Gal4p is required for proper expression of this gene through the UAS_G site (Figure 4, Table 4.1.5). The absence of the Gal4 binding site alone leads to a 3-fold drop in reporter expression; the combined absence of both Gal4p and its binding site leads to a 5-fold drop of *TLO*-driven reporter expression. This suggests that Gal4p transactivates the expression of *TLO* through a classical UAS_G. Because the absence of the Gal4p binding site in the wild type allows greater β-galactosidase activity than does the absence of Gal4p, it is possible that there is additional Gal4p-dependent regulation of *TLO* expression.

The promoters of a number of the Gal4p-upregulated glycolysis genes (e.g. *LAT1, LSC2*) contain a potential regulatory site $CGG(N_{11})CCA$ that is conserved in the promoters of the same genes in *C. dubliniensis* and *C. tropicalis*, and which closely resembles a UAS_G site. The removal of this site from the *LAT1* promoter, the loss of Gal4p, or the combined loss of both Gal4p and the regulatory site all lower expression by about 2.5 fold, suggesting Gal4p works through this site to regulate *LAT1* expression (Figure 4, Table 4.1.5). This implies that the proper expression of TCA cycle genes is dependent on both the presence of a functional Gal4p and the Gal4p binding sites within their promoters.

To test whether Gal4p is responsible for the differential expression of the genes whose transcription changes in response to a carbon source of either dextrose or galactose (Figure 2, Tables 4.1.1 and 4.1.2), we used microarrays to compare the transcriptional behavior of the *gal4* strain to that of its wild type equivalent grown on dextrose-

containing media (YPD). We observed that Gal4p is needed for the expression of the same genes in the presence of either dextrose or galactose.

The regulators of the expression of *GAL10*:

To identify the *trans*-acting regulator(s) of galactolysis, we transformed the promoter of *GAL10* fused to a *lacZ* reporter into a *C. albicans* mutant library (229). We observed that in the *rmd5* background *GAL10*-driven *lacZ* yields two times less β-galactosidase when grown on galactose compared to growth on dextrose (data not shown). In addition, we found that in the *rim101* strain *GAL10*-driven *lacZ* yields three times lower β-galactosidase levels when grown on dextrose than when grown on galactose (data not shown). In *C. albicans* Rim13p activates Rim101p by a proteolytic cleavage (184), and as expected, the *rim13* strain shows the same pattern of *GAL10*-driven *lacZ* expression as does the *rim101* strain (data not shown).

Extended incubation of *C. albicans* with macrophage cells leads to a differential expression of *GAL* genes through a process that may involve the MAP kinase pathway acting through Cph1p (273). The half-palindrome sequence TGTAAC is similar to the Ste12p/Cph1p binding site TGAAAC, so we analyzed the expression of the *GAL10* promoter in a *cph1* knock out strain (188). In order to exclude the influence of the upstream *GAL10* promoter regulatory sequence on the actions of the palindrome, we examined the expression patterns of the minimal promoter containing the palindrome. The absence of the palindrome alone leads to 3-fold drop in the reporter expression; the combined absence of both Cph1p and palindrome leads to a 5-fold drop of *GAL10*-driven reporter expression, suggesting that Cph1p transactivates the expression of *GAL10* by

interacting with the palindrome. Because the absence of the palindrome in the wild type strain permits greater β-galactosidase activity than does the absence of Cph1p, it is possible that there is additional Cph1p-dependent regulation of *GAL10* expression. The removal of the Cph1p had a very mild effect on *GAL10* expression in the absence of the palindrome. These results suggest that in the presence of galactose Cph1p acts as an activator of the *GAL10* expression through the palindrome sequence.

4.1.5 Discussion:

Evolution of gene regulation is an important component of phenotypic diversity. In this study we show the same biochemical process, the induction of the Leloir pathway genes by galactose, is regulated by different transcriptional circuits in *S. cerevisiae* and in *C. albicans*. This transcriptional rewiring has resulted in the *C. albicans* protein that is the closest homologue of the *S. cerevisiae* transcriptional regulator Gal4p not controlling the expression of *GAL* genes in the pathogen. Since the presumptive *C. albicans GAL4* has a different function from its *S. cerevisiae* homologue, we suggest the name *CGF1*, for Candida Gal Four homolog, for this transcription regulator, to reduce the confusion of the functional implications of the name *GAL4*.

Our observations show that the mechanism of regulation of transcription of *GAL* genes is fundamentally different in *S. cerevisiae* and *C. albicans*. A previous study showed that unlike the case in *S. cerevisiae*, *C. albicans* Mig1p does not repress *GAL* genes (312). In *S. cerevisiae* the promoters of *GAL* genes contain a UAS_G, which functions as a galactose - inducible enhancer. We observed that the regulation of the *GAL10* promoter in *C. albicans* is controlled by at least 2 regulatory sites: an enhancer

and a galactose - dextrose responsive element. Our study shows that the latter is regulated by Cph1p, a homologue of Ste12p of *S. cerevisiae*. The difference in the regulation of the Leloir pathway genes may be due to the fact that galactose plays important roles in adhesion and biofilm formation of *C. albicans*, processes that contribute to the virulence of this pathogen and are absent in *S. cerevisiae* (144).

To investigate when the changes of the *cis*-regulatory element in the *GAL* promoters occurred during the evolution of the yeast species, we used available genomic data of the ascomycota that have *GAL* genes conserved in a cluster in their genome: *Schizosaccharomyces pombe, S. cerevisiae, S. paradoxus, S. mikatae, S. bayanus, S. castellii, Kluyveromyces lactis, Debaryomyces hansenii, C. parapsilosis, C. dubliniensis,* and *C. albicans* (Figure 4.1.6). *Saccharomyces* species contain classic tandem Gal4p binding sites and lack Cph1p-like binding motifs, while *Candida* species contain Cph1p-like binding sites and lack any Gal4p-like binding motifs in their *GAL* promoters. Interestingly, the intermediate species, *S. castellii* and *K. lactis*, contain both Gal4p-like and Cph1p-like motifs. *S. pombe*, which lies outside of the Hemiascomycetes lineage, lacks a strong Gal4p homolog and lacks Gal4p binding motifs, but contains Cph1p-like motifs upstream of its *GAL* genes, which suggests a Gal4p role in galactolysis may be specific to *Saccharomyces* species. Recent studies show that the evolution of the regulatory circuits that control cell type regulation coincide with a whole-genome duplication event in yeast (298). We observed that evolution of metabolic *GAL* regulation appears to have similar evolutionary pattern; the *Saccharomyces* lineage controls circuits differently from other fungi, with *K. lactis* representing an intermediate having both regulatory motifs (Figure 4.1.6).

Since Gal4p is present only in fungi it would be interesting to know what regulates *GAL* genes in humans, where the absence of those enzymes results in galactosemia (26). It is interesting to note that the core of the MAPK controlled AP1 binding site, tgACGTca (145), shares similarity to the *GAL* palindrome, tgtaACGTtaca, and human *GAL* genes contain potential AP1 binding sites.

Table 4.1.1. Genes whose expression was up-regulated in response to galactose.

Systematic	Normalized	Common	Description
orf19.3668	0.129	GAL2	hexose transporter
orf19.6169	0.178		conserved hypothetical gene, DNA metabolism
orf19.3672	0.181	GAL10	UDP glucose-4-epimerase
orf19.7094	0.196	HGT12	glucose sensor or transporter protein
orf19.2849	0.209	AQY1	aquaporin
orf19.3670	0.239	GAL1	galactokinase
orf19.3675	0.242	GAL7	galactose-1-phosphate uridyl transferase
orf19.4384	0.266	HXT5	fructose symporter
orf19.7310	0.300	MSC1	Meiotic Sister-Chromatid recombination
orf19.4527	0.301	HGT1	hexose transporter
orf19.4591	0.305	CTN2	carnitine acetyltransferase, amino acid metabolism
orf19.3359	0.320	ARP8	actin-related protein
orf19.5307	0.328	JEN2	carboxylic acid transporter
orf19.734	0.331	GLK1	aldohexose specific glucokinase
orf19.4555	0.348	ALS4	agglutinin-like protein 4
orf19.1408	0.370	GLK2	glucokinase
orf19.4980	0.376	HSP70	heat shock protein 70
orf19.2882	0.380		highly conserved hypothetical protein, putative purine permease
orf19.2496	0.387	FUN34	transporter involved in nitrogen utilization
orf19.6491	0.389		hypothetical protein
orf19.13	0.399	GLK1	glucokinase, aldohexose specific
orf19.111	0.408	CAN2	arginine permease
orf19.3160	0.415	HSP12	heat shock protein;molecular chaperone
orf19.6116	0.416	GLK4	aldohexose specific glucokinase
orf19.3637	0.417		hypothetical protein
orf19.7021	0.435	GPH1	glycogen phosphorylase
orf19.1097	0.455	ALS2	cell wall protein
orf19.4539	0.457		putative rho GDP dissociation inhibitor
orf19.1794	0.457		hypothetical protein
orf19.4735	0.461		putative ornithine cyclodeaminase
orf19.7514	0.462	PCK1	Phosphoenolpyruvate carboxykinase gluconeogenesis
orf19.2896	0.473	SOU1	peroxisomal 2,4- dienoyl-CoA reductase, and sorbitol utilization
orf19.2515	0.479		hypothetical protein
orf19.3131	0.479	OYE32	NADPH dehydrogenase
orf19.3684	0.481	SPS20	peroxisomal 2,4-dienoyl-CoA reductase
orf19.802	0.487	UGA12	4-aminobutyrate aminotransferase (nitrogen)
orf19.4274	0.487	PUT1	proline oxidase
orf19.1847	0.492	ARO10	Protein described as pyruvate decarboxylase, LEU catabolism
orf19.2121	0.496	ALS4	cell wall protein
orf19.4551	0.498	CTN1	mitochondrial carnitine acetyltransferase
orf19.744	0.498	GDB1	glycogen debranching enzyme
orf19.3070.1	0.500	FOX3	hydratase-dehydrogenase- epimerase
orf19.4317	0.518	GRE3	aldose reductase
orf19.1743	0.518	ACS1	acetyl-coenzyme A synthetase
orf19.3548.1	0.521	WH11	heat shock protein induced during entry into stationnary phase
orf19.4682	0.524	HGT17	quinate permease
orf19.5037	0.524		conserved hypothetical protein
orf19.682	0.525		hypothetical protein
orf19.5525	0.525		conserved hypothetical protein
orf19.1862	0.545		conserved hypothetical protein

orf19.7566	0.550	GNP1	high-affinity glutamine permease
orf19.338	0.561		conserved hypothetical protein
orf19.272	0.574	FAA2-2	long-chain fatty acid CoA ligase
orf19.1974	0.578	TFS1	carboxypeptidase Y inhibitor
orf19.7520	0.584	POT1	peroxysomal 3-ketoacyl-CoA thiolase A
orf19.5604	0.587	MDR1	benomyl/methotrexate resistance protein
orf19.7531	0.590	YMR90	conserved hypothetical protein
orf19.6083	0.594		hypothetical protein
orf19.1152	0.600		hypothetical protein
orf19.3278	0.600	GSY1	glycogen(starch) synthase
orf19.1788	0.602	XKS1	Xylulokinase
orf19.4477	0.604	CSH1	aryl-alcohol dehydrogenase
orf19.532	0.608	RBR2	Cell wall protein
orf19.5000	0.610	CYB2	cytochrome b2 precursor
orf19.7281	0.612	PDK2	pyruvate dehydrogenase kinase
orf19.1149	0.620	ETR1	mitochondrial 2-enoyl thioester reductase, respiration
orf19.4287	0.623		alcohol dehydrogenase (glucose catabolism to butanediol)
orf19.3325	0.624	GLG1	self-glucosylating initiator of glycogen synthesis
orf19.345	0.627	UGA2	succinate semialdehyde dehydrogenase
orf19.1180	0.627		conserved hypothetical protein
orf19.1809	0.629	FOX2	peroxisomal trifunctional hydratase-dehydrogenase-epimerase
orf19.2613	0.638	ECM4	involved in cell wall biogenesis and architecture
orf19.7676	0.640	SOR1	sorbitol dehydrogenase
orf19.6443	0.641	ECI12	enoyl-CoA isomerase;delta3-cis-delta2-trans-enoyl- CoA isomerase
orf19.6082	0.644	GSF2	ER localized promote secretion of GAL2
orf19.5640	0.645	PEX5	peroxisomal protein receptor
orf19.2737	0.646		Kinase
orf19.3029	0.651	EHD3	enpyl-CoA hydratase
orf19.6637	0.652		predicted glycosilase
orf19.1325	0.656	ECM38	gamma-glutamyltransferase, cell wall organization and biogenesis
orf19.854	0.660	UGA1	4-aminobutyrate aminotransferase (GABA transaminase)
orf19.1652	0.662	POX1-3	fatty-acyl coenzyme A oxidase

Table 4.1.2. Genes whose expression was up-regulated in response dextrose.

Systematic	Normalized	Common	Description
orf19.4737	4.88	TPO3	membrane transporter of the MFS-MDR family
orf19.1354	3.77	YER67	conserved hypothetical protein
orf19.6656	2.74	DUR3	urea transport protein
orf19.4941	2.57	TYE7	basic helix-loop-helix transcription factor
orf19.3707	2.55	YHB1	flavohemoglobin;dihydropteridine reductase
orf19.2659	2.52		conserved hypothetical protein
orf19.4716	2.40	GDH3	NADP-glutamate dehydrogenase
orf19.5288	2.26	IFE2	Zn-containing alcohol dehydrogenase
orf19.85	2.15	GPX2	glutathione peroxidase
orf19.5437	2.13	RHR2	DL-glycerol-3-phosphatase(glycewrol biosynthesis)
orf19.5626	2.07		conserved hypothetical protein
orf19.978	2.03	BDF1	transcription factor
orf19.889	1.96	THI20	thiamine biosynthesis;phosphomethylpyrimidine kinase
orf19.3448	1.96		hypothetical protein
orf19.1415	1.94	FRE10	ferric reductase
orf19.6249	1.93	HAK1	high affinity potassium transporter
orf19.5193	1.93	FMA1	benzil reductase
orf19.670.2	1.88		hypothetical protein

orf19.909	1.77	STP4	TF
orf19.5305	1.75	RHD3	Putative GPI-anchored protein
orf19.3712	1.73		hypothetical protein
orf19.3441	1.73	FRP6	putative transporter
orf19.2158	1.72	NAG3	related to MDR family transporters
orf19.6586	1.71		conserved hypothetical protein
orf19.4779	1.71		multidrug-resistance transporter
orf19.740	1.71	HAP41	CCAAT-binding factor complex
orf19.23	1.70	RTA3	putative transporter upregulated during the acquisition of azole resistance
orf19.2475	1.69	PGA26	Putative GPI-anchored protein
orf19.5759	1.65	SNQ2	ABC transporter
orf19.5025	1.65	MET3	ATP sulfurlyas
orf19.5431	1.65		hypothetical protein
orf19.3676	1.61	ABP140	hypothetical protein
orf19.1855	1.60	SEO2	putative permease
orf19.339	1.56	NDE1	mitochondria directed NADH dehydrogenase
orf19.3088	1.56		TF
orf19.6770	1.56	ENT4	epsin N-terminal homology- containing protein
orf19.6766	1.56	NOP13	Nucleolar Protein 13
orf19.1264	1.55	CFL2	ferric reductase
orf19.183	1.55	HIS3	imidazoleglycerol-phosphate dehydratase
orf19.5992	1.55	ZCF33	zinc finger transcription factor
orf19.5305	1.55	RHD3	GPI-anchored protein
orf19.3393	1.55	DBP9	dead-box protein 9
orf19.449	1.55		possible phosphatidyl synthase
orf19.4033	1.54	PRP22	RNA-dependent ATPase
orf19.7657	1.53	POP3	RNase P and RNase MRP subunit
orf19.6902	1.53	DBP7	Dead-box protein;RNA helicase
orf19.2308	1.53	PFK26	6-phosphofructose-2-kinase
orf19.6727	1.53	RIT1	ribosyltransferase of initiator tRNA methionine
orf19.3040	1.53	EHT1	alcohol acyl transferase, synthesis of mitochondrial outer membrane
orf19.5010	1.52	DIM1	dimethyladenosine transferase, rRNA
orf19.409	1.52		conserved hypothetical protein, membrane
orf19.4459	1.51	YNL234W	weakly simiular to to mammalian globins
orf19.6014	1.50	RRS1	regulator of ribosome synthesis
orf19.3406	1.50		highly conserved hypothetical protein

Table 4.1.3. Genes whose expression was down-regulated in *gal4* strain.

Systematic	Normalized	Common	Description
orf19.6191	0.286	CTA2	hypothetical protein
orf19.3074	0.308	CTA2	hypothetical protein
orf19.3386	0.425	CTA2	transcription factor
orf19.5467	0.430	CTA2	transcription factor
orf19.710	0.453	LSC2	succinate-CoA ligase beta subunit (TCA)
orf19.4054	0.461	CTA24	transcriptional regulation
orf19.5383	0.471	PMA1	plasma membrane H+-ATPase (phosphate)
orf19.3182	0.475	GIS2	nucleic acid binding protein
orf19.6112	0.481	CTA21	putative transcriptional activator
orf19.3145.4	0.507	PGI12	glucose-6-phosphate isomerase (glycolysis)
orf19.2762	0.517	AHP1	alkyl hydroperoxide reductase
orf19.362	0.526	CTA2	transcriptional activation
orf19.7544	0.553	CTA2	hypothetical protein
orf19.4885	0.556	MIR1	mitochondrial phosphate transport protein

orf19.2777	0.566	TOS3	target of SBF (glucose repression)
orf19.7680	0.581	CTA26	transcriptional activation
orf19.5338	0.594	GAL4	potential fungal Zn(2)-Cys(6) binuclear cluster domain sim to that of GAL4
orf19.6561	0.596	LAT1	dihydrolipoamide acetyltransferase component (TCA)
orf19.7127	0.596	CTA2.5	hypothetical protein
orf19.6435	0.600		conserved hypothetical protein
orf19.655	0.606	PHO842	high-affinity inorganic phosphate/H+ symporter
orf19.6337	0.609	CTA28	transcriptional activator
orf19.4155.12	0.614	IDH1	similar to protion of isocitrate dehydrogenase 1 alpha-4-beta-4 subunit (TCA)
orf19.7231	0.616	FTR2	plasma membrane iron permease
orf19.1860	0.616	LSC2	hypothetical protein
orf19.631	0.617	CTA2	transcription factor
orf19.6165	0.619	KGD1	alpha-ketoglutarate dehydrogenase (TCA)
orf19.3642	0.625	SUN4	Putative cell wall protein
orf19.5451	0.625		hypothetical protein
orf19.2685	0.639	PGA54	GPI-anchored protein
orf19.7635	0.641	DRS1	ATP dependent RNA helicase
orf19.7276.1	0.642	CTA24	transcriptional activator, 3- prime end
orf19.1124.2	0.644	DPH52	diphthamide methyltransferase
orf19.7127.1	0.662	CTA2.3	hypothetical protein
orf19.4211	0.663	FET3	multicopper ferro-O2- oxidoreductase involved in high-affinity iron uptake
orf19.7424	0.664	NSA2	Killer toxin Resistant;ribosomal biogenesis
orf19.5193	0.668		benzil reductase
orf19.1105.2	0.669	PGA56	Regulator of sorbose utilization; putative GPI-anchor
orf19.1536	0.669	ZRC1	zinc- and cadmium resistance protein
orf19.968	0.673	PGA14	GPI-anchored protein
orf19.7443	0.681		hypothetical protein
orf19.5305	0.683	PGA29	GPI-anchored protein
orf19.6724	0.686	FUM12	mitochondrial fumarate hydratase, 5-prime end (TCA)
orf19.1591	0.686	ERG10	acetyl-CoA acetyltransferase;acetoacetyl CoA thiolase
orf19.903	0.689	GPM1	phosphoglycerate mutase (glycolysis)
orf19.5892	0.691	HUL4	ubiquitin-protein ligase
orf19.3223	0.697	ATP3	proton-transporting ATP synthase, central stalk
orf19.6676	0.701	DPH5	diphthamide (modified histidine residue) biosynthesis methyltransferase
orf19.4309	0.701	GRP2	induced by osmotic stress
orf19.932	0.703	DNF1	phopholipid transporting ATPase
orf19.4263	0.706		hypothetical protein
orf19.3475	0.710		Gag-related protein; hyphal induced
orf19.7125	0.711		hypothetical protein
orf19.10	0.713	ALK8	cytochrome p450
orf19.3713	0.715		hypothetical protein
orf19.7673	0.715	SMD1	snRNA-associated protein, snRNP
orf19.6189	0.715		conserved hypothetical protein
orf19.339	0.722	NDE1	mitochondria directed NADH dehydrogenase (glycolysis)
orf19.3829	0.723	PHR1	pH regulated GPI-anchored membrane protein needed for morphogenesis
orf19.5021	0.723	PDX1	pyruvate dehydrogenase complex protein X (TCA)
orf19.7534	0.723	MIS1	mitochondrial C1- tetrahydrofolate synthase precursor
orf19.7219	0.727	FTR1	plasma membrane iron permease
orf19.717	0.727	HSP60	mitochondrial groEL-type heat shock protein
orf19.5684	0.730	MRPL14	mitochondrial ribosomal protein L14
orf19.3097	0.731	PDA1	alpha subunit of pyruvate dehydrogenase (TCA)
orf19.5052	0.733		conserved hypothetical protein
orf19.6948	0.734	CCC1	transmembrane Ca2+ transporter
orf19.4051	0.735	HTS1	histidine tRNA synthetase

orf19.301	0.735		hypothetical protein, likely cell wall localized and GPI-anchored
orf19.6090	0.735	NSR1	ribosomal small subunit assembly and maintenance
orf19.2346	0.737		highly conserved hypothetical protein
orf19.962	0.738		hypothetical protein
orf19.6105	0.739	MVD	mevalonate pyrophosphate decarboxylase
orf19.5112	0.740	TKL1	transketolase 1 (pentose)
orf19.327	0.742	HTA3	histone variant involved in chromatin and transcriptional control
orf19.252	0.742		conserved hypothetical protein
orf19.5943.1	0.743	STM1	maintain telomere structure
orf19.929	0.747		hypothetical protein
orf19.5989	0.748	HRP1	nuclear polyadenylated-RNA- binding protein
orf19.930	0.751	PET9	ATP/ADP translocator
orf19.1329	0.752		hypothetical protein
orf19.5444	0.753	TIM44	inner membrane translocase component
orf19.6190	0.753	SRB1	GDP-mannose pyrophosphorylase cell wall
orf19.3038	0.756	TPS2	threalose-6-phosphate phosphatase
orf19.2200	0.756	VIP36	vesicular integral-membrane protein VIP36-like
orf19.2953	0.757	TOM20	mitochondrial outer membrane translocase complex
orf19.4826	0.759	IDH1	mitochondrial isocitrate dehydrogenase 1 (TCA)

Table 4.1.4. Genes whose expression was up-regulated in *gal4* strain.

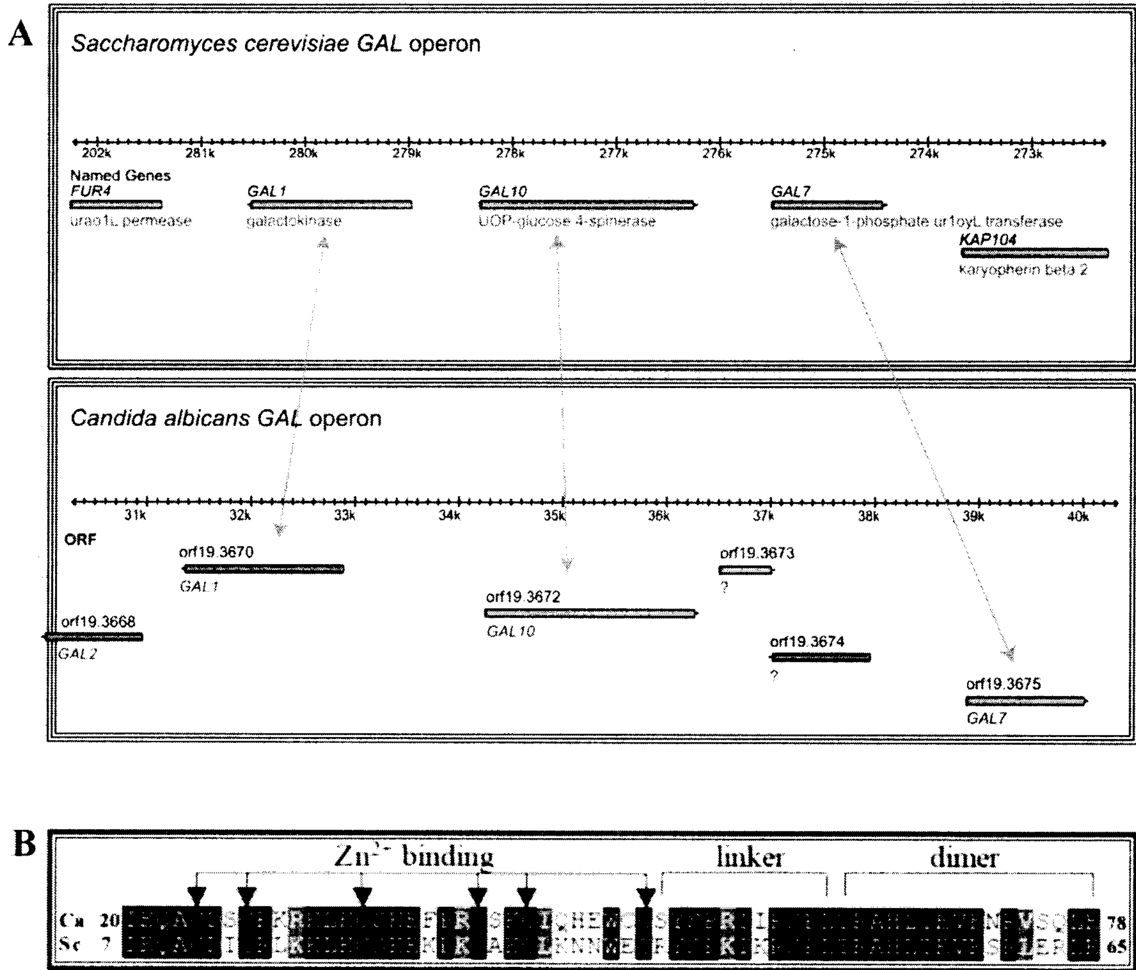
Systematic	Normalized	Common	Description
orf19.6139	2.42	FRE7	ferric reductase
orf19.3646	1.93	CTR1	copper transport protein
orf19.2107.1	1.85	STF2	ATP synthase regulatory factor
orf19.7111.1	1.84	SOD3	superoxide dismutase
orf19.1862	1.83		conserved hypothetical protein
orf19.2344	1.68	ASR1	similar to heat shock protein
orf19.2020	1.63	HGT6	hexose transporter
orf19.638	1.62	FDH1	formate dehydrogenase, glycine catabolism
orf19.4555	1.60	ALS4	agglutinin-like protein 4
orf19.3548.1	1.60	WH11	heat shock protein potentially induced during entry into stationnary phase
orf19.2048	1.58		hypothetical protein
orf19.670.2	1.58		hypothetical protein
orf19.1996	1.57	CHA1	catabolic serine/threonine dehydratase
orf19.7077	1.53	FRE7	transmembrane subunit of ferric reductase
orf19.5626	1.51		conserved hypothetical protein
orf19.2121	1.51	ALS4	cell wall protein
orf19.7310	1.49	MSC1	Meiotic Sister-Chromatid recombination
orf19.1354	1.47	YER67	conserved hypothetical protein
orf19.4941	1.46	TYE7	basic helix-loop-helix transcription factor
orf19.5288	1.45	YAL60	Zn-containing alcohol dehydrogenase
orf19.2849	1.42	AQY1	Aquaporin
orf19.6816	1.42		aldehyde reductase
orf19.4682	1.42	HGT17	quinate permease
orf19.5514	1.42		conserved hypothetical protein
orf19.5876	1.39		hypothetical protein
orf19.3682	1.38	CWH8	generation of mannoprotein layer of the cell wall;putative phosphatase
orf19.3932	1.38		conserved hypothetical protein
orf19.1331	1.37	HSM3	MutS family (putative);mismatch repair
orf19.946	1.37	MET14	adenylylsulfate kinase
orf19.5000	1.37	CYB2	cytochrome b2 precursor
orf19.2292	1.36	OPT4	oligopeptide transporter protein

orf19.2584	1.35	OPT9	oligopeptide transporter specific for tetra- and pentapeptides;possible pseudogene
orf19.1974	1.35	TFS1	carboxypeptidase Y inhibitor
orf19.2693	1.35	URE1	nitrogen catabolite repression regulator and prion generator
orf19.7284	1.34	ASR2	conserved hypothetical protein
orf19.3749	1.34	IFC3	oligopeptide transporter protein
orf19.2173	1.33	MAF1	nuclear-cytoplasmic transport;negative effector of Pol III synthesis

Table 4.1.5: β-galactosidase activities in Miller units of *C. albicans* promoters in *C. albicans* strains.

	WT	WT	*cph1*	*cph1*	*gal4*	*gal4*
Construct	**Dextrose**	**Galactose**	**Dextrose**	**Galactose**	**Dextrose**	**Galactose**
plac-poly	0±0	0±0	0±0	0±0	0±0	0±0
p*LAT1-lacZ*-270	379±34				150±10	
p*LAT1-lacZ*-253	146±11				152±11	
p*CTA2-lacZ*-126	430±33				76±6	
p*CTA2-lacZ*-114	138±9				81±6	
p*GAL10-lacZ*	503±37	505±36				
p*GAL10-lacZ*-840	501±30	499±31			504±35	489±31
p*GAL10-lacZ*-840Δ176-188	393±22	96±6				
p*GAL10-lacZ*-188	7±1	99±7	9±1	9±2		
p*GAL10-lacZ*-176	98±8	30±3	43±3	7±1		

A



B

Zn^{2+} binding | linker | dimer

Ca 20 … 78

Sc 7 … 65

Figure 4.1.1. *Candida albicans GAL* regulon.

A. The structural members of the *S. cerevisiae GAL* regulon are all conserved in *C. albicans* and are also organized in a cluster. *GAL2* is a part of the *C. albicans GAL* cluster.

B. ScGal4p acts as the activator of the *S. cerevisiae GAL* regulon. The putative CaGal4p protein, encoded by *C. albicans ORF19.5338*, shares strong sequence similarity (86%) with *S. cerevisiae* Gal4p only within the DNA binding domain; the DNA binding domain has the six cysteine residues, the linker region, and the dimerization region all well conserved.

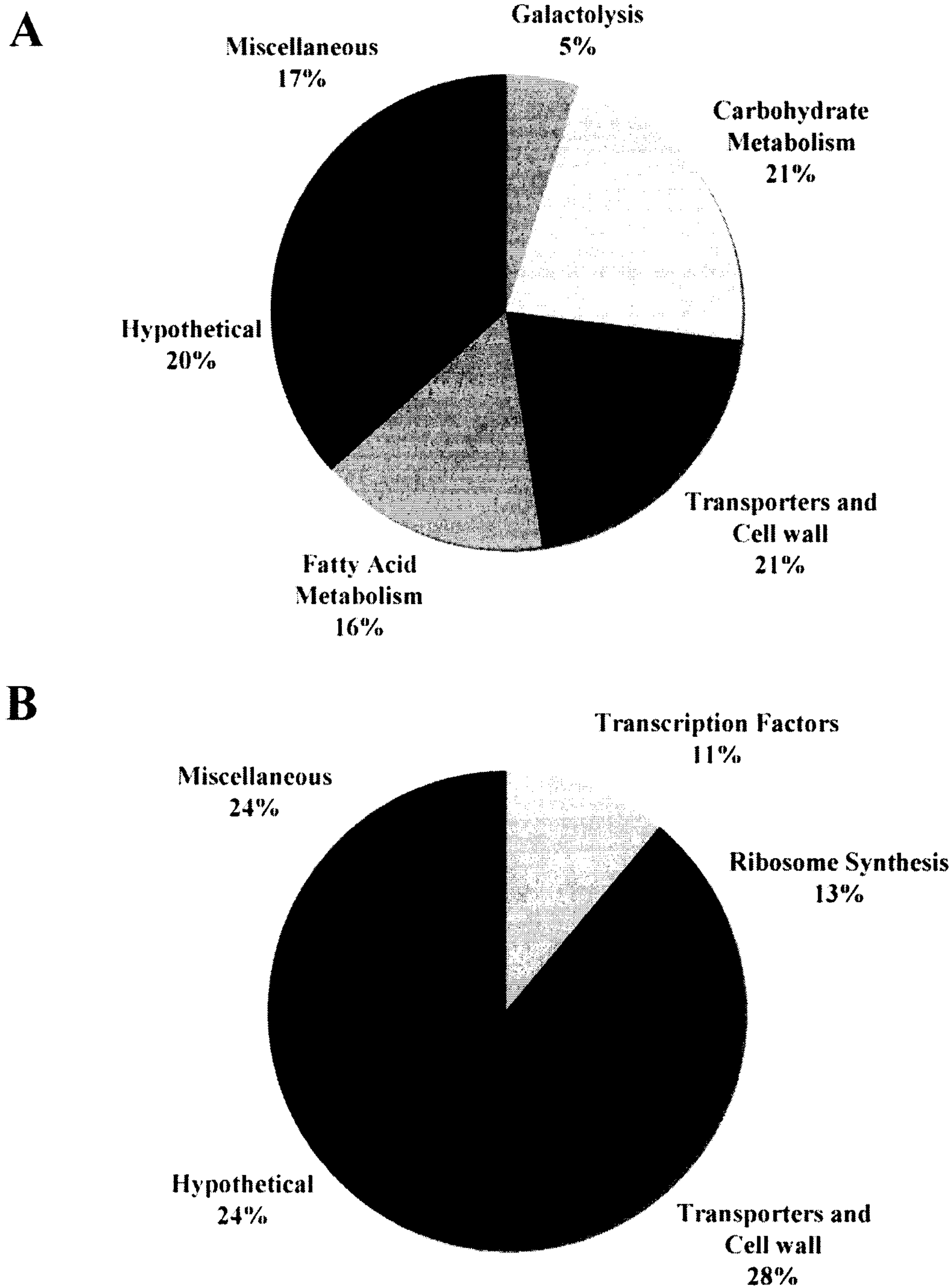


Figure 4.1.2. Galactose leads to elevated expression of *CaGAL* genes.

Galactose leads to higher expression of 82 genes including *GAL1, 2, 7,* and *10* (A), while dextrose leads to higher expression of 54 genes in *C. albicans* (B).

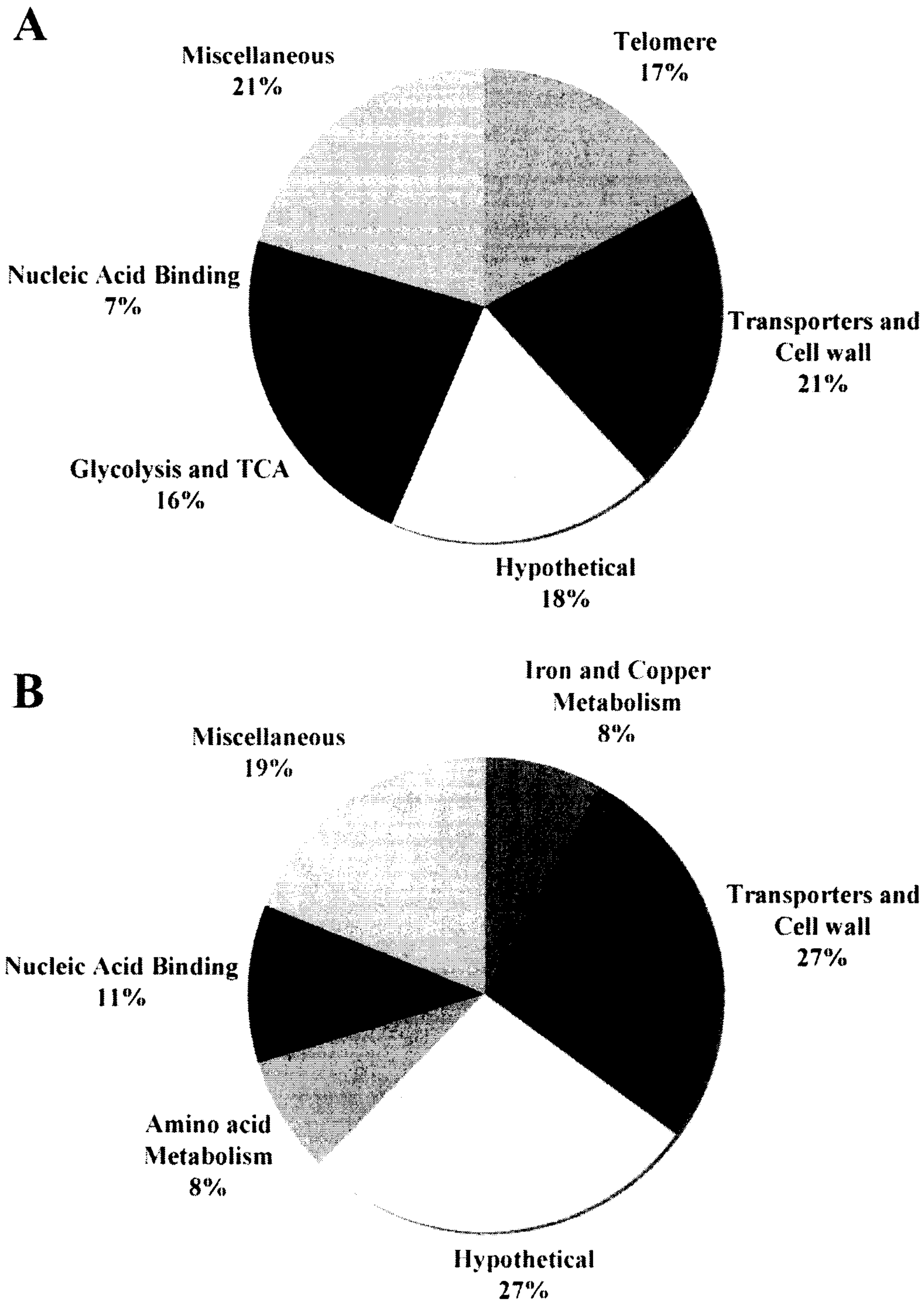


Figure 4.1.3. The function of *C. albicans* Gal4p.

The transcriptional gene targets of *C. albicans* Gal4p were analyzed by microarrays. Loss of *C. albicans GAL4* led to reduced expression of 87 genes (A), and the increased expression of 37 genes (B).

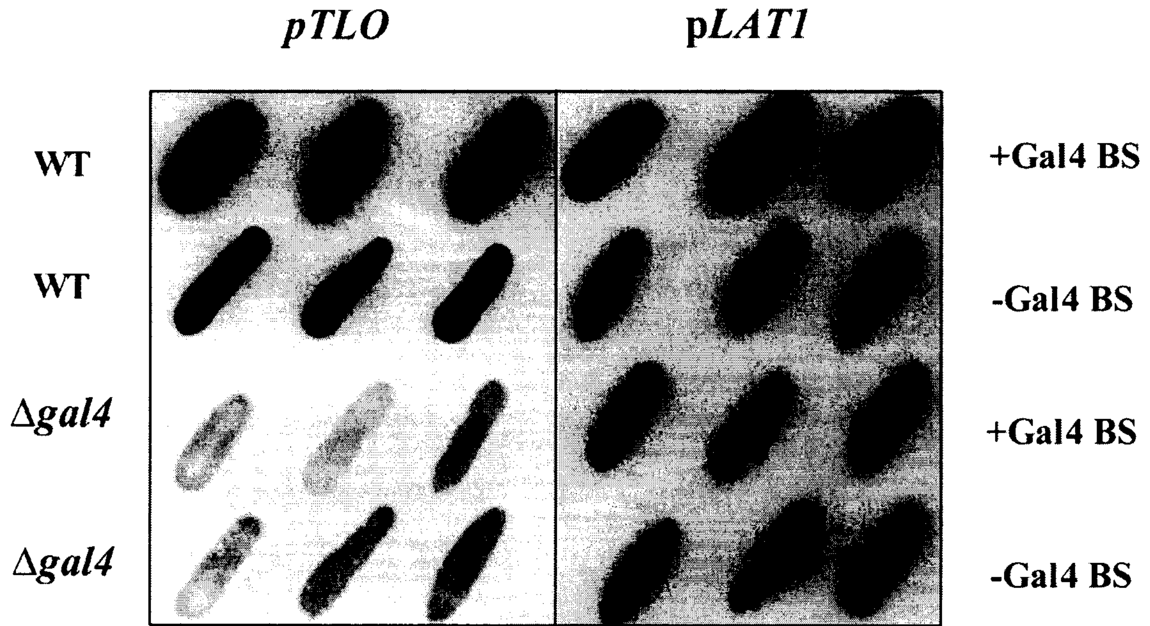


Figure 4.1.4. Gal4 regulates expression through Gal4 DNA binding site.

The promoter of *TLO*, a member of the *C. albicans* telomeric gene family, and of *LAT1*, a member of the TCA cycle, were connected to drive the expression of the downstream *lacZ* reporter. The promoters either contained or lacked a potential Gal4p binding site (CGG(N_{11})CCG/A), and their resulting β-galactosidase activities were analyzed in wild type and *gal4* strains.

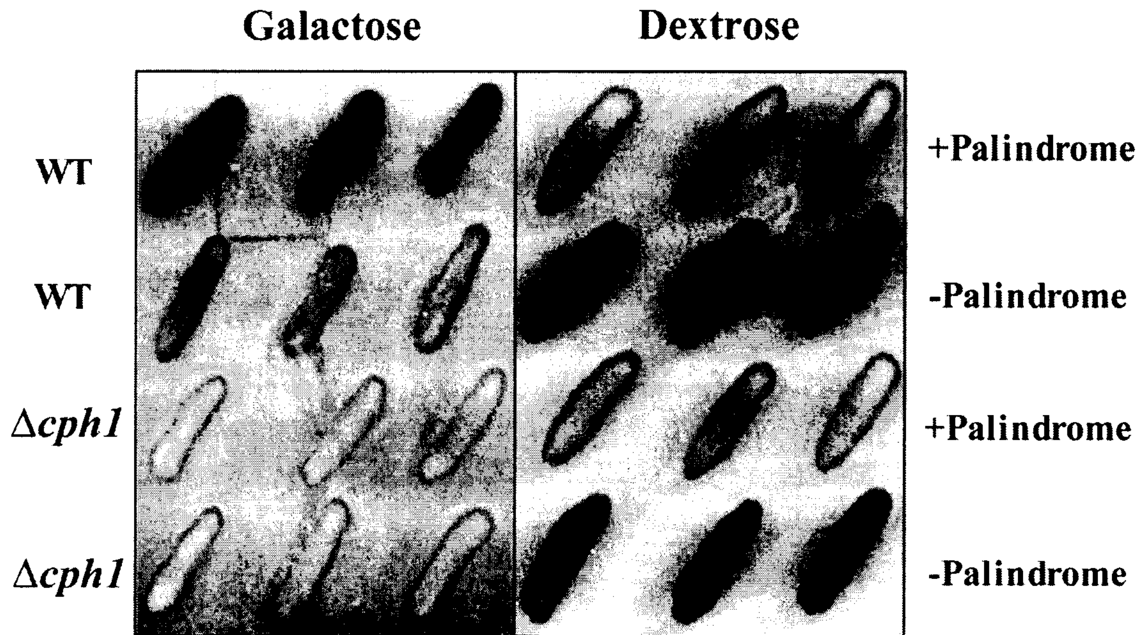


Figure 4.1.5. Cph1p regulates the expression of *GAL10*.

The promoter of *GAL10*, a member of the *C. albicans* Leloir pathway, was fused to a downstream *lacZ* reporter. The promoter either contained or lacked a potential *GAL*-specific palindrome (TGTAACGTTACA). Their resulting β-galactosidase activities were analyzed in wild type and *cph1* strains grown either in galactose or dextrose.

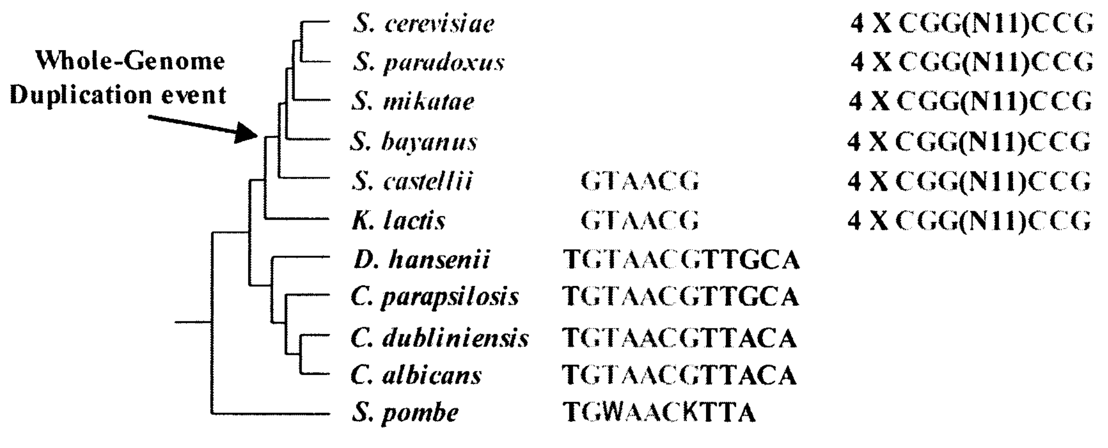


Figure 4.1.6. Analysis of *cis-GAL* regulation across Ascomycota.

The primary sequences of *GAL* promoters were analyzed for the presence of UAS_G's in species only if those genes were arranged in a cluster. W stands for either T or A, and K stands for either G or T. The phylogenetic relationship between the fungi is shown as in (74).

CHAPTER 5: FINAL CONCLUSION AND SUMMARY

The genome comparison to *S. cerevisiae* revealed an increased number of *C. albicans* superoxide dismutase genes. We analyzed the transcriptional regulation and the function of *SOD5*, whose role is to protect the pathogen against the extracellularly produced, neutrophil-generated superoxide radicals. The search for sequences similar to CuZnSOD in genomes of ascomycota revealed that all of these organisms have *SOD1* homologs, but only filamentous fungi, such as *N. crassa* and *A. fumigatus*, have additional sequences homologous to *SOD5*. It is possible that filamentous fungi have more sources of production of harmful radicals, which applies more pressure on these fungi to possess extra isoforms of the detoxifiers of such radicals.

Comparative genomics showed that although many of the *C. albicans* transcription factors, such as Gal4p and Gcn4p, have homologues in *S. cerevisiae*, the sequence similarities occur only in the DNA binding motifs of those proteins. Deletion analysis of the CaGcn4p protein shows that the N terminus is needed for transactivation ability; deletion analysis of CaGal4p identifies a C-terminal transactivation domain. These two transactivation regions show no sequence similarity to the respective domains in their *S. cerevisiae* homologues, and the two *C. albicans* transactivating domains themselves show little similarity. The comparative analysis of the transcriptional machinery between *C. albicans* and *S. cerevisiae* showed a poor homology of the mediator complex which bridges activation domains of transcription factors to the RNA polymerase II complex, which might explain the poor homology between *S. cerevisiae* and *C. albicans* activation domains. As previously stated, one of the most important virulence factors of *C. albicans* as a human pathogen is its ability to switch between

different cell shapes. We now know a number of transcription factors involved in this dimorphic switch, and the deletion of transcription factors that participate in the yeast to hyphal transitions leads to reduced virulence of this fungus (69, 190). The different structure of the mediator is of interest for a number of reasons: 1) the difference in transcriptional machinery elements, such as mediator, between fungi such as *C. albicans* and *S. cerevisiae* and metazoans presents a potential target for antifungal drugs; 2) different mediators and their interacting activation domains are involved in different cellular biology processes such as transcriptional rewiring of signaling pathways and different posttranslational modifications of transcriptional activators; 3) the ability of the cell to change its mediator structure may suggest that the plasticity of the mediator plays a key role of the cell's capacity to evolve in response to a changing environment.

We also performed comparative genomics of intergenic DNA regions to identify the *cis*-regulatory elements from three *Candida* species and four *Saccharomyces* species to show the reorganization of the transcriptional regulatory network between these two groups of yeasts. We observed that *C. albicans GAL* genes lack Gal4p binding sites, and that Gal4p regulated the expression of telomeric genes and genes involved in glycolysis, all of which contain UAS_G. We observed that galactose activates the expression of the *Candida* homologs of the *GAL1*, *2*, *7*, and *10* genes. Our results show that the regulation of the *C. albicans GAL10* promoter is controlled by at least two regulatory sites: an enhancer and a galactose - dextrose responsive element. We show that the latter site is regulated by Cph1p, a homologue of Ste12p of *S. cerevisiae*. In the future we will assess the regions of the *GAL10* promoter, through which Rim101p and Rmd5p regulate *GAL10* expression.

S. cerevisiae responds to mating pheromone by activating a mitogen-activated protein kinase (MAPK) signaling pathway (119). This MAPK pathway leads to the activation of Ste12p transcription factor that in turn binds, along with Mcm1p, the pheromone response elements (PREs) upstream of genes involved in mating, resulting in their expression (80). In addition, Ste12p regulates, in concert with Tec1p, the genes involved in pseudohyphal growth by cooperatively binding filamentation response elements (FREs) upstream of genes involved in filamentous growth (189). The *C. albicans* Ste12p homologue, Cph1p, is also required for mating and hyphal growth (188, 204), and in this work we show that Cph1p also regulates the expression of *GAL* genes. In future work it would be necessary to address the nature of the Cph1p-biniding partners in the presence of different carbon sources.

In conclusion, sequencing the genome of any given organism proves to be a powerful tool in studying its biology. The comparison of the genome of any species to that of its close homologue to search for differences in both intergenic and intragenic DNA regions will reveal clues about its evolutionary path and will tell us a great deal about its environmental conditions.

APPENDIX A: Tables

Table 1: Strains used in this study.

Strain	Genotype	Source
SC5314	Wild-type isolate	(88)
RM1000	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG	(226)
RM1000-HU	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG SOD5/SOD5::[p37H]::[p37U]	This study
H10	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG sod5Δ::HIS1/SOD5	This study
UH16	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG sod5Δ::hisG-URA3-hisG/sod5Δ::HIS1	This study
UH16-1	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG sod5Δ::hisG/sod5Δ::HIS1	This study
UH16-2	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG sod5Δ::hisG/sod5Δ::HIS1::[p37U]	This study
UH16-3	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG sod5Δ::HIS1/sod5Δ::hisG::[pSOD5-110]	This study
HLC52	ura3Δ::λimm434/ura3Δ::λimm434 efg1Δ::hisG/efg1Δ::hisG-URA3-hisG	(190)
DAY25	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG rim101:ARG4/rim101::URA3	(69)
DAY185	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG ARG4::URA3::arg4::hisG/arg4::hisG	(69)
GKO6	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG mds3::Tn7-UAU1/mds3::Tn7-URA3	(71)
BWP17	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG	(305)
CMM1	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG GAL4/GAL4::[pHis1-GAL4]::[pARG4-GAL4]	This study
CMM2	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::HIS1/GAL4	This study
CMM3	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::ARG4/gal4Δ::HIS1	This study
CMM4	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG GAL4/GAL4::[pHis1-GAL4]::[pARG4-GAL4] RPS1/RPS1::[pGAL10-lacZ]	This study
CMM5	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::ARG4/gal4Δ::HIS1 RPS1/RPS1::[pGAL10-lacZ]	This study
rmd5	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG rmd5::Tn7-UAU1/rmd5::Tn7-URA3	(229)
CMM6	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG rmd5::Tn7-UAU1/rmd5::Tn7-URA3 RPS1/RPS1::[pGAL10-lacZ-840-HIS1]	This study
rim101	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG rim101::Tn7-UAU1/rim101::Tn7-URA3	(229)
CMM7	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG rim101::Tn7-UAU1/rim101::Tn7-URA3 RPS1/RPS1::[pGAL10-lacZ-840-HIS1]	This study
rim13	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG rim13::Tn7-UAU1/rim13::Tn7-URA3	(71)
CMM8	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG rim13::Tn7-UAU1/rim13::Tn7-URA3 RPS1/RPS1::[pGAL10-lacZ-840-HIS1]	This study
gcn4	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gcn4::Tn7-UAU1/gcn4::Tn7-URA3	(229)
gcn4-ura3	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gcn4::Tn7-UAU1/gcn4::Tn7	This study
DAY286	ura3Δ::λimm434/ura3Δ::λimm434 arg4::URA3::arg4::hisG/arg4::hisG his1::hisG/his1::hisG	(71)
CMM9	ura3Δ::λimm434/ura3Δ::λimm434 arg4::URA3::arg4::hisG/arg4::hisG his1::hisG/his1::hisG RPS1/RPS1::[pGAL10-lacZ-750-HIS1]	This study
CAI8	ura3Δ::λimm434/ura3Δ::λimm434 ade2::hisG/ade2::hisG	(88)
CRC103	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pCRlacZ]	(256)
CRC106	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ]	(256)
CMM10	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA]	This study
CMM11	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GAL4]	This study
CMM12	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GAL4-GCN4AD]	This study
CMM13	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GAL4Δ188-261]	This study
CMM60	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GAL4Δ10-187]	This study
CMM14	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4]	This study
CMM62	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ247-323]	This study
CMM15	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ161-323]	This study
CMM16	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ123-323]	This study
CMM17	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ82-323]	This study
CMM18	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ69-323]	This study
CMM19	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ56-323]	This study
CMM20	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ42-323]	This study
CMM21	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ1-81]	This study
CMM64	ura3Δ::λimm434/ura3Δ::λ, ade2::hisG/ade2::hisG::[pOPlacZ] RPS1/RPS1::[CIp-lexA-GCN4Δ1-81Δ247-323]	This study
CMM25	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1]	This study
CMM26	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA]	This study
CMM27	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GAL4]	This study
CMM28	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GAL4-GCN4AD]	This study
CMM29	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GAL4Δ188-261]	This study
CMM61	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1]	This study

	RPS1/RPS1::[CIp-lexA-GAL4Δ10-187]	
CMM30	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4]	This study
CMM63	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ247-323]	This study
CMM31	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ161-323]	This study
CMM32	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ123-323]	This study
CMM33	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ82-323]	This study
CMM34	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ69-323]	This study
CMM35	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ56-323]	This study
CMM36	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ42-323]	This study
CMM37	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ1-81]	This study
CMM65	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG::[plexA-HIS1] RPS1/RPS1::[CIp-lexA-GCN4Δ1-81Δ247-323]	This study
CAI4	ura3Δ::λimm434/ura3Δ::λimm434	(88)
CMM41	ura3Δ::λimm434/ura3Δ::λimm434 RPS1/RPS1::[plac-poly]	This study
CMM43	ura3Δ::λimm434/ura3Δ::λimm434 RPS1/RPS1::[pGAL10-lacZ]	This study
CMM44	ura3Δ::λimm434/ura3Δ::λimm434 RPS1/RPS1::[pGAL10-lacZ-840]	This study
CMM99	ura3Δ::λimm434/ura3Δ::λimm434 RPS1/RPS1::[pGAL10-lacZ-188]	This study
CMM100	ura3Δ::λimm434/ura3Δ::λimm434 RPS1/RPS1::[pGAL10-lacZ-176]	This study
CMM101	ura3Δ::λimm434/ura3Δ::λimm434 RPS1/RPS1::[pGAL10-lacZ-840Δ176-188]	This study
JKC19	ura3Δ::λimm434/ura3Δ::λimm434 cph1::hisG/cph1::hisG-URA3-hisG	(188)
CMM102	ura3Δ::λimm434/ura3Δ::λimm434 cph1::hisG/cph1::hisG	This study
CMM103	ura3Δ::λimm434/ura3Δ::λimm434 cph1::hisG/cph1::hisG RPS1/RPS1::[pGAL10-lacZ-840]	This study
CMM104	ura3Δ::λimm434/ura3Δ::λimm434 cph1::hisG/cph1::hisG RPS1/RPS1::[pGAL10-lacZ-840Δ176-188]	This study
CMM105	ura3Δ::λimm434/ura3Δ::λimm434 cph1::hisG/cph1::hisG RPS1/RPS1::[pGAL10-lacZ-188]	This study
CMM106	ura3Δ::λimm434/ura3Δ::λimm434 cph1::hisG/cph1::hisG RPS1/RPS1::[pGAL10-lacZ-176]	This study
CMM54	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG GAL4/GAL4::[pH]::[pA] RPS1/RPS1::[pTLO-lacZ-126]	This study
CMM55	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::ARG4/gal4Δ::HIS1 RPS1/RPS1::[pTLO-lacZ-126]	This study
CMM56	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG GAL4/GAL4::[pH]::[pA] RPS1/RPS1::[pTLO-lacZ-114]	This study
CMM57	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::ARG4/gal4Δ::HIS1 RPS1/RPS1::[pTLO-lacZ-114]	This study
CMM65	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG GAL4/GAL4::[pH]::[pA] RPS1/RPS1::[pLAT1-lacZ-270]	This study
CMM66	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::ARG4/gal4Δ::HIS1 RPS1/RPS1::[pLAT1-lacZ-270]	This study
CMM67	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG GAL4/GAL4::[pH]::[pA] RPS1/RPS1::[pLAT1-lacZ-253]	This study
CMM68	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::ARG4/gal4Δ::HIS1 RPS1/RPS1::[pLAT1-lacZ-253]	This study
CMM58	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::HIS1/GAL4 CIp-lexa-GAL4	This study
CMM59	ura3Δ::λimm434/ura3Δ::λimm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG gal4Δ::HIS1/GAL4 CIp-lexa-GAL4-GCN4AD	This study
CMM85	ura3Δ::λimm434/ura3Δ::λimm434 arg4::hisG/arg4::hisG his1::hisG/his1::hisG::[pOPlacZ-HIS1] gcn4::Tn7-UAU1/gcn4::Tn7	This study
CMM86	ura3Δ::λimm434/ura3Δ::λimm434 arg4::hisG/arg4::hisG his1::hisG/his1::hisG::[pOPlacZ-HIS1] gcn4::Tn7-UAU1/gcn4::Tn7 RPS1/RPS1::[CIp-lexA]	This study
CMM87	ura3Δ::λimm434/ura3Δ::λimm434 arg4::hisG/arg4::hisG his1::hisG/his1::hisG::[pOPlacZ-HIS1] gcn4::Tn7-UAU1/gcn4::Tn7 RPS1/RPS1::[CIp-lexA-GCN4]	This study
CMM88	ura3Δ::λimm434/ura3Δ::λimm434 arg4::hisG/arg4::hisG his1::hisG/his1::hisG::[pOPlacZ-HIS1] gcn4::Tn7-UAU1/gcn4::Tn7 RPS1/RPS1::[CIp-lexA-GCN4Δ1-81]	This study
CMM89	ura3Δ::λimm434/ura3Δ::λimm434 arg4::hisG/arg4::hisG his1::hisG/his1::hisG::[pOPlacZ-HIS1] gcn4::Tn7-UAU1/gcn4::Tn7 RPS1/RPS1::[CIp-lexA-GCN4Δ1-81Δ247-323]	This study

Table 2: Plasmids used in this study.

Plasmids	Description	Source
pVEC	CaARS CaURA3	(205)
p37	CaARS upstream flank 1kb CaSOD5 downstream flank 1kb	This study
p37H	CaARS upstream flank 1kb CaHIS1downstream flank 1kb	This study
p37U	CaARS upstream flank 1kb hisG-CaURA3-hisG downstream flank 1kb	This study
pSOD5-110	CaARS CaURA3 upstream flank 1kb CaSOD5 downstream flank 1kb	This study

p5921	*hisG-CaURA3-hisG*	(88)
pRM100	*CaARS CaHIS1*	(239)
pFA-*HIS1*	pFA6a backbone, *CaHIS1*	(100)
pFA-*ARG4*	pFA6a backbone, *CaARG4*	(100)
p*HIS1-GAL4*	pFA6a backbone, upstream flan 600bp *CaHIS1* downstream flank 500bp	This study
p*ARG4-GAL4*	pFA6a backbone, upstream flan 600bp *CaARG4* downstream flank 500bp	This study
pOP*lacZ*	*CaADE2*, *lacZ*, *lexA* operon, *CaADH1* basal promoter	(256)
pOP*lacZ-HIS1*	*CaADE2*, *lacZ*, *lexA* operon, *CaADH1* basal promoter, *CaHIS1*	This study
pCR*lacZ*	*CaADE2*, *lacZ*, *CaADH1* basal promoter	(256)
p*lexA-HIS1*	pFA-*ARG4* backbone, promoterless *CaHIS1*, *lexA* operon, *ADH2* TATA box	This study
CIp-*lexA*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*	(256)
CIp-*lexA-GCN4*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4*	(256)
CIp-*lexA-GAL4*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGAL4*	This study
CIp-*lexA-GCN4Δ161-323*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ161-323*	This study
CIp-*lexA-GCN4Δ247-323*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ247-323*	This study
CIp-*lexA-GCN4Δ123-323*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ123-323*	This study
CIp-*lexA-GCN4Δ82-323*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ82-323*	This study
CIp-*lexA-GCN4Δ69-323*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ69-323*	This study
CIp-*lexA-GCN4Δ56-323*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ56-323*	This study
CIp-*lexA-GCN4Δ42-323*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ42-323*	This study
CIp-*lexA-GCN4Δ1-81*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ1-81*	This study
CIp-*lexA-GCN4Δ1-81Δ247-323*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGCN4Δ1-81Δ247-323*	This study
CIp-*lexA- GAL4Δ188-261*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGAL4Δ188-261*	This study
CIp-*lexA- GAL4Δ10-187*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGAL4Δ10-188*	This study
CIp-*lexA- GAL4-GCN4AD*	*CaURA3*, *CaRPS1*, promoter of *CaACT1*, *lexA*, *CaGAL4-GCN4AD*	This study
plac-poly	*CaURA3*, *RPS1*, *lacZ*	(Brown A, not published)
p*TLO-lacZ-126*	*CaURA3*, *RPS1*, *lacZ*, 126 bp promoter of *CaCTA2*	This study
p*TLO-lacZ-109*	*CaURA3*, *RPS1*, *lacZ*, 114 bp promoter of *CaCTA2*	This study
p*LAT1-lacZ-270*	*CaURA3*, *RPS1*, *lacZ*, 270 bp promoter of *CaLAT1*	This study
p*LAT1-lacZ-253*	*CaURA3*, *RPS1*, *lacZ*, 253 bp promoter of *CaLAT1*	This study
p*GAL10-lacZ*	*CaURA3*, *RPS1*, *lacZ*, full length promoter of *CaGAL10*	This study
p*GAL10-lacZ*-840	*CaURA3*, *RPS1*, *lacZ*, 840 bp promoter of *CaGAL10*	This study
p*GAL10-lacZ*-188	*CaURA3*, *RPS1*, *lacZ*, 180 bp promoter of *CaGAL10*	This study
p*GAL10-lacZ*-176	*CaURA3*, *RPS1*, *lacZ*, 176 bp promoter of *CaGAL10*	This study
p*GAL10-lacZ*-840Δ176-188	*CaURA3*, *RPS1*, *lacZ*, 840 bp Δ176-188 promoter of *CaGAL10*	This study

Table 3: Oligos used in this study.

Oligos	Sequence	Used to create
ODH157(*Eco*RI)	ggaattcAAACCTGGTCTACCAGAAGTT	*SOD5* + 1kb flank F
ODH158(*Eco*RI)	ggaattccGGTTATGTTAACGGTTAGAG	*SOD5* + 1kb flank R
OMM1(*Bam*HI)	cgggatccGATGAATGGTAAGTTAGATTGATAT	Diverge *SOD5* F
OMM2(*Bam*HI)	cgggatccGAAAAATAAAATAGATGAGCCATTT	Diverge *SOD5* R
OMM3(*Xba*I)	gctctagaAAAGAAAAATAAAATAGATGAGCC	Diverge *SOD5* R
OMM4(*Xba*I)	gctctagaGGTACCTGGAGGATGAGGAG	Ca*HIS1* F
CB20R(*Bam*HI)	cgggatccAATATTTATGAGAAACTATCACTTC	Ca*HIS1* R
OMM8	GGAGCAGTAGAAGCCATACTA	*SOD5* R Northern
OMM9	ATGGGCGAGTCCTACAAAACT	*SOD5* F Northern
OMM10	ACCAGTGAATCATTTGAAGTT	*SOD4* F Northern
OMM11	AGAAGCACTAGTTGATGAACC	*SOD4* R Northern
OMM14	TCCAGAAGATGATGAAAGACATG	*SOD1* F Northern
OMM15	GTCTAGCACCAGCATGACCAG	*SOD1* R Northern
OMM16	AACACAACTTGCTTTGAGTTGAAA	*SOD6* F Northern
OMM17	ACAATCACTAGAGATTCCGTGGAT	*SOD6* R Northern
OMM18(*Bam*HI)	cgggatccCGGTTATGTTAACGGTTAGAG	*SOD5* + 1kb flank R
OMM19(*Xba*I)	gctctagaAAACCTGGTCTACCAGAAGTT	*SOD5* + 1kb flank F
UO106	AAGCCGGTTTTGCCGGTG	*ACT1* probe F
UO107	TGGTGAACAATGGATGGAC	*ACT1* probe R
OMM46(*SalI*)	caggtcgacGAGAATGCCTATTGACTTTAAAGG	p*lexa-HIS1* R
OMM47(*SunI*)	cttcgtacgCGAACAAATATTCGATATACAAGCACTACACATAATGGATTTA GTCAATCATTTACCAG	p*lexa-HIS1* F
OMM50(*SunI*)	gtacgATACGAACAAATATTCGCAAATACGAACAAATATTCGCAAcccgggc	p*lexa-HIS1* R
OMM51(*SunI*)	gtacgcccgggTTGCGAATATTTGTTCGTATTTGCGAATATTTGTTCGTATc	p*lexa-HIS1* F
OMM56(*MluI*)	GTAATTGGTTTGTACCGCGAAATG	CIp-*lexA-GCN4*'s F
OMM57(*PstI*)	CACCGTACGGACTGCAGCTC	CIp-*lexA-GCN4Δ1-81* R
OMM58(*PstI*)	aactgcagttaAGTATCTTCTCTTGGAACAAGAGCA	CIp-*lexA-GCN4Δ161-323* R
OMM59(*PstI*)	aactgcagttaAGCTATAGCAACACCATCATCATTG	CIp-*lexA-GCN4Δ123-323* R
OMM60(*PstI*)	aactgcagttaATCATTGACTCCATCCAAATTGGTG	CIp-*lexA-GCN4Δ82-323* R
OMM61(*PstI*)	aactgcagttaACCAATTTCGTTTTTGGTAGCAACA	CIp-*lexA-GCN4Δ42-323* R
OMM62(*PstI*)	aactgcagttaTACACTGGAATGGATTTGAAATGGA	CIp-*lexA-GCN4Δ69-323* R
OMM63(*PstI*)	aactgcagttaTGAAGCCTTTTGCATTTCTGGTAAA	CIp-*lexA-GCN4Δ56-323* R
OMM71(*PstI*)	ccacgcgtatgACCAATTTGGATGGAGTCAATGATA	CIp-*lexA-GCN4Δ1-81* F
OMM66(*MluI*)	ccacgcgtATGTCTGAAACTAATGAAATCAAATTGG	CIp-*lexA-GAL4* F
OMM67(*PstI*)	aactgcagTCAAACGTTAACAGTTTCATCGACC	CIp-*lexA-GAL4* R
OMM125(*AatII*)	ttgacgtctATGCCTGCTACTACTCCTATTATTTATGAAGA	CIp-*lexA-GAL4-GCN4AD*
OMM87(*SunI*)	ttcgtacgTGTTGGATTACCAAGCACTTAAGCAT	*gal4* knock out

OMM88(*BamHI*)	acggatccGGCTGTTGGTAGTTTGGCGAA	*gal4* knock out
OMM89(*PmeI*)	ctgtttaaacTGAGTTTGTTTTTTATCGTTTAGTT	*gal4* knock out
OMM90(*SacI*)	acgagctcAACTAACCCAAAAAAAAATCGT	*gal4* knock out
OMM95	CAGATCTTTATACCCATGGTTG	*gal4* knock out
OMM96	AAACATACTATGGAAATAAATACCACC	*gal4* knock out
OMM99(*BamHI*)	ggggatccGGTATAACTCTTTCTTATAAAAATCGG	p*GAL10-lacZ*
OMM100(*BamHI*)	ggggatccGTTTAAGTTTTTATTATTATGAGTTGTATATG	p*GAL10-lacZ*
OMM115(*PstI*)	ggctgcagTTGGTTCCTCTTCGTCAATAT	p*GAL10-lacZ*-840Δ176-188
OMM116(*PstI*)	ggctgcagCTGAAGAAAACACATAATAAACGG	p*GAL10-lacZ*-840Δ176-188
OMM123(*PstI+BsrGI*)	gAGTGTAACGTTACATTGGTTCCTCTTCGTCAATATCAGATCTTCCCTCACt	p*GAL10-lacZ*-188
OMM124(*PstI+BsrGI*)	gtacaGTGAGGGAAGATCTGATATTGACGAAGAGGAACCAATGTAACGTTACACTctgca	p*GAL10-lacZ*-188
OMM131(*PstI+BsrGI*)	gctgcagTTGGTTCCTCTTCGTCAATATCAGATCTTCCCTCACt	p*GAL10-lacZ*-176
OMM132(*PstI+BsrGI*)	gtacaGTGAGGGAAGATCTGATATTGACGAAGAGGAACCAActgcagctgca	p*GAL10-lacZ*-176
OMM133(*PstI*)	ggctgcaGGGCCCGTCGACAAG	p*GAL10-lacZ*'s
OMM202(*PstI+BamHI*)	gCGGAGGTCGTGTCTCCGTCTACACTCTGGGTGTAGATGTCCCTATATATCAAGGTGG TATTTCCCTGACACAAACGTCGCATAAACCAACAAGAATAATTTTATCACACCTTTAT TTCCCCCCACCg	p*TLO-lacZ-126*
OMM203(*PstI+BamHI*)	gatccGGTGGGGGGAAATAAAGGTGTGATAAAATTATTCTTGTTGGTTTATGCGACGTT TGTGTCAGGGAAATACCACCTTGATATATAGGGACATCTACACCCAGAGTGTAGACG GAGACACGACCTCCGctgca	p*TLO-lacZ-126*
OMM204(*PstI+BamHI*)	gTCTACACTCTGGGTGTAGATGTCCCTATATATCAAGGTGGTATTTCCCTGACACAAA CGTCGCATAAACCAACAAGAATAATTTTATCACACCTTTATTTCCCCCCACCg	p*TLO-lacZ-109*
OMM205(*PstI+BamHI*)	gatccGGTGGGGGGAAATAAAGGTGTGATAAAATTATTCTTGTTGGTTTATGCGACGTT TGTGTCAGGGAAATACCACCTTGATATATAGGGACATCTACACCCAGAGTGTAGActgca	p*TLO-lacZ-109*
OMM239(*ZraI*)	ttgacgtcTCCAATTTGATTTCATTAGTTTCAGA	CIp-*lexA-GAL4AD* F
OMM240(*ZraI*)	CATTGACGTCTCCTTCATCGTTATTA	CIp-*lexA-GAL4AD* R
OMM241(*PstI*)	ggctgcagTGGAAAACACCTCCCCG	p*LSC2-lacZ-504* F
OMM242(*PstI*)	ggctgcagCAGCCCGGAAACAACAGG	p*LSC2-lacZ-487* F
OMM243(*BamHI*)	ggggatccTGTTGAAAAATATATAGATTATGAATATGTAATATCAG	p*LSC2-lacZ* R

APPENDIX B: Animal research and radioactive certificates

National Research Council Canada
Local Animal Care Committee

Form No. LACC 95
(revised 10/01/2001)

APPLICATION FOR USE OF LIVE ANIMALS IN RESEARCH
- ANIMAL USE PROTOCOL -

ACC use	Protocol I.D.# **02-011**	
	Start Avril 2002	End March 2003

Part I. *COVER PAGE*

1a. TITLE OF PROJECT **Please give a title that generally describes the project.**
Role of the TLR-2 receptor in the resistance to *Candida albicans* infection.

1b. SUMMARY OF PROCEDURES **Describe the procedures to which the animals will be subjected (Use key words. See example at annexe A.**
Injection (intravenous), peritoneal lavage, fungal infection, tissue collection, breeding, euthanasia

2. PRINCIPAL INVESTIGATOR

Name (degree/title)	Group/Section & Institute		Telephone (work & home)
Anne-Marie Alarco (Ph.D.)/	Health sector/BRI	w	496-6365
Post-doctoral fellow		h	748-7008

3. NRC GROUP/SECTION AFFILIATION OF PROJECT

Group or Section	Group Leader or Section Head	Telephone
Health, BRI	Malcolm Whiteway	496-6146

4. ANIMALS **(anticipated requirements)**

	Species	Strain	Age/Weight	Sex	No. per treatment group	Total Number per year
1.	Mus musculus	TLR2 KO	8-10 weeks	F/M	10	100
2.	Mus musculus	129/Sv	8-10 weeks	F/M	10	100
3.	Mus musculus	C57BL/6	8-10 weeks	F/M	10	100

5. CLASS OF ANIMAL USE & MAXIMUM CATEGORY OF INVASIVENESS **(See Annexe I)**

Acute (Non-survival)	Chronic	Maximum Duration of Survival, Chronic	Max. Invasiveness (A-E)
	X	30 days	D

6. PROJECT PERIOD ANTICIPATED	from	April 2002	to	March 2003

7. PROJECT STATUS

New	Continuing	Previous Protocol No., if continuing
X		

BIBLIOGRAPHY:

1. **Abouheif, E., and G. A. Wray.** 2002. Evolution of the gene network underlying wing polyphenism in ants. Science **297:**249-52.
2. **Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A. Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. H. Wei, C. Ibegwam, et al.** 2000. The genome sequence of Drosophila melanogaster. Science **287:**2185-95.
3. **Akache, B., K. Wu, and B. Turcotte.** 2001. Phenotypic analysis of genes encoding yeast zinc cluster proteins. Nucleic Acids Res **29:**2181-90.
4. **Alarco, A. M., I. Balan, D. Talibi, N. Mainville, and M. Raymond.** 1997. AP1-mediated multidrug resistance in *Saccharomyces cerevisiae* requires FLR1 encoding a transporter of the major facilitator superfamily. J Biol Chem **272:**19304-13.
5. **Altschul, S. F.** 1997. Evaluating the statistical significance of multiple distinct local alignments, p. 1-14. *In* S. Suhai (ed.), Theoretical and Computational Methods in Genome Research. Plenum, New York, NY.
6. **Arnaud, M. B., M. C. Costanzo, M. S. Skrzypek, G. Binkley, C. Lane, S. R. Miyasato, and G. Sherlock.** 2005. The Candida Genome Database (CGD), a community resource for Candida albicans gene and protein information. Nucleic Acids Res **33:**D358-63.
7. **Arndt, K., and G. R. Fink.** 1986. GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. Proc Natl Acad Sci U S A **83:**8516-20.
8. **Attardi, L. D., and R. Tjian.** 1993. Drosophila tissue-specific transcription factor NTF-1 contains a novel isoleucine-rich activation motif. Genes Dev **7:**1341-53.
9. **Ballance, D. J.** 1986. Sequences important for gene expression in filamentous fungi. Yeast **2:**229-36.

10. **Barberis, A., J. Pearlberg, N. Simkovich, S. Farrell, P. Reinagel, C. Bamdad, G. Sigal, and M. Ptashne.** 1995. Contact with a component of the polymerase II holoenzyme suffices for gene activation. Cell **81:**359-68.
11. **Barreto, L., A. Garcera, K. Jansson, P. Sunnerhagen, and E. Herrero.** 2006. A peroxisomal glutathione transferase of Saccharomyces cerevisiae is functionally related to sulfur amino acid metabolism. Eukaryot Cell **5:**1748-59.
12. **Bassel, J., and R. Mortimer.** 1971. Genetic order of the galactose structural genes in Saccharomyces cerevisiae. J Bacteriol **108:**179-83.
13. **Beauchamp, C., and I. Fridovich.** 1971. Superoxide dismutase: improved assays and an assay applicable to acrylamide gels. Anal Biochem **44:**276-87.
14. **Beck-Sague, C., and W. R. Jarvis.** 1993. Secular trends in the epidemiology of nosocomial fungal infections in the United States, 1980-1990. National Nosocomial Infections Surveillance System. J Infect Dis **167:**1247-51.
15. **Bennett, R. J., M. A. Uhl, M. G. Miller, and A. D. Johnson.** 2003. Identification and characterization of a Candida albicans mating pheromone. Mol Cell Biol **23:**8189-201.
16. **Berger, S. L.** 1999. Gene activation by histone and factor acetyltransferases. Curr Opin Cell Biol **11:**336-41.
17. **Bergmann, S., J. Ihmels, and N. Barkai.** 2004. Similarities and differences in genome-wide expression data of six organisms. PLoS Biol **2:**E9.
18. **Berman, J., and P. E. Sudbery.** 2002. Candida Albicans: a molecular revolution built on lessons from budding yeast. Nat Rev Genet **3:**918-30.
19. **Bernstein, B. E., J. K. Tong, and S. L. Schreiber.** 2000. Genomewide studies of histone deacetylase function in yeast. Proc Natl Acad Sci U S A **97:**13708-13.
20. **Berriman, M., and K. Rutherford.** 2003. Viewing and annotating sequence data with Artemis. Brief Bioinform **4:**124-32.
21. **Bevan, P., and H. C. Douglas.** 1969. Genetic control of phosphoglucomutase variants in Saccharomyces cerevisiae. J Bacteriol **98:**532-5.
22. **Bilinski, T., Z. Krawiec, A. Liczmanski, and J. Litwinska.** 1985. Is hydroxyl radical generated by the Fenton reaction in vivo? Biochem Biophys Res Commun **130:**533-9.
23. **Biswas, K., K. J. Rieger, and J. Morschhauser.** 2003. Functional analysis of CaRAP1, encoding the Repressor/activator protein 1 of Candida albicans. Gene **307:**151-8.
24. **Biswas, K., K. J. Rieger, and J. Morschhauser.** 2003. Functional characterization of CaCBF1, the Candida albicans homolog of centromere binding factor 1. Gene **323:**43-55.
25. **Boeke, J. D., F. LaCroute, and G. R. Fink.** 1984. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet **197:**345-6.
26. **Bosch, A. M.** 2006. Classical galactosaemia revisited. J Inherit Metab Dis **29:**516-25.
27. **Boube, M., L. Joulia, D. L. Cribbs, and H. M. Bourbon.** 2002. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. Cell **110:**143-51.

28. **Boucher, H. W., A. H. Groll, C. C. Chiou, and T. J. Walsh.** 2004. Newer systemic antifungal agents: pharmacokinetics, safety and efficacy. Drugs **64:**1997-2020.
29. **Boveris, A.** 1978. Production of superoxide anion and hydrogen peroxide in yeast mitochondria. In: Bacila M, Horecker BL, Stoppane AOM (eds) Biochemistry and genetics of yeast. Academic Press, New York:66-79.
30. **Braun, B. R., W. S. Head, M. X. Wang, and A. D. Johnson.** 2000. Identification and characterization of TUP1-regulated genes in Candida albicans. Genetics **156:**31-44.
31. **Braun, B. R., and A. D. Johnson.** 1997. Control of filament formation in Candida albicans by the transcriptional repressor TUP1. Science **277:**105-9.
32. **Braun, B. R., D. Kadosh, and A. D. Johnson.** 2001. NRG1, a repressor of filamentous growth in C.albicans, is down-regulated during filament induction. Embo J **20:**4753-61.
33. **Braun, B. R., van het Hoog, M., d'Enfert, C., Martchenko, M., Dungan, J., and A. Kuo, Inglis, D.O., Uhl, M.A., Hogues, H., Berriman, M., Lorenz, M., Levitin, A., Oberholzer, U., BAchewich, C., Harcus, D., Marcil, A., Dignard, D., Iouk, T., Zito, R., Frangeul, L., Tekaia, F., Rutherford, K., Wang, E., Munro, C., Bates, S., Gow, N.A., Hoyer, L., Kohler, G., Morschhauser, J., Newport, G., Znaidi, S., Raymond, M., Turcotte, B., Sherlock, G., Constanzo, M., Ihmels, J., Berman, J., Sanglard, D., Agabian, N., Mitchell, A., Johnson, A.D., Whiteway, M., Nantel, A.** 2005. A human-curated annotation of the Candida albicans genome. PLOS Genetics **1:**36-57.
34. **Breitling, R., O. Sharif, M. L. Hartman, and S. K. Krisans.** 2002. Loss of compartmentalization causes misregulation of lysine biosynthesis in peroxisome-deficient yeast cells. Eukaryot Cell **1:**978-86.
35. **Brown, A. J., and N. A. Gow.** 1999. Regulatory networks controlling Candida albicans morphogenesis. Trends Microbiol **7:**333-8.
36. **Brown, D. H., Jr., A. D. Giusani, X. Chen, and C. A. Kumamoto.** 1999. Filamentous growth of Candida albicans in response to physical environmental cues and its regulation by the unique CZF1 gene. Mol Microbiol **34:**651-62.
37. **Brownell, J. E., and C. D. Allis.** 1996. Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. Curr Opin Genet Dev **6:**176-84.
38. **Bruhn, L., J. J. Hwang-Shum, and G. F. Sprague, Jr.** 1992. The N-terminal 96 residues of MCM1, a regulator of cell type-specific genes in Saccharomyces cerevisiae, are sufficient for DNA binding, transcription activation, and interaction with alpha 1. Mol Cell Biol **12:**3563-72.
39. **Bruno, V. M., and A. P. Mitchell.** 2004. Large-scale gene function analysis in Candida albicans. Trends Microbiol **12:**157-61.
40. **Bryant, G. O., and M. Ptashne.** 2003. Independent recruitment in vivo by Gal4 of two complexes required for transcription. Mol Cell **11:**1301-9.
41. **Burley, S. K., and R. G. Roeder.** 1998. TATA box mimicry by TFIID: autoinhibition of pol II transcription. Cell **94:**551-3.

42. **Carlioz, A., and D. Touati.** 1986. Isolation of superoxide dismutase mutants in Escherichia coli: is superoxide dismutase necessary for aerobic life? Embo J **5:**623-30.
43. **Carlson, M.** 1999. Glucose repression in yeast. Curr Opin Microbiol **2:**202-7.
44. **Carlson, M., and D. Botstein.** 1982. Two differentially regulated mRNAs with different 5' ends encode secreted with intracellular forms of yeast invertase. Cell **28:**145-54.
45. **Carroll, S. B.** 2000. Endless forms: the evolution of gene regulation and morphological diversity. Cell **101:**577-80.
46. **Casteilla, L., M. Rigoulet, and L. Penicaud.** 2001. Mitochondrial ROS metabolism: modulation by uncoupling proteins. IUBMB Life **52:**181-8.
47. **Chance, B., H. Sies, and A. Boveris.** 1979. Hydroperoxide metabolism in mammalian organs. Physiol Rev **59:**527-605.
48. **Chen, D. C., B. C. Yang, and T. T. Kuo.** 1992. One-step transformation of yeast in stationary phase. Curr Genet **21:**83-4.
49. **Chen, X., B. B. Magee, D. Dawson, P. T. Magee, and C. A. Kumamoto.** 2004. Chromosome 1 trisomy compromises the virulence of Candida albicans. Mol Microbiol **51:**551-65.
50. **Chen, X., B. Zhang, P. M. Harmon, W. Schaffner, D. O. Peterson, and D. P. Giedroc.** 2004. A novel cysteine cluster in human metal-responsive transcription factor 1 is required for heavy metal-induced transcriptional activation in vivo. J Biol Chem **279:**4515-22.
51. **Cherry, J. M.** 1995. Genetic nomenclature guide. Saccharomyces cerevisiae. Trends Genet**:**11-2.
52. **Chibana, H., E. L. Heinecke, J. L. Beckerman, and P. T. Magee.** 2001. A system of rapid isolation of end-DNA from a small amount of fosmid DNA, with vector-based PCR for chromosome walking. Genome **44:**305-8.
53. **Chibana, H., S. Iwaguchi, M. Homma, A. Chindamporn, Y. Nakagawa, and K. Tanaka.** 1994. Diversity of tandemly repetitive sequences due to short periodic repetitions in the chromosomes of Candida albicans. Journal of Bacteriology **176:**3851-8.
54. **Chibana, H., B. B. Magee, S. Grindle, Y. Ran, S. Scherer, and P. T. Magee.** 1998. A physical map of chromosome 7 of Candida albicans. Genetics **149:**1739-52.
55. **Chibana, H., B. B. Magee, S. Grindle, Y. Ran, S. Scherer, and P. T. Magee.** 1998. A physical map of Chromosome 7 of Candida albicans. Genetics **149:**1739-1752.
56. **Chibana, H., N. Oka, H. Nakayama, T. Aoyama, B. B. Magee, P. T. Magee, and Y. Mikami.** 2005. Sequence finishing and gene mapping for Candida albicans chromosome 7 and syntenic analysis against the Saccharomyces cerevisiae genome. Genetics **170:**1525-37.
57. **Chibana, H., N. Oka, H. Nakayama, T. Aoyama, B. B. Magee, P. T. Magee, and Y. Mikami.** 2005. Sequence finishing and gene mapping for Candida albicans chromosome 7, and syntenic analysis against Saccharomyces cerevisiae genome. Genetics **170:**1525-37.

58. **Chindamporn, A., Y. Nakagawa, M. Homma, H. Chibana, M. Doi, and K. Tanaka.** 1995. Analysis of the chromosomal localization of the repetitive sequences (RPSs) in Candida albicans. Microbiology **141:**469-76.

59. **Chindamporn, A., Y. Nakagawa, I. Mizuguchi, H. Chibana, M. Doi, and K. Tanaka.** 1998. Repetitive sequences (RPSs) in the chromosomes of Candida albicans are sandwiched between two novel stretches, HOK and RB2, common to each chromosome. Microbiology **144:**849-857.

60. **Christie, K. R., S. Weng, R. Balakrishnan, M. C. Costanzo, K. Dolinski, S. S. Dwight, S. R. Engel, B. Feierbach, D. G. Fisk, J. E. Hirschman, E. L. Hong, L. Issel-Tarver, R. Nash, A. Sethuraman, B. Starr, C. L. Theesfeld, R. Andrada, G. Binkley, Q. Dong, C. Lane, M. Schroeder, D. Botstein, and J. M. Cherry.** 2004. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. Nucleic Acids Res **32 Database issue:**D311-4.

61. **Cleary, M. A., S. Stern, M. Tanaka, and W. Herr.** 1993. Differential positive control by Oct-1 and Oct-2: activation of a transcriptionally silent motif through Oct-1 and VP16 corecruitment. Genes Dev **7:**72-83.

62. **Cliften, P., P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton, J. Majors, R. Waterston, B. A. Cohen, and M. Johnston.** 2003. Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science **301:**71-6.

63. **Cliften, P. F., L. W. Hillier, L. Fulton, T. Graves, T. Miner, W. R. Gish, R. H. Waterston, and M. Johnston.** 2001. Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis. Genome Res **11:**1175-86.

64. **Corner, B. E., and P. T. Magee.** 1997. *Candida* pathogenesis: unravelling the threads of infection. Curr Biol **7:**R691-4.

65. **Cote, J., J. Quinn, J. L. Workman, and C. L. Peterson.** 1994. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science **265:**53-60.

66. **d'Enfert, C., S. Goyard, S. Rodriguez-Arnaveilhe, L. Frangeul, L. Jones, F. Tekaia, O. Bader, A. Albrecht, L. Castillo, A. Dominguez, J. F. Ernst, C. Fradin, C. Gaillardin, S. Garcia-Sanchez, P. de Groot, B. Hube, F. M. Klis, S. Krishnamurthy, D. Kunze, M. C. Lopez, A. Mavor, N. Martin, I. Moszer, D. Onesime, J. Perez Martin, R. Sentandreu, E. Valentin, and A. J. Brown.** 2005. CandidaDB: a genome database for Candida albicans pathogenomics. Nucleic Acids Res **33:**D353-7.

67. **Davidson, J. F., B. Whyte, P. H. Bissinger, and R. H. Schiestl.** 1996. Oxidative stress is involved in heat-induced cell death in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A **93:**5116-21.

68. **Davies, B. S., H. S. Wang, and J. Rine.** 2005. Dual activators of the sterol biosynthetic pathway of Saccharomyces cerevisiae: similar activation/regulatory domains but different response mechanisms. Mol Cell Biol **25:**7375-85.

69. **Davis, D., J. E. Edwards, Jr., A. P. Mitchell, and A. S. Ibrahim.** 2000. Candida albicans RIM101 pH response pathway is required for host-pathogen interactions. Infect Immun **68:**5953-9.

70. **Davis, D., R. B. Wilson, and A. P. Mitchell.** 2000. RIM101-dependent and-independent pathways govern pH responses in Candida albicans. Mol Cell Biol **20:**971-8.
71. **Davis, D. A., V. M. Bruno, L. Loza, S. G. Filler, and A. P. Mitchell.** 2002. Candida albicans Mds3p, a conserved regulator of pH responses and virulence identified through insertional mutagenesis. Genetics **162:**1573-81.
72. **De Freitas, J. M., A. Liba, R. Meneghini, J. S. Valentine, and E. B. Gralla.** 2000. Yeast lacking Cu-Zn superoxide dismutase show altered iron homeostasis. Role of oxidative stress in iron metabolism. J Biol Chem **275:**11645-9.
73. **de Groot, P. W., A. D. de Boer, J. Cunningham, H. L. Dekker, L. de Jong, K. J. Hellingwerf, C. de Koster, and F. M. Klis.** 2004. Proteomic analysis of Candida albicans cell walls reveals covalently bound carbohydrate-active enzymes and adhesins. Eukaryot Cell **3:**955-65.
74. **Diezmann, S., C. J. Cox, G. Schonian, R. J. Vilgalys, and T. G. Mitchell.** 2004. Phylogeny and evolution of medical species of Candida and related taxa: a multigenic analysis. J Clin Microbiol **42:**5624-35.
75. **Douglas, H. C.** 1961. A mutation in saccharomyces that affects phosphoglucomutase activity and galactose utilization. Biochim Biophys Acta **52:**209-11.
76. **Douglas, H. C., and D. C. Hawthorne.** 1964. Enzymatic Expression And Genetic Linkage Of Genes Controlling Galactose Utilization In Saccharomyces. Genetics **49:**837-44.
77. **Douglas, H. C., and D. C. Hawthorne.** 1966. Regulation of genes controlling synthesis of the galactose pathway enzymes in yeast. Genetics **54:**911-6.
78. **Drysdale, C. M., E. Duenas, B. M. Jackson, U. Reusser, G. H. Braus, and A. G. Hinnebusch.** 1995. The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. Mol Cell Biol **15:**1220-33.
79. **El Barkani, A., O. Kurzai, W. A. Fonzi, A. Ramon, A. Porta, M. Frosch, and F. A. Muhlschlegel.** 2000. Dominant active alleles of RIM101 (PRR2) bypass the pH restriction on filamentation of Candida albicans. Mol Cell Biol **20:**4635-47.
80. **Elion, E. A., J. A. Brill, and G. R. Fink.** 1991. FUS3 represses CLN1 and CLN2 and in concert with KSS1 promotes signal transduction. Proc Natl Acad Sci U S A **88:**9392-6.
81. **Ellenberger, T. E., C. J. Brandl, K. Struhl, and S. C. Harrison.** 1992. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. Cell **71:**1223-37.
82. **Enjalbert, B., A. Nantel, and M. Whiteway.** 2003. Stress induced gene expression in *Candida albicans*: Absence of a general stress response. Mol. Biol. Cell. in press.
83. **Estruch, J. J., L. Crossland, and S. A. Goff.** 1994. Plant activating sequences: positively charged peptides are functional as transcriptional activation domains. Nucleic Acids Res **22:**3983-9.
84. **Ferre-D'Amare, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley.** 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature **363:**38-45.

85. **Fields, S., and O. Song.** 1989. A novel genetic system to detect protein-protein interactions. Nature **340:**245-6.

86. **Fink, R. C., and J. G. Scandalios.** 2002. Molecular evolution and structure--function relationships of the superoxide dismutase gene families in angiosperms and their relationship to other eukaryotic and prokaryotic superoxide dismutases. Arch Biochem Biophys **399:**19-36.

87. **Flores, C. L., C. Rodriguez, T. Petit, and C. Gancedo.** 2000. Carbohydrate and energy-yielding metabolism in non-conventional yeasts. FEMS Microbiol Rev **24:**507-29.

88. **Fonzi, W. A., and M. Y. Irwin.** 1993. Isogenic strain construction and gene mapping in Candida albicans. Genetics **134:**717-28.

89. **Forche, A., G. May, and P. T. Magee.** 2005. Demonstration of loss of heterozygosity by single-nucleotide polymorphism microarray analysis and alterations in strain morphology in Candida albicans strains during infection. Eukaryot Cell **4:**156-65.

90. **Fradin, C., P. De Groot, D. MacCallum, M. Schaller, F. Klis, F. C. Odds, and B. Hube.** 2005. Granulocytes govern the transcriptional response, morphology and proliferation of Candida albicans in human blood. Mol Microbiol **56:**397-415.

91. **Fradin, C., and B. Hube.** 2003. Tissue infection and site-specific gene expression in Candida albicans. Adv Appl Microbiol **53:**271-90.

92. **Fridovich, I.** 1978. The biology of oxygen radicals. Science **201:**875-80.

93. **Fridovich, I.** 1995. Superoxide radical and superoxide dismutases. Annu Rev Biochem **64:**97-112.

94. **Galagan, J. E., S. E. Calvo, K. A. Borkovich, E. U. Selker, N. D. Read, D. Jaffe, W. FitzHugh, L. J. Ma, S. Smirnov, S. Purcell, B. Rehman, T. Elkins, R. Engels, S. Wang, C. B. Nielsen, J. Butler, M. Endrizzi, D. Qui, P. Ianakiev, D. Bell-Pedersen, M. A. Nelson, M. Werner-Washburne, C. P. Selitrennikoff, J. A. Kinsey, E. L. Braun, A. Zelter, U. Schulte, G. O. Kothe, G. Jedd, W. Mewes, C. Staben, E. Marcotte, D. Greenberg, A. Roy, K. Foley, J. Naylor, N. Stange-Thomann, R. Barrett, S. Gnerre, M. Kamal, M. Kamvysselis, E. Mauceli, C. Bielke, S. Rudd, D. Frishman, S. Krystofova, C. Rasmussen, R. L. Metzenberg, D. D. Perkins, S. Kroken, C. Cogoni, G. Macino, D. Catcheside, W. Li, R. J. Pratt, S. A. Osmani, C. P. DeSouza, L. Glass, M. J. Orbach, J. A. Berglund, R. Voelker, O. Yarden, M. Plamann, S. Seiler, J. Dunlap, A. Radford, R. Aramayo, D. O. Natvig, L. A. Alex, G. Mannhaupt, D. J. Ebbole, M. Freitag, I. Paulsen, M. S. Sachs, E. S. Lander, C. Nusbaum, and B. Birren.** 2003. The genome sequence of the filamentous fungus Neurospora crassa. Nature **422:**859-68.

95. **Garcera, A., L. Barreto, L. Piedrafita, J. Tamarit, and E. Herrero.** 2006. Saccharomyces cerevisiae cells have three Omega class glutathione S-transferases acting as 1-Cys thiol transferases. Biochem J **398:**187-96.

96. **Gasch, A. P., A. M. Moses, D. Y. Chiang, H. B. Fraser, M. Berardini, and M. B. Eisen.** 2004. Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol **2:**e398.

97. **Gaur, M., D. Choudhury, and R. Prasad.** 2005. Complete inventory of ABC proteins in human pathogenic yeast, Candida albicans. J Mol Microbiol Biotechnol **9:**3-15.
98. **Geiduschek, E. P., and G. A. Kassavetis.** 1992. RNA polymerase III transcription complexes. In Transcriptional Regulation (S.L. McKnight, K.R. Yamamoto, eds.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
99. **Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver.** 1996. Life with 6000 genes. Science **274:**546, 563-7.
100. **Gola, S., R. Martin, A. Walther, A. Dunkler, and J. Wendland.** 2003. New modules for PCR-based gene targeting in Candida albicans: rapid and efficient gene targeting using 100 bp of flanking homology region. Yeast **20:**1339-47.
101. **Gompel, N., B. Prud'homme, P. J. Wittkopp, V. A. Kassner, and S. B. Carroll.** 2005. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature **433:**481-7.
102. **Goodwin, T. J., and R. T. Poulter.** 1998. The CARE-2 and rel-2 repetitive elements of Candida albicans contain LTR fragments of a new retrotransposon. Gene **218:**85-93.
103. **Goodwin, T. J., and R. T. Poulter.** 2000. Multiple LTR-retrotransposon families in the asexual yeast Candida albicans. Genome Res **10:**174-91.
104. **Goshorn, A. K., S. M. Grindle, and S. Scherer.** 1992. Gene isolation by complementation in Candida albicans and applications to physical and genetic mapping. Infection & Immunity **60:**876-84.
105. **Gow, N. A., A. J. Brown, and F. C. Odds.** 2002. Fungal morphogenesis and host invasion. Curr Opin Microbiol **5:**366-71.
106. **Green, M. R.** 2000. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. Trends Biochem Sci **25:**59-63.
107. **Griggs, D. W., and M. Johnston.** 1991. Regulated expression of the GAL4 activator gene in yeast provides a sensitive genetic switch for glucose repression. Proc Natl Acad Sci U S A **88:**8597-601.
108. **Grunstein, M.** 1997. Histone acetylation in chromatin structure and transcription. Nature **389:**349-52.
109. **Gunasekaran, U., R. Yang, and M. Gunasekaran.** 1998. Regulation of superoxide dismutase synthesis in Candida albicans. Mycopathologia **141:**59-63.
110. **Hagman, J., M. J. Gutch, H. Lin, and R. Grosschedl.** 1995. EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. Embo J **14:**2907-16.
111. **Halliwell, B., and J. M. Gutteridge.** 1999. Free Radicals in Biology and Medicine. Oxford Science Publications, Oxford:105-245.
112. **Halliwell, B., and J. M. Gutteridge.** 1990. Role of free radicals and catalytic metal ions in human disease: an overview. Methods Enzymol **186:**1-85.
113. **Hampsey, M.** 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev **62:**465-503.

114. **Han, S. J., Y. C. Lee, B. S. Gim, G. H. Ryu, S. J. Park, W. S. Lane, and Y. J. Kim.** 1999. Activator-specific requirement of yeast mediator proteins for RNA polymerase II transcriptional activation. Mol Cell Biol **19:**979-88.
115. **Harbison, C. T., D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, and R. A. Young.** 2004. Transcriptional regulatory code of a eukaryotic genome. Nature **431:**99-104.
116. **Heckman, D. S., D. M. Geiser, B. R. Eidell, R. L. Stauffer, N. L. Kardos, and S. B. Hedges.** 2001. Molecular evidence for the early colonization of land by fungi and plants. Science **293:**1129-33.
117. **Hedges, S. B., J. E. Blair, M. L. Venturi, and J. L. Shoe.** 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol **4:**2.
118. **Hellauer, K., M. H. Rochon, and B. Turcotte.** 1996. A novel DNA binding motif for yeast zinc cluster proteins: the Leu3p and Pdr3p transcriptional activators recognize everted repeats. Mol Cell Biol **16:**6096-102.
119. **Herskowitz, I.** 1995. MAP kinase pathways in yeast: for mating and more. Cell **80:**187-97.
120. **Himmelfarb, H. J., J. Pearlberg, D. H. Last, and M. Ptashne.** 1990. GAL11P: a yeast mutation that potentiates the effect of weak GAL4-derived activators. Cell **63:**1299-309.
121. **Hinman, V. F., A. T. Nguyen, R. A. Cameron, and E. H. Davidson.** 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. Proc Natl Acad Sci U S A **100:**13356-61.
122. **Hinnebusch, A. G.** 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. Proc Natl Acad Sci U S A **81:**6442-6.
123. **Hinnebusch, A. G., and G. R. Fink.** 1983. Positive regulation in the general amino acid control of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A **80:**5374-8.
124. **Hinnebusch, A. G., and K. Natarajan.** 2002. Gcn4p, a master regulator of gene expression, is controlled at multiple levels by diverse signals of starvation and stress. Eukaryot Cell **1:**22-32.
125. **Hope, I. A., S. Mahadevan, and K. Struhl.** 1988. Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. Nature **333:**635-40.
126. **Hope, I. A., and K. Struhl.** 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. Embo J **6:**2781-4.
127. **Hoyer, L. L.** 2001. The ALS gene family of Candida albicans. Trends Microbiol **9:**176-80.
128. **Hoyer, L. L., and J. E. Hecht.** 2001. The ALS5 gene of Candida albicans and analysis of the Als5p N-terminal domain. Yeast **18:**49-60.
129. **Hoyer, L. L., T. L. Payne, M. Bell, A. M. Myers, and S. Scherer.** 1998. Candida albicans ALS3 and insights into the nature of the ALS gene family. Curr Genet **33:**451-9.

130. **Hube, B., and J. Naglik.** 2001. Candida albicans proteinases: resolving the mystery of a gene family. Microbiology **147:**1997-2005.

131. **Hube, B., and J. Naglik.** 2001. Candida albicans proteinases: resolving the mystery of a gene family. Microbiology **147:**1997-2005.

132. **Hube, B., F. Stehr, M. Bossenz, A. Mazur, M. Kretschmar, and W. Schafer.** 2000. Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members. Arch Microbiol **174:**362-74.

133. **Hwang, C. S., Y. U. Baek, H. S. Yim, and S. O. Kang.** 2003. Protective roles of mitochondrial manganese-containing superoxide dismutase against various stresses in Candida albicans. Yeast **20:**929-41.

134. **Hwang, C. S., G. Rhie, S. T. Kim, Y. R. Kim, W. K. Huh, Y. U. Baek, and S. O. Kang.** 1999. Copper- and zinc-containing superoxide dismutase and its gene from Candida albicans. Biochim Biophys Acta **1427:**245-55.

135. **Hwang, C. S., G. E. Rhie, J. H. Oh, W. K. Huh, H. S. Yim, and S. O. Kang.** 2002. Copper- and zinc-containing superoxide dismutase (Cu/ZnSOD) is required for the protection of *Candida albicans* against oxidative stresses and the expression of its full virulence. Microbiology **148:**3705-13.

136. **Hwang, D., J. J. Smith, D. M. Leslie, A. D. Weston, A. G. Rust, S. Ramsey, P. de Atauri, A. F. Siegel, H. Bolouri, J. D. Aitchison, and L. Hood.** 2005. A data integration methodology for systems biology: experimental verification. Proc Natl Acad Sci U S A **102:**17302-7.

137. **Ihmels, J., G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv, and N. Barkai.** 2002. Revealing modular organization in the yeast transcriptional network. Nat Genet **31:**370-7.

138. **Imlay, J. A., and I. Fridovich.** 1991. Isolation and genetic analysis of a mutation that suppresses the auxotrophies of superoxide dismutase-deficient *Escherichia coli* K12. Mol Gen Genet **228:**410-6.

139. **Ito-Kuwa, S., K. Nakamura, S. Aoki, T. Osafune, V. Vidotto, and K. Pienthaweechai.** 1999. Oxidative stress sensitivity and superoxide dismutase of a wild-type parent strain and a respiratory mutant of Candida albicans. Med Mycol **37:**307-14.

140. **Iwaguchi, S., M. Homma, H. Chibana, and K. Tanaka.** 1992. Isolation and characterization of a repeated sequence (RPS1) of Candida albicans. Journal of General Microbiology **138:**1893-900.

141. **Jackson, B. M., C. M. Drysdale, K. Natarajan, and A. G. Hinnebusch.** 1996. Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation. Mol Cell Biol **16:**5557-71.

142. **Jeong, C. J., S. H. Yang, Y. Xie, L. Zhang, S. A. Johnston, and T. Kodadek.** 2001. Evidence that Gal11 protein is a target of the Gal4 activation domain in the mediator. Biochemistry **40:**9421-7.

143. **Jeong, J. H., E. S. Kwon, and J. H. Roe.** 2001. Characterization of the manganese-containing superoxide dismutase and its gene regulation in stress response of Schizosaccharomyces pombe. Biochem Biophys Res Commun **283:**908-14.

144. **Jin, Y., L. P. Samaranayake, Y. Samaranayake, and H. K. Yip.** 2004. Biofilm formation of Candida albicans is variably affected by saliva and dietary sugars. Arch Oral Biol **49:**789-98.
145. **John, M., R. Leppik, S. J. Busch, M. Granger-Schnarr, and M. Schnarr.** 1996. DNA binding of Jun and Fos bZip domains: homodimers and heterodimers induce a DNA conformational change in solution. Nucleic Acids Res **24:**4487-94.
146. **Johnson, P. F., E. Sterneck, and S. C. Williams.** 1993. Activation domains of transcriptional regulatory proteins. J. Nutr. Biochem. **4:**386-98.
147. **Johnston, M.** 1987. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. Microbiol Rev **51:**458-76.
148. **Jones, T., N. A. Federspiel, H. Chibana, J. Dungan, S. Kalman, B. B. Magee, G. Newport, Y. R. Thorstenson, N. Agabian, P. T. Magee, R. W. Davis, and S. Scherer.** 2004. The diploid genome sequence of Candida albicans. Proc Natl Acad Sci U S A **101:**7329-34.
149. **Kadosh, D., and A. D. Johnson.** 2001. Rfg1, a protein related to the Saccharomyces cerevisiae hypoxic regulator Rox1, controls filamentous growth and virulence in Candida albicans. Mol Cell Biol **21:**2496-505.
150. **Kaiser, B., T. Munder, H. P. Saluz, W. Kunkel, and R. Eck.** 1999. Identification of a gene encoding the pyruvate decarboxylase gene regulator CaPdc2p from Candida albicans. Yeast **15:**585-91.
151. **Kanematsu, S., and K. Asada.** 1991. Chloroplast and cytosol isozymes of CuZn-superoxide dismutase: their characteristic amino acid sequences. Free Radic Res Commun **12-13 Pt 1:**383-90.
152. **Keegan, L., G. Gill, and M. Ptashne.** 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science **231:**699-704.
153. **Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander.** 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423:**241-54.
154. **Kelly, R., and K. J. Kwon-Chung.** 1992. A zinc finger protein from Candida albicans is involved in sucrose utilization. J Bacteriol **174:**222-32.
155. **Keyer, K., A. S. Gort, and J. A. Imlay.** 1995. Superoxide and the production of oxidative DNA damage. J Bacteriol **177:**6782-90.
156. **Kim, U. J., H. Shizuya, P. J. de Jong, B. Birren, and M. I. Simon.** 1992. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Research **20:**1083-5.
157. **Kohler, J. R., and G. R. Fink.** 1996. Candida albicans strains heterozygous and homozygous for mutations in mitogen-activated protein kinase signaling components have defects in hyphal development. Proc Natl Acad Sci U S A **93:**13223-8.
158. **Krebs, J. E., C. J. Fry, M. L. Samuels, and C. L. Peterson.** 2000. Global role for chromatin remodeling enzymes in mitotic gene expression. Cell **102:**587-98.
159. **Kullberg, B. J., and S. G. Filler.** 2002. Candidemia, p. 327-340. *In* R. Calderone (ed.), Candida and Candidiasis. ASM Press, Washington, DC.
160. **Kunze, D., I. Melzer, D. Bennett, D. Sanglard, D. MacCallum, J. Norskau, D. C. Coleman, F. C. Odds, W. Schafer, and B. Hube.** 2005. Functional analysis

of the phospholipase C gene CaPLC1 and two unusual phospholipase C genes, CaPLC2 and CaPLC3, of Candida albicans. Microbiology **151:**3381-94.

161. **Kurtzman, C., and P. Blanz.** 2000. Ribisomal RNA/DNA sequence comparisons for assessing phylogenetic relationships. Kutzman CP, Fell JW (Eds) The Yeasts: a Taxonomic Study, 4th edition.:69-74.

162. **Lamarre, C., J. D. LeMay, N. Deslauriers, and Y. Bourbonnais.** 2001. *Candida albicans* expresses an unusual cytoplasmic manganese-containing superoxide dismutase (SOD3 gene product) upon the entry and during the stationary phase. J Biol Chem **276:**43784-91.

163. **Lamb, T. M., W. Xu, A. Diamond, and A. P. Mitchell.** 2001. Alkaline response genes of *Saccharomyces cerevisiae* and their relationship to the RIM101 pathway. J Biol Chem **276:**1850-6.

164. **Lan, C. Y., G. Newport, L. A. Murillo, T. Jones, S. Scherer, R. W. Davis, and N. Agabian.** 2002. Metabolic specialization associated with phenotypic switching in Candida albicans. Proc Natl Acad Sci U S A **99:**14907-12.

165. **Lan, C. Y., G. Newport, L. A. Murillo, T. Jones, S. Scherer, R. W. Davis, and N. Agabian.** 2002. Metabolic specialization associated with phenotypic switching in Candidaalbicans. Proc Natl Acad Sci U S A **99:**14907-12.

166. **Lan, C. Y., G. Rodarte, L. A. Murillo, T. Jones, R. W. Davis, J. Dungan, G. Newport, and N. Agabian.** 2004. Regulatory networks affected by iron availability in Candida albicans. Mol Microbiol **53:**1451-69.

167. **Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al.** 2001. Initial sequencing and analysis of the human genome. Nature **409:**860-921.

168. **Lane, S., C. Birse, S. Zhou, R. Matson, and H. Liu.** 2001. DNA array studies demonstrate convergent regulation of virulence factors by Cph1, Cph2, and Efg1 in Candida albicans. J Biol Chem **276:**48988-96.

169. **Lane, S., C. Birse, S. Zhou, R. Matson, and H. Liu.** 2001. DNA array studies demonstrate convergent regulation of virulence factors by Cph1, Cph2, and Efg1 in Candida albicans. J Biol Chem **276:**48988-96.

170. **Laprade, L., V. L. Boyartchuk, W. F. Dietrich, and F. Winston.** 2002. Spt3 plays opposite roles in filamentous growth in Saccharomyces cerevisiae and Candida albicans and is required for C. albicans virulence. Genetics **161:**509-19.
171. **Lasker, B. A., L. S. Page, T. J. Lot, and G. S. Kobayashi.** 1992. Isolation, characterization, and sequencing of Candida albicans repetitive sequence element 2. Gene **116:**51-57.
172. **Laughon, A., and R. F. Gesteland.** 1982. Isolation and preliminary characterization of the GAL4 gene, a positive regulator of transcription in yeast. Proc Natl Acad Sci U S A **79:**6827-31.
173. **Leberer, E., D. Harcus, I. D. Broadbent, K. L. Clark, D. Dignard, K. Ziegelbauer, A. Schmidt, N. A. Gow, A. J. Brown, and D. Y. Thomas.** 1996. Signal transduction through homologs of the Ste20p and Ste7p protein kinases can trigger hyphal formation in the pathogenic fungus Candida albicans. Proc Natl Acad Sci U S A **93:**13217-22.
174. **Lee, J., E. S. Kwon, D. W. Kim, J. Cha, and J. H. Roe.** 2002. Regulation and the role of Cu,Zn-containing superoxide dismutase in cell cycle progression of Schizosaccharomyces pombe. Biochem Biophys Res Commun **297:**854-62.
175. **Lee, T. I., H. C. Causton, F. C. Holstege, W. C. Shen, N. Hannett, E. G. Jennings, F. Winston, M. R. Green, and R. A. Young.** 2000. Redundant roles for the TFIID and SAGA complexes in global transcription. Nature **405:**701-4.
176. **Lee, T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young.** 2002. Transcriptional regulatory networks in Saccharomyces cerevisiae. Science **298:**799-804.
177. **Lee, T. I., and R. A. Young.** 1998. Regulation of gene expression by TBP-associated proteins. Genes Dev **12:**1398-408.
178. **Lee, T. I., and R. A. Young.** 2000. Transcription of eukaryotic protein-coding genes. Annu Rev Genet **34:**77-137.
179. **Leloir, L. F.** 1951. The enzymatic transformation of uridine diphosphate glucose into a galactose derivative. Arch Biochem **33:**186-90.
180. **Lemieux, K., and L. Gaudreau.** 2004. Targeting of Swi/Snf to the yeast GAL1 UAS G requires the Mediator, TAF IIs, and RNA polymerase II. Embo J **23:**4040-50.
181. **Lenaz, G.** 2001. The mitochondrial production of reactive oxygen species: mechanisms and implications in human pathology. IUBMB Life **52:**159-64.
182. **Leng, P., P. R. Lee, H. Wu, and A. J. Brown.** 2001. Efg1, a morphogenetic regulator in Candida albicans, is a sequence-specific DNA binding protein. J Bacteriol **183:**4090-3.
183. **Lephart, P. R., H. Chibana, and P. T. Magee.** 2005. Effect of the major repeat sequence on chromosome loss in Candida albicans. Eukaryot Cell **4:**733-41.
184. **Li, M., S. J. Martin, V. M. Bruno, A. P. Mitchell, and D. A. Davis.** 2004. Candida albicans Rim13p, a protease required for Rim101p processing at acidic and alkaline pHs. Eukaryot Cell **3:**741-51.

185. **Limjindaporn, T., R. A. Khalaf, and W. A. Fonzi.** 2003. Nitrogen metabolism and virulence of Candida albicans require the GATA-type transcriptional activator encoded by GAT1. Mol Microbiol **50:**993-1004.
186. **Liochev, S. I., and I. Fridovich.** 1994. The role of O2.- in the production of HO.: in vitro and in vivo. Free Radic Biol Med **16:**29-33.
187. **Liu, H.** 2001. Transcriptional control of dimorphism in Candida albicans. Curr Opin Microbiol **4:**728-35.
188. **Liu, H., J. Kohler, and G. R. Fink.** 1994. Suppression of hyphal formation in Candida albicans by mutation of a STE12 homolog. Science **266:**1723-6.
189. **Liu, H., C. A. Styles, and G. R. Fink.** 1993. Elements of the yeast pheromone response pathway required for filamentous growth of diploids. Science **262:**1741-4.
190. **Lo, H. J., J. R. Kohler, B. DiDomenico, D. Loebenberg, A. Cacciapuoti, and G. R. Fink.** 1997. Nonfilamentous C. albicans mutants are avirulent. Cell **90:**939-49.
191. **Loftus, B. J., E. Fung, P. Roncaglia, D. Rowley, P. Amedeo, D. Bruno, J. Vamathevan, M. Miranda, I. J. Anderson, J. A. Fraser, J. E. Allen, I. E. Bosdet, M. R. Brent, R. Chiu, T. L. Doering, M. J. Donlin, C. A. D'Souza, D. S. Fox, V. Grinberg, J. Fu, M. Fukushima, B. J. Haas, J. C. Huang, G. Janbon, S. J. Jones, H. L. Koo, M. I. Krzywinski, J. K. Kwon-Chung, K. B. Lengeler, R. Maiti, M. A. Marra, R. E. Marra, C. A. Mathewson, T. G. Mitchell, M. Pertea, F. R. Riggs, S. L. Salzberg, J. E. Schein, A. Shvartsbeyn, H. Shin, M. Shumway, C. A. Specht, B. B. Suh, A. Tenney, T. R. Utterback, B. L. Wickes, J. R. Wortman, N. H. Wye, J. W. Kronstad, J. K. Lodge, J. Heitman, R. W. Davis, C. M. Fraser, and R. W. Hyman.** 2005. The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans. Science **307:**1321-4.
192. **Lohr, D., P. Venkov, and J. Zlatanova.** 1995. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. Faseb J **9:**777-87.
193. **Longo, V. D., E. B. Gralla, and J. S. Valentine.** 1996. Superoxide dismutase activity is essential for stationary phase survival in *Saccharomyces cerevisiae*. Mitochondrial production of toxic oxygen species in vivo. J Biol Chem **271:**12275-80.
194. **Ludwig, M. Z., C. Bergman, N. H. Patel, and M. Kreitman.** 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature **403:**564-7.
195. **Ludwig, M. Z., N. H. Patel, and M. Kreitman.** 1998. Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development **125:**949-58.
196. **Luikenhuis, S., G. Perrone, I. W. Dawes, and C. M. Grant.** 1998. The yeast Saccharomyces cerevisiae contains two glutaredoxin genes that are required for protection against reactive oxygen species. Mol Biol Cell **9:**1081-91.
197. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res **26:**1107-15.
198. **Lundberg, M., C. Johansson, J. Chandra, M. Enoksson, G. Jacobsson, J. Ljung, M. Johansson, and A. Holmgren.** 2001. Cloning and expression of a

novel human glutaredoxin (Grx2) with mitochondrial and nuclear isoforms. J Biol Chem **276**:26269-75.

199. **Lutfiyya, L. L., V. R. Iyer, J. DeRisi, M. J. DeVit, P. O. Brown, and M. Johnston.** 1998. Characterization of three related glucose repressors and genes they regulate in Saccharomyces cerevisiae. Genetics **150**:1377-91.
200. **Ma, J., and M. Ptashne.** 1988. Converting a eukaryotic transcriptional inhibitor into an activator. Cell **55**:443-6.
201. **Ma, J., and M. Ptashne.** 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. Cell **48**:847-53.
202. **Ma, J., and M. Ptashne.** 1987. A new class of yeast transcriptional activators. Cell **51**:113-9.
203. **Magee, B. B., B. Hube, R. J. Wright, P. J. Sullivan, and P. T. Magee.** 1993. The genes encoding the secreted aspartyl proteinases of Candida albicans constitute a family with at least three members. Infect Immun **61**:3240-3.
204. **Magee, B. B., M. Legrand, A. M. Alarco, M. Raymond, and P. T. Magee.** 2002. Many of the genes required for mating in Saccharomyces cerevisiae are also required for mating in Candida albicans. Mol Microbiol **46**:1345-51.
205. **Magee, B. B., and P. T. Magee.** 1997. WO-2, a stable aneuploid derivative of *Candida albicans* strain WO-1, can switch from white to opaque and form hyphae. Microbiology **143 (Pt 2)**:289-95.
206. **Malik, S., and R. G. Roeder.** 2000. Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. Trends Biochem Sci **25**:277-83.
207. **Malloy, J. M., J. P. Rieker, and C. F. Rizzo.** 1984. Quantitation of proteins on Coomassie blue-stained polyacrylamide gels based on spectrophotometric determination of electroeluted dye. Anal Biochem **141**:503-9.
208. **Marcil, A., D. Harcus, D. Y. Thomas, and M. Whiteway.** 2002. Candida albicans killing by RAW 264.7 mouse macrophage cells: effects of Candida genotype, infection ratios, and gamma interferon treatment. Infect Immun **70**:6319-29.
209. **Maris, A. F., A. L. Assumpcao, D. Bonatto, M. Brendel, and J. A. Henriques.** 2001. Diauxic shift-induced stress resistance against hydroperoxides in *Saccharomyces cerevisiae* is not an adaptive stress response and does not depend on functional mitochondria. Curr Genet **39**:137-49.
210. **Marmorstein, R., M. Carey, M. Ptashne, and S. C. Harrison.** 1992. DNA recognition by GAL4: structure of a protein-DNA complex. Nature **356**:408-14.
211. **Martchenko, M., A. M. Alarco, D. Harcus, and M. Whiteway.** 2004. Superoxide dismutases in Candida albicans: transcriptional regulation and functional characterization of the hyphal-induced SOD5 gene. Mol Biol Cell **15**:456-67.
212. **Martchenko, M., A. Levitin, and M. Whiteway.** 2006. Transcriptional activation domains of the Candida albicans Gcn4p and Gal4p homologs. Eukaryot Cell.
213. **Matthews, J. M., and M. Sunde.** 2002. Zinc fingers--folds for many occasions. IUBMB Life **54**:351-5.
214. **McCarroll, S. A., C. T. Murphy, S. Zou, S. D. Pletcher, C. S. Chin, Y. N. Jan, C. Kenyon, C. I. Bargmann, and H. Li.** 2004. Comparing genomic expression

patterns across species identifies shared transcriptional profile in aging. Nat Genet **36:**197-204.

215. **Meneghini, R.** 1997. Iron homeostasis, oxidative stress, and DNA damage. Free Radic Biol Med **23:**783-92.
216. **Miller, L. G., R. A. Hajjeh, and J. E. Edwards, Jr.** 2001. Estimating the cost of nosocomial candidemia in the united states. Clin Infect Dis **32:**1110.
217. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245:**371-8.
218. **Moreno, I., Y. Pedreno, S. Maicas, R. Sentandreu, E. Herrero, and E. Valentin.** 2003. Characterization of a Candida albicans gene encoding a putative transcriptional factor required for cell wall integrity. FEMS Microbiol Lett **226:**159-67.
219. **Murad, A. M., P. R. Lee, I. D. Broadbent, C. J. Barelle, and A. J. Brown.** 2000. CIp10, an efficient and convenient integrating vector for Candida albicans. Yeast **16:**325-7.
220. **Murad, A. M., P. Leng, M. Straffon, J. Wishart, S. Macaskill, D. MacCallum, N. Schnell, D. Talibi, D. Marechal, F. Tekaia, C. d'Enfert, C. Gaillardin, F. C. Odds, and A. J. Brown.** 2001. NRG1 represses yeast-hypha morphogenesis and hypha-specific gene expression in Candida albicans. Embo J **20:**4742-52.
221. **Myers, L. C., and R. D. Kornberg.** 2000. Mediator of transcriptional regulation. Annu Rev Biochem **69:**729-49.
222. **Naglik, J. R., S. J. Challacombe, and B. Hube.** 2003. Candida albicans secreted aspartyl proteinases in virulence and pathogenesis. Microbiol Mol Biol Rev **67:**400-28, table of contents.
223. **Nantel, A.** 2006. The long hard road to a completed Candida albicans genome. Fungal Genet Biol **43:**311-5.
224. **Nantel, A., D. Dignard, C. Bachewich, D. Harcus, A. Marcil, A. P. Bouin, C. W. Sensen, H. Hogues, M. van het Hoog, P. Gordon, T. Rigby, F. Benoit, D. C. Tessier, D. Y. Thomas, and M. Whiteway.** 2002. Transcription profiling of Candida albicans cells undergoing the yeast-to-hyphal transition. Mol Biol Cell **13:**3452-65.
225. **Nantel, A., T. Rigby, H. Hogues, and M. Whiteway.** 2006. Microarrays for studying pathology in *Candida albicans*., vol. In Press. Wiley Press, K. Kavanaugh Eds.
226. **Negredo, A., L. Monteoliva, C. Gil, J. Pla, and C. Nombela.** 1997. Cloning, analysis and one-step disruption of the ARG5,6 gene of Candida albicans. Microbiology **143 (Pt 2):**297-302.
227. **Nehlin, J. O., M. Carlberg, and H. Ronne.** 1991. Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. Embo J **10:**3373-7.
228. **Nicholls, S., M. Straffon, B. Enjalbert, A. Nantel, S. Macaskill, M. Whiteway, and A. J. Brown.** 2004. Msn2- and Msn4-like transcription factors play no obvious roles in the stress responses of the fungal pathogen Candida albicans. Eukaryot Cell **3:**1111-23.

229. **Nobile, C. J., and A. P. Mitchell.** 2005. Regulation of cell-surface genes and biofilm formation by the C. albicans transcription factor Bcr1p. Curr Biol **15:**1150-5.
230. **Nogi, Y., and T. Fukasawa.** 1989. Functional domains of a negative regulatory protein, GAL80, of Saccharomyces cerevisiae. Mol Cell Biol **9:**3009-17.
231. **Odds, F. C.** 1994. Pathogenesis of Candida infections. J Am Acad Dermatol **31:**S2-5.
232. **Ohama, T., T. Suzuki, M. Mori, S. Osawa, T. Ueda, K. Watanabe, and T. Nakase.** 1993. Non-universal decoding of the leucine codon CUG in several Candida species. Nucleic Acids Res **21:**4039-45.
233. **Orphanides, G., T. Lagrange, and D. Reinberg.** 1996. The general transcription factors of RNA polymerase II. Genes Dev **10:**2657-83.
234. **Panyim, S., and R. Chalkley.** 1969. High resolution acrylamide gel electrophoresis of histones. Arch Biochem Biophys **130:**337-46.
235. **Park, J. M., H. S. Kim, S. J. Han, M. S. Hwang, Y. C. Lee, and Y. J. Kim.** 2000. In vivo requirement of activator-specific binding targets of mediator. Mol Cell Biol **20:**8709-19.
236. **Pedrajas, J. R., E. Kosmidou, A. Miranda-Vizuete, J. A. Gustafsson, A. P. Wright, and G. Spyrou.** 1999. Identification and functional characterization of a novel mitochondrial thioredoxin system in Saccharomyces cerevisiae. J Biol Chem **274:**6366-73.
237. **Peng, G., and J. E. Hopper.** 2002. Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. Proc Natl Acad Sci U S A **99:**8548-53.
238. **Piskur, J., and R. B. Langkjaer.** 2004. Yeast genome sequencing: the power of comparative genomics. Mol Microbiol **53:**381-9.
239. **Pla, J., R. M. Perez-Diaz, F. Navarro-Garcia, M. Sanchez, and C. Nombela.** 1995. Cloning of the Candida albicans HIS1 gene by direct complementation of a C. albicans histidine auxotroph using an improved double-ARS shuttle vector. Gene **165:**115-20.
240. **Platt, A., and R. J. Reece.** 1998. The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex. Embo J **17:**4086-91.
241. **Prather, D. M., E. Larschan, and F. Winston.** 2005. Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in Saccharomyces cerevisiae. Mol Cell Biol **25:**2650-9.
242. **Qui, H. F., E. Dubois, and F. Messenguy.** 1991. Dissection of the bifunctional ARGRII protein involved in the regulation of arginine anabolic and catabolic pathways. Mol Cell Biol **11:**2169-79.
243. **Ramanan, N., and Y. Wang.** 2000. A high-affinity iron permease essential for Candida albicans virulence. Science **288:**1062-4.
244. **Ramon, A. M., and W. A. Fonzi.** 2003. Diverged binding specificity of Rim101p, the Candida albicans ortholog of PacC. Eukaryot Cell **2:**718-28.
245. **Reece, R. J., and M. Ptashne.** 1993. Determinants of binding-site specificity among yeast C6 zinc cluster proteins. Science **261:**909-11.
246. **Reeder, R. H.** 1992. Regulation of Transcription by RNA polymerase I. In Transcriptional Regulation (S.L. McKnight, K.R. Yamamoto, eds.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

247. **Reeves, E. P., H. Lu, H. L. Jacobs, C. G. Messina, S. Bolsover, G. Gabella, E. O. Potma, A. Warley, J. Roes, and A. W. Segal.** 2002. Killing activity of neutrophils is mediated through activation of proteases by K+ flux. Nature **416:**291-7.
248. **Reeves, W. M., and S. Hahn.** 2005. Targets of the Gal4 transcription activator in functional transcription complexes. Mol Cell Biol **25:**9092-102.
249. **Rhie, G. E., C. S. Hwang, M. J. Brady, S. T. Kim, Y. R. Kim, W. K. Huh, Y. U. Baek, B. H. Lee, J. S. Lee, and S. O. Kang.** 1999. Manganese-containing superoxide dismutase and its gene from Candida albicans. Biochim Biophys Acta **1426:**409-19.
250. **Rifkin, S. A., J. Kim, and K. P. White.** 2003. Evolution of gene expression in the Drosophila melanogaster subgroup. Nat Genet **33:**138-44.
251. **Riggle, P. J., K. A. Andrutis, X. Chen, S. R. Tzipori, and C. A. Kumamoto.** 1999. Invasive lesions containing filamentous forms produced by a Candida albicans mutant that is defective in filamentous growth in culture. Infect Immun **67:**3649-52.
252. **Rikhvanov, E. G., N. N. Varakina, T. M. Rusaleva, E. I. Rachenko, and V. K. Voinikov.** 2002. Sodium azide reduces the thermotolerance of respiratively grown yeasts. Curr Microbiol **45:**394-9.
253. **Roeder, R. G., and W. J. Rutter.** 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature **224:**234-7.
254. **Roemer, T., B. Jiang, J. Davison, T. Ketela, K. Veillette, A. Breton, F. Tandia, A. Linteau, S. Sillaots, C. Marta, N. Martel, S. Veronneau, S. Lemieux, S. Kauffman, J. Becker, R. Storms, C. Boone, and H. Bussey.** 2003. Large-scale essential gene identification in Candida albicans and applications to antifungal drug discovery. Mol Microbiol **50:**167-81.
255. **Rupp, S.** 2002. LacZ assays in yeast. Methods Enzymol **350:**112-31.
256. **Russell, C. L., and A. J. Brown.** 2005. Expression of one-hybrid fusions with Staphylococcus aureus lexA in Candida albicans confirms that Nrg1 is a transcriptional repressor and that Gcn4 is a transcriptional activator. Fungal Genet Biol **42:**676-83.
257. **Rustici, G., J. Mata, K. Kivinen, P. Lio, C. J. Penkett, G. Burns, J. Hayles, A. Brazma, P. Nurse, and J. Bahler.** 2004. Periodic gene expression program of the fission yeast cell cycle. Nat Genet **36:**809-17.
258. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944-5.
259. **Samaranayake, Y. H., R. S. Dassanayake, J. A. Jayatilake, B. P. Cheung, J. Y. Yau, K. W. Yeung, and L. P. Samaranayake.** 2005. Phospholipase B enzyme expression is not associated with other virulence attributes in Candida albicans isolates from patients with human immunodeficiency virus infection. J Med Microbiol **54:**583-93.
260. **Sander, C. S., U. C. Hipler, U. Wollina, and P. Elsner.** 2002. Inhibitory effect of terbinafine on reactive oxygen species (ROS) generation by Candida albicans. Mycoses **45:**152-5.

261. **Sanglard, D., F. Ischer, M. Monod, and J. Bille.** 1997. Cloning of Candida albicans genes conferring resistance to azole antifungal agents: characterization of CDR2, a new multidrug ABC transporter gene. Microbiology **143 (Pt 2):**405-16.
262. **Sanglard, D., K. Kuchler, F. Ischer, J. L. Pagani, M. Monod, and J. Bille.** 1995. Mechanisms of resistance to azole antifungal agents in Candida albicans isolates from AIDS patients involve specific multidrug transporters. Antimicrob Agents Chemother **39:**2378-86.
263. **Santos, M. A., and M. F. Tuite.** 1995. The CUG codon is decoded in vivo as serine and not leucine in Candida albicans. Nucleic Acids Res **23:**1481-6.
264. **Sanyal, K., M. Baum, and J. Carbon.** 2004. Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique. Proc Natl Acad Sci U S A **101:**11374-9.
265. **Schjerling, P., and S. Holmberg.** 1996. Comparative amino acid sequence analysis of the C6 zinc cluster family of transcriptional regulators. Nucleic Acids Res **24:**4599-607.
266. **Schroter, C., U. C. Hipler, A. Wilmer, W. Kunkel, and U. Wollina.** 2000. Generation of reactive oxygen species by *Candida albicans* in relation to morphogenesis. Arch Dermatol Res **292:**260-4.
267. **Segal, E., M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman.** 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat Genet **34:**166-76.
268. **Selmecki, A., S. Bergmann, and J. Berman.** 2005. Comparative genome hybridization reveals widespread aneuploidy in Candida albicans laboratory strains. Mol Microbiol **55:**1553-65.
269. **Selmecki, A., A. Forche, and J. Berman.** 2006. Aneuploidy and isochromosome formation in drug-resistant Candida albicans. Science **313:**367-70.
270. **Seyler, R. W., Jr., J. W. Olson, and R. J. Maier.** 2001. Superoxide dismutase-deficient mutants of Helicobacter pylori are hypersensitive to oxidative stress and defective in host colonization. Infect Immun **69:**4034-40.
271. **Sigler, P. B.** 1988. Transcriptional activation. Acid blobs and negative noodles. Nature **333:**210-2.
272. **Sil, A. K., S. Alam, P. Xin, L. Ma, M. Morgan, C. M. Lebo, M. P. Woods, and J. E. Hopper.** 1999. The Gal3p-Gal80p-Gal4p transcription switch of yeast: Gal3p destabilizes the Gal80p-Gal4p complex in response to galactose and ATP. Mol Cell Biol **19:**7828-40.
273. **Singh, V., I. Sinha, and P. P. Sadhale.** 2005. Global analysis of altered gene expression during morphogenesis of Candida albicans in vitro. Biochem Biophys Res Commun **334:**1149-58.
274. **Smith, F. C., S. P. Davies, W. A. Wilson, D. Carling, and D. G. Hardie.** 1999. The SNF1 kinase complex from Saccharomyces cerevisiae phosphorylates the transcriptional repressor protein Mig1p in vitro at four sites within or near regulatory domain 1. FEBS Lett **453:**219-23.
275. **Smith, M. W., and R. F. Doolittle.** 1992. A comparison of evolutionary rates of the two major kinds of superoxide dismutase. J Mol Evol **34:**175-84.

276. **Spector, D., J. Labarre, and M. B. Toledano.** 2001. A genetic investigation of the essential role of glutathione: mutations in the proline biosynthesis pathway are the only suppressors of glutathione auxotrophy in yeast. J Biol Chem **276:**7011-6.
277. **Struhl, K.** 1997. Selective roles for TATA-binding-protein-associated factors in vivo. Genes Funct **1:**5-9.
278. **Stuart, J. M., E. Segal, D. Koller, and S. K. Kim.** 2003. A gene-coexpression network for global discovery of conserved genetic modules. Science **302:**249-55.
279. **Sudarsanam, P., and F. Winston.** 2000. The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. Trends Genet **16:**345-51.
280. **Svejstrup, J. Q., P. Vichi, and J. M. Egly.** 1996. The multiple roles of transcription/repair factor TFIIH. Trends Biochem Sci **21:**346-50.
281. **Talibi, D., and M. Raymond.** 1999. Isolation of a putative Candida albicans transcriptional regulator involved in pleiotropic drug resistance by functional complementation of a pdr1 pdr3 mutation in Saccharomyces cerevisiae. J Bacteriol **181:**231-40.
282. **Tanay, A., A. Regev, and R. Shamir.** 2005. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. Proc Natl Acad Sci U S A **102:**7203-8.
283. **Tanay, A., R. Sharan, M. Kupiec, and R. Shamir.** 2004. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. Proc Natl Acad Sci U S A **101:**2981-6.
284. **Tekaia, F., G. Blandin, A. Malpertuy, B. Llorente, P. Durrens, C. Toffano-Nioche, O. Ozier-Kalogeropoulos, E. Bon, C. Gaillardin, M. Aigle, M. Bolotin-Fukuhara, S. Casaregola, J. de Montigny, A. Lepingle, C. Neuveglise, S. Potier, J. Souciet, M. Wesolowski-Louvel, and B. Dujon.** 2000. Genomic exploration of the hemiascomycetous yeasts: 3. Methods and strategies used for sequence analysis and annotation. FEBS Lett **487:**17-30.
285. **Thoma, F., T. Koller, and A. Klug.** 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol **83:**403-27.
286. **Thrash-Bingham, C., and J. A. Gorman.** 1993. Identification, characterization and sequence of Candida albicans repetitive DNAs Rel-1 and Rel-2. Current Genetics **23:**455-62.
287. **Timson, D. J., H. C. Ross, and R. J. Reece.** 2002. Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry. Biochem J **363:**515-20.
288. **Todd, R. B., and A. Andrianopoulos.** 1997. Evolution of a fungal regulatory gene family: the Zn(II)2Cys6 binuclear cluster DNA binding motif. Fungal Genet Biol **21:**388-405.
289. **Toledano, M., A. Delaunay, B. Biteau, D. Spector, and D. Azevedo.** 2003. Oxidative stress responses in yeast. In: Hohmann S, Mager PWH (Eds) Topics in CurrentcGenetics, vol. 1, Yeast StresscResponses.:241-303.
290. **Tournu, H., G. Tripathi, G. Bertram, S. Macaskill, A. Mavor, L. Walker, F. C. Odds, N. A. Gow, and A. J. Brown.** 2005. Global role of the protein kinase Gcn2 in the human pathogen Candida albicans. Eukaryot Cell **4:**1687-96.

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

292. **Triezenberg, S. J.** 1995. Structure and function of transcriptional activation domains. Curr Opin Genet Dev **5:**190-6.

293. **Tripathi, G., C. Wiltshire, S. Macaskill, H. Tournu, S. Budge, and A. J. Brown.** 2002. Gcn4 co-ordinates morphogenetic and metabolic responses to amino acid starvation in Candida albicans. Embo J **21:**5448-56.

294. **Trotter, E. W., and C. M. Grant.** 2005. Overlapping roles of the cytoplasmic and mitochondrial redox regulatory systems in the yeast Saccharomyces cerevisiae. Eukaryot Cell **4:**392-400.

295. **Trotter, E. W., and C. M. Grant.** 2002. Thioredoxins are required for protection against a reductive stress in the yeast Saccharomyces cerevisiae. Mol Microbiol **46:**869-78.

296. **True, J. R., and S. B. Carroll.** 2002. Gene co-option in physiological and morphological evolution. Annu Rev Cell Dev Biol **18:**53-80.

297. **Tschopp, J. F., S. D. Emr, C. Field, and R. Schekman.** 1986. GAL2 codes for a membrane-bound subunit of the galactose permease in Saccharomyces cerevisiae. J Bacteriol **166:**313-8.

298. **Tsong, A. E., B. B. Tuch, H. Li, and A. D. Johnson.** 2006. Evolution of alternative transcriptional circuits with identical logic. Nature **443:**415-20.

299. **Tzung, K. W., R. M. Williams, S. Scherer, N. Federspiel, T. Jones, N. Hansen, V. Bivolarevic, L. Huizar, C. Komp, R. Surzycki, R. Tamse, R. W. Davis, and N. Agabian.** 2001. Genomic evidence for a complete sexual cycle in Candida albicans. Proc Natl Acad Sci U S A **98:**3249-53.

300. **Uhl, M. A., M. Biery, N. Craig, and A. D. Johnson.** 2003. Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen C.albicans. Embo J **22:**2668-78.

301. **Van Hoy, M., K. K. Leuther, T. Kodadek, and S. A. Johnston.** 1993. The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. Cell **72:**587-94.

302. **Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson,**

M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature **420:**520-62.

303. **West, R. W., Jr., R. R. Yocum, and M. Ptashne.** 1984. Saccharomyces cerevisiae GAL1-GAL10 divergent promoter region: location and function of the upstream activating sequence UASG. Mol Cell Biol **4:**2467-78.
304. **White, T. C., S. H. Miyasaki, and N. Agabian.** 1993. Three distinct secreted aspartyl proteinases in Candida albicans. J Bacteriol **175:**6126-33.
305. **Wilson, R. B., D. Davis, and A. P. Mitchell.** 1999. Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. J Bacteriol **181:**1868-74.
306. **Wolberger, C., A. K. Vershon, B. Liu, A. D. Johnson, and C. O. Pabo.** 1991. Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell **67:**517-28.
307. **Wolfe, K. H., and D. C. Shields.** 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature **387:**708-13.
308. **Wong, S., G. Butler, and K. H. Wolfe.** 2002. Gene order evolution and paleopolyploidy in hemiascomycete yeasts. Proc Natl Acad Sci U S A **99:**9272-7.
309. **Wood, V., R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabbinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schafer, S. Muller-Auer, C. Gabel, M. Fuchs, C. Fritzc, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dreano, S. Gloux, V. Lelaure, et al.** 2002. The genome sequence of Schizosaccharomyces pombe. Nature **415:**871-80.
310. **Wray, G. A., M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer, M. V. Rockman, and L. A. Romano.** 2003. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol **20:**1377-419.
311. **Yoo, H. Y., S. S. Kim, and H. M. Rho.** 1999. Overexpression and simple purification of human superoxide dismutase (SOD1) in yeast and its resistance to oxidative stress. J Biotechnol **68:**29-35.
312. **Zaragoza, O., C. Rodriguez, and C. Gancedo.** 2000. Isolation of the MIG1 gene from Candida albicans and effects of its disruption on catabolite repression. J Bacteriol **182:**320-6.
313. **Zhou, K. M., and G. B. Kohlhaw.** 1990. Transcriptional activator LEU3 of yeast. Mapping of the transcriptional activation function and significance of activation domain tryptophans. J Biol Chem **265:**17409-12.