Title
The regulation and implications of white-opaque switching in *Candida albicans*

Author
Lohse, Matthew B.

Publication Date
2010

Peer reviewed|Thesis/dissertation

The regulation and implications of white-opaque switching in *Candida albicans*

by

Matthew B. Lohse

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements

I would like to start by thanking my advisor, Sandy Johnson. For the past five and a half years, Sandy has done an amazing job as an advisor and a mentor both scientifically and professionally. Working with him on four papers has taught me a great deal about the hardest part of science… communicating your results to the scientific community as a whole. Sandy has always been available to talk and has provided countless suggestions that helped shape my research. I am grateful for the support and guidance he has provided throughout my graduate career.

My thesis committee, Hana El-Samad, Hao Li, Wendell Lim, and Hiten Madhani, have also been a great source of knowledge and advice during the research process and made my committee meetings an enjoyable and informative event. Their questions have made me think about many issues that I might otherwise have overlooked, and their suggestions have led to several informative lines of experimentation.

There are many current and former members of the Johnson lab that contributed to my research. First and foremost, Rebecca Zordan not only laid the foundation for much of my research with her work on *WOR1*, she happily answered countless questions over the years. Both Becky and Shannon Stroschein-Stevenson were excellent mentors during my rotation and provided invaluable help during the phagocytosis studies presented in Chapter 2. Aaron Hernday also was a great source of knowledge, helping with topics as far ranging as microarrays and gel shifts, and served as an excellent sounding board for many of my ideas. Chris Cain also was another willing sounding board as well as an excellent source of input on a variety of scientific topics. Chris's work with YEL007 played an important part in the development of the Wor1 story

presented in Chapter 3, making it far more comprehensive than it might otherwise have been. Oliver Homann kindly put up with hundreds of questions about microarrays and experimental design, without his input the temperature shift studies would not have been nearly as elegant or informative. Cihan Oguz, from Hana El-Samad's lab group, was always a pleasure to collaborate with on the modeling of the white-opaque regulatory circuit and got me to think about switching in ways I had never anticipated. Other lab members, including but not limited to Lauren Booth, Chris Baker, Jeanselle Dea, Sarah Elson, Emily Fox, Brad Green, Lisa Kohn, Quinn Mitrovich, Clarissa Nobile, Suzanne Noble, Christian Perez, and Brian Tuch, have been a pleasure to work with and have always made lab meetings enjoyable. Sudarsi Desta and Jorge Mendoza made much of this work possible, by ensuring that I always had the media and glassware I needed even if I requested it at the last minute.

Graduate school would have been much more trying without the support of my friends. Chris Rivera and Jason Huff, who I lovingly call the existential crisis twins, as well as Erik Lontok, Katherine Sorber, and Chris Brown have been the participants in many an enjoyable conversation, outing, or meal. They were always willing to give advice, listen to a complaint, and to share a drink or five. Plus, Erik and Jason hosted the most important party of my life... for which I am forever grateful.

I have had far too many great teachers over the years to list them all, but I will always be especially grateful for having known Mrs. Hatchett, Mrs. Houston, Mrs. Johnston, Mr. Townley, Dr. Sakon, Dr. Henry, Dr. Coon, and Dr. Sloan. They played an important role in developing my love of both science and history as well as learning in general.

Many thanks go out to my family, who were always supportive of me, never asked when I was going to be done with school, and never said "why aren't you doing something with cancer." My parents, William and Cindy Lohse, instilled me with a love of learning and a sense of curiosity that have been critical to my success through the years. They also tolerated my strange sense of humor, possibly to the regret of everyone who has known me over the years. What I am, I am because of them. My sister, Alison, is truly an awesome sister (as determined by the PBR correlation statistical test, $p<.01$) who will soon be a doctor herself and I think has forgiven me for rolling her down the hall as a kid.

Finally, I really cannot adequately describe the support I have received from my wife, Sarah Foss. Over the past several years, Sarah has kindly proofed countless drafts of documents I was working on and listened to me repeatedly practice numerous talks. More importantly, Sarah has been my solemate and brought endless joy and humor into my existence. I really cannot describe how important Sarah has been to me, and this really doesn't seem like the place for a Simpsons or a Lord North reference, so I shall close by noting that I look forward to spending my life with her.

Chapter 1 has been slightly modified from its original format, where it was originally published as "White-opaque switching in *Candida albicans*." Matthew B. Lohse and Alexander D. Johnson (2009) *Current Opinion in Microbiology* 12:650-654.

and electronic formats. A copy of this license is included at the end of this section. This review was written by both Matthew B. Lohse and Alexander D. Johnson.

The research described in Chapter 2 was originally published with open access as "Differential phagocytosis of white versus opaque *Candida albicans* by *Drosophila* and mouse phagocytes." Matthew B. Lohse and Alexander D. Johnson (2008) *PLoS One* 3:e1473. PLoS publications are freely available through the Creative Commons Attributions License, which allows reprints so long as the original source and authors are cited. All work described was performed by Matthew Lohse. Alexander Johnson oversaw the work.

The research described in Chapter 3 was originally published as "Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in Candida albicans." Matthew B. Lohse, Rebecca E. Zordan, Christopher W. Cain, and Alexander D. Johnson (2010) *Proceedings of the National Academy of Sciences U S A* 107:14105-14110. Copyright for this paper is retained by the authors, although PNAS retains an exclusive license to publish these articles. Authors need not obtain specific permission to reprint their work for inclusion in thesis dissertations. All work described was performed by Matthew Lohse with the following exceptions. Rebecca Zordan decided on the regions of Wor1 to use and designed and cloned the MBP-based Wor1 fusions. Christopher Cain deleted *YEL007* and *YHR177* from *S. cerevisiae* as well as building the Wor1 ectopic expression plasmid and optimizing the procedure used for activation assays. Alexander Johnson oversaw the work.

Chapter 4 was originally published as "Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans." Matthew B. Lohse and

Alexander D. Johnson (2010) *Molecular Microbiology* 78:331-343. ©2010 by Blackwell Publishing Ltd. Blackwell Publishing is part of John Wiley and Sons, which has granted a license (#2499570525197, August 31, 2010) to me (Matthew B. Lohse) to include the full article as a part of my dissertation in both print and electronic formats. A copy of this license is included at the end of this section. All work described was performed by Matthew Lohse. Alexander Johnson oversaw the work.

The research described in Appendix 1 was performed by Matthew Lohse. Alexander Johnson oversaw the work.

The research described in Appendix 2 was performed by Matthew Lohse. During this project, valuable help and advice was provided by several members of Robert Stroud's lab at UCSF, namely Akram Alian, Anirban Adhikari, Louis Metzger, John Pak, and Sarah Griner. This work would not have been possible without their help and guidance. Alexander Johnson oversaw the work.

The research described in Appendix 3 was performed by Matthew Lohse. Alexander Johnson oversaw the work.

The research described in Appendix 4 was performed by Matthew Lohse. Valuable input was provided during the course of research by Cihan Oguz, Christopher Cain, and Aaron Hernday. Alexander Johnson and Hana El-Samad oversaw the work.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Nov 18, 2010

This is a License Agreement between Matthew B Lohse ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Matthew B Lohse
Customer address	MB Genentech Hall RM N374 San Francisco, CA 94143
License number	2552130856647
License date	Nov 18, 2010
Licensed content publisher	Elsevier
Licensed content publication	Current Opinion in Microbiology
Licensed content title	White–opaque switching in *Candida albicans*
Licensed content author	Matthew B Lohse, Alexander D Johnson
Licensed content date	December 2009
Licensed content volume number	12
Licensed content issue number	6
Number of pages	5
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	
Title of your thesis/dissertation	The regulation and implications of white-opaque switching in Candida albicans
Expected completion date	Nov 2010
Estimated size (number of pages)	285

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Nov 17, 2010

This is a License Agreement between Matthew B Lohse ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2499570525197
License date	Aug 31, 2010
Licensed content publisher	John Wiley and Sons
Licensed content publication	Molecular Microbiology
Licensed content title	Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans
Licensed content author	Matthew B. Lohse,Alexander D. Johnson
Licensed content date	Sep 1, 2010
Start page	no
End page	no
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Order reference number	
Total	0.00 USD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one if its group companies (each a "Wiley Company") or a society for whom a Wiley Company has exclusive publishing rights in relation to a particular journal (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your Rightslink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

1. The materials you have requested permission to reproduce (the "Materials") are protected by copyright.

2. You are hereby granted a personal, non-exclusive, non-sublicensable, non-transferable,

The regulation and implications of white-opaque switching in *Candida albicans*

Matthew B. Lohse

Abstract

The human fungal pathogen *Candida albicans* undergoes a switch between two distinct cell types, referred to as white and opaque. These cell types differ in cell and colony morphologies, ability to mate, metabolic preferences, preferred niches in the host, and interactions with the host innate immune system. Both cell types are stable through many generations; switching between them is rare, stochastic, and occurs without any known changes in the primary sequence of the genome; thus the switch is epigenetic. When I started my work, the circuit regulating this switch had just been identified and recently published results suggested that the innate immune system recognized the two cell types differently.

My work determined that both the *Drosophila* hemocyte-derived S2 cell line and the mouse macrophage-derived RAW264.7 cell lines preferentially phagocytose white cells over opaque cells. This difference is observed for both the percentage of S2 or RAW cells that phagocytose *C. albicans* cells as well as the average number of *C. albicans* cells phagocytosed by individual S2 or RAW cells.

A second line of research focused on the transcriptional regulator Wor1, a master regulator of white-opaque switching. Wor1 belongs to a family of proteins lacking similarity to any known DNA-binding protein but conserved across the fungal kingdom. My work determined that Wor1 binds DNA directly, identified the DNA sequence

recognized by Wor1, and demonstrated that this sequence is sufficient for Wor1-dependent activation of transcriptional *in vivo*. This conserved domain, which we have termed the WOPR box, represents a distinct and previously unidentified family of DNA-binding proteins.

Finally, we investigated the order of regulatory changes during the switch from the opaque to the white cell type. Surprisingly, changes in key transcriptional regulators occur gradually, extending over several cell divisions with little cell-to-cell variation. Additional experiments, including perturbations to regulator concentrations, refine the signature of the commitment point. Transcriptome analysis reveals that opaque cells begin to globally resemble white cells well before they irreversibly commit to switching. We propose that these characteristics of the switching process permit *C. albicans* to "test the waters" before making an all-or-none decision.

Table of Contents

List of Tables

List of Figures

Chapter 1:

White-opaque switching in *Candida albicans*

White-opaque switching in *Candida albicans*

Matthew B. Lohse[1#] and Alexander D. Johnson[1,2*]

[1]Departments of Biochemistry and Biophysics and [2]Microbiology and Immunology, University of California San Francisco, San Francisco, California, United States of America

Summary

The human commensal yeast *Candida albicans* undergoes an epigenetic switch between two distinct types of cells, referred to as white and opaque. These two cell types differ in many respects, including their cell and colony morphologies, their metabolic states, their mating behaviors, their preferred niches in the host, and their interactions with the host immune system. Each of the two cell types is heritable for many generations and switching between them appears stochastic; however, environmental cues can significantly alter the frequency of switching. We review recent work on white-opaque switching, including the establishment of the transcriptional circuit underlying this switch, the identification of environmental signals that affect switching rates, newly discovered differences between the two types of cells, and the involvement of white-opaque switching in biofilm formation. We also review recent speculation on the evolution and adaptive value of white-opaque switching.

Introduction

Candida albicans, the most common human fungal pathogen, resides asymptomatically in the gastrointestinal tract of most healthy humans. However, when the host immune system is compromised due to age, disease, or medical treatment, *C. albicans* can cause a range of medical problems; including disseminated blood-born infections with mortality rates exceeding 40% (1). *C. albicans* differs from the great majority of fungal species (and even from some very closely related species) by possessing the ability to switch between two distinctive types of cells, white and opaque. Each cell type is heritable for many generations and switching occurs without a change in the DNA sequence of the genome (hence our use of the term 'epigenetic') (2-6). Under standard laboratory conditions switching is rare (approximately every 10^4 cell divisions) although, as will be discussed, environmental factors can strongly influence the switching rate (7). White and opaque cell types have distinct properties which result largely from the differential regulation of approximately 400 genes, about seven percent of the genome (8, 9). Metabolic preferences (8), environmental responses (10), biofilm interactions (11), host-cell immune interactions (12, 13), and the ability to mate (14) show significant differences between white and opaque cells. It is an attractive hypothesis that white-opaque switching evolved during *C. albicans*' long association with its warm-blooded host and that it plays a crucial role in the pathogen-host relationship.

White-opaque switching is closely linked with mating

White-opaque switching was discovered in 1987 by David Soll and colleagues (2). Its key role in the mating cycle of *C. albicans* was established some 15 years later

(15, 16). In brief, *C. albicans*' mating is controlled by transcriptional regulators encoded at the Mating Type Like (MTL) locus. There are two versions of this locus, referred to as **a** and α, which contain transcriptional regulators specifying **a** and α-type mating, respectively. In cells containing both versions of this locus, referred to as **a**/α cells, two homeodomain proteins (one from the **a** locus and one from the α locus) form the **a**1-α2 heterodimer which represses mating functions as well as white-opaque switching. In order to mate, an **a** cell must encounter an α cell but both cells must be in the opaque form for mating to culminate. These observations help to explain why *C. albicans* mating was so difficult to detect: more than 95% of clinical *C. albicans* isolates are **a**/α strains (17) (and hence are mating and switching incompetent) and, in order to mate, **a** and α cells both have to undergo the rare switch to the opaque state [Figure 1].

Multiple feedback loops contribute to stability of the two cell types

One key characteristic of the white and opaque cell types is their stability over thousands of generations under normal laboratory conditions; that is, upon cell division, white cells almost always give rise to white cell progeny and opaque cells to opaque progeny. It has been proposed that this behavior results from the topology of the transcriptional circuit underlying the switch; this circuit consists largely of interlocking positive feedback loops [Figure 2a] (18).

At the core of this circuit is a protein called WOR1, the first major regulator of the opaque state to be identified (19-21). WOR1 expression is required to switch from the white to the opaque state and ectopic expression of WOR1 can drive an entire population of white cells to the opaque state. WOR1 expression produces a direct positive feedback loop by binding its own promoter and turning on its own expression

(19-21). Activation of this feedback loop produces a forty-fold increase in WOR1 transcript levels in opaque cells compared to white cells. The WOR1 promoter is directly repressed by the **a**1-α2 heterodimer, thus explaining the inability of **a**/α cells to switch from white to opaque [Figure 2b] (9, 14, 20).

Three additional transcriptional regulators, EFG1, WOR2, and CZF1, complete the known regulatory circuit [Figure 2a] (18, 22-25), and it has been proposed that this circuit can account for the major characteristics of the white-opaque switch. According to this model, the circuit is largely inactive in the white state; this is the default state [Figure 2c]. Switching occurs when the circuit becomes excited; because of the series of positive feedback loops, the circuit can remain excited for many generations [Figure 2d]. It has been hypothesized that inheritance of the opaque state results from molecules of the regulators being passed on to daughter cells following cell division; the concentrations of these regulators in the daughter cells would then be sufficiently high to re-excite the circuit and retain the opaque state. What triggers white-to-opaque switching in the first place? It has been proposed that when components of the switching circuit reach a critical threshold concentration, the circuit becomes excited. In principle, this could occur through random fluctuations in the levels of a critical molecule such as WOR1. The reverse process, opaque-to-white switching, would then occur when insufficient quantities of the regulators are passed on to a daughter cell and the circuit would wind down.

Layered on this core transcriptional circuit are several chromatin modifying factors whose presence also affects switching. For example, deletion of the histone deacetylase HDA1 slightly increases switching rates from white to opaque while deletion

of RPD3, another histone deacetylase, increases switching in both directions (26). Recently, the acetyltransferase NAT4 and histone deacetylases HST2, SET3, and HOS2 have also been implicated in the regulation of switching. Deletion of each of these genes reduced rates of white-to-opaque switching and all but HST2 increased switching from opaque to white (25). It is not yet clear precisely what role these chromatin remodeling factors play in the switch; one simple model is that they serve to reinforce the characteristics of the transcriptional circuit described above.

Growth rate affects switching frequency

Although the circuit described above can explain much of the behavior of the two cell types, it does not readily account for how environmental cues can affect the rate of switching. It has long been known that oxidative stress increases switching from white to opaque (10) while elevated temperatures drive switching from opaque to white (7). Recent studies have identified several new signals that affect switching frequencies and have led to new hypotheses linking these signals to the regulatory circuit

Exposure to the DNA damaging agent methyl methane sulfonate (MMS) and deletions of the DNA repair genes RAD51 and RAD52 both increase switching from white to opaque (27). Hydroxyurea (HU), which delays cells in S phase of the cell cycle, also increased the white-to-opaque switching rate. These results suggest a direct link between progression though the cell cycle and white-opaque switching. According to one model, an extended cell cycle in slowly growing cells would facilitate switching by allowing WOR1 to accumulate to higher levels before being diluted by cell division; this would result in a greater proportion of cells with sufficient WOR1 to excite the regulatory circuit shown in Figure 2. In support of this hypothesis, depletion of the mitotic cell

cycle regulator CLB4 slows cell growth and also produces increased switching from white to opaque (27).

Anaerobic growth stabilizes the opaque cell type

Recent work has also revealed new insights into the interplay between temperature and white-opaque switching. The observation of *in vivo* mating of *C. albicans* in a mouse host (28) had long presented a paradox. *C. albicans* must be opaque in order to mate (14) and yet opaque cells were not believed to be stable at the internal body temperature of 37°C. However, it has recently been discovered that opaque cells are stable and capable of mating at 37°C if also exposed to anaerobic conditions (29). Further work revealed that anaerobic conditions not only stabilize opaque cells but also stimulate white to opaque switching (30). Studies in the mouse support this view: when feces were collected from mice injected with a switching competent white strain, it was discovered that between 4 and 10 percent of cells were in the opaque form. Thus white-to-opaque switching can indeed occur at 37°C in a mouse model.

What is the basis for anaerobic growth stimulating white-to-opaque switching? Growth under anaerobic conditions leads to ergosterol depletion, and it has been hypothesized that this metabolic cue may signal cells to switch. In support of this view, addition of the ergosterol biosynthesis inhibitors lovastatin or ketoconazole produced an increase in switching to opaque similar to that seen under anaerobic conditions (30).

An independent set of experiments has shown that elevated CO_2 levels can also stimulate white-to-opaque switching and can stabilize opaque cells at 37°C (31). Because CO_2 levels vary dramatically among different tissues of the host, this

observation strongly suggests that rates of white-opaque switching are tuned to different niches within the host.

Opaque cells may evade the innate immune system

Differences between white and opaque cells significantly affect *C. albicans*' interactions with the host. For example, white cells appear better suited to internal infections while opaque cells thrive in skin infections (32-34). Opaque and white cells also differ in their interactions with specific components of the host immune system. For example, opaque cells, unlike white cells, do not secrete a chemoattractant for leukocytes (12). As a possible consequence of this difference, opaque cells are significantly less susceptible than white cells to phagocytosis by macrophage cell lines *in vitro* (13). This ability to avoid the innate immune system may help *C. albicans* colonize internal environments where opaque cells are particularly stable. It should be stressed, however, that we know relatively little about the complete relationship between the two cell types and the host.

White cells respond to pheromone by forming biofilms

In response to the α mating pheromone, opaque **a** cells up-regulate several hundred genes, many of which are involved in preparing the cells for the subsequent steps of mating (35). Although they appear unable to mate, white **a** cells also respond to the α pheromone (11, 36). Both white and opaque **a** cells sense α-factor through the same α-factor receptor, STE2 (37), although white cells use a novel signaling loop not required for detection in opaque cells (38).

The response of white cells to mating pheromone is markedly different from the response observed in opaque cells. Many fewer genes are induced and, rather than

preparing to mate, white cells undergo a different type of response: they form a biofilm (11). What could be the purpose of this response? One hypothesis is that biofilm formation facilitates subsequent mating. According to this idea, detection of pheromone allows white cells to recognize the presence of a small population of opaque cells, and by forming a biofilm, concentrates cells of opposite mating types to promote mating. Regardless of its precise role, the observation that white and opaque cells respond very differently to mating pheromone is a particularly fascinating difference between the two types of cells.

Conservation of the switch components

Although white-opaque switching appears unique to *C. albicans* and its very close relatives (39), the components that regulate the switch are conserved among many fungal species. For example, WOR1, the central regulator of white-opaque switching, is a member of a large family of fungal proteins distinguished by a highly conserved N-terminal region (21). Although nearly every species of fungi for which a genome sequence is available has a WOR1 homolog, the structure and biochemical function of this conserved region is not known. In all species where they have been studied, the WOR1 homologs appear to play some role in regulating morphological changes. For example, the *Histoplasma capsulatum* WOR1 ortholog RYP1 is a key regulator of the yeast-to-mycelial switch. RYP1 is differentially expressed in the two types of cells, binds its own promoter, and is necessary for the establishment and maintenance of the yeast cell type (40). At this point, we can only guess how these deeply conserved transcriptional regulators became "wired-up" to catalyze white-opaque switching in *C. albicans*.

Conclusion

Despite many advances in our understanding of the regulation and role of white-opaque switching in *C. albicans*, we still possess only a tentative grasp of many important aspects of this switching system. It remains to be determined whether random fluctuations in regulator levels can explain the stochastic nature of switching. The role of opaque cells in the host and their interaction with the innate immune system remain among the most pressing questions in this field. Specific host niches that stabilize opaque cells or promote switching from white to opaque remain largely unidentified. The exact role of biofilm formation in facilitating *in vivo* mating remains to be determined, as does the *in vivo* role of anaerobic growth and CO_2. Finally, the selective pressures under which the white-opaque switch evolved remain obscure.

Acknowledgements

The authors thank Chris Cain, Sarah Foss, Aaron Hernday, and Oliver Homann for their valuable comments on this manuscript. Work of the authors was supported by grants from the National Institutes of Health (AI49187) and the Ellison Foundation (ID-SS-0628-04).

References

1. Noble SM, Johnson AD (2007) Genetics of *Candida albicans*, a diploid human fungal pathogen. *Annu Rev Genet* 41:193-211.
2. Slutsky B, et al. (1987) "White-opaque transition": a second high-frequency switching system in *Candida albicans*. *J Bacteriol* 169:189-197.
3. Soll DR, Morrow B, Srikantha T (1993) High-frequency phenotypic switching in *Candida albicans*. *Trends Genet* 9:61-65.
4. Johnson A (2003) The biology of mating in *Candida albicans*. *Nat Rev Microbiol* 1:106-116.
5. Lockhart SR, Daniels KJ, Zhao R, Wessels D, Soll DR (2003) Cell biology of mating in *Candida albicans*. *Eukaryot Cell* 2:49-61.
6. Whiteway M, Bachewich C (2007) Morphogenesis in *Candida albicans*. *Annu Rev Microbiol* 61:529-553.
7. Rikkerink EH, Magee BB, Magee PT (1988) Opaque-white phenotype transition: a programmed morphological transition in *Candida albicans*. *J Bacteriol* 170:895-899.
8. Lan CY, et al. (2002) Metabolic specialization associated with phenotypic switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 99:14907-14912.
9. Tsong AE, Miller MG, Raisner RM, Johnson AD (2003) Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115:389-399.
10. Kolotila MP, Diamond RD (1990) Effects of neutrophils and in vitro oxidants on survival and phenotypic switching of *Candida albicans* WO-1. *Infect Immun* 58:1174-1179.

11. Daniels KJ, Srikantha T, Lockhart SR, Pujol C, Soll DR (2006) Opaque cells signal white cells to form biofilms in *Candida albicans*. *EMBO J* 25:2240-2252.

12. Geiger J, Wessels D, Lockhart SR, Soll DR (2004) Release of a potent polymorphonuclear leukocyte chemoattractant is regulated by white-opaque switching in *Candida albicans*. *Infect Immun* 72:667-677.

13. Lohse MB, Johnson AD (2008) Differential phagocytosis of white versus opaque *Candida albicans* by *Drosophila* and mouse phagocytes. *PLoS One* 3:e1473.

14. Miller MG, Johnson AD (2002) White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* 110:293-302.

15. Magee PT, Magee BB (2004) Through a glass opaquely: the biological significance of mating in *Candida albicans*. *Curr Opin Microbiol* 7:661-665.

16. Bennett RJ, Johnson AD (2005) Mating in *Candida albicans* and the search for a sexual cycle. *Annu Rev Microbiol* 59:233-255.

17. Lockhart SR, et al. (2002) In *Candida albicans*, white-opaque switchers are homozygous for mating type. *Genetics* 162:737-745.

18. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* 5:e256.

19. Huang G, et al. (2006) Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 103:12813-12818.

20. Srikantha T, et al. (2006) TOS9 regulates white-opaque switching in *Candida albicans*. *Eukaryot Cell* 5:1674-1687.

21. Zordan RE, Galgoczy DJ, Johnson AD (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci U S A* 103:12807-12812.

22. Vinces MD, Kumamoto CA (2007) The morphogenetic regulator Czf1p is a DNA-binding protein that regulates white opaque switching in *Candida albicans*. *Microbiology* 153:2877-2884.

23. Srikantha T, Tsai LK, Daniels K, Soll DR (2000) EFG1 null mutants of *Candida albicans* switch but cannot express the complete phenotype of white-phase budding cells. *J Bacteriol* 182:1580-1591.

24. Sonneborn A, Tebarth B, Ernst JF (1999) Control of white-opaque phenotypic switching in *Candida albicans* by the Efg1p morphogenetic regulator. *Infect Immun* 67:4655-4660.

25. Hnisz D, Schwarzmüller T, Kuchler K (2009) Transcriptional loops meet chromatin: a dual-layer network controls white-opaque switching in *Candida albicans*. *Mol Microbiol* 74:1-15.

26. Srikantha T, Tsai L, Daniels K, Klar AJ, Soll DR (2001) The histone deacetylase genes HDA1 and RPD3 play distinct roles in regulation of high-frequency phenotypic switching in *Candida albicans*. *J Bacteriol* 183:4614-4625.

27. Alby K, Bennett RJ (2009) Stress-induced phenotypic switching in *Candida albicans*. *Mol Biol Cell* 20:3178-3191.

28. Hull CM, Raisner RM, Johnson AD (2000) Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. *Science* 289:307-310.

29. Dumitru R, et al. (2007) In vivo and in vitro anaerobic mating in *Candida albicans*. *Eukaryot Cell* 6:465-472.

30. Ramírez-Zavala B, Reuss O, Park YN, Ohlsen K, Morschhäuser J (2008) Environmental induction of white-opaque switching in *Candida albicans*. *PLoS Pathog* 4:e1000089.

31. Huang G, Srikantha T, Sahni N, Yi S, Soll DR (2009) CO(2) regulates white-to-opaque switching in *Candida albicans*. *Curr Biol* 19:330-334.

32. Kvaal CA, Srikantha T, Soll DR (1997) Misexpression of the white-phase-specific gene WH11 in the opaque phase of *Candida albicans* affects switching and virulence. *Infect Immun* 65:4468-4475.

33. Kvaal C, et al. (1999) Misexpression of the opaque-phase-specific gene PEP1 (SAP1) in the white phase of *Candida albicans* confers increased virulence in a mouse model of cutaneous infection. *Infect Immun* 67:6652-6662.

34. Lachke SA, Lockhart SR, Daniels KJ, Soll DR (2003) Skin facilitates *Candida albicans* mating. *Infect Immun* 71:4970-4976.

35. Zhao R, et al. (2005) Unique aspects of gene expression during *Candida albicans* mating and possible G(1) dependency. *Eukaryot Cell* 4:1175-1190.

36. Lockhart SR, Zhao R, Daniels KJ, Soll DR (2003) Alpha-pheromone-induced "shmooing" and gene regulation require white-opaque switching during *Candida albicans* mating. *Eukaryot Cell* 2:847-855.

37. Yi S, et al. (2008) The same receptor, G protein, and mitogen-activated protein kinase pathway activate different downstream regulators in the alternative white and opaque pheromone responses of *Candida albicans*. *Mol Biol Cell* 19:957-970.

38. Yi S, et al. (2009) A *Candida albicans*-specific region of the alpha-pheromone receptor plays a selective role in the white cell pheromone response. *Mol Microbiol* 71:925-947.

39. Pujol C, et al. (2004) The closely related species *Candida albicans* and *Candida dubliniensis* can mate. *Eukaryot Cell* 3:1015-1027.

40. Nguyen VQ, Sil A (2008) Temperature-induced switch to the pathogenic yeast form of *Histoplasma capsulatum* requires Ryp1, a conserved transcriptional regulator. *Proc Natl Acad Sci U S A* 105:4880-4885.

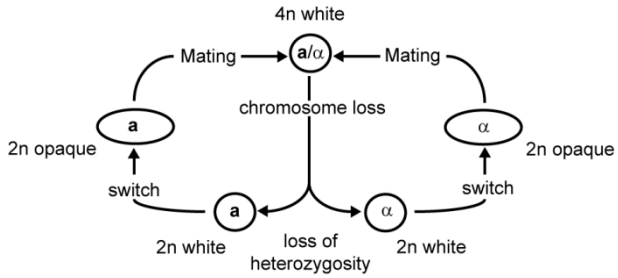


Figure 1: Outline of the parasexual cycle of *C. albicans*. **a** cells mate with α cells, but both types of cells must be in the opaque form for this to happen. *C. albicans* is diploid so mating produces a tetraploid **a**/α cell. This cell can lose chromosomes to return to the diploid state; a conventional meiosis has not been observed in *C. albicans*.

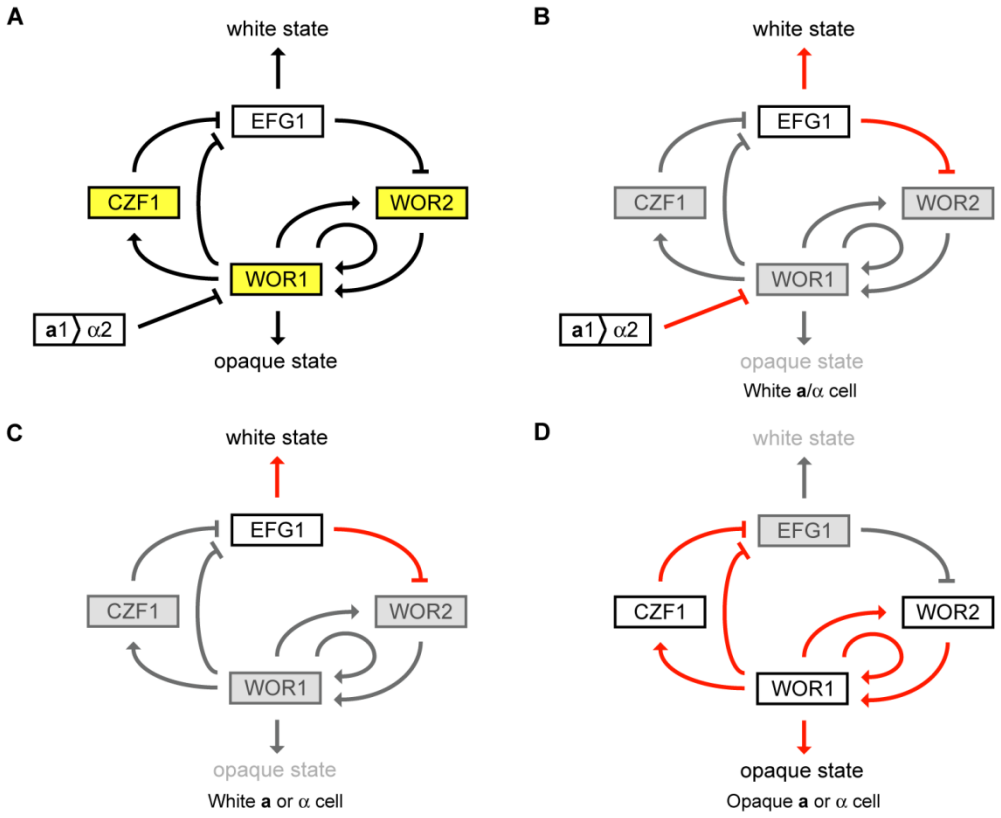


Figure 2: Model of the white-opaque regulatory circuit and its activity in different cell types. (a) White boxes represent factors that are enriched in white cells compared with opaque cells, and yellow boxes represent opaque-enriched factors. Lines with arrows represent positive control and lines with bars, negative control. (b) White **a**/α cell, with the **a**1/α2 heterodimer repressing WOR1. This cell is locked in the white form. (c) White **a** or α cell. (d) Opaque **a** or α cell. WOR1, WOR2, and CZF1 are up-regulated relative to white cells, and EFG1 is down regulated. In panels b-d, down-regulated genes and inactive interactions are shown in gray while up-regulated genes are shown in black and active interactions are shown in red. Adapted from Zordan et al 2007 (18).

Chapter 2:

Differential Phagocytosis of White versus Opaque *Candida albicans* by *Drosophila* and Mouse Phagocytes

Differential Phagocytosis of White versus Opaque *Candida albicans* by *Drosophila* and Mouse Phagocytes

Matthew B. Lohse[1] and Alexander D. Johnson[1,2*]

[1]Departments of Biochemistry and Biophysics and [2]Microbiology and Immunology, University of California San Francisco, San Francisco, California, United States of America

Abstract:

The human fungal pathogen *Candida albicans* resides asymptomatically in the gut of most healthy people but causes serious invasive diseases in immunocompromised patients. Many *C. albicans* strains have the ability to stochastically switch between distinct white and opaque cell types, but it is not known with certainty what role this switching plays in the physiology of the organism. Here, we report a previously undescribed difference between white and opaque cells, namely their interaction with host phagocytic cells. We show that both *Drosophila* hemocyte-derived S2 cells and mouse macrophage-derived RAW264.7 cells preferentially phagocytose white cells over opaque cells. This difference is seen both in the overall percentage of cultured cells that phagocytose white versus opaque *C. albicans* and in the average number of *C. albicans* taken up by each phagocytic cell. We conclude that susceptibility to phagocytosis by cells of the innate immune system is an important distinction between white and opaque *C. albicans*, and propose that one role of switching from the prevalent white form into the rarer opaque form may be to allow *C. albicans* to escape phagocytosis.

Introduction

Many strains of the commensal yeast *Candida albicans* undergo rare, stochastic switching between two distinct types of cells, white and opaque (1). The opaque phase was first noted due to its distinct colony morphology; at the single cell level, opaque cells are elongated, while white cells are nearly spherical (2-4). Approximately 400 genes are differentially expressed in white versus opaque cells, resulting in a wide variety of phenotypic differences (5, 6). For example, white and opaque cells differ markedly in their ability to mate: mating between opaque cells of the opposite mating type is at least 10^6 times more efficient than mating between white cells (7). White and opaque cells also differ in their suitability for different host environments; opaque cells appear better at colonizing the skin while white cells appear better suited to the bloodstream and internal environments (8-10).

In addition to the differences discussed above, white and opaque cells appear to interact differentially with the host immune system. Kolotila *et al*., 1990, reported that opaque cells were more susceptible than white cells to killing by oxidative stress and by neutrophils (11). These stresses were observed to increase the frequency of switching to the opaque phase, raising the possibility that Candida cells may respond to the innate immune system by modulating their rates of white-opaque switching (11). More recently, Geiger *et al*. 2004 showed that leukocytes respond differently to white and opaque cells. Both types of cells induce chemokinesis in leukocytes, but only white cells release a chemoattractant promoting chemotaxis towards a *C. albicans* source (12). These results suggest that opaque cells may be less "visible" to the innate immune

system, an idea which could have important implications for pathogenesis, commensalism, and mating.

In this paper, we examine whether phagocytic cells derived from the innate immune system differ in their ability to phagocytose white and opaque cells. We carried out this analysis first using *Drosophila melanogaster* S2 cells, a hemocyte-derived line. The use of *D. melanogaster* S2 cells for studying aspects of the innate immune response to *C. albicans* and other pathogens has been documented by numerous studies (13-15). We also examined differences in the response of *Mus musculus* RAW 264.7 cells, a macrophage-derived line, to *C. albicans* white and opaque cells. The use of this cell line to study the innate immune response to pathogens, including *C. albicans*, is also well-documented (16). In both the fly and mouse systems, we found that opaque cells were phagocytosed at significantly lower rates than were white cells. These results reveal an additional difference between white and opaque cells, one with implications for pathogenesis, and provide independent evidence that opaque cells are "less visible" than white cells to the innate immune system.

Results

Differential phagocytosis of white and opaque *C. albicans* by S2 cells

Recent studies have established *D. melanogaster* as a model system for studying many human pathogens, including *C. albicans*. Several RNAi screens have revealed a core set of genes required for phagocytosis of many microbes as well as additional genes needed specifically for individual pathogens (13, 15, 17-19). These studies have also revealed differences in the efficiency of phagocytosis among fungi. For example, Stroschein-Stevenson *et al*., 2006 showed that *C. albicans* is much more efficiently

phagocytosed than the non-pathogenic yeast *Saccharomyces cerevisiae* (15). To compare phagocytosis of *C. albicans* white and opaque cells, *C. albicans* cells were co-incubated with S2 cells and after variable time intervals the samples were lightly formaldehyde fixed, and phagocytosis was scored using immunofluoresence. For this analysis, the fixed cells were stained with rabbit anti-*Candida* and anti-rabbit Cy3-labeled antibodies followed by DAPI. Since membranes were not permeabilized until after antibody staining, internalized *C. albicans* lack Cy3 and are visually distinguishable from the cells that were not phagocytosed. This assay, optimized to study *C. albicans* white cells (15) (FIG. 1A), also efficiently monitors opaque cell phagocytosis (FIG. 1B).

Using this assay, we compared the ability of S2 cells to phagocytose white and opaque *C. albicans* cells of exactly the same genotype. S2 cells were co-incubated with either white or opaque *C. albicans* at a multiplicity of infection (MOI) of 5. After various time periods, the samples were fixed and stained as described in Materials and Methods. From each infection, four separate sets of 100 S2 cells were scored for phagocytosis of *C. albicans* cells and the values averaged. After a 2 hour co-incubation, approximately 5-fold more S2 cells had taken up at least one white cell than had taken up at least one opaque cell, even though the S2 cells were exposed to equal numbers of *C. albicans* cells (FIG. 2A). Similar differences were also seen after 3.5 hours of co-incubation. This difference was independent of mating type: **a**/**a** whites were phagocytosed similarly to **a**/α (FIG. 2B) and α/α whites, while **a**/**a** opaques were phagocytosed similarly to α/α opaques (data not shown). This difference between whites and opaques repeated on multiple days and in multiple *C. albicans* strain backgrounds despite fluctuations in overall levels of phagocyosis. Previous studies have shown that S2 cells preferentially

phagocytose *C. albicans* white cells over non-pathogenic *S. cerevisiae* (15); the level of opaque cell phagocytosis we observed in this study is approximately equal to that of *S. cerevisisae* (data not shown).

In addition to scoring differences in the percentages of S2 cells that phagocytosed white versus opaque cells, we examined the total number of *C. albicans* cells phagocytosed by each set of 100 S2 cells scored. Since this value incorporates both differences in the number of S2 cells that had taken up at least one *C. albicans* cell and the average number of *C. albicans* cells phagocytosed per S2 cell, it more accurately reflects differences in the efficiency of phagocytosis. The value for white *C. albicans* cells is 5 to 10-fold greater after 2 hours and approximately 10-fold greater after 3.5 hours than the equivalent value for opaque cells (FIG. 2B).

We independently confirmed the preference of S2 cells for white over opaque cells by presenting the S2 cells with mixed populations of white and opaque *C. albicans* cells. In the experiments shown in Figure 3, S2 cells were presented with a 50:50 mixture of white and opaque cells whose combined cell number was equal to the number of pure white or opaque cells used in the previous assays. White and opaque cells were easily distinguished microscopically and were counted separately using the described Cy3 antibody based strategy described in Materials and Methods. As before, sets of 100 S2 cells were counted at a time, but in this case six data sets were collected for each experiment and combined to calculate average values and standard deviations. When presented with a mixed *C. albicans* culture, S2 cells were 6-fold more likely to have phagocytosed at least one white cell than at least one opaque cell after 3.5 hours (FIG. 3A). Similarly, the number of *C. albicans* cells phagocytosed per S2 cell scored was

approximately 10-fold higher for white cells than opaque cells (FIG. 3B). Although most S2 cells that phagocytosed multiple *C. albicans* cells took up either exclusively white or opaque cells, an occasional S2 cell was observed to have phagocytosed a mixture of white and opaque cells. We believe that switching between the white and the opaque phases during the assay is minimal, based on the short duration of the assay as well as plating assays which indicate that the switching rate of *C. albicans* does not change significantly on exposure to S2 cells (data not shown).

RAW cells are also more efficient at phagocytosing white than opaque *C. albicans*

We next tested whether a mouse macrophage line exhibited the same preference for white cells over opaque cells. We modified the procedure described above to accommodate cultured mammalian cells (see Materials and Methods) and established that the procedure could detect uptake of both white (FIG. 1C) and opaque (FIG. 1D) cells of *C. albicans*. The combination of 10% serum and 37ºC temperature needed for growth of the macrophage cell line complicated the experiment at later time points, as *C. albicans* extensively filaments under these conditions. Moreover, opaque cells are not stable during prolonged exposure at 37 ºC and eventually revert to white cells (20). For these reasons, we concentrated on early (less than 2 hours) timepoints for this analysis.

After a 1 hour co-incubation, the percentage of RAW cells that had phagocytosed at least one white *C. albicans* was roughly 5 to 10-fold greater than for opaques (FIG. 4A). As with S2 cells, **a**/α and **a**/**a** whites were phagocytosed at similar rates (FIG. 4A) as were equivalent α/α strains (data not shown). We also monitored the total number of *C. albicans* phagocytosed and used this to calculate the number of *C. albicans* cells

phagocytosed per RAW cell scored. The value for white cells was approximately 10-fold greater than opaques after one hour (FIG. 4B).

When presented with a mixed population of *C. albicans* cells, the RAW cells showed a trend similar to the results using S2 cells: at 1 hour a given macrophage-derived cell was 4 to 6-fold more likely to have phagocytosed at least one white cell than at least one opaque cell (FIG. 4C). In these mixed experiments, we observed the number of *C. albicans* cells phagocytosed per RAW cell scored was 4 to 8-fold higher for white cells than for opaque cells (FIG. 4D). The majority of RAW cells phagocytosed exclusively white or opaque *C. albicans*, but as with S2 cells there were a small number of cases where a RAW cell was observed to have phagocytosed both cell types.

To determine whether the difference in phagocytosis rates of white and opaque cells reflects a difference in adherence, we modified the assay for use at 4ºC, a temperature that allows adherence but blocks phagocytosis (see Materials and Methods). We found that adherence levels were roughly 2 to 3-fold greater for white cells than for opaque cells (data not shown). This difference is significantly lower than the difference in phagocytosis and indicates that, although contributory, adherence is unlikely to be responsible for the entire effect.

Discussion

White-opaque switching is one of the most enigmatic features of *Candida albicans*. Switching occurs, on average, every 10^3-10^4 generations, and the white and opaque forms are each heritable for many generations- until a new switching event occurs. It has been proposed that the heritability of the two states is based on a self-perpetuating feedback loop which is excited in the opaque state but which remains

broken in the white state (21-23). Regardless of the exact mechanism, the phenomenon appears to be epigenetic, that is, switching from the white form to the opaque form and back occurs without any changes in the primary DNA sequence of the *Candida* genome.

Despite having identical genomes, white and opaque cells show enormous phenotypic differences. The two types of cells are easily distinguished under the microscope and each gives rise to a specific type of colony easily distinguished by the naked eye (2-4). The two types of cells appear to favor different niches in the host: opaque cells are more suited to skin infection, while white cells appear more stable in a systemic mode of infection (8). The two types of cells also differ in their mating behavior; while opaque cells of opposite mating types readily mate, white cells do not. Indeed, white-opaque switching is itself controlled by the mating type locus- **a** and α cells can undergo switching but **a**/α cells cannot. Finally, approximately 400 genes are differentially regulated between white and opaque cells; although a few of these genes make conceptual sense (for example some mating genes are upregulated in opaque cells), most do not and, instead, point to our incomplete understanding of the fundamental differences between these two types of cells (24, 25).

In the present study, we examined whether white and opaque cells differ in the extent to which they are phagocytosed by cells derived from the innate immune system. We chose this property because phagocytosis of microorganisms is an especially important component of the immune response (26). We show that two distinct cell lines derived from the innate immune system, *D. melanogaster* S2 cells and *M. musculus* RAW cells, phagocytose white *C. albicans* cells much more efficiently than they do opaque cells. This difference is seen in both the percentage of S2 or RAW cells that have

phagocytosed at least one *C. albicans* and by the average number of *C. albicans* phagocytosed per S2 or RAW cell. These results are statistically highly significant and repeated in multiple assays in multiple strain backgrounds carried out over a period of six months. Experiments carried out with mixed white and opaque *C. albicans* populations produced similar results: from this mixture, S2 and RAW cells both selectively phagocytosed white cells. Taken together, these results suggest that the difference in phagocytosis is due to an intrinsic difference between white and opaque cells. For example, the mixed population experiments show that white cells do not secrete a signal that stimulates phagocytosis of opaque cells. This result is consistent with the results seen by Behnsen *et al.* 2007 where the efficient phagocytosis of *Aspergillus fumigatus* did not result in an increase of *C. albicans* phagocytosis by polymorphonuclear neutrophils in conditions unfavorable to *C. albicans* phagocytosis (27). It is worth pointing out that while opsonization of *C. albicans* cells prior to exposure to macrophages would be expected to change the overall rate of phagocytosis, it is not clear whether the distinction between white and opaque cells would be increased or decreased.

Although this is the first study to examine phagocytosis of white and opaque cells, it is among several to indicate a differential response of the innate immune system to white and opaque cells. As described in the introduction, opaque cells are more sensitive than whites to oxidative stress and white cells, but not opaque cells, secrete a chemoattractant for leukocytes. Although it may be years before we understand the molecular basis of any of the differences in the way the host deals with white versus opaque cells, it is tempting to imagine that white-opaque switching in *C. albicans* plays a role conceptually similar to the many switching systems in bacteria, one of whose roles is

to present the immune system with multiple "identities." Although many bacterial pathogens typically use programmed DNA rearrangements to generate these differences, *C. albicans*- a eukaryote- has evolved an epigenetic mechanism that may serve the same basic function.

Materials and Methods

Strains

The *C. albicans* strains used for these experiments were all derivatives of CAF2-1 (URA3/ura3::λimm434). The white **a**/α strain was CAF2-1. Construction of the sorbose selected CAF2-1 **a**/**a** white **(**RZY12) and opaque (RZY491) strains is described in Zordan *et al*., 2006 (22). Equivalent α/α strains were sorbose selected and derived from the same parent strain. All strains were plated on synthetic complete medium complemented with 2% glucose and 100μg/mL Uridine (SD+aa+Uri) at 25ºC and grown in liquid SD+aa+Uri at room temperature prior to assays.

Cell Lines

D. melanogaster S2 cells were cultured in Schneider's medium (Invitrogen, Carlsbad, California, United States) supplemented with 10% Heat Inactivated fetal bovine serum (FBS), penicillin, and streptomycin (pen/strep)(.1mg/mL, 100U/mL). *M. musculus* RAW 264.7 cells were cultured in RPMI 1640 Media (UCSF Cell Culture Facility, San Francisco, California, United States) supplemented with 10% Heat Inactivated/Refiltered FBS, 10mM HEPES, .11mg/mL Sodium Pyruvate, .002M L-glutamine, and pen/strep (.1mg/mL, 100U/mL). Jeff Cox and Erik Lontok (UCSF, San Francisco, California, United States) kindly provided the RAW 264.7 cells.

S2 Phagocytosis Assays

The phagocytosis assay was based on the method described in Stroschein-Stevenson *et al*., 2006 (15). Briefly, *C. albicans* liquid cultures were started from single colonies and grown overnight at room temperature. Cultures were diluted back to OD 0.2-0.5 then allowed to grow back to OD 1.0 at room temperature. $8x10^4$ S2 cells per well were plated in 96 well plastic tissue culture plates. In each well, conditioned media was added to the cell containing solution to bring the volume to 75μL, to which 75 μL fresh media was added. *C. albicans* cultures were PBS washed and 50μL of $8x10^6$ cells/mL PBS stock was added to each S2 well for a MOI of 5. For mixed white-opaque assays, the stock had $4x10^6$ cells/mL of each type for MOIs of 2.5 each for a total MOI of 5. Incubations proceeded at 25ºC for 30min, 1 or 2.5 hours. Cells were then transferred to Concanavalin-A coated glass bottom 96 well microplates (Greiner Bio-One, Longwood, Florida, United States) and incubated for a further hour (or 30 minutes for 1 hour time points). Fixing, staining, and examination are described below. All cell concentrations were determined by hematocytometer.

RAW Phagocytosis Assays

RAW cell phagocytois assays were based on the S2 cell procedure with the following modifications. $4x10^4$ RAW cells were directly plated onto untreated 96 well glass bottom plates and allowed to settle for 18 hours at 37ºC and 5%CO_2. Prior to the experiment, old media was aspirated off and 150 μL fresh media was added. *C. albicans* stocks were made at $4x10^6$ cells/mL for a MOI of 5, or $2x10^6$ cells/mL of both whites and opaques for mixed culture assays. Incubations were 1 hour at 37ºC and were not transferred to new wells before fixing, staining, and examination as described below.

RAW Adherence Assays

RAW cell adherence assays were based on the RAW cell phagocytosis procedure with the following modifications. $1.2x10^5$ RAW cells were plated into the 96 well plates, and allowed to settle overnight. Following addition of fresh media, the tray was incubated for 30 minutes at 4ºC. *C. albicans* stocks were made at $1.2x10^7$/mL for a MOI of 5. Incubations were for 2 hours at 4ºC, after which wells were washed 4 times with PBS before formaldehyde fixing cells. Cells for scoring were not further stained. Additional control wells were stained using the normal procedure described below, phagocytosis levels in these wells decreased to almost zero for both whites and opaques. Scoring was conducted using a Zeiss Axiovert 200M microscope (Carl Zeiss, Oberkochen, Germany), the total number of *C. albicans* cells in a field of vision was manually determined for at least three distinct spots in at least three distinct wells for each condition. Assays were performed on two distinct days.

Immunofluorescence and Microscopy

After incubations, media was removed and wells allowed to air dry for 2 minutes. Wells were fixed for 5 minutes in a 1% Formaldehyde/PBS solution, washed once with PBS, and incubated for 1 hour in PBS plus 5% FBS. Unphagocytosed *C. albicans* was detected with a 4ºC overnight exposure to 1:2000 dilutions of primary rabbit anti-Candida (Biodesign, Saco, Maine, United States, Cat# B65881). Secondary incubation was with a 1:2000 dilution of Cy3-labeled goat anti-rabbit antibody (Jackson ImmunoResearch, West Grove Pennsylvania, United States) for two hours at room temperature. Two PBS washes were performed after each incubation. As phagocytic cell walls had not been permeabilized at this point, the antibodies could only bind unphagocytosed *C. albicans*. A 15 minute incubation with DAPI in PBS/0.1% Triton X-

100 solution allowed for DNA visualization to localize S2 or RAW cells. Wells were placed in Fluoromount-G (Southern Biotech, Birmingham, Alabama, United States) for storage pending examination. Wells were scored using a Zeiss Axiovert 200M microscope (Carl Zeiss, Oberkochen, Germany). *C. albicans* cells that did not light up with Cy3 were considered to have been phagocytosed. Filaments were not scored for either assay. Batches of 100 S2 or RAW cells were scored for phagocytosis of one or more *C. albicans* cells and for the total number of *C. albicans* cells phagocytosed. At least 2 sets of 100 cells were scored for each of at least two wells from each of at least two experiments conducted on different days for each condition. Each experiment was thereby repeated at least 4 times. The number of *C. albicans* cells phagocytosed per S2 or RAW cell scored, referred to as the phagocytic index in the figures, was calculated by dividing the total number of *C. albicans* cells phagocytosed by the number of S2 or RAW cells scored, 100. Charts in this paper represent the average of 4 or 6 data sets from the multiple wells for each condition on a single day; error bars represent the standard deviation. Statistical significance of the difference between white and opaque **a**/**a** strains was determined using the t-test assuming unequal variations data analysis function in Microsoft Excel (Microsoft Corporation, Redmond, Washington, United States), setting a p-value of less than .001 for the one-tailed distribution as the significance threshold. Charts were prepared in Microsoft Excel and edited in Adobe Illustrator. Images for this paper were processed automatically using AxioVision software (Carl Zeiss, Oberkochen, Germany). The black and white images for the different color channels were automatically converted to grey (DIC), Orange (Cy3) and Blue (DAPI). Figures were assembled in Adobe Photoshop (Adobe Systems, San Jose, California, United States).

Acknowledgements

This work was supported by Ellison Medical Foundation grant ID-SS-0628-04 and National Institutes of Health RO1 grant A149187. MBL was supported by a NSF graduate research fellowship.

We thank Rebecca Zordan and Shannon Stroschein-Stevenson for valuable advice, strains, and cell lines, as well as for comments on this manuscript. We thank Annie Tsong, Richard Bennett, Aaron Hernday, and other Johnson lab members for input, reagents, and discussions. We thank Erik Lontok, Michael Shiloh, and Jeff Cox for advice in handling the RAW264.7 cell line and Jeff Cox for valuable comments on the manuscript. We thank Sarah Foss for comments on the manuscript.

References

1. Slutsky B, et al. (1987) "White-opaque transition": a second high-frequency switching system in *Candida albicans*. *J Bacteriol* 169:189-197.
2. Soll DR, Morrow B, Srikantha T (1993) High-frequency phenotypic switching in *Candida albicans*. *Trends Genet* 9:61-65.
3. Lockhart SR, Daniels KJ, Zhao R, Wessels D, Soll DR (2003) Cell biology of mating in *Candida albicans*. *Eukaryot Cell* 2:49-61.
4. Johnson A (2003) The biology of mating in *Candida albicans*. *Nat Rev Microbiol* 1:106-116.
5. Lan CY, et al. (2002) Metabolic specialization associated with phenotypic switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 99:14907-14912.
6. Tsong AE, Miller MG, Raisner RM, Johnson AD (2003) Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115:389-399.
7. Miller MG, Johnson AD (2002) White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* 110:293-302.
8. Kvaal CA, Srikantha T, Soll DR (1997) Misexpression of the white-phase-specific gene WH11 in the opaque phase of *Candida albicans* affects switching and virulence. *Infect Immun* 65:4468-4475.
9. Kvaal C, et al. (1999) Misexpression of the opaque-phase-specific gene PEP1 (SAP1) in the white phase of *Candida albicans* confers increased virulence in a mouse model of cutaneous infection. *Infect Immun* 67:6652-6662.

10. Lachke SA, Lockhart SR, Daniels KJ, Soll DR (2003) Skin facilitates *Candida albicans* mating. *Infect Immun* 71:4970-4976.

11. Kolotila MP, Diamond RD (1990) Effects of neutrophils and in vitro oxidants on survival and phenotypic switching of *Candida albicans* WO-1. *Infect Immun* 58:1174-1179.

12. Geiger J, Wessels D, Lockhart SR, Soll DR (2004) Release of a potent polymorphonuclear leukocyte chemoattractant is regulated by white-opaque switching in *Candida albicans*. *Infect Immun* 72:667-677.

13. Pearson AM, et al. (2003) Identification of cytoskeletal regulatory proteins required for efficient phagocytosis in Drosophila. *Microbes Infect* 5:815-824.

14. Cheng LW, Portnoy DA (2003) Drosophila S2 cells: an alternative infection model for Listeria monocytogenes. *Cell Microbiol* 5:875-885.

15. Stroschein-Stevenson SL, Foley E, O'Farrell PH, Johnson AD (2006) Identification of Drosophila gene products required for phagocytosis of Candida albicans. *PLoS Biol* 4:e4.

16. Marcil A, Harcus D, Thomas DY, Whiteway M (2002) Candida albicans killing by RAW 264.7 mouse macrophage cells: effects of Candida genotype, infection ratios, and gamma interferon treatment. *Infect Immun* 70:6319-6329.

17. Rämet M, Manfruelli P, Pearson A, Mathey-Prevot B, Ezekowitz RA (2002) Functional genomic analysis of phagocytosis and identification of a Drosophila receptor for E. coli. *Nature* 416:644-648.

18. Agaisse H, et al. (2005) Genome-wide RNAi screen for host factors required for intracellular bacterial infection. *Science* 309:1248-1251.

19. Philips JA, Rubin EJ, Perrimon N (2005) Drosophila RNAi screen reveals CD36 family member required for mycobacterial infection. *Science* 309:1251-1253.

20. Rikkerink EH, Magee BB, Magee PT (1988) Opaque-white phenotype transition: a programmed morphological transition in *Candida albicans*. *J Bacteriol* 170:895-899.

21. Huang G, et al. (2006) Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 103:12813-12818.

22. Zordan RE, Galgoczy DJ, Johnson AD (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci U S A* 103:12807-12812.

23. Srikantha T, et al. (2006) TOS9 regulates white-opaque switching in *Candida albicans*. *Eukaryot Cell* 5:1674-1687.

24. Soll DR (2004) Mating-type locus homozygosis, phenotypic switching and mating: a unique sequence of dependencies in Candida albicans. *Bioessays* 26:10-20.

25. Bennett RJ, Johnson AD (2005) Mating in *Candida albicans* and the search for a sexual cycle. *Annu Rev Microbiol* 59:233-255.

26. Elrod-Erickson M, Mishra S, Schneider D (2000) Interactions between the cellular and humoral immune responses in Drosophila. *Curr Biol* 10:781-784.

27. Behnsen J, et al. (2007) Environmental dimensionality controls the interaction of phagocytes with the pathogenic fungi Aspergillus fumigatus and Candida albicans. *PLoS Pathog* 3:e13.

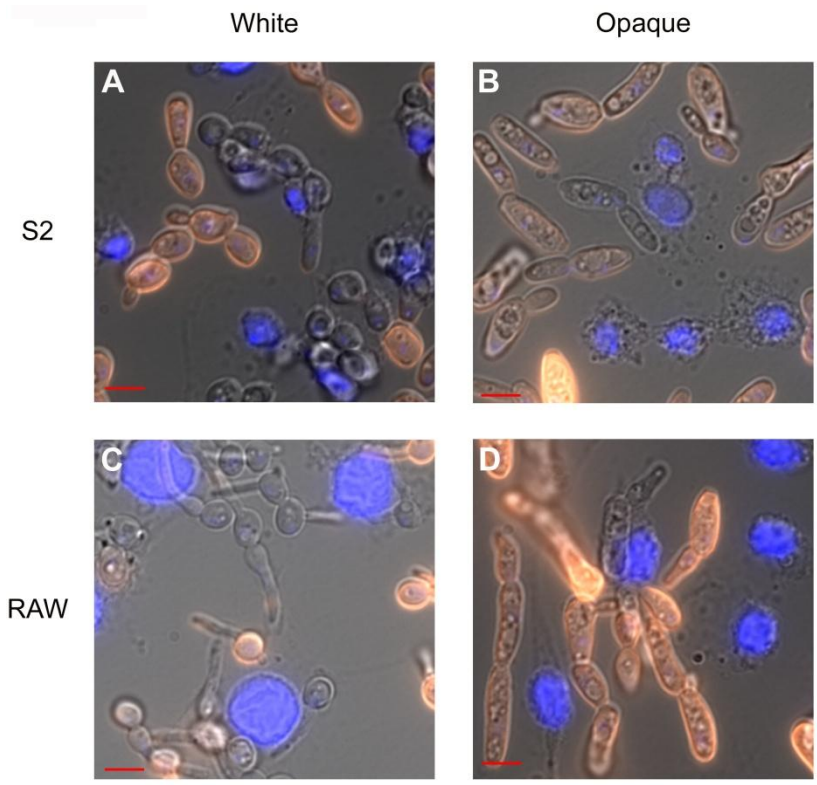


Figure 1: Phagocytosis of white and opaque *C. albicans* by *D. melanogaster* S2 and *M. musculus* RAW cells. White (A,C) or opaque (B,D) *C. albicans* cells were co-incubated with S2 cells for 3.5 hours (A,B) or RAW cells for 1 hour (C,D), lightly fixed with formaldehyde, stained with rabbit anti-Candida and Cy3-labeled anti-rabbit antibodies. Cells were then stained with a DAPI solution (blue) to localize S2 or RAW cells. *C. albicans* cells that were not phagocytosed appear orange in these figures.

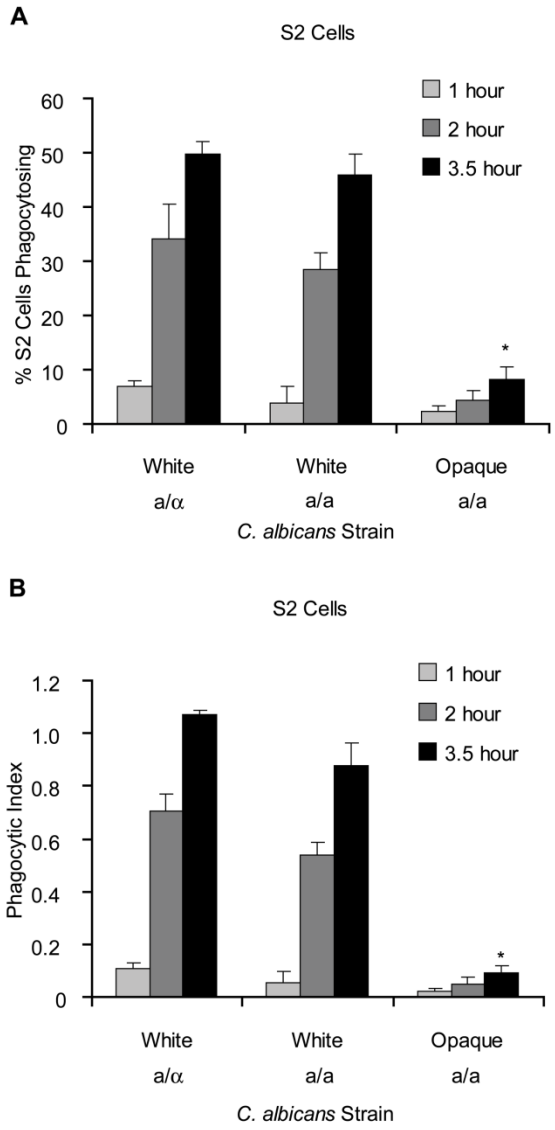


Figure 2: *D. melanogaster* S2 cells more efficiently phagocytose white than opaque *C. albicans*. (A) *D. melanogaster* S2 cells were co-incubated with white or opaque *C. albicans* for 1, 2, or 3.5 hours, fixed, and stained as described in the Materials and Methods. The number of S2 cells phagocytosing one or more *C. albicans* cells was determined. (B) For the same sets of S2 cells, the number of *C. albicans* cells phagocytosed by S2 cells was quantified and the total number of *C. albicans* cells phagocytosed divided by the number of S2 cells scored, referred to as the phagocytic index, was plotted. Values plotted are the averages from four data sets from a given day

and error bars represent the standard deviation. 100 S2 cells were counted for each data set. For the 3.5 hour time point, statistical significance of differences from the white **a**/**a** strain was evaluated using a t-test assuming unequal variance, sets with $p<.001$ are marked with an asterisk.

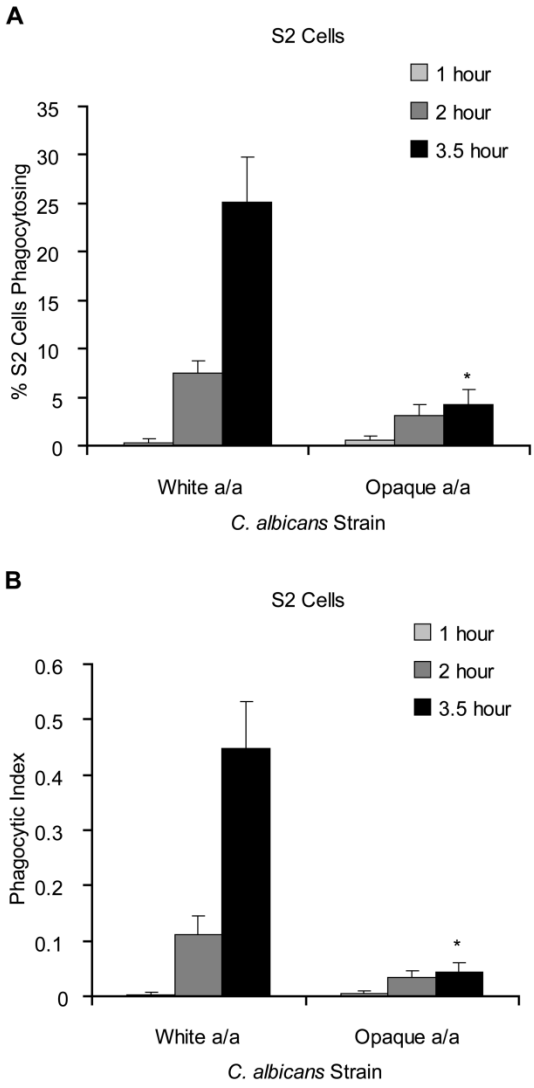


Figure 3: S2 cells preferentially phagocytose white *C. albicans* from a mixed white-opaque population. (A) *D. melanogaster* S2 cells were co-incubated with equal numbers of white and opaque *C. albicans* for 1, 2, and 3.5 hours. The number of S2 cells phagocytosing one or more *C. albicans* cells was determined. (B) For the same S2 cells, the number of *C. albicans* cells phagocytosed was quantified and the total number of *C. albicans* cells phagocytosed divided by the number of S2 cells scored, referred to as the phagocytic index, was plotted. 100 S2 cells were counted for each data set; values reflect the average of six data sets. For the 3.5 hour time point, statistical significance of

differences from the **a**/**a** whites was determined using a t-test and differences with $p<.001$ are marked with an asterisk.

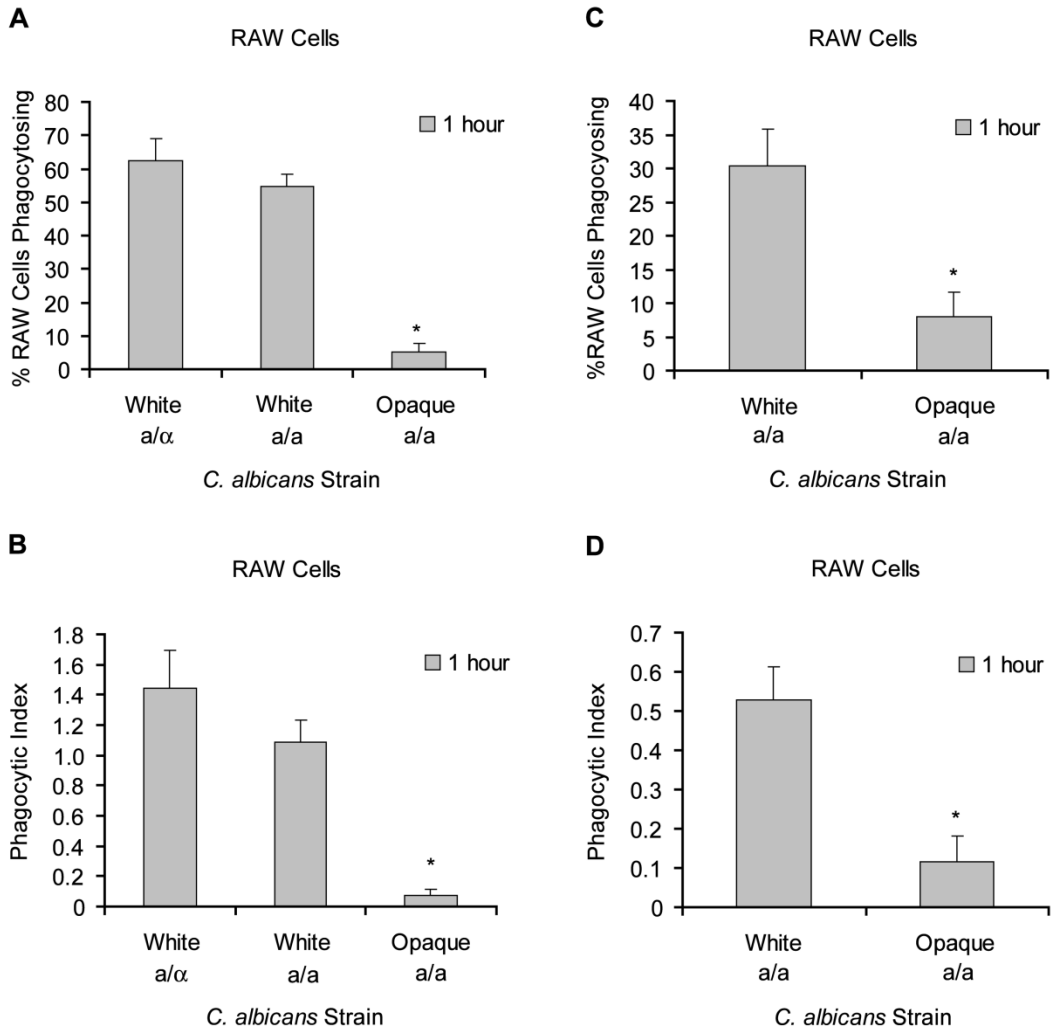


Figure 4: *M. musculus* RAW cells more efficiently phagocytose white than opaque *C. albicans*. (A) *M. musculus* RAW cells were co-incubated with white or opaque *C. albicans* for 1 hour then stained using the same methods as for S2 cells. The number of RAW cells phagocytosing one or more *C. albicans* cells was determined. (B) The number of *C. albicans* cells phagocytosed was quantified and used to calculate and plot the total number of *C. albicans* cells phagocytosed divided by the number of RAW cells scored, referred to as the phagocytic index. (C) *M. musculus* RAW cells were co-incubated with equal numbers of white and opaque *C. albicans* for 1 hour. The number of RAW cells phagocytosing one or more *C. albicans* cells was determined. (D) For the

same RAW cells, the number of *C. albicans* cells phagocytosed was quantified and the phagocytic index plotted. 100 RAW cells were counted for each data set and values plotted are the average of four (A,B) or six (C,D) data sets with error bars representing standard deviation. Statistical significance of differences from the **a**/**a** whites was determined using a t-test and differences with $p<.001$ are marked with an asterisk.

Chapter 3:

A distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in *Candida albicans*

A distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in *Candida albicans*

Matthew B. Lohse[a,b], Rebecca E. Zordan[a,b,c], Christopher W. Cain[a,b], and Alexander D. Johnson[a,b,*]

[a]Departments of Biochemistry and Biophysics and [b]Microbiology and Immunology, 600 16th Street, University of California San Francisco, San Francisco, CA, 94158

[c]Department of Molecular Biology and Genetics, 725 North Wolfe Street, Johns Hopkins School of Medicine, Baltimore, MD, 21205

Abstract:

Among the most important classes of regulatory proteins are the sequence-specific DNA-binding proteins that control transcription through the occupancy of discrete DNA sequences within genomes. Currently, this class of proteins encompasses at least 37 distinct structural superfamilies and more than 100 distinct structural motifs. In this paper, we examine the transcriptional regulator Wor1, a master regulator of white-opaque switching in the human fungal pathogen *Candida albicans*. As assessed by a variety of algorithms, this protein has no sequence or structural similarity to any known DNA-binding protein. It is, however, conserved across the vast fungal lineage, with a 300aa region of sequence conservation. Here, we show that this 300aa region of Wor1 exhibits sequence-specific DNA binding and therefore represents a new superfamily of DNA-binding proteins. We identify the 14-nucleotide-pair DNA sequence recognized by Wor1, characterize the site through mutational analysis, and demonstrate that this sequence is sufficient for the Wor1-dependent activation of transcription *in vivo*. Within the 300aa DNA-binding conserved region, which we have termed the WOPR box, are

two domains (WOPRa and WOPRb), dissimilar to each other but especially well-conserved across the fungal lineage. We show that the WOPR box binds DNA as a monomer and that neither domain, when expressed and purified separately, exhibits sequence-specific binding. DNA binding is restored, however, when the two isolated domains are added together. These results indicate that the WOPR family of DNA-binding proteins involves an unusual coupling between two dissimilar, covalently-linked domains.

Introduction:

Changes in gene expression are responsible for many aspects of cell and molecular biology, ranging from the adaptation of a bacterium to a new food source to the orchestration of the development of a complex organism from a single cell. One of the most important classes of proteins that regulate gene expression are the sequence-specific DNA-binding proteins that control transcription by occupying discrete DNA sequences within genomes. This large group of proteins (5-10% of the coding capacity of most genomes) can be divided into different structural classes, based on the way they recognize double-stranded DNA. For instance, terms like homeodomain, leucine zipper, and Cys2-His2 Zinc finger each describe a particular structural domain found in many transcriptional regulators in many different organisms.

It is difficult to rigorously pinpoint the number of distinct structural classes of DNA-binding proteins, because many of the known structures are related to one another (for example, see (1)). Nonetheless, several taxonomies have been developed, and the DBD database currently lists 37 distinct superfamilies and around 150 PFAM domains, with most domains encompassing several variants (2, 3). Although the majority of these domains are now represented by multiple three-dimensional structures, nearly all were identified as sequence-specific DNA-binding motifs prior to their detailed structural characterization. With the increased use of genome sequencing, the number of individual members assigned to each domain family is greatly increasing.

In this paper, we describe the WOPR box, a previously uncharacterized class of motif (likely representing a distinct superfamily) responsible for the sequence-specific recognition of duplex DNA. We studied the WOPR box in the transcriptional regulator

Wor1 from *Candida albicans*, the most prevalent fungal pathogen of humans. Wor1 is a master regulator of a phenomenon called white-opaque switching whereby two distinct, heritable, cell types (white and opaque) arise from the same genome (reviewed in (4)). In white cells, Wor1 is expressed at low levels; in opaque cells, it is upregulated approximately 40-fold (5-7). Wor1, along with several other regulatory proteins, activates its own synthesis, and this positive feedback loop is believed to maintain the opaque state through many cell generations (6). In addition, Wor1 controls the transcription of many target genes which give opaque cells their specialized characteristics, including mating ability (8), host-tissue preferences (9-12), and metabolic preferences (13).

Although white-opaque switching appears to be confined to *C. albicans* and its very close pathogenic relatives, Wor1 is broadly conserved across the entire fungal domain of life, which encompasses over a billion years of evolutionary history. The portion of Wor1 that is conserved across fungi consists of about 300 amino acids (out of 785 for Wor1) at the N-terminus of the protein (Figure 1a). This conserved region (the WOPR box) consists of two blocks of sequence (in Wor1 roughly positions 1 to 100 (Supplemental Figure S1a) and 185 to 265 (Supplemental Figure S1b)) separated by a less conserved region of variable length. Homology detection programs such as PSI-BLAST (14, 15) fail to detect family members outside the fungal kingdom, although this approach cannot rigorously rule out the existence of this domain in other kingdoms. Structural prediction programs suggest that the two conserved blocks each code for globular domains (Figure 1b), but even advanced structure homology searches like Phyre (16) fail to find a match between either region and any published structures.

Although found in every sequenced fungal genome, homologs of Wor1 have been studied in only a few species besides *C. albicans*. The Ryp1 protein from *Histoplasma capsulatum* is a key regulator of the yeast-to-mycelial transition, and like Wor1 it affects transcription of hundreds of genes (17). Gti1 from *Schizosaccharomyces pombe* plays a role in the regulation of gluconate uptake in response to starvation conditions, although its biochemical mechanism has not been investigated (18). Most recently, Sge1 from the plant pathogen *Fusarium oxysporum* has been shown to be required for parasitic growth, possibly through the regulation of several effector genes (19). Many fungal species include a second, distinct member of this family, that has been separately maintained over long periods of evolution. This related family is even less characterized than the Wor1 family and includes Pac2 which regulates mating in *S. pombe* (20) and Pth2 in *C. albicans* which has no known function. The term "WOPR box" is based on family members Wor1, Pac2, and Ryp1.

Full genome chromatin immunoprecipitation experiments (ChIP-Chip) have shown that Wor1 is bound to nearly 200 regions in the *C. albicans* genome (6). In some cases, (e.g. the *WOR1* gene itself) association of DNA by Wor1 appears to activate transcription; in other cases, the effect appears to be negative. Chromatin immunoprecipitation experiments have also revealed that Ryp1 from *H. capsulatum* is bound to specific regions of the genome (17, 21). In principle, Wor1 could either recognize DNA sequences directly or it could do so indirectly through associations with other DNA-bound proteins. To distinguish between the possibilities, and to determine the function of the region of Wor1 conserved across the fungal lineage, we expressed, purified, and studied a series of Wor1-derived proteins.

Results:

Expression and Purification of the Conserved Domains of Wor1

Protein structure prediction programs suggest that the WOPR box contains two globular domains roughly encompassing the 5-101aa and 196-321aa regions of Wor1. These predicted globular domains superimpose on the two regions of high conservation among species (Figure 1b). The amino acid stretches that connect these domains are poorly conserved and of variable length among species.

We expressed the entire 300aa WOPR box (1-321aa) as well as the individual domains (WOPRa, 1-101aa and WOPRb, 196-321aa). In order to express these proteins in *E.coli*, the six CUG codons in Wor1 were converted to other serine encoding codons to compensate for the alternative genetic code in *C. albicans* (22). Codon-changed *WOR1* was then cloned into a modified pET28b expression plasmid containing N-terminal 6xHis and Maltose Binding Protein (MBP) tags separated from the Wor1 constructs by a PreScission protease site. We purified the 6xHis-MBP-Wor1 1-321aa protein (henceforth referred to as MBP-Wor1 1-321) on a Ni-NTA column. An aliquot of this purified protein was cleaved overnight with PreScission protease and further purified over Amylose resin and Glutathione Sepharose to remove the 6xHis and MBP tags. As judged by coomassie blue staining, the Wor1 fragment was by far the predominant protein component of the purified fraction. MBP-tagged WOPRa and WOPRb constructs were purified using a similar approach. We also expressed and purified His tagged (without MBP) 1-321aa, WOPRa, and WOPRb constructs using the PET28b vector (see Supplemental Table S1 for a list of constructs expressed).

Wor1 Binds DNA Directly

To determine whether Wor1 binds DNA directly, we performed electrophoretic mobility shift assays (EMSAs) using the purified MBP-Wor1 1-321. As target DNAs, we used three 200bp DNA fragments selected from the *C. albicans* genome as regions bound by Wor1 as determined by ChIP (6). As shown in Figure 2, MBP-Wor1 1-321 binds 200bp DNA fragments from the promoters of *MDR1* (orf19.5604) (Figure 2a), *WOR1* (orf19.4884) (Figure 2b), and orf19.4394 (Figure 2b). Control experiments using DNA from promoter fragments that were not enriched in the Wor1 ChIP-Chip experiments showed that the binding to *MDR1*, *WOR1* itself, and orf19.4394 promoter sequences was sequence-specific (Supplemental Figure S2a). We also performed DNA-binding experiments using a MBP-Wor1 1-321 derivative from which the MBP had been cleaved (Wor1 1-321aa) as well as the 6xHis tagged Wor1 1-321aa version (referred to as 6xHis-Wor1 1-321). Both of these proteins also exhibited sequence-specific binding but produced smaller sized mobility shifts than MBP-Wor1 1-321 (Figure 2a). These experiments show that the activity responsible for sequence-specific DNA binding is due to Wor1 rather than to MBP or a bacterial contaminant in the preparation. Based on additional DNA-binding experiments using a 20bp DNA fragment and omitting the competitor Poly(dI-dC), we estimate the nominal affinity of MBP-Wor1 1-321 for DNA (expressed as a K_D) to be 4-8nM (Figure 2c). Since we do not know whether the Wor1 preparation from bacteria is fully active, this value should be regarded as an upper limit; the binding could, in principle, be stronger but not weaker. This affinity is in the range exhibited by many other sequence-specific DNA-binding proteins (for example, see (23)).

WOR1 binds a specific DNA sequence

We next examined the sequences of DNA specifically recognized by Wor1. We modified the EMSA experiments to include unlabeled competitor DNA corresponding to 20bp sub-regions of the three previously shifted 200bp fragments. We identified six 20bp regions (Supplemental Table S2) that efficiently competed for Wor1 binding, one from the *MDR1* promoter (Figure 2d), two from the orf19.4394 promoter, and three from the *WOR1* promoter.

We submitted these six 20bp fragments to MEME (24) to develop a preliminary motif recognized by Wor1. This resulted in a 14 bp motif, with most of the information content located in positions 6 through 14 (Figure 3a). For the rest of this study, we will focus on positions 6 through 14; we will refer to this as the core motif. To test the relevance of this core motif to Wor1 binding, we made all possible single base pair substitutions and examined the ability of a 20bp fragment containing each substitution to compete for binding of MBP-Wor1 1-321 to the 200bp *MDR1* promoter fragment. These results, summarized in Figure 3b, verify the basic motif derived from MEME and demonstrate the influence of each position on Wor1's binding affinity. The second member of this protein family in *C. albicans*, Pth2, also recognizes this motif (Supplemental Figure S2b) although we have not determined if the Wor1 motif represents the optimal sequence for Pth2 binding to DNA.

Compared to the motifs recognized by some sequence-specific DNA-binding proteins, the Wor1 position-specific weight matrix is relatively information poor. However, the motif can account for approximately 54% of the Wor1-bound sites determined by full-genome chromatin immunoprecipitation (6) using a stringency level

that produces false positive hits in 14% of unbound promoters (see Supplementary Materials).

The WOR1 core motif is sufficient for transcriptional activation by Wor1

We next determined whether the 9-nucleotide core motif was sufficient for Wor1 to function *in vivo*. We introduced a 20bp fragment containing the consensus Wor1 core motif (TTAAAGTTT) flanked by linker DNA into a plasmid containing a version of the *CYC1* promoter driving expression of LacZ but lacking an upstream activation sequence (enhancer). We ectopically expressed *C. albicans* Wor1 in strains of *S. cerevisiae* (instead of *C. albicans*) where the two Wor1 homologs (*YEL007* and *YHR177*) had been deleted, and then performed β-galactosidase assays to determine whether this motif was sufficient to drive transcription in a Wor1-dependent manner. The experiment was performed in *S. cerevisiae* so we could monitor to the effect of ectopic Wor1 expression without triggering the entire white-to-opaque switch and its associated large-scale changes in the gene expression profile. Thus, the experimental strategy allowed us to monitor the effect of Wor1 on the promoter without the concern of indirect effects.

Using this strategy, we determined that the presence of the core motif from Figure 3a was sufficient to activate transcription more than 10-fold in a Wor1 dependent manner (Figure 4a). Mutations in the motif shown to affect binding (Figure 3b) strongly reduced this Wor1 dependent activation (Figure 4b). This analysis shows that the *in vitro* biochemical activity of Wor1 is mirrored by its *in vivo* activity.

We repeated the activation assays using the first 321aa of Wor1 (the WOPR box) instead of the entire protein. Wor1 1-321 also activated transcription from the core motif, although the level of activation is less than that observed for full length Wor1 (Figure 4c).

This experiment shows that that the WOPR box is sufficient for DNA binding (Figure 2) and at least some degree of transcriptional activation.

Wor1 binds DNA as a monomer

To determine whether Wor1 binds DNA as a monomer or a dimer, we performed additional DNA-binding experiments and gel filtration chromatography. Taking advantage of the different sized Wor1 1-321aa constructs available, we co-incubated DNA with a mixture of the ~80kDa MBP-Wor1 1-321 and the ~40kDa 6xHis-Wor1 1-321 constructs. We did not observe any hybrid shifts (25), instead obtaining only the two distinct shifts we observed when the DNA was incubated separately with each Wor1 derivative (Supplemental Figure S2c). This result shows that Wor1 binds either as a monomer or as a very stable oligomer.

To distinguish between these possibilities, we further purified 6xHis-Wor1 1-321 and passed it over a Superdex 200 gel filtration column. We observed that the ~40kDa 6xHis-Wor1 1-321 eluted from the column at approximately the same position as a 44kDa standard protein (Chicken Ovalbumin, Supplemental Figures S3a, S3b), indicating that it exists as a monomer in solution. Taken together, the mixing and chromatography experiments show that Wor1 1-321 binds DNA as a monomer.

Wor1 Binding Requires the Presence of Both Globular Domains

Thus far, we have shown that the WOPR box of Wor1 is sufficient to bind DNA. To further narrow down the portion of Wor1 required for sequence-specific DNA binding, we separately purified the two domains that make up the WOPR box, as both N-terminal MBP- and N-terminal 6xHis-tagged fusions. Neither the 6xHis- or MBP-tagged versions of either domain was observed to shift DNA in a sequence specific

manner, even when added at concentrations several hundred times the K_D of the 300aa construct. However, when the DNA was co-incubated with both the 6xHis WOPRa domain (1-101) and the 6xHis WOPRb domain (196-321), restoration of sequence-specific DNA binding was observed (Figure 5, Supplemental Figures S2d, S2e). This result demonstrates that both globular domains of the WOPR box are needed for efficient sequence-specific DNA binding, although the two domains do not need to be present within the same protein molecule.

Discussion:

In this paper, we identify and study a previously uncharacterized DNA-binding domain found in a widely conserved group of fungal proteins, exemplified by the *C. albicans* Wor1 protein. We experimentally determined that Wor1 is a sequence-specific DNA-binding protein and that the portion of Wor1 which is conserved is sufficient for DNA binding *in vitro* and *in vivo*. The sequence of DNA recognized by this conserved region is given as a position specific weight matrix in Figure 3. We have called the conserved region the WOPR box.

Unanticipated Features of Wor1s ability to bind DNA

The WOPR box is predicted to include two globular domains (each of which is very highly conserved among fungi) separated by a variably-spaced and poorly-conserved linker. We showed that neither individual domain showed sequence-specific DNA-binding activity on its own, but when the individually-expressed and purified domains were mixed, DNA-binding specificity was restored. Two different models of Wor1 binding to DNA can explain these results. Both models rely on the nonconserved region between the two conserved globular domains functioning as a tether. Consistent

with this idea, the stretch of amino acids between the two domains varies from less than 50 to nearly 200 amino acids when observed across fungal species. Structural prediction programs suggest that this region is largely unstructured. In the first model, each of the domains forms a part of the interface with DNA but neither domain alone produces sufficient affinity or specificity to be detected in our DNA-binding experiments (Figure 6a). The second model holds that only one of the two domains contacts DNA, but, in order to do so, contact with the other domain is needed to induce a conformational change (Figure 6b). Although precedents exist for different aspects of each model, we are unaware of any examples which exactly match this property of the WOPR domain.

We believe that the WOPR box represents a distinct superfamily of DNA-binding proteins. Although conserved across all fungi it has thus far been studied in detail in only a few species. In the two species where WOPR box proteins have been studied most extensively, they function as master regulators of distinctive cell morphologies. In *C. albicans*, Wor1 regulates the white-opaque transition, and in *H. capsulatum* Ryp1 regulates the yeast-mycelial transition. Both proteins regulate hundreds of genes and the transitions involve changes in cell shape, preferred environmental niches, and interactions with the host immune system. Although previously unrecognized as sequence-specific DNA-binding proteins, WOPR box proteins are deeply conserved across the fungal lineage and, in the cases studied in most detail, they are key for regulating host-pathogen interactions.

Materials and Methods:

Bioinformatic analysis

The Wor1 protein sequence was submitted to online structure prediction and homology comparison programs, including Phyre(16), ELM(26), and GlobPlot(27). Estimated domain regions are based on the predictions from ELM. The full Wor1 sequence and predicted globular domains were compared versus other sequences using the PSI-BLAST tool (14, 15). The Wor1 motif was developed by submitting 6 20bp binding sites we identified to MEME (24), using the one site per sequence option with minimum and maximum motif size constraints of 10 and 20bp. Comparisons between the Wor1 motif and previously published Wor1 ChIP-Chip data were performed using MochiView v1.43 and are described in detail in the Supplementary Materials (28). A list of Wor1 peaks used for the motif-binding site comparison has been provided in Table S3.

Cloning

Primers used in this study are included in Supplemental Table S4. Plasmids used in this study are included in Supplemental Table S5. Strains used in this study are included in Supplemental Table S6.

The six CUG codons in *WOR1* were changed using several rounds of PCR with primers corresponding to the CUG codon to be mutated. The Wor1 sequence used contains a silent mutation of one other codon, however the mutation is to a more common variant. The seven fragments containing the altered *WOR1* sequence were combined using fusion PCR. Codon-changed *WOR1* was amplified to add NheI and XhoI sites at the 5' and 3' ends, as well as a stop codon at the end of the ORF and introduced into the pCR-BluntII TOPO (Invitrogen, Carlsbad, California) backbone and sequenced. From this full-length Wor1 construct, truncated 1-101, 196-321, and 1-321 amino acid versions

were cloned with the same placement of the NheI and XhoI sites and C-terminal stop codon.

Constructs were ligated into a modified version of BHM1092, itself a derivative of pET28b (Novagen, Gibbstown, NJ), with a N-terminal copy of both a 6xHis tag and MBP separated from Wor1 by a PreScission protease site. After verifying ligation, plasmids were transformed into BL21 cells for expression. Similar 6xHis tagged contrasts were made by ligating NheI/XhoI digested Wor1 constructs into pET28b.

Activation assays used derivatives of the BES146 plasmid (29) which uses the lacZ system first described by Guarente *et al*. (30) . Oligo pairs were ordered (Integrated DNA Technologies, Coralville, IA) containing the putative motif, or varients thereof, flanked by 5 or 6 bp on each end and sticky ends corresponding to a XhoI digest site. Oligos were phosphorylated using ATP and T4 PNK (New England Biolabs, Ipswich, MA), ligated (Epicentre Biotechnologies, Madison, WI) into XhoI digested, phosphatase treated (Epicenter Biotechnologies), and gel extracted (Zymo Research Corporation, Orange, CA) BES146 plasmid. Plasmids were sequenced to verify the insert sequence and orientation. Unmodified BES146 was used as a negative control.

Constitutively active Wor1 plasmids were constructed in the p413TEF plasmid (31). Codon changed full length Wor1 or Wor1 1-321 was cut out of the plasmid described above using the NheI and XhoI sites and ligated into pTEF413 between the SpeI and XhoI sites. p413TEF with no insert was used as a negative control.

Strain construction

Assays were performed in the Sigma 2000 *S. cerevisiae* background. *WOR1* homologs *YEL007w* and *YHR177w* were deleted using Leu and KanMx markers PCR-

amplified with 50bp flanks matching the gene ORF flanks using a standard homologous recombination protocol (32). Strains were transformed with either full length Wor1 or 1-321aa Wor1 ectopic expression plasmids or the Wor1-less control. Activation plasmid constructs were then transformed into each background. Strains were grown on -Ura/-His media to maintain selection for both plasmids.

Protein Expression and Purification

Expression was performed in 2xYT media supplemented with 1mM $MgSO_4$ and 0.15% glucose. 5mL cultures were inoculated from frozen stocks, grown at 37°C overnight, diluted back 200-fold in fresh media, and grown to an approximate OD of 0.6 where they were moved to 25°C and induced with 0.1mM IPTG. Induced cultures were grown for 4 hours, pelleted, and frozen in liquid nitrogen before storing at -80°C.

For purification, cells were resuspended in lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM imidazole, pH 8.0) supplemented with 10mM β-mercaptoethanol (βME), 1mg/mL lysozyme, and 1 Complete Mini Protease inhibitor cocktail tablet, EDTA-free (Roche, Basel, Switzerland) per 10mL volume. Resuspended cells were incubated at 4°C for 20 minutes, sonicated, and incubated with 50U/mL DNaseI for 15-30 minutes at 4°C. The lysate was centrifuged and the soluble fraction passed over a .45µm SFCA filter (Nalge Company, Rochester, NY). The filtered fraction was co-incubated with at least 1mL of Ni-NTA agarose beads (QIagen, Valencia, CA) for 1 hour at 4°C. Beads were then batch washed five times with 10 column volumes of wash media (50mM NaH_2PO_4, 300mM NaCl, 20mM imidazole, pH 8.0) followed by elution with 3x5 column volumes of elution media (50mM NaH_2PO_4, 300mM NaCl, 250mM imidazole, pH 8.0). Elution fractions were pooled and concentrated on an Amicon-Ultra ultracel 10k centrifugal filter

(Millipore, Billerica, MA) and passed over a Illustra NAP-5 or a NAP-25 column (GE Healthcare) into storage (10mM Tris pH 7.4, 100mM NaCl, 5mM DTT, 50% glycerol) or Amylose resin binding buffer (20mM Tris pH 7.4, 200mM NaCl, 10mM βME). For cleavage, protein was bound to Amylose resin (NEB) for 90 minutes at 4°C, washed, equilibrated with cleavage buffer (50mM Tris pH7.5, 150mM NaCl, 1mM EDTA, 1mM DTT), then incubated with PreScission protease (GE Healthcare) for 4hr at 4°C. Cleaved WOR1 was then eluted, passed over a Glutathione Separose 4 Fast Flow (GE Healthcare) slurry to remove PreScission protease, concentrated, and stored.

Purity of MBP-Wor1 or cleaved Wor1 was verified on SDS-PAGE gels. Protein concentrations were determined based on comparison to a BSA dilution series on SDS-PAGE gels as well as a BCA Assay versus BSA standards. Similar methods were used for the expression and purification of the 6xHis tagged constructs.

Gel Shifts

Gel shifts were performed according to a standard protocol (33). Probes were labeled with P32 gamma-ATP (MP Biomedicals, Solon, OH) using T4 PNK (New England Biolabs). Binding conditions were 20mM Tris pH8, 100mM NaCl, 5% Glycerol, 5mM MgCl2, 1mM DTT, 0.1% NP40, 1µg/µL BSA, 25µg/mL Poly(dI-dC) (Sigma-Aldrich, St. Louis, MO). For K_D determinations and shifts with 20bp probes, binding reaction conditions were modified to 50mM NaCl and no Poly(dI-dC). When used, unlabeled competitor DNA was added prior to addition of protein. Binding reactions were incubated for 30 minutes at room temperature (23-25°C). Samples were run on 6% arcylamide, 0.5X TGE, 2.5% glycerol gels at ~130mA for 90 minutes. Gels were dried and exposed to Phosphor imaging screens (GE Healthcare). Imaging was

conducted on a Storm or Typhoon imagers (GE Healthcare). Image analysis for quantitation was performed using ImageQuant 5.1 (GE Healthcare) followed by further analysis in Microsoft Excel.

Competition assays to identify specific Wor1 binding sites were performed at Wor1 concentrations that resulted in shifting of roughly half of labeled probe. Competitor concentrations were 5µM/500nM/50nM or 6µM/600nM/60nM. For motif mutation competitor analysis, all assays used labeled 200bp *MDR1* promoter probe and a fixed concentration of MBP-WOR1. Unlabeled 20bp competitors were identical except for the single base pair substitutions. Unlabeled competitors concentrations were 12µM, 6µM, 3µM, 1.2µM, 600nM, 300nM; no non-specific competitor was used. The fraction of labeled probe shifted was calculated for each concentration. Data were rounded to the nearest tenth, and minimum competitor concentration need to reduce the amount of shifting by half (0.5) was determined for each competitor. Concentrations were normalized versus the concentration of competitor containing the unmutated motif required for a similar effect, with the wild type concentration equaling 1. Data shown in Figure 3b reflect the inverse of this normalized concentration.

β-galactosidase Activation Assays

β-galactosidase assays were performed using a standard protocol (34). Strains were grown in SD-Ura-His media to maintain selection for both plasmids. For each strain three colonies were grown overnight, diluted back, and allowed to reach log phase. Cells were harvested, permeablized, and activation assays performed. Data in any figure panel are from the same day.

Acknowledgements:

Work of the authors was supported by grants from the National Institutes of Health (AI49187) and the Ellison Foundation (ID-SS-0628-04). MBL was also supported by a NSF graduate research fellowship. The authors thank the Hiten Madhani lab for plasmids and strains and the Robert Stroud lab, particularly Akram Alian and Sarah Griner, for help with chromatography. The authors thank Oliver Homann and Aaron Hernday for help with data analysis and valuable comments on the manuscript.

Supplemental Materials

Wor1 Motif-Binding Site Comparison

The existing Wor1 ChIP-Chip experiments (6) were reexamined using MochiView's (28) "Peak Extraction" function. All datasets were based on the previously published C-terminal Wor1 peptide antibody ChIP-Chip data (6), using two replicates each of wild type opaque and *wor1/wor1* cells. Default settings were used, except as follows. Under "smoothing/peak settings" the "significance sampling settings" was placed at 100,000 for "number of random samples" and 100 for "maximum number of random samples". Under "Control Settings", "minimum value for control peak consideration was set to 0.27, "number of random samples" was set to 100,000, and "minimum number of control samples" was set to 501. This reanalysis resulted in the identification of 337 distinct 500bp Wor1 bound peaks of Wor1 binding. A list of these peaks has been provided in Supplemental Table S6.

This Wor1 peak set was then analyzed as follows using the MochiView program (v1.43) and a hand curated version of the Assembly 21 gene annotation list. A list of all promoters was created using the "Build promoter set function" (size limits were set to 50 and 20,000bp) with the aforementioned gene list, which excludes intergenic regions

between convergent ORFs. The "Merge Location Set (Union)" function (with "Only keep locations intersected by all contributing location sets" selected) was then used to create a location set of all promoters that contained one or more Wor1 peaks. A location set of all promoters without a Wor1 peak was then created by using the "Merge Location Set (Subtraction)" function to remove Wor1 bound promoters from the starting Promoter list. The "Sample Fixed Length Locations from Location Set" function was then used to create a set of 2309 500bp regions from within the set of all promoters not bound by Wor1 to use as a control. Motif comparisons are described below. To control against the effects of eliminating all promoters bound by Wor1 from the development of the control set, a second control set was developed as follows. Starting with the same initial promoter set, the "Merge Location Set- Subtraction" function was used to directly remove the 337 Wor1 peaks from the promoter location set, but otherwise leaving those promoters intact. The "Sample Fixed Length Locations from Location Set" function was then used as before to create a random set of 2704 500bp regions from within the set of all promoters (excluding only the 500bp Wor1 binding sites).

The Wor1 position-specific weight matrix (PSWM) used to create Figure 3a was modified as follows. Positions 1-5 were discarded to focus on the core of the motif. Positions 7, 12, and 13 were not changed due to the strong preference for T at these positions. For positions 6, 8-11, and 14, a 1% pseduocount was added to bases with a starting value of zero and bases with non-zero values were adjusted accordingly. This revised motif had a maximum Logarithm of the Odds (LOD) score of 4.048 (against background frequencies of 16.7% C/G and 33.2% A/T).

Motif analysis was performed using the motif "Enrichment plot" function, with location scoring set to the maximum LOD value for each location, using the set of 337 Wor1 peaks as the "Query" and the random set of non-bound 500bp regions set as the "Control". For the 2309 region control set, 38.3% of the Wor1 bound regions had an instance of the motif with a LOD score of 3.74 or greater (out of 4.048), while only 8.1% of the control regions had a similarly strong motif hit. This corresponds to 129 of the 337 Wor1 peaks. Looking at different cutoffs, 44.2% of the Wor1 bound regions had an instance of the motif with a LOD score of 3.58 or greater, only 9.7% of the control regions had a similarly strong motif hit. 54% of the Wor1 bound regions had an instance of the motif with a LOD score of 3.27 or greater, only 14.2% of the control regions had a similarly strong motif hit. 62.6% of the Wor1 bound regions had an instance of the motif with a LOD score of 2.86 or greater, only 18.2% of the control regions had a similarly strong motif hit. When the 2704 region control was used, the instances of motif occurrence in the control regions were 8.2%, 9.6%, 13.8%, and 17.8% at the thresholds used above.

Pth2 Cloning and Expression

Primers and plasmids used for cloning full length *PTH2* are included in Supplemental Tables S3 and S4 respectively. The one CUG codon in *PTH2* was changed using one rounds of PCR with primers corresponding to the CUG codon to be mutated. The two fragments containing the altered *PTH2* sequence were combined using fusion PCR. Codon-changed *PTH2* was amplified to add BamHI /NheI and HindIII/XhoI sites at the 5' and 3' ends, as well as a stop codon at the end of the ORF, introduced into the pUC19 plasmid (New England Biolabs, Ipswich, MA) between the BamHI and HindIII sites, and

sequenced. The sequenced construct was subcloned into pET28b between the NheI/XhoI sites to create 6xHis-Pth2. After verifying ligation, plasmids were transformed into BL21 cells for expression. Protein purification was performed as with the Wor1 constructs. Gel shift conditions were the same used for the Wor1 K_D determinations on 20bp DNA fragments.

References

1. MacPherson S, Larochelle M, Turcotte B (2006) A Fungal Family of Transcriptional Regulators: the Zinc Cluster Proteins. *Microbiol Mol Biol Rev* 70:583-604.
2. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA (2008) DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res* 36:D88-D92.
3. Shelest E (2008) Transcription factors in fungi. *FEMS Microbiology Letters* 286:145-151.
4. Lohse MB, Johnson AD (2009) White-opaque switching in *Candida albicans*. *Curr Opin Microbiol* 12:650-654.
5. Huang G, et al. (2006) Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 103:12813-12818.
6. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* 5:e256.
7. Srikantha T, et al. (2006) TOS9 regulates white-opaque switching in *Candida albicans*. *Eukaryot Cell* 5:1674-1687.
8. Miller MG, Johnson AD (2002) White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* 110:293-302.

9. Kvaal CA, Srikantha T, Soll DR (1997) Misexpression of the white-phase-specific gene WH11 in the opaque phase of *Candida albicans* affects switching and virulence. *Infect Immun* 65:4468-4475.

10. Kvaal C, et al. (1999) Misexpression of the opaque-phase-specific gene PEP1 (SAP1) in the white phase of *Candida albicans* confers increased virulence in a mouse model of cutaneous infection. *Infect Immun* 67:6652-6662.

11. Geiger J, Wessels D, Lockhart SR, Soll DR (2004) Release of a potent polymorphonuclear leukocyte chemoattractant is regulated by white-opaque switching in *Candida albicans*. *Infect Immun* 72:667-677.

12. Lohse MB, Johnson AD (2008) Differential phagocytosis of white versus opaque *Candida albicans* by *Drosophila* and mouse phagocytes. *PLoS One* 3:e1473.

13. Lan CY, et al. (2002) Metabolic specialization associated with phenotypic switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 99:14907-14912.

14. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.

15. Altschul SF, et al. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS Journal* 272:5101-5109.

16. Kelley LA, Sternberge MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc* 4:363-371.

17. Nguyen VQ, Sil A (2008) Temperature-induced switch to the pathogenic yeast form of *Histoplasma capsulatum* requires Ryp1, a conserved transcriptional regulator. *Proc Natl Acad Sci U S A* 105:4880-4885.

18. Caspari T (1997) Onset of gluconate-H+ symport in *Schizosaccharomyces pombe* is regulated by the kinases Wis1 and Pka1, and requires the gti1+ gene product. *J Cell Sci* 110:2599-2608.

19. Michielse CM, et al. (2009) The Nuclear Protein Sge1 of Fusarium oxysporum Is Required for Parasitic Growth. *PLoS Pathog* 5:e1000637.

20. Kunitomo H, Sugimoto A, Wilkinson CR, Yamamoto M (1995) *Schizosaccharomyces pombe* pac2+ controls the onset of sexual development via a pathway independent of the cAMP cascade. *Curr Genet* 28:32-38.

21. Webster RH, Sil A (2008) Conserved factors Ryp2 and Ryp3 control cell morphology and infectious spore formation in the fungal pathogen Histoplasma capsulatum. *Proc Natl Acad Sci U S A* 105:14573-14578.

22. Santos MA, Tuite MF (1995) The CUG codon is decoded in vivo as serine and not leucine in Candida albicans. *Nucleic Acids Res* 29:1481-1486.

23. Reece RJ, Ptashne M (1993) Determinants of Binding-Site Specificity Among Yeast C6 Zinc Cluster Proteins. *Science* 261:909-911.

24. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, (AAAI Press), pp 28-36.

25. Hope IA, Struhl K (1987) GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J* 6:2781-2784.

26. Puntervoll P, et al. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res* 31:3625-3630.

27. Linding R, Russell RB, Neduva V, Gibson TJ (2003) GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 31:3701-3708.

28. Homann OR, Johnson AD (2010) MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol* 8:49.

29. Rupp S, Summers E, Lo H, Madhani H, Fink G (1999) MAP kinase and cAMP filamentation signaling pathways converge on the unusually large promoter of the yeast FLO11 gene. *EMBO J* 18:1257-1269.

30. Guarente L, Ptashne M (1981) Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 78:2199-2203.

31. Mumberg D, Müller R, Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156:119-122.

32. Johnston M, Riles L, Hegemann JH (2002) Gene disruption. *Methods in Enzymology* 350:290-315.

33. Chodosh L (1988) Mobility Shift DNA-Binding Assay Using Gel Electrophoresis. *Current Protocols in Molecular Biology*, (John Wiley & Sons, Inc., Hoboken, NJ), Vol 2, pp 12.12.11-12.12.10.

34. Rupp S (2002) *LacZ* assays in yeast. *Methods in Enzymology* 350:112-131.

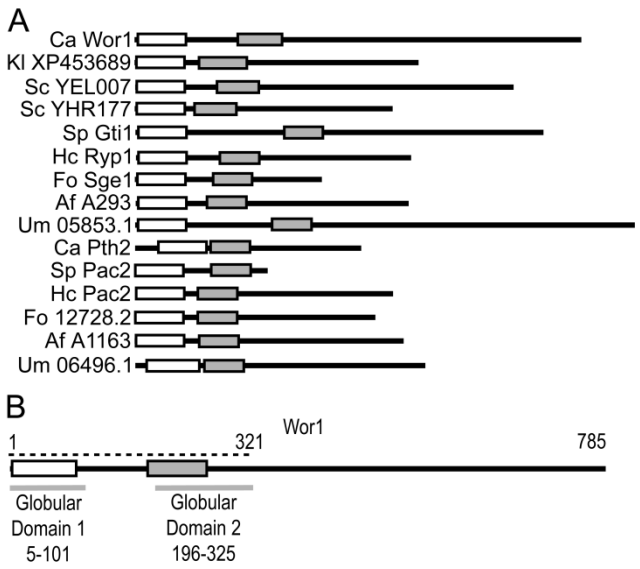


Figure 1: Wor1 is a member of a conserved family of fungal proteins. (a) Alignment of Wor1 homologs across 8 fungal species for 15 representative members of this protein family, *C. albicans* Wor1 and Pth2, *S. cerevisiae* YEL007 and YHR177, *F. oxysporum* Sge1 and 12728.2, *H. capsulatum* Ryp1 and Pac2, *S. pombe* Gti1 and Pac2, *Kluyveromyces lactis* XP-453689, *Ustilago mayis* UM05853.1 and UM06496.1, and *Aspergillus fumigatus* A293 and A1163. Wor1, Gti1, YEL007, YHR177, Ryp1, A293, UM05853.1, XP_453689, and Sge1 all represent one distinct set of this protein family while Pth2, Pac2, HcPac2, A1163, UM06496.1, and 12728.2 represent the second set. Wor1 is 785aa long, other proteins are drawn to scale. The two conserved domains in this family of proteins are indicated by clear (WOPRa) and gray (WOPRb) filled boxes respectively. Family members are representative of a wide range of fungal species and include those that have been experimentally characterized. Amino acid sequences are given in Figure S1 (b) Location of the two globular domains of Wor1 predicted by ELM and their positions relative to the WOPRa (clear) and WOPRb (gray) domains conserved

in this family of proteins. The region encompassed by the Wor1 1-321aa construct is also illustrated (dashed line).

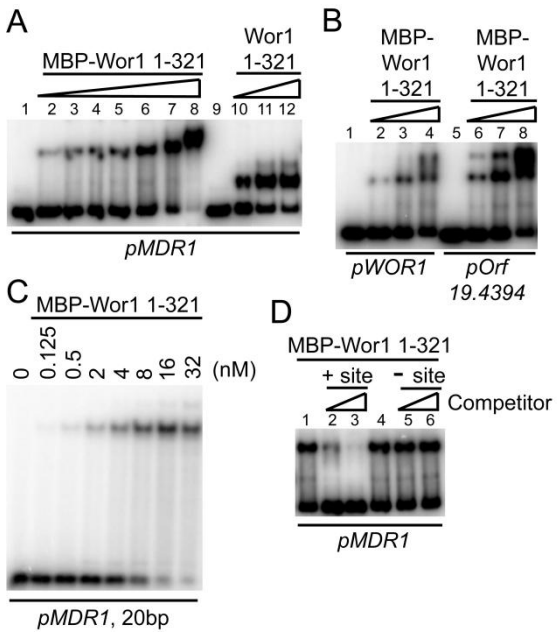


Figure 2: Wor1 binds to specific sequences of DNA, as monitored by mobility shift assays. (a) MBP-Wor1 1-321 [lanes 2-8] and Wor1 1-321 [lanes 10-12] bind to a 200bp fragment from the *MDR1* promoter. Protein concentrations are 0nM [lanes 1,9], 32nM [lanes 2,10], 128nM [lanes 3,10], 200nM [lanes 4,12], 256nM [lane 5], 512nM [lane 6], 1024nM [lane 7] and 2048nM [lane 8]. (b) MBP-Wor1 1-321 binds 200bp fragments from the *WOR1* promoter [lanes 1-4] and the orf19.4394 [lanes 5-8] promoter. Protein concentrations are 0nM [lanes 1, 5], 32nM [lanes 2, 6], 128nM [lanes 3,7], and 512nM [lanes 4, 8]. Both of these 200bp regions have multiple Wor1 binding sites. (c) Wor1 binds with a K_D in the 4-8nM range. Increasing concentrations of MBP Wor1 1-321 were incubated with a 20bp fragment from *pMDR1* that contains the Wor1 binding site. Half of the probe shifted in the 4-8nM range. (d) Binding of MBP-Wor1 1-321 to the *MDR1* promoter fragment can be competed away by increasing concentrations of a specific 20bp fragment. Unlabeled competitors correspond to base pairs 120-140 [lanes 4-6] and 140-160 [lanes 1-3] from the larger 200bp *MDR1* promoter fragment. Positions

140-160bp contain a Wor1 binding site, unlike positions 120-140bp. Protein concentrations are 512nM in each lane, competitor concentrations are 0nM [lanes 1, 4], 600nM [lanes 2, 5], and 6µM [lanes 3,6].

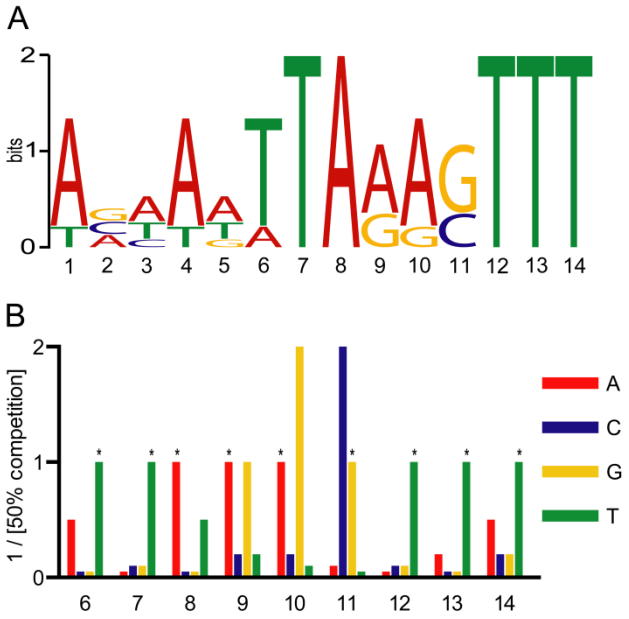


Figure 3: DNA sequence recognized by Wor1. (a) Motif recognized by Wor1, developed from six 20bp regions shown to compete for Wor1 binding. Motif and logo were developed using MEME. (b) Motif single base pair substitution screen for effects on Wor1 binding. All 27 possible single base pair substitutions to positions 6-14 from the motif in (a) were made to a 20bp fragment containing the Wor1 motif. This unlabeled 20bp mutation library was then screened for the ability to compete for binding with the labeled 200bp pMDR1 fragment. For each fragment, the concentration of competitor needed to achieve a 50% reduction in binding to *pMDR1* was determined. Plot shows the inverse of the concentration needed, with the wild type concentration normalized to 1 so values greater than 1 correspond to stronger and values less than 1 to weaker binding. For samples that did not produce a 50% reduction in binding at any concentration, we used the value of the maximum concentration used in the experiment. Asterisks mark the most commonly occurring nucleotide at each position.

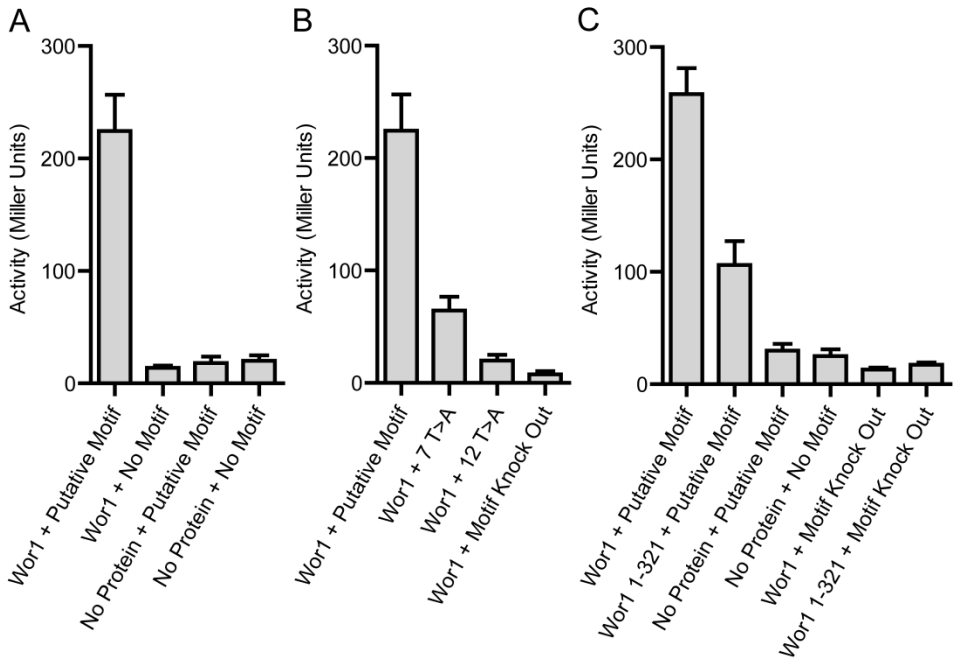


Figure 4: Wor1-dependent *in vivo* transcriptional activation. (a) Transcriptional activation of a UAS-less *CYC1* promoter is dependent on both the presence of the putative Wor1 motif in the *CYC1* promoter and the ectopic expression of Wor1. (b) Mutations in the putative Wor1 motif that affected binding *in vitro* also result in a reduction in transcriptional activation *in vivo* when inserted into the UAS-less *CYC1* promoter. The Wor1 motif knock out consists of four combined mutations, 7 T>A, 8 A>C, 12 T>A, and 13 T>C. (c) Ectopic expression of Wor1 1-321aa does not produce the level of transcriptional activation seen for ectopic expression of full length Wor1. Activation assays were performed in triplicate on the same day for each strain type, data in each panel reflect the mean of the three values for each strain on one day, error bars represent the standard deviation. The two Wor1 homologs in *S. cerevisiae* (YEL007 and YHR177) were deleted from all strains to avoid potential cross-activation.

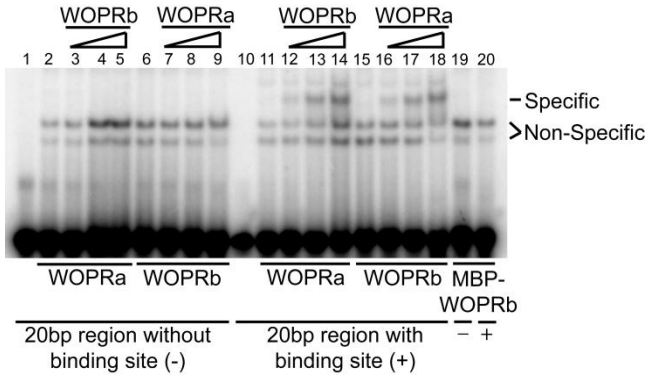


Figure 5: Wor1 sequence-specific binding to DNA requires both conserved domains. Fixed amounts of either WOPRa [50nM, lanes 2-5 and 11-14] or WOPRb [50nM, lanes 6-9 and 15-18] were present and increasing concentrations of the other domain were added in. Lanes 2 and 11 contain only WOPRa, lanes 6 and 15 contain only WOPRb. Titrated WOPRa concentrations were 10nM [lanes 7, 16], 50nM [lanes 8, 17], or 250nM [lanes 9, 18]. Titrated WOPRb concentrations were 10nM [lanes 3, 12], 50nM [lanes 4, 13], or 250nM [lanes 5, 14]. DNA probes are either a 20bp DNA sequence containing a Wor1 binding site [140-160bp, lanes 10-18, 20] or lacking a Wor1 binding site [120-140bp, lanes 1-9, 19], both taken from *pMDR1*. Wor1 binding is only seen in the presence of both domains and the binding sites [lanes 12-14, 16-18]. The non-specific band appears at the same place in lanes where either DNA fragment was co-incubated with the larger 6xHis-MBP-Wor1 WOPRb construct that was purified using the same protocol as the 6xHis WOPRa or WOPRb constructs, indicating that it is not due to either domain binding the DNA but rather some *E. coli* proteins that were not removed by the purification process. No non-specific competitor was included with these reactions.

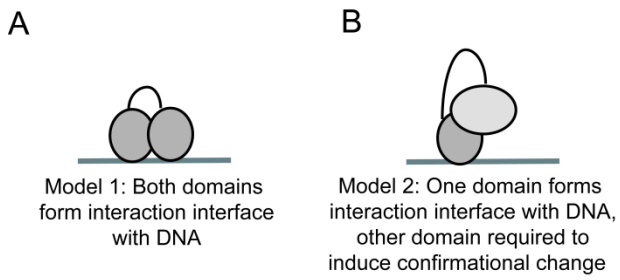
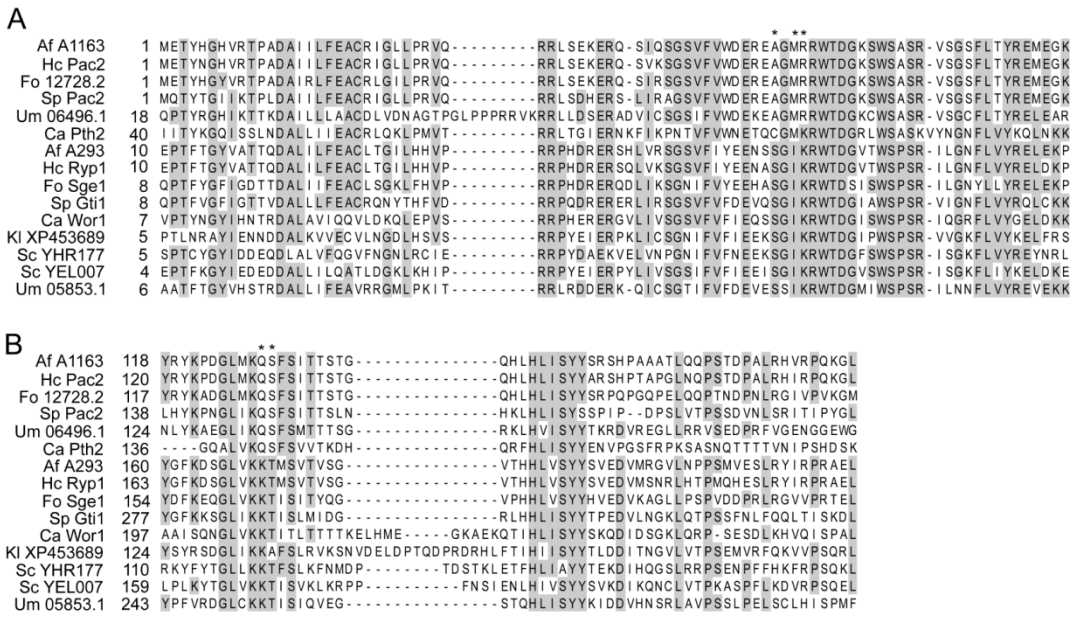
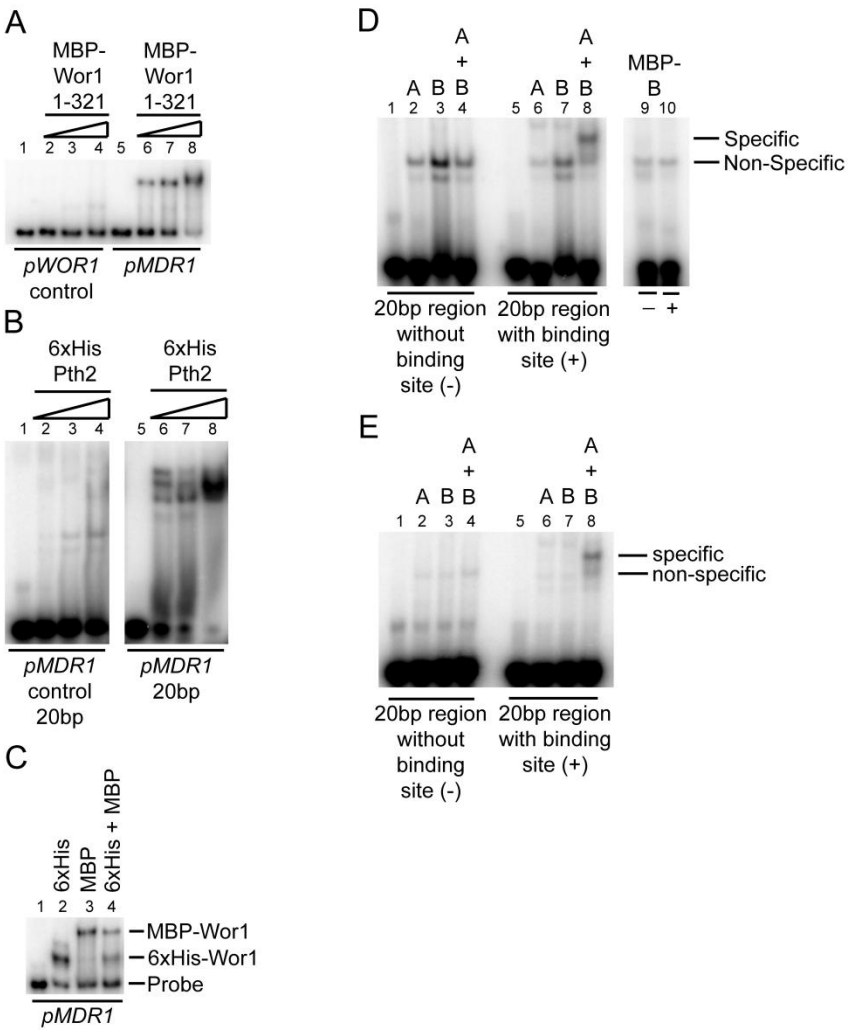


Figure 6: Models for DNA-binding by Wor1. (a) One model is that both domains contribute directly to the interaction with the DNA molecule. (b) A second model is that only one domain directly interacts with DNA, with the other domain inducing a conformational change stabilizing this binding.

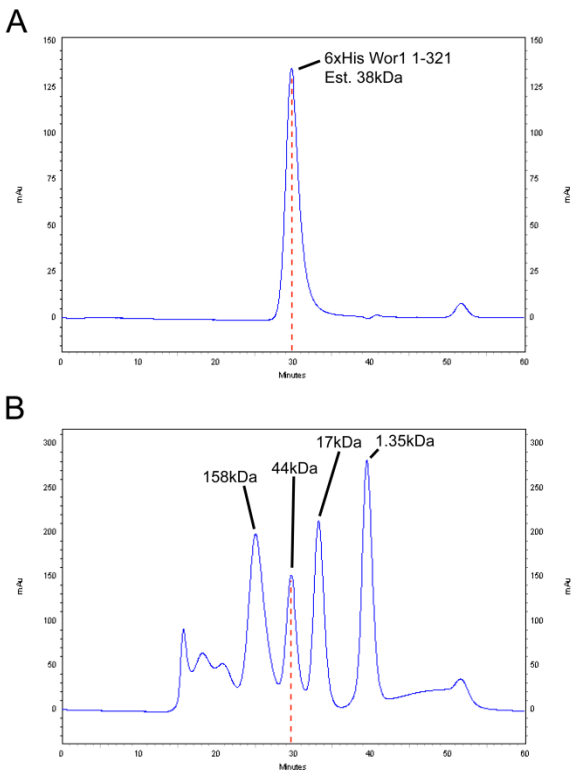


Supplemental Figure S1: The Wor1 family of fungal proteins is defined by two regions of conservation. (a) Alignment of WOPRa of 15 members of this family taken from 8 fungal species. These are *C. albicans* Wor1 and Pth2, *S. cerevisiae* YEL007 and YHR177, *F. oxysporum* Sge1 and 12728.2, *H. capsulatum* Ryp1 and Pac2, *S. pombe* Gti1 and Pac2, *Kluyveromyces lactis* XP-453689, *Ustilago mayis* UM05853.1 and UM06496.1, and *Aspergillus fumigatus* A293 and A1163. This region corresponds to residues 7-90 of Wor1. (b) Alignment of WOPRb from the same members of this family. This region corresponds to residues 197-262 of Wor1. In both panels, residues that differ between the two distinct family members are indicated with an asterisk. Proteins were chosen to incorporate a range of fungal species and family members to include those that have been previously characterized.



Supplemental Figure S2: Wor1 and Pth2 bind to specific sequences of DNA as monitored by mobility shift assays, Wor1 binds DNA as a monomer and sequence-specific DNA binding requires the presence of both conserved domains. (a) MBP-Wor1 1-321 [lanes 2-4 and 6-8] binds to a 200bp fragment from the *MDR1* promoter [lanes 5-8] that was enriched for the presence of Wor1 in a previously published ChIP-Chip experiment (6) but not to a 200bp fragment of the *WOR1* [lanes 1-4] promoter that was not enriched for the presence of Wor1 in the same experiment. Protein concentrations are 0nM [lanes 1,5], 100nM [lanes 2,6], 500nM [lanes 3,7], and 1.5µM [lanes 4,8]. (b) The Wor1 ortholog Pth2 binds to a 20bp fragment from *pMDR1* that contains the Wor1 binding site

[base pairs 140-160, lanes 5-8] but not to a 20bp fragment from *pMDR1* lacking the binding site [base pairs 120-140, lanes 1-4]. Protein concentrations are 0nM [lanes 1,5], 3nM [lanes 2,6], 14nM [lanes 3,7], 70nM [lanes 4,8]. (c) The 200bp fragment from p*MDR1* was incubated with either an 80kDa MBP-Wor1 1-321aa construct [lanes 3, 4] or a 40kDa 6xHis-Wor1 1-321aa construct [lanes 2, 4]. When DNA is co-incubated with both Wor1 constructs [lane 4], we observe shifted bands identical to those seen with the individual proteins but not a new intermediate sized band, indicating that Wor1 binds DNA as a monomer rather than a dimer. (d) Fixed amounts of 6xHis tagged WOPRa (A) [240nM, lanes 2, 4, 6, 8], 6xHis tagged WOPRb (B) [240nM, lanes 3, 4, 7, 8], or 6xHis and MBP tagged WOPRb (MBP-B) [1 µM, lanes 9-10] were incubated with a 20bp DNA sequences from *pMDR1* containing a Wor1 binding site [140-160bp, lanes 5-8, 10] or lacking a Wor1 binding site [120-140bp, lanes 1-4, 9]. Wor1 binding is only seen in the presence of both domains and the Wor1 binding site [lane 8], the sequence independent shift seen in lanes 2-4 and 6-8 also appears when a 6xHis and MBP tagged WOPRb is present in lanes 9 and 10. As the band equivalent shifts are seen in lanes 6-8 and 9-10 despite the ~40kDa size difference in the two WOPRb constructs, it does not reflect binding by either WOPRa or WOPRb but rather an *E. coli* protein that was not removed during the purification process. Non-specific competitor was not included with these reactions to more clearly show the non-specific band. Lanes 9 and 10 are from the same gel as lanes 1-8; both portions of the panel were treated in the same manner during creation of this image. (e) Lanes 1-8 as in (d) but with non-specific competitor included in the binding reactions to reduce non-specific binding. As in (d), Wor1 binding is only seen in the presence of both domains and the Wor1 binding site [lane 8].



Supplemental Figure S3: 6xHis Wor1 1-321 is a monomer in solution. (a) Purified 6xHis Wor1 1-321 (estimated weight 38.2kDa) was run over a Superdex 200 gel filtration column, eluting at the 29-30 minute mark. (b) Standards of known weight were run over the same Superdex 200 gel filtration column under the same conditions. Standards are bovine Gamma globulin (158kDa), chicken Ovalbumin (44kDa), horse Myoglobin (17kDa), and Vitamin B-12 (1.35kDa). The ~38kDa 6xHis-Wor1 1-321 eluted from the column at approximately the same position as the 44kDa Chicken Ovalbumin standard (also 29-30 minutes). Both plots are of the absorbance at 280nm, red dashed lines indicate the center of the 6xHis Wor1 1-321 and chicken Ovalbumin peaks.

Name	Expanded Description
MBP-Wor1 1-321	6xHis-MBP-PreScission Protease Site- Wor1 1-321aa
Wor1 1-321	PreScission Protease Cleaved MBP-Wor1 1-321 to give Wor1 1-321aa only
6xHis Wor1 1-321	6xHis-Wor1 1-321aa
6xHis WOPRa	6xHis-Wor1 1-101aa
6xHis WOPRb	6xHis-Wor1 196-321aa
MBP-WOPRa	6xHis-MBP-PreScission Protease Site- Wor1 1-101aa
MBP-WOPRb	6xHis-MBP-PreScission Protease Site- Wor1 196-321aa

Supplemental Table S1: Wor1 constructs used in this study.

Binding Site	Oligonucleotide Sequence
MDR1 Site	tacttttaaggttttgttt
orf19.4394 site 1	aaaaatttaaagtttaatta
orf19.4394 site 2	tattaccaattaaactttat
WOR1 site 1	gttaaaaaactctattttca
WOR1 site 2	aagaagttaaacttttttga
WOR1 site 3	ggaataattagagttttaca

Supplemental Table S2: List of binding sites used to generate the Wor1 motif.

Supplemental Table S3: List of 337 Wor1 peaks called by Mochiview.

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chr1	142365	142864	+	2.06
Ca21chr1	147402	147901	+	2.12
Ca21chr1	149395	149894	+	4.04
Ca21chr1	230409	230908	+	2.43
Ca21chr1	232285	232784	+	3.29
Ca21chr1	238668	239167	+	4.05
Ca21chr1	293834	294333	+	4.05
Ca21chr1	297852	298351	+	3.02
Ca21chr1	298702	299201	+	2.53
Ca21chr1	371137	371636	+	3.29
Ca21chr1	439891	440390	+	3.32
Ca21chr1	447898	448397	+	2.43
Ca21chr1	497363	497862	+	2.36
Ca21chr1	534643	535142	+	3.74
Ca21chr1	539283	539782	+	4.05
Ca21chr1	571820	572319	+	4.05
Ca21chr1	664765	665264	+	1.93
Ca21chr1	712541	713040	+	3.74
Ca21chr1	843923	844422	+	4.05
Ca21chr1	920947	921446	+	2.12
Ca21chr1	991257	991756	+	2.06
Ca21chr1	1077409	1077908	+	3.74
Ca21chr1	1081939	1082438	+	1.93
Ca21chr1	1089884	1090383	+	2.56
Ca21chr1	1156924	1157423	+	2.88
Ca21chr1	1225413	1225912	+	3.02
Ca21chr1	1233308	1233807	+	4.05
Ca21chr1	1250637	1251136	+	-0.23
Ca21chr1	1285810	1286309	+	2.06
Ca21chr1	1334795	1335294	+	1.78
Ca21chr1	1499272	1499771	+	1.91
Ca21chr1	1500624	1501123	+	2.35
Ca21chr1	1522346	1522845	+	2.57
Ca21chr1	1525905	1526404	+	2.88
Ca21chr1	1573142	1573641	+	2.13
Ca21chr1	1577546	1578045	+	3.02
Ca21chr1	1676018	1676517	+	2.43

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chr1	1738666	1739165	+	2.36
Ca21chr1	1743388	1743887	+	2.36
Ca21chr1	1745138	1745637	+	2.06
Ca21chr1	1849559	1850058	+	2.35
Ca21chr1	1979083	1979582	+	2.06
Ca21chr1	1989879	1990378	+	3.74
Ca21chr1	2003130	2003629	+	4.04
Ca21chr1	2052101	2052600	+	3.73
Ca21chr1	2053895	2054394	+	2.43
Ca21chr1	2099145	2099644	+	2.36
Ca21chr1	2100994	2101493	+	2.43
Ca21chr1	2105311	2105810	+	2.57
Ca21chr1	2194109	2194608	+	1.92
Ca21chr1	2224342	2224841	+	4.05
Ca21chr1	2226263	2226762	+	3.60
Ca21chr1	2227811	2228310	+	3.74
Ca21chr1	2233463	2233962	+	4.05
Ca21chr1	2248886	2249385	+	3.74
Ca21chr1	2274525	2275024	+	2.53
Ca21chr1	2291635	2292134	+	3.33
Ca21chr1	2345015	2345514	+	2.57
Ca21chr1	2347476	2347975	+	2.35
Ca21chr1	2348642	2349141	+	3.32
Ca21chr1	2402472	2402971	+	1.98
Ca21chr1	2474727	2475226	+	2.88
Ca21chr1	2520758	2521257	+	4.04
Ca21chr1	2623108	2623607	+	0.75
Ca21chr1	2630806	2631305	+	4.05
Ca21chr1	2826356	2826855	+	4.05
Ca21chr1	2939069	2939568	+	4.05
Ca21chr1	2940804	2941303	+	3.73
Ca21chr1	2947729	2948228	+	4.04
Ca21chr1	2992746	2993245	+	2.05
Ca21chr1	2998968	2999467	+	2.36
Ca21chr1	3071151	3071650	+	2.57
Ca21chr1	3109489	3109988	+	0.76
Ca21chr1	3150848	3151347	+	3.29
Ca21chr2	18168	18667	+	3.02

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chr2	28627	29126	+	3.74
Ca21chr2	129629	130128	+	3.74
Ca21chr2	161323	161822	+	2.12
Ca21chr2	163018	163517	+	3.74
Ca21chr2	203076	203575	+	3.74
Ca21chr2	221545	222044	+	2.06
Ca21chr2	316356	316855	+	4.05
Ca21chr2	373188	373687	+	2.11
Ca21chr2	453275	453774	+	2.35
Ca21chr2	473916	474415	+	2.12
Ca21chr2	475682	476181	+	3.60
Ca21chr2	523682	524181	+	4.05
Ca21chr2	593952	594451	+	3.33
Ca21chr2	608478	608977	+	2.05
Ca21chr2	732864	733363	+	3.73
Ca21chr2	735500	735999	+	3.74
Ca21chr2	743902	744401	+	3.74
Ca21chr2	779394	779893	+	3.02
Ca21chr2	781048	781547	+	0.37
Ca21chr2	785245	785744	+	4.05
Ca21chr2	884021	884520	+	4.05
Ca21chr2	933725	934224	+	2.88
Ca21chr2	936956	937455	+	4.05
Ca21chr2	1081538	1082037	+	4.05
Ca21chr2	1149203	1149702	+	1.81
Ca21chr2	1201224	1201723	+	1.82
Ca21chr2	1202911	1203410	+	4.05
Ca21chr2	1220900	1221399	+	4.05
Ca21chr2	1282439	1282938	+	2.12
Ca21chr2	1316394	1316893	+	4.05
Ca21chr2	1333863	1334362	+	3.33
Ca21chr2	1339364	1339863	+	3.32
Ca21chr2	1380820	1381319	+	1.78
Ca21chr2	1422052	1422551	+	4.05
Ca21chr2	1576555	1577054	+	4.05
Ca21chr2	1652888	1653387	+	2.36
Ca21chr2	1741483	1741982	+	3.74
Ca21chr2	1743393	1743892	+	2.23

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chr2	1746209	1746708	+	1.74
Ca21chr2	1805241	1805740	+	3.60
Ca21chr2	1808004	1808503	+	3.74
Ca21chr2	1837709	1838208	+	4.05
Ca21chr2	1840186	1840685	+	2.05
Ca21chr2	1866360	1866859	+	0.43
Ca21chr2	1870527	1871026	+	3.32
Ca21chr2	1871960	1872459	+	3.60
Ca21chr2	1934248	1934747	+	2.36
Ca21chr2	1993700	1994199	+	2.06
Ca21chr2	2002547	2003046	+	2.06
Ca21chr2	2033762	2034261	+	3.74
Ca21chr2	2055045	2055544	+	1.51
Ca21chr3	105727	106226	+	2.05
Ca21chr3	128342	128841	+	3.74
Ca21chr3	152725	153224	+	1.61
Ca21chr3	246643	247142	+	4.04
Ca21chr3	338735	339234	+	4.05
Ca21chr3	380441	380940	+	4.05
Ca21chr3	401939	402438	+	2.53
Ca21chr3	432144	432643	+	3.74
Ca21chr3	521920	522419	+	3.74
Ca21chr3	564653	565152	+	3.59
Ca21chr3	566024	566523	+	3.02
Ca21chr3	573485	573984	+	2.42
Ca21chr3	609158	609657	+	4.05
Ca21chr3	881578	882077	+	3.01
Ca21chr3	957292	957791	+	1.78
Ca21chr3	972921	973420	+	4.05
Ca21chr3	984306	984805	+	2.06
Ca21chr3	1110917	1111416	+	4.05
Ca21chr3	1116075	1116574	+	2.05
Ca21chr3	1117913	1118412	+	3.59
Ca21chr3	1121431	1121930	+	4.04
Ca21chr3	1123335	1123834	+	3.33
Ca21chr3	1145105	1145604	+	2.36
Ca21chr3	1154080	1154579	+	3.60
Ca21chr3	1167201	1167700	+	2.35

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chr3	1175202	1175701	+	4.05
Ca21chr3	1212837	1213336	+	4.04
Ca21chr3	1218344	1218843	+	2.36
Ca21chr3	1221519	1222018	+	3.74
Ca21chr3	1222559	1223058	+	2.05
Ca21chr3	1244595	1245094	+	3.60
Ca21chr3	1254056	1254555	+	2.06
Ca21chr3	1356880	1357379	+	1.60
Ca21chr3	1360547	1361046	+	2.05
Ca21chr3	1362532	1363031	+	3.74
Ca21chr3	1365162	1365661	+	2.88
Ca21chr3	1413122	1413621	+	3.74
Ca21chr3	1472691	1473190	+	1.74
Ca21chr3	1474838	1475337	+	3.01
Ca21chr3	1500368	1500867	+	3.29
Ca21chr3	1510085	1510584	+	3.60
Ca21chr3	1588917	1589416	+	4.05
Ca21chr3	1605800	1606299	+	3.32
Ca21chr3	1672730	1673229	+	4.04
Ca21chr3	1722519	1723018	+	2.05
Ca21chr3	1725042	1725541	+	3.29
Ca21chr3	1726940	1727439	+	3.74
Ca21chr3	1728977	1729476	+	1.92
Ca21chr3	1758452	1758951	+	2.23
Ca21chr3	1760833	1761332	+	0.74
Ca21chr4	92813	93312	+	3.74
Ca21chr4	172780	173279	+	4.05
Ca21chr4	200875	201374	+	4.05
Ca21chr4	214776	215275	+	3.60
Ca21chr4	277357	277856	+	4.05
Ca21chr4	288945	289444	+	1.61
Ca21chr4	289825	290324	+	1.92
Ca21chr4	469882	470381	+	3.74
Ca21chr4	693270	693769	+	3.02
Ca21chr4	696715	697214	+	3.74
Ca21chr4	704272	704771	+	3.02
Ca21chr4	799514	800013	+	4.05
Ca21chr4	831669	832168	+	3.74

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chr4	863890	864389	+	4.05
Ca21chr4	867281	867780	+	4.05
Ca21chr4	869847	870346	+	3.02
Ca21chr4	875903	876402	+	2.88
Ca21chr4	893701	894200	+	1.06
Ca21chr4	935662	936161	+	2.13
Ca21chr4	942037	942536	+	3.33
Ca21chr4	1059074	1059573	+	2.36
Ca21chr4	1165301	1165800	+	3.74
Ca21chr4	1317441	1317940	+	3.60
Ca21chr4	1457074	1457573	+	4.05
Ca21chr4	1470121	1470620	+	2.12
Ca21chr4	1523138	1523637	+	4.05
Ca21chr4	1524478	1524977	+	1.20
Ca21chr4	1526274	1526773	+	3.28
Ca21chr4	1535038	1535537	+	3.74
Ca21chr5	100078	100577	+	3.73
Ca21chr5	111368	111867	+	2.13
Ca21chr5	115846	116345	+	3.73
Ca21chr5	117592	118091	+	4.05
Ca21chr5	123919	124418	+	3.60
Ca21chr5	163386	163885	+	2.06
Ca21chr5	167846	168345	+	3.02
Ca21chr5	205555	206054	+	3.33
Ca21chr5	228246	228745	+	4.05
Ca21chr5	302645	303144	+	2.36
Ca21chr5	308020	308519	+	3.74
Ca21chr5	496652	497151	+	2.53
Ca21chr5	500024	500523	+	3.74
Ca21chr5	592568	593067	+	3.02
Ca21chr5	599029	599528	+	3.74
Ca21chr5	628088	628587	+	2.06
Ca21chr5	655467	655966	+	3.33
Ca21chr5	695264	695763	+	3.60
Ca21chr5	708297	708796	+	3.74
Ca21chr5	720572	721071	+	3.02
Ca21chr5	733057	733556	+	4.05
Ca21chr5	740562	741061	+	2.42

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chr5	751644	752143	+	3.60
Ca21chr5	774158	774657	+	3.73
Ca21chr5	827580	828079	+	3.32
Ca21chr5	852452	852951	+	4.05
Ca21chr5	855847	856346	+	3.74
Ca21chr5	867295	867794	+	1.33
Ca21chr5	875030	875529	+	3.01
Ca21chr5	895320	895819	+	4.05
Ca21chr5	922241	922740	+	4.05
Ca21chr5	924140	924639	+	4.05
Ca21chr5	932793	933292	+	3.74
Ca21chr5	1057206	1057705	+	2.08
Ca21chr5	1115536	1116035	+	4.05
Ca21chr5	1117217	1117716	+	2.36
Ca21chr6	36858	37357	+	1.93
Ca21chr6	56605	57104	+	3.02
Ca21chr6	147845	148344	+	3.74
Ca21chr6	156518	157017	+	4.04
Ca21chr6	163753	164252	+	4.05
Ca21chr6	208632	209131	+	1.90
Ca21chr6	348284	348783	+	3.74
Ca21chr6	361222	361721	+	3.60
Ca21chr6	362493	362992	+	3.33
Ca21chr6	473577	474076	+	1.75
Ca21chr6	582279	582778	+	4.05
Ca21chr6	668343	668842	+	3.60
Ca21chr6	752002	752501	+	3.02
Ca21chr6	765412	765911	+	4.05
Ca21chr6	829383	829882	+	4.05
Ca21chr6	1027397	1027896	+	0.07
Ca21chr7	23142	23641	+	3.60
Ca21chr7	40893	41392	+	1.32
Ca21chr7	134815	135314	+	1.34
Ca21chr7	145311	145810	+	2.57
Ca21chr7	147043	147542	+	3.32
Ca21chr7	279657	280156	+	2.06
Ca21chr7	319674	320173	+	4.04
Ca21chr7	327352	327851	+	2.06

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chr7	364829	365328	+	2.12
Ca21chr7	478407	478906	+	1.82
Ca21chr7	673673	674172	+	4.05
Ca21chr7	725103	725602	+	4.04
Ca21chr7	770473	770972	+	3.74
Ca21chr7	839516	840015	+	3.01
Ca21chr7	912197	912696	+	3.33
Ca21chr7	913774	914273	+	2.06
Ca21chr7	917837	918336	+	3.28
Ca21chr7	923461	923960	+	2.35
Ca21chrR	34875	35374	+	3.74
Ca21chrR	35749	36248	+	3.29
Ca21chrR	51729	52228	+	3.29
Ca21chrR	79031	79530	+	3.02
Ca21chrR	114115	114614	+	1.75
Ca21chrR	175260	175759	+	4.05
Ca21chrR	265156	265655	+	3.73
Ca21chrR	266819	267318	+	2.23
Ca21chrR	295499	295998	+	3.74
Ca21chrR	398268	398767	+	3.29
Ca21chrR	403544	404043	+	1.93
Ca21chrR	444633	445132	+	4.05
Ca21chrR	488144	488643	+	1.21
Ca21chrR	555946	556445	+	4.05
Ca21chrR	586770	587269	+	2.36
Ca21chrR	591544	592043	+	4.04
Ca21chrR	596368	596867	+	4.05
Ca21chrR	598363	598862	+	2.23
Ca21chrR	604117	604616	+	2.36
Ca21chrR	631192	631691	+	2.36
Ca21chrR	779760	780259	+	4.05
Ca21chrR	799295	799794	+	3.33
Ca21chrR	845214	845713	+	4.05
Ca21chrR	872691	873190	+	2.36
Ca21chrR	874642	875141	+	4.05
Ca21chrR	876947	877446	+	3.28
Ca21chrR	878014	878513	+	3.29
Ca21chrR	881881	882380	+	3.74

SEQ_NAME	START	END	STRAND	Maximum LOD Score, Wor1 Motif
Ca21chrR	981401	981900	+	2.88
Ca21chrR	1042851	1043350	+	2.36
Ca21chrR	1093267	1093766	+	3.74
Ca21chrR	1145695	1146194	+	3.01
Ca21chrR	1196212	1196711	+	4.05
Ca21chrR	1259373	1259872	+	2.36
Ca21chrR	1265177	1265676	+	2.05
Ca21chrR	1288324	1288823	+	3.73
Ca21chrR	1289982	1290481	+	3.74
Ca21chrR	1296959	1297458	+	1.82
Ca21chrR	1313593	1314092	+	4.05
Ca21chrR	1314937	1315436	+	2.05
Ca21chrR	1374536	1375035	+	3.29
Ca21chrR	1410800	1411299	+	3.74
Ca21chrR	1413121	1413620	+	1.93
Ca21chrR	1478697	1479196	+	2.06
Ca21chrR	1525914	1526413	+	2.53
Ca21chrR	1528631	1529130	+	1.93
Ca21chrR	1637154	1637653	+	3.60
Ca21chrR	1704362	1704861	+	3.74
Ca21chrR	1717727	1718226	+	3.29
Ca21chrR	1997692	1998191	+	3.02
Ca21chrR	2058490	2058989	+	3.60
Ca21chrR	2077057	2077556	+	4.05
Ca21chrR	2081165	2081664	+	2.53
Ca21chrR	2084219	2084718	+	0.44
Ca21chrR	2086456	2086955	+	3.74
Ca21chrR	2087966	2088465	+	3.29
Ca21chrR	2099972	2100471	+	4.05
Ca21chrR	2110821	2111320	+	2.88
Ca21chrR	2116837	2117336	+	3.33
Ca21chrR	2160202	2160701	+	4.05
Ca21chrR	2168885	2169384	+	3.60
Ca21chrR	2204627	2205126	+	3.74

Supplemental Table S3: List of 337 Wor1 peaks called by Mochiview. Sequence positions based on CGD Assembly 21, 11/25/2008 version. Maximum LOD score for this background is ~4.048

Supplemental Table S4: List of primers used in this study.

Name	Description	Primer sequence 5'-3'
Wor1 CUG Codon Changes		
wor1.1 ClaI	Wor1 5' for (ClaI)	CGGatcgatATGTCTAATTCAAGTATAGTCCCTACATATAATGGG
wor1.2 for	Wor1 CUG#1 for	GTTTATTGAACAATCTTCGGGAATC
wor1.2 rev	Wor1 CUG#1 rev	GATTCCCGAAGATTGTTCAATAAAC
wor1.3 for	Wor1 CUG#2 for	CAGCTGCTATATCACAAAACGGATTAG
wor1.3 rev	Wor1 CUG#2 rev	CTAATCCGTTTTGTGATATAGCAGCTG
wor1.4 for	Wor1 CUG#3 for	GTTTCTCAACTATCCTATATGTTGCCACCTC
wor1.4 rev	Wor1 CUG#3 rev	GAGGTGGCAACATATAGGATAGTTGAGAAAC
wor1.5 for	Wor1 CUG#4 for	CAATTAATCACTCGCATACTTCATCATATG
wor1.5 rev	Wor1 CUG#4 rev	CATATGATGAAGTATGCGAGTGATTAATTG
wor1.6 for	Wor1 CUG#5 for	GCATCAATACATCTTCCCAACAACAAC
wor1.6 rev	Wor1 CUG#5 rev	GTTGTTGTTGGGAAGATGTATTGATGC
wor1.7 for	Wor1 CUG#6 for	CAACAACATCAATCCCAGCAGCAAGTG
wor1.7 rev	Wor1 CUG#6 rev	CACTTGCTGCTGGGATTGATGTTGTTG
wor1.8 XhoI	Wor1 3' rev (XhoI)	CGGctcgagCTAAGTACCGGTGTAATACGACCCAGAAG
Cloning Wor1		
RZO373	Wor1 5' NheI	GgctagcATGTCTAATTCAAGTATAGTCCCTAC
RZO379	Wor1 (1-101) stop 3' XhoI	GctcgagTTACTTTTTCTTCTTCTTGTCTTTATCAATTAG
RZO377	Wor1 (1-321) stop 3' XhoI	GctcgagTTAACTTCCTGCAACTGAAGAAGAATG
MBL1015	Wor1 (196-321) 5' NheI	Gttaca gct agc TCA GCT GCT ATA TCA CAA AAC GG

Name	Description	Primer sequence 5'-3'
MBL1016	Wor1 (196-321) stop 3' XhoI	gttaca ctc gag TTAACTTCCTGCAACTGAAG AAGAAT
200bp EMSA Fragments		
MBL721	*pWOR1* 5'	GTT AAA AAA CTC TAT TTT CAC AAT GAC CC
MBL722	*pWOR1* 3'	TAA TTA TTC CCT TGT TTT TTT TTA TGT TAC AGT
MBL729	*pWOR1* control 5'	TTT TTG GTT TTG GCT TTG ACT TTT
MBL730	*pWOR1* control 3'	TGA AAC AGT TTC AGT CTA ATA TAT ACT ACA AG
MBL735	pOrf19.4394 5'	ATC AAA CAA AAT TAT ATT AAG TTA ATC AAC GTT A
MBL736	pOrf19.4394 3'	TTT TCG GGG AAA AAG AGT AGT TT
MBL739	*pMDR1* 5'	TCA ATT TTC ATT TTA GGA AAT TTA CCG AG
MBL740	*pMDR1* 3'	TAA TTG GTA GCA ACA CCC ATA ATT G
20bp competition primers [only forward strand listed]		
pMDR1		
MBL771	120-140bp	CCG TGG CAT TTT TCC GTG GC
MBL773	130-150bp	TTT CCG TGG CTA CTT TTT AA
MBL775	140-160bp	TAC TTT TTA AGG TTT TGT TT
MBL777	150-170bp	GGT TTT GTT TTG TGT TTT TG
MBL779	160-180bp	TGT GTT TTT GTA CAA CAA TT
pOrf19.4394		
MBL837	0-20bp	ATC AAA CAA AAT TAT ATT AA
MBL839	10-30bp	ATT ATA TTA AGT TAA TCA AC
MBL841	20-40bp	GTT AAT CAA CGT TAT ATA AT
MBL843	30-50bp	GTT ATA TAA TAA AAA TTT AA
MBL845	40-60bp	AAA AAT TTA AAG TTT AAT

Name	Description	Primer sequence 5'-3'
		TA
MBL847	50-70bp	AGT TTA ATT ATA TTA CCA AT
MBL849	60-80bp	TAT TAC CAA TTA AAC TTT AT
MBL851	70-90bp	TAA ACT TTA TGG TTA ATA GC
MBL853	80-100bp	GGT TAA TAG CCA TGA AAC CC
MBL855	120-140bp	GAC AAT CTT AAA CCA AGT TT
MBL857	130-150bp	AAC CAA GTT TAG TTG CCA TC
MBL859	140-160bp	AGT TGC CAT CAA CAT AAA TT
pWOR1		
MBL861	minus 20-0bp	TGT TGT TGT TAG TGT ATA AT
MBL863	minus 10-10bp	AGT GTA TAA TGT TAA AAA AC
MBL865	0-20bp	GTT AAA AAA CTC TAT TTT CA
MBL867	10-30bp	TCT ATT TTC ACA ATG ACC CA
MBL869	20-40bp	CAA TGA CCC AAA TAA AAC CA
MBL871	30-50bp	AAT AAA ACC AAA AAA ACA CT
MBL873	40-60bp	AAA AAA CAC TAA GAA GTT AA
MBL875	50-70bp	AAG AAG TTA AAC TTT TTT GA
MBL877	60-80bp	ACT TTT TTG AAA CTT TTC AA
MBL887	180-200bp	AAA AAA CAA GGG AAT AAT TA
MBL889	190-210bp	GGA ATA ATT AGA GTT TTA CA
MBL891	200-220bp	GAG TTT TAC AAG AAA TTT AT
Motif Mutation Scan		
MBL893	2 T>C	gata tCaaagtttttt acta
MBL895	2 T>A	gata tAaaagtttttt acta
MBL897	3 A>G	gata ttGaagtttttt acta

Name	Description	Primer sequence 5'-3'
MBL899	3 A>T	gata ttTaagtttttt acta
MBL901	6 G>A	gata ttaaaAtttttt acta
MBL903	6 G>C	gata ttaaaCtttttt acta
MBL905	7 T>C	gata ttaaagCttttt acta
MBL907	7 T>A	gata ttaaagAttttt acta
MBL909	8 T>C	gata ttaaagtCtttt acta
MBL911	8 T>A	gata ttaaagtAtttt acta
MBL913	9 T>C	gata ttaaagttCttt acta
MBL915	9 T>A	gata ttaaagttAttt acta
MBL917	6 G>T	gata ttaaaTtttttt acta
MBL919	9 T>G	gata ttaaagttGttt acta
MBL921	WT Control	gata ttaaagtttttt acta
MBL925	3 A>C	gata ttCaagtttttt acta
MBL927	4 A>G	gata ttaGagtttttt acta
MBL929	5 A>G	gata ttaaGgtttttt acta
MBL935	2 T>G	gata tGaaagtttttt acta
MBL939	4 A>C	gata ttaCagtttttt acta
MBL941	4 A>T	gata ttaTagtttttt acta
MBL943	5 A>C	gata ttaaCgtttttt acta
MBL945	5 A>T	gata ttaaTgtttttt acta
MBL947	7 T>G	gata ttaaagGttttt acta
MBL949	8 T>G	gata ttaaagtGtttt acta
MBL951	1 T>A	gata Ataaagtttttt acta
MBL953	1 T>C	gata Ctaaagtttttt acta
MBL955	1 T>G	gata Gtaaagtttttt acta
S. cerevisiae cloning		
ΔYEL007 KO 5'	KO Forward	TTCCACAAAACAACATTTCA ACTAGACACAAGGACAACG TAAATTTCTAAATGAGTAAA GGAGAAGAACT
ΔYEL007 KO 3'	KO Reverse	AGGCAAGGAGAAGAAAAGA AAATCGGAGTACTTTTTTAA AATATATTTACGAATTCGAG CTCGTTTAAAC
ΔYHR177KO 5'	KO Forward	AATTAATCAAACAATAGAT AGGAAAAGGAGATCGACGA GTAAGCGTAAGACGGATCC CCGGGTTAATTAA
ΔYHR177 KO 3'	KO Reverse	TGAGCGGTACTTTAGGTGAA CCTTTCAGCAATTGAAAGGA TGCTGATCTTGAATTCGAGC TCGTTTAAAC

Name	Description	Primer sequence 5'-3'
Motifs for Activation		
MBL1039	Putative Motif Forward	Tcgag AAA AA TTAAA GTTTT TT TAT c
MBL1040	Putative Motif Reverse	Tcgag ATA AAA AAA CTT TAA TTT TT c
MBL1045	Motif KO Forward [7T>A, 8A>C,12T>A,13 T>C]	Tcgag AAA AA T**AC**AA G**AC**TT TT TAT c
MBL1046	Motif KO Reverse [7T>A, 8A>C,12T>A,13 T>C]	Tcgag ATA AAA AGT CTT GTA TTT TT c
MBL1048	Motif 7 T>A Forward	Tcgag AAA AA T**A**AAA GTTTT TT TAT c
MBL1049	Motif 7 T>A Reverse	Tcgag ATA AAA AAA CTT TTA TTT TT c
MBL1054	Motif 12 T>A Forward	Tcgag AAA AA TTAAA G**A**TTT TT TAT c
MBL1055	Motif 12 T>A Reverse	Tcgag ATA AAA AAT CTT TAA TTT TT c
CYC990	External Check Primer	tcaaaggtgtgttcttcgtc
Pth2 Codon Changes		
MBL1072	Pth2 5' [BamHI/NheI]	gttaca GGATCC GCTAGC ATG TTA ATC ATC TCA GAA TTC AAT CCT C
MBL1075	Pth2 CUG#1 5'	GGT TTG AAG CGG ATT TTG GTC TA
MBL1074	Pth2 CUG#1 3'	TAG ACC AAA ATC CGC TTC AAA CC
MBL1073	Pth2 3' [HindIII/XhoI]	Ctcgga AAGCTT CTCGAG CTA TTT TAA CCA GGT CTT ATT CAA TGC A

Supplemental Table S4: List of primers used in this study.

Name in Text	Name	Description	Reference
Activation Assays			
No Protein	413TEF	pTEF CEN His3 overexpression construction	Reference 1
Wor1	pCC35	pTEF Wor1 CEN His3	This Study
Wor1 1-321	pMBL236	pTEF Wor1 1-321aa CEN His3	This Study
No Motif	BES146	CYC1ΔUAS:LacZ with no insert	Reference 1
Putative Motif	pMBL228	CYC1ΔUAS:LacZ with 20bp motif fragment	This Study
Motif KO	pMBL230	CYC1ΔUAS:LacZ with 20bp motif fragment KO, 4 mutations	This Study
7 T>A	pMBL231	CYC1ΔUAS:LacZ with 20bp motif fragment 7 T>A	This Study
12 T>A	pMBL233	CYC1ΔUAS:LacZ with 20bp motif fragment 12 T>A	This Study
Expression Plasmids			
MBP Wor1 1-321	bRZ94	Modified pET28 with 6xHis-MBP-Wor1 1-321	This Study
MBP Wor1 1-101	bRZ95	Modified pET28 with 6xHis-MBP-Wor1 1-101	This Study
MBP Wor1 196-321	pMBL222	Modified pET28 with 6xHis-MBP-Wor1 196-321	This Study
6xHis Wor1 1-321	pMBL217	pET28b with 6xHis-Wor1 1-321	This Study
6xHis Wor1 1-101	pMBL218	pET28b with 6xHis-Wor1 1-101	This Study
6xHis Wor1 196-321	pMBL234	pET28b with 6xHis-Wor1 196-321	This Study
6xHis Wor1 Full	pMBL223	pET28b with full length Wor1	This Study
6xHis Pth2	pMBL245	pET28b with full length Pth2	This Study

References

1. Mumberg D, Müller R, Funk M. *Gene* (1995) 156: 119-122.

2. Rupp S, Summers E, Lo H, Madhani H, Fink G. *EMBO J* (1999) 18: 1257-1269.

Supplemental Table S5: List of plasmids used in this study.

Supplemental Table S6: List of strains used in this study.

Description	Strain	Genotype
Sigma 2000 Starting Background	MY1404	MATa, *ura3*, *trp1*::hisG, *his3*::hisG
ΔYEL007 ΔYHR177	CC150	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx
Wor1 + Putative Motif	MLY780	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF-Wor1::CEN:His3, CYC1ΔUAS 20bp Motif:LacZ::2μ:URA3
Wor1 + No Motif	MLY768	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF-Wor1::CEN:His3, CYC1ΔUAS:LacZ::2μ:URA3
No Protein + Putative Motif	MLY773	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF::CEN:His3, CYC1ΔUAS 20bp Motif:LacZ::2μ:URA3
No Protein + No Motif	MLY766	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF::CEN:His3, CYC1ΔUAS :LacZ::2μ:URA3
Wor1 + 7 T>A	MLY783	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF-Wor1::CEN:His3, CYC1ΔUAS 20bp Motif 7T>A:LacZ::2μ:URA3
Wor1 + 12 T>A	MLY785	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF-Wor1::CEN:His3, CYC1ΔUAS 20bp Motif 12T>A:LacZ::2μ:URA3
Wor1 + Motif Knock Out	MLY782	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF-Wor1::CEN:His3, CYC1ΔUAS 20bp Motif 4mutations:LacZ::2μ:URA3
No Protein + 7 T>A	MLY776	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF::CEN:His3, CYC1ΔUAS 20bp Motif 7T>A:LacZ::2μ:URA3
No Protein + 12 T>A	MLY778	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF::CEN:His3, CYC1ΔUAS 20bp Motif 12T>A:LacZ::2μ:URA3

Description	Strain	Genotype
No Protein + Motif Knock Out	MLY775	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF::CEN:His3, CYC1ΔUAS 20bp Motif 4mutations:LacZ::2μ:URA3
Wor1 1-321 + Putative Motif	MLY789	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF-Wor1 1-321aa::CEN:His3, CYC1ΔUAS 20bp Motif:LacZ::2μ:URA3
Wor1 1-321 + No Motif	MLY791	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF-Wor1 1-321aa::CEN:His3, CYC1ΔUAS :LacZ::2μ:URA3
Wor1 1-321 + Motif Knock Out	MLY790	MATa, *ura3*, *trp1*::hisG, *his3*::hisG, *yel007*::LeuGFP, *yhr177*::KanMx, pTEF-Wor1 1-321aa::CEN:His3, CYC1ΔUAS 20bp Motif 4mutations:LacZ::2μ:URA3

Supplemental Table S6: List of strains used in this study.

Chapter 4:

Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of *C. albicans*.

Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of *C. albicans*.

Matthew B. Lohse[1,2] and Alexander D. Johnson[1,2*]

[1]Departments of Biochemistry and Biophysics and [2]Microbiology and Immunology, University of California San Francisco, San Francisco, California, United States of America

Abstract:

The human pathogen *Candida albicans* undergoes a well-defined switch between two distinct cell types, named "white" and "opaque." White and opaque cells differ in metabolic preferences, mating behaviors, cellular morphologies, and host interactions. Each cell type is stable through many generations; switching between them is rare, stochastic, and occurs without any known changes in the primary sequence of the genome; thus the switch is epigenetic. The white-opaque switch is regulated by a transcriptional circuit, composed of four regulators arranged in a series of interlocking feedback loops. To understand how switching occurs, we investigated the order of regulatory changes that occur during the switch from the opaque to the white cell type. Surprisingly, changes in key transcriptional regulators occur gradually, extending over several cell divisions with little cell-to-cell variation. Additional experiments, including perturbations to regulator concentrations, refine the signature of the commitment point. Transcriptome analysis reveals that opaque cells begin to globally resemble white cells well before they irreversibly commit to switching. We propose that these characteristics of the switching process permit *C. albicans* to "test the waters" before making an all-or-none decision.

Introduction

The architecture of transcriptional circuits often determines how organisms orchestrate developmental programs and how they respond to environmental cues. For example, circuit characteristics such as cooperativity and feedback can determine whether responses are graded or bistable and whether a transcriptional pattern can be directly inherited by descendent cells. Understanding the behavior of complex transcriptional circuits has important implications for many areas of biology, including the differentiation of stem cells into adult tissues, the response of cells to environmental perturbations, and the ability of cells to "remember" their cell type through repeated cell divisions.

A well defined system for examining how transcriptional patterns can be inherited is found in the human commensal yeast *Candida albicans.* Although a normal resident of the gastrointestinal tract of healthy humans, *C. albicans* is also the most prevalent fungal pathogen in humans, causing a variety of infections, particularly in immunocompromised individuals. *C. albicans* can switch between two genetically identical but phenotypically distinct types of cells, each of which is inherited through many generations (1-8). These two cell types, referred to as white and opaque, are distinguished by the differential regulation of approximately one-tenth of the genome (9, 10) resulting in distinct cellular and colony morphologies (1), metabolic preferences (9), mating behaviors (11), and interactions with the host immune system (12-15). Each cell type is stable through many generations, with switching between the two cell types estimated to occur every 10^4 generations (16). Thus the switch is epigenetic in the classical sense of the word: it produces a heritable change of phenotype without a change in the nucleotide sequence of

the genome. Although switching is stochastic, environmental cues can affect the frequency of switching events in one direction or the other. These cues include temperature (16), oxidative stress (17), anaerobic conditions (18, 19), and carbon dioxide (20).

White-opaque switching offers many experimental advantages for studying the inheritance of transcriptional patterns. First, it takes place on well defined laboratory medium and requires no input from other cells or tissue. Second, each cell type is stable through many generations, so pure cultures of each cell type can easily be obtained. Third, the switching is reversible. Finally, *C. albicans* can be easily manipulated genetically (for example, genes can be easily deleted, over-expressed, or tagged), and many of the key regulators of switching have been identified. White-opaque switching is controlled by a core circuit of four transcriptional regulators arranged in multiple interlocking feedback loops, shown in Figure 1 (21-25). It has been hypothesized that this transcriptional network determines many of the characteristics of switching: according to the model, the circuit is largely inactive in the white state and this constitutes the default state (Figure 1a). Switching from the white to the opaque state is postulated to occur when the circuit becomes excited, and the series of positive feedback loops ensures that the circuit remains active (Figure 1b). According to the hypothesis, molecules of the regulators are passed on following cell division; the concentrations of these regulators in the progeny cells would then be sufficiently high to keep the circuit active, and the progeny cells would thereby stay opaque for many generations. Switching from the opaque to the white state would occur when the activity of the circuit decays,

perhaps due to a stochastic decrease of one or more of the key regulators below a critical threshold level.

Circuits composed of a series of transcriptional regulators arranged in feedback loops are common in biology. For example, eye development in flies ((26), reviewed in (27)) and muscle development in mammals ((28) reviewed in (29)) are specified by such circuits. Although the circuit summarized in Figure 1 can account, at least in broad outline, for the stability of the white and opaque states, it does not describe how a switching event occurs. This is an important general question because the characteristics of switching determine the stability of the two states; for example, if the two states are heritable through many generations (as they are in *C. albicans*), there must be a barrier to frequent switching. In *C. albicans*, it is known that concentrations of the four principle regulators of Figure 1 must change during a switching event: *EFG1* (orf19.610) transcript levels are higher in white cells than in opaque cells and *WOR1* (orf19.4884), *WOR2* (orf19.5992), and *CZF1* (orf19.3127) transcript levels are higher in opaque cells than white cells (9, 10); however, we do not know the order of these changes or the concentration of the regulators at the commitment point, the point at which a switching event becomes irreversible. In addition, on a single cell basis, it is not known whether white-opaque switching is a gradual process-- in which different sets of genes are activated sequentially-- or whether it is an abrupt, irreversible phenomenon. Finally, hundreds of genes are differentially expressed between white and opaque cells, and we do not know the order of changes in expression of these genes, or how they correspond to changes in the key regulators depicted in Figure 1.

Using a combination of fluorescence microscopy, western blotting, RT-qPCR, and microarray analysis, we describe the order of events during a switching event and develop a model of switching, one that has implications for other switching systems and for *C. albicans*' ability to survive in a human host.

Results

Synchronized opaque-to-white switching

To effectively study the opaque-to-white switch using biochemical approaches, a majority of cells must be made to undergo this switch at approximately the same time. We chose to examine the opaque-to-white switch using a physiological cue: a temperature shift from 25°C to 37°C. Although several other experimental manipulations can cause an entire population of cells to switch (ectopic expression of a key transcriptional regulator, for example), we chose temperature for several reasons. The effect of temperature has been well characterized (16) and it likely represents a physiological signal, one encountered in *C. albicans'* warm-blooded animal hosts**.** Although the average internal body temperature of *C. albicans*' human host is 37°C, *C. albicans* experiences a range of temperatures in different niches; for example, skin surfaces are significantly cooler. Rikkerink *et al*. showed that the temperature induced switch was not instantaneous: following a shift from 20°C to 34°C of a suspended culture of opaque cells, approximately 6 hours passed before a significant number of opaque cells had committed to the white cell type; before that time if the temperature was lowered, the cells could be "rescued" and would remain opaque. The commitment point was determined experimentally by plating cells from the 34°C culture and incubating the plates at 25°C, the temperature where opaque cells are stable. Approximately 6 hours

after the temperature was raised, roughly half the cells (all of which started as opaque) had committed to the white state; that is, they formed white colonies even if the temperature was lowered back to 25°C. At later times, higher fractions of white colonies were observed and by approximately 28 hours the transition was complete with all plated cells giving rise to white colonies (16). We will use the term "50% commitment point" to refer to the time by which half the opaque cells in a culture had committed to the white state following the temperature shift.

We compared the concentrations of the four regulators of the white-opaque regulatory circuit shown in Figure 1 to the commitment point. We also followed mRNA levels across the entire genome over the time course of switching. Briefly, a population of pure opaque cells was grown in suspension at 25°C, shifted to 37°C, and then plated at 25°C at various time points to monitor the percent commitment at each time point. Aliquots of these samples were harvested for protein and mRNA measurements. In the experiments reported here, we observed significant numbers of white colonies by 4.5 hours. By 6 hours the majority of cells plated gave rise to white cells, and by 7.5 hours virtually all colonies were white (Figure 2a). The 50% commitment point is therefore 4.5 to 5 hours for the conditions used in this work. Similar results were observed in multiple independent experiments (see error bars in Figure 2a).

Graded-- not switch-like-- changes in regulatory proteins and transcripts

To follow levels of the regulatory proteins of Figure 1 on a population level over the time course of the opaque-to-white switch, cells were harvested at different time points after the temperature shift, lysed, and analyzed by western blotting (Figure 2b and Supplemental Figure S1). We first consider the activators of the opaque state (Wor1,

Wor2, Czf1), which are expressed at higher levels in opaque cells than white cells. Levels of these proteins showed clear decreases within 90 minutes of the 25°C to 37°C temperature shift. By 4.5 hours-- the 50% commitment point-- levels of Wor1 had decreased to approximately 20% of its starting concentration and those of Wor2 and Czf1 to approximately 40% of their starting levels. At later time points, well past the commitment point, the levels of these proteins had further decreased to approximately the same levels observed in stable white cells. Efg1, the white-enriched regulator, showed a nearly reciprocal but delayed response. Its levels had increased less than 2-fold by the 4.5 hour commitment point, with significant increases in Efg1 levels observed only when the majority of cells were past the commitment point.

Western blots of cell cultures give the average characteristics of the population but do not reveal the behavior of single cells. To observe switching at the single-cell level, we tagged each of the four regulatory proteins with GFP and observed individual cells undergoing switching by fluorescence microscopy. *C. albicans* is a diploid, so both copies of a given regulator were tagged in order to ensure that the protein being followed was actively participating in the switch. Cells containing the tagged constructs exhibited switching behavior comparable to wild-type strains, although the tagged Wor1 resulted in a slight delay, relative to the untagged control strain, in reaching the commitment point (Figure 2c). Fluorescence levels were quantitated using a custom MATLAB script. Due to low fluorescence levels, it was not possible to reliably track Czf1 levels. For the other three proteins, we observed gradual, uniform changes in the levels in individual cells rather than all-or-none behaviors (Figures 2d and 2e, Supplemental Figure S2). In other words, the cells showed intermediate levels of the regulatory proteins across the time

course as opposed to mixed populations of low and high expressors. Overall, the trends were similar to those observed in the bulk culture by western blotting. These results clearly show that individual cells undergo a gradual and uniform transition from the opaque to the white state. As noted above, in the Wor1-tagged strain the commitment point was slightly delayed (from 4.5 to 6 hours); we do not believe this delay complicates our interpretation because the results are very similar to those observed by western blotting of an untagged strain.

White and opaque cells exhibit numerous observable differences; many of these are attributable to the large number of genes (approximately 400) that are differentially regulated at least two-fold between the two states (9, 10). To examine the temporal order of the large change in gene expression that occurs during the opaque-to-white switch, we isolated mRNA at different time points following the 25°C to 37°C temperature shift of a culture of opaque cells. In the same experiment, aliquots of cells were also analyzed for the commitment point as described above. Four independent experiments were performed to ensure the reproducibility and statistical significance of the results. As controls for this experiment, four independent time courses were performed with populations of white cells that were subjected to the same 25°C to 37°C temperature shift. Switching does not occur here, allowing us to separate changes induced solely by temperature from those resulting from the opaque to white switch.

We first consider changes in the mRNA levels of *WOR1*, *WOR2*, *CZF1*, and *EFG1* (Figure 3a). For *WOR1* and *CZF1*, the mRNA and protein levels declined simultaneously. For *WOR2*, a lag in the mRNA decrease relative to the protein decrease was observed. For *EFG1*, the mRNA levels increased before the protein levels, as might

be expected for an increase in expression levels. To confirm these observations, we monitored mRNA levels of *WOR1*, *WOR2*, and *EFG1* in opaque cells by RT-qPCR (Figure 3b). Although the magnitude in changes observed using RT-qPCR data was greater (likely due to the more limited dynamic range of microarray analysis), the trends remained the same. Thus, both the mRNA and protein measurements indicate that changes to the core regulatory circuit during the opaque to white switch are relatively slow and gradual, occurring over a period of several hours.

Transcriptome reprogramming

We now turn to the behavior of the remainder of the *C. albicans* transcriptome during the temperature-induced opaque-to-white switch. We first note that a few transcripts annotated as part of the heat shock or stress responses changed in both white and opaque cells after the temperature shift, but most did not. The lack of general heat shock and stress responses probably reflects the use of 37°C, a modest change in temperature. We also noted additional expression changes that occurred in both the white and opaque populations, likely involved in a general response to a subtle temperature change. The data for these genes are available in supplemental materials (Supplemental Table S1), and we will not discuss them further as they are not specific to the switching event. It is also worth noting that the 25°C to 37°C temperature shift can trigger hyphal growth (30), this response can explain some of the gene expression changes observed in both white and opaque cells.

Based on the four matched time courses, we developed a high-confidence set of 64 white- and 145 opaque-enriched genes that we followed during the switch from opaque to white (Figure 4; Supplemental Table S2). Our criteria for inclusion in this set

was a greater than two-fold mRNA difference between white and opaque cells at 25°C (the T=0 time point) in at least 3 of the 4 replicate time courses, with the differences all in the same direction. The smaller number of white and opaque enriched genes relative to previous sets (9, 10) largely reflects the increased stringency of our criteria. For example, many additional white and opaque-enriched genes consistently differed by less than two-fold in expression; because it is difficult to meaningfully track such subtle changes across a time course we excluded them from our analysis, focusing instead on higher-magnitude changes. Significance Analysis of Microarrays for this gene set is discussed in the supplemental materials.

The opaque-enriched genes in our analysis cluster into distinct groups based on the timing of their expression change during the opaque-to-white switch (Figure 4a). For example, a large group of 63 opaque-enriched genes were fully down-regulated within the first 90 minutes following the temperature shift, well before the 50% commitment point was reached. These mRNAs remained at this level for the remainder of the time course (Figure 4a; Set1, Supplemental Table S3). A second group of 29 genes was down-regulated in a similarly rapid manner but, subsequently, these genes were slightly up-regulated (Figure 4a; Set2, Supplemental Table S3). These two sets of genes comprise roughly 65% of the opaque-enriched gene set; thus opaque cells transition to a "white-like" expression state well before the 50% commitment point (4.5 hours in these experiments) is reached. Significantly, the mRNAs of the key opaque regulators *WOR1*, *WOR2*, and *CZF1* decrease hours later than these 63- and 29-gene sets. Other classes of opaque-enriched genes are discussed in the Supplemental Materials and Supplemental Table S4, as is a comparison of these genes.

White-enriched genes also fall into several temporal clusters (Figure 4b) which, in many respects, show reciprocal behavior to the opaque clusters. Mirroring the rapid down-regulation of many opaque-enriched genes, at least a quarter of the white enriched genes ('early genes') approach their final levels well before the commitment point is reached (Figure 4b; Set4). A second cluster of genes ("middle" genes), including *EFG1*, change more slowly but are also up-regulated before the cells reach the 50% commitment point (Figure 4b; Set5). Finally, a third group of 11 genes ("late" genes) is not up-regulated until after the commitment point is reached (Figure 4b; Set6). A summary of the behavior of each set of genes is included in Figure 4c.

In summary, the behavior of the transcriptome during the temperature-induced switch from opaque to white reveals that large changes occur well before the cells reach the commitment point. Not only do the cells lose much of their opaque signature prior to commitment, but they also up-regulate expression of many white-enriched genes. Levels of the key transcriptional regulators themselves change only after this initial transformation has taken place, thus lagging behind most of the transcriptome changes.

Perturbing the commitment point

Thus far we have described, during the course of switching, the changes in levels of the four known switching regulators (Wor1, Wor2, Czf1, Efg1) and the behavior of the transcriptome relative to the 50% commitment point-- the time at which 50% of the cells have passed the "point of no return." We now consider the molecular events that coincide with commitment. Although post-translational modifications could play important roles in the switch (31), the simplest model holds that the commitment point is caused by changes in Wor1, Wor2, Czf1, and/or Efg1 levels. To examine this possibility,

we perturbed levels of each regulator and monitored the effect on the commitment point. Due to the extensive feedback loops in the switching circuit (see Figure 1), it is not possible to drastically change the level of one regulator without changing the others as well. We therefore turned to more subtle changes; more specifically, we lowered the gene dosage of each regulator by knocking out one of the two gene copies. The commitment point in these heterozygous strains was then determined experimentally.

The largest effect was observed with the *WOR1*/*wor1* heterozygote: the 50% commitment point was reached more quickly than with the control strain (less than 3 hours for the heterozygote versus 5 hours for the wild type) (Figure 5a). Heterozygous strains constructed for the other regulators showed less pronounced effects: removing one copy of *CZF1* had very little effect on the commitment point while removing one copy of *WOR2* decreased the time to commitment by about one hour (Figure 5a). Removal of one copy of *EFG1* delayed the commitment point by about one hour, from 5 to 6 hours (Figure 5a). It is worth noting that each of the heterozygous strains had altered frequencies of white-to-opaque switching (data not shown), suggesting that small changes in regulator levels affect stochastic switching as well as temperature induced switching.

We further investigated the one hour delay in the commitment point in the *EFG1/efg1* strain by tracking the levels of Wor1 in this strain. As shown in Figure 5b, the time required for Wor1 to decline by 80% is increased by roughly one hour in the *EFG1*/*efg1* strain. Deleting both copies of *EFG1* resulted in even slower decreases in Wor1, whose levels dropped by less than 50% after 9 hours (Supplemental Figure S3a). We also performed the complementary experiment, tracking Efg1 levels in a *WOR1*/*wor1*

heterozygote. Although deleting one copy of *WOR1* advanced the commitment point by two hours, we observed only a small increase (<10%) in Efg1 at the commitment point (Supplemental figure S3b). We note that this increase in Efg1 levels is much less than that observed in a wild type cell at commitment. Thus, the commitment point shows a higher correlation with Wor1 levels than with Efg1 levels. To determine the effect of deleting one copy of Wor1 on its absolute levels, we used normalized western blotting and flow cytometery. We found that the level of Wor1 in the *WOR1*/*wor1* opaque cells was roughly 70-80% that of the *WOR1*/*WOR1* control strain (data not shown). We also note that Wor1 levels had dropped by more than 80% in less than 3 hours in this strain (Figure 5b).

These data, taken together, are consistent with the decline in Wor1 levels triggering commitment, although we cannot rule out more complicated models. We note that the commitment point in these different strains corresponds to an 80% decline in Wor1 levels.

Transcriptional Signature of Commitment

What are the characteristics of the commitment point on a genome-wide scale? To answer this question, we compared the transcriptional profiles of a wild-type strain and a *WOR1/wor1* strain at the commitment point. As described above, the *WOR1/wor1* strain reaches commitment more than 2 hours earlier than the wild type strain, so, in principle, this experiment distinguishes commitment *per se* from other effects (e.g. nutrient depletion, heat shock, ect.) of the experimental manipulations.

This analysis, described in greater detail in the Supplemental Text, identified a set of 36 genes whose mRNA expression levels track with the commitment point in both

strains (Supplemental Figure S4, Supplemental Table S5). Since levels of these genes change more rapidly in the *WOR1* heterozygote compared to wild type cells, these genes appear linked to the commitment process itself. Most of these genes (27/36) are white or opaque-enriched genes that appear to be structural genes, transporters, or genes of unknown function. Significantly, this gene set includes three transcriptional regulators—*WOR1*, *CZF1*, and *EFG1*—thus supporting the model that changes in concentration of these regulators are responsible for the switching commitment point. Of these three, *WOR1* levels have changed the most at commitment, consistent with the idea that a decline in this protein underlies commitment.

Discussion

The human commensal yeast *C. albicans* can switch between two distinct cell types, white and opaque. Switching between the two cell types is rare, with each cell type being stable for hundreds of generations. A transcriptional regulatory circuit responsible for the regulation of the two cell types has been recently described (Figure 1) which can account, in part, for the stability of the two cell types but does not illustrate the order of events during a switch. In this paper we address the following questions: Is the switch gradual or does it display rapid, all or none behavior? What is the order of events when cells switch from opaque to white? What features of switching stabilized the two states (i.e. what prevents rapid switching back and forth)? What triggers the switch? Do properties of the switch provide specific advantages to the organism?

Temperature induced opaque-to-white switching is a gradual and synchronous process

It has previously been established that temperature-induced opaque-to-white switching is a slow process, requiring approximately 5-6 hours for 50% of cells to reach the commitment point-- the point at which switching has become irreversible-- and approximately 12 hours for full commitment of the entire population (16). But what happens during this time? We found that the levels of the four regulatory proteins shown in Figure 1 (Wor1, Wor2, Czf1, and Efg1) change in a gradual, ordered manner during this time period (Figure 2b and Supplemental Figure S1). These gradual changes were observed in both bulk populations of cells and in individual cells. A summary of events following the temperature shift is presented in Figure 6.

We were initially surprised by the smooth nature of the opaque-to-white transition and by the lack of variation in regulatory protein levels between individual cells. Based on the bistable nature of the switch circuitry and the existence of time points where only a fraction of cells had committed to switching, we expected to observe two clearly differentiated sub-populations of cells and few transitional intermediates in a switching population. Instead, the analysis revealed that the levels of regulatory proteins changed gradually in individual cells with very few outliers, let alone more than one distinct population at any time point. It is possible that the gradual decline in the regulators prevents frequent switching and thereby contributes to the stability of the two types of cells.

Given the gradual changes observed for the regulators, how can the all-or-none nature of the commitment point on a single cell level be explained? We found that small perturbations of the levels of Wor1, Wor2, and Efg1 (but not Czf1) changed the commitment point. Because (1) perturbation of Wor1 had the largest effects, (2) it is the

regulator whose concentration changes the most between white and opaque cells, and (3) changes in its levels appear to track the best with commitment in the various strains examined, we propose that the decline in Wor1 levels is the critical event in switching. However, because the transcriptional network controlling switching is highly interconnected (that is, each regulator controls every other regulator), it is currently difficult to definitively test this proposal.

We do not know why Wor1 levels decrease slowly in response to a temperature shift, but a simple model can account for this: if the temperature shift blocks new *WOR1* synthesis, then Wor1 levels gradually decline simply because new mRNA synthesis and translation fail to keep up with degradation rates and with dilution due to cell growth and division. This idea is supported by two observations described in this work: (1) the decline in Wor1 protein and mRNA levels track together (Figures 2b and 3) and (2) the 25°C to 37°C temperature shift reduces *WOR1* mRNA levels in both white and opaque cells (Figure 3a) even though the level of expression of *WOR1* is much greater in opaque cells. As first noted by Bergen *et al.* (32) and Srikantha *et al*. (22), following the temperature shift, the cells continue to divide for at least two rounds of cell division before the commitment point is reached. Multiple rounds of cell division following the temperature shift also occurred in the experiments reported here (data not shown). Thus, reducing new *WOR1* synthesis, coupled with dilution due to cell growth and division, can easily account for the gradual decrease in *WOR1* levels following the temperature shift. In addition, Wor1 is known to activate its own transcription in opaque cells (21, 23) and so a continual decline in levels could eventually break this positive feedback loop and lead irreversibly to the white state. Regardless of its exact mechanism, the slow decline

in Wor1 documented here would create a time window of several hours during which the cells are clearly changing in response to temperature, but have not yet committed to switching. As described below, we believe this time window has important implications for the switching program.

Opaque cells phenotypically transition to "white-like" cells before the switch becomes heritable

We were surprised to find that a massive transformation of the opaque cell transcriptional program into a "white-like" program occurred shortly after the temperature shift and several hours before the commitment point was reached. Srikantha *et al*. previously reported that the transcript levels of two opaque-enriched genes (*OP4* (orf19.4934) and *SAP1* (orf19.5714)) decreased prior to the commitment point (33); our results show that this response encompasses approximately 65% of all opaque-enriched genes (which are down-regulated) and 25% of the white-enriched genes (which are up-regulated). Thus, immediately following the temperature shift, opaque cells switch their transcriptome program to that resembling white cells and, only if the temperature shift is maintained for several more hours, do they make these changes irreversible by committing to the switch.

What advantage might be conferred by changing the transcriptome in advance of the switch? One possibility, of course, is that this strategy is not adaptive at all; that is, this feature of the switch was a non-selected by-product of the evolutionary history of the switch. However, changing the transcriptome in advance of the switching event may, at least on paper, offer certain advantages. The white-opaque switch has long been hypothesized to allow *C. albicans* to better colonize a range of environments in the host;

consistent with this idea, a major class of genes that is differentially regulated between the two cell types encodes metabolic enzymes (9). Rapidly changing from an opaque transcriptional program to a white program before the switching mechanism locks in could allow opaque cells to "test the waters" before switching. For example, if a temperature shift is only temporary, the cells can quickly return to the opaque state. Such a temporary separation of the bulk of the transcriptome from its control circuit may allow opaque cells to sample a new environment, committing to switching only when a favorable match occurs between the gene expression profile and the environment. This idea can account for why, in the absence of oxygen and/or the presence of carbon dioxide, increased temperature fails to induce the opaque-to-white switch (18-20). It remains to be seen if this temporary separation of a regulatory circuit and the transcriptome is a strategy used by other microbial pathogens to respond to changes in the host environment.

Similarity to other cell differentiation programs

On the surface, white-opaque switching resembles other examples of cell differentiation where cells remember their differentiated state through many rounds of cell division without environmental cues. For example, the gradual nature of the switch and the relatively long time lag before the commitment point is reached are reminiscent of certain features of the formation of induced pluripotent stem (iPS) cells from differentiated tissues (see (34) and (35) for reviews). Both processes appear to be controlled by a series of transcriptional regulators, each of which controls the others, arranged in a complex network that includes multiple feedback loops. Moreover, each

has a relatively long time delay (up to 2 weeks for iPS cells) with the transcriptome changing in a stepwise fashion.

The slow, gradual nature of iPS cell reprogramming is often attributed to slow changes in methylation/demethylation or in chromatin structure. However, a systematic investigation of histone modifying enzymes and chromatin remodelers in *C. albicans* indicated relatively small changes in white-opaque switching rates in comparison to the effects of the transcriptional regulators (31). We have proposed that the slow decay of a key transcriptional regulator is sufficient to explain the delay and its drop below a threshold level is sufficient to explain the commitment point of the switch. Whether this type of framework applies to other cases of cell differentiation remains to be determined.

Acknowledgements

The authors thank Aaron Hernday, Rebecca Zordan, and Quinn Mitrovich for strains. The authors thank Chris Cain, Sarah Foss, Aaron Hernday, Chris Baker, Anita Sil, Hiten Madhani, and Oliver Homann for valuable comments on this manuscript. MBL was supported by a National Science Foundation Graduate Research Fellowship. Research was supported by grants from the National Institutes of Health (RO1 AI49187) and the Ellison Foundation (ID-SS-0628-04) to ADJ.

Experimental procedures

Strains and primers used in this paper are listed in Supplemental Tables S6 and S7 respectively.

Media

Cells were grown in synthetic complete media supplemented with 2% glucose and 100μg/mL uridine (SD+aa+Urd).

Plasmids

Plasmids for GFP tagging were based on the SAT1 flipper cassette (36). Briefly, the last 500bp of a gene (excluding the stop codon) was PCR amplified with a 5' SphI site. *C. albicans* optimized GFP was PCR amplified with a 3x Gly linker and a 3' XhoI site. The first 500bp immediately 3' of the open reading frame (ORF) was PCR amplified between NotI and AatII sites. These three fragments were combined in a further round of PCR, resulting in a SphI-ORF-GFP-XhoI-NotI-3' flank-AatII fragment that was digested with SphI and AatII and ligated into pUC19 (NEB). Plasmids were sequenced to ensure matching to the sequence listed at the *Candida* Genome Database (http://www.candidagenome.org/). Plasmids with correct sequences were digested with XhoI and NotI, then ligated to a similarly digested SAT1 flipper cassette from pSFS2A. Plasmids were then digested with SphI and AatII for transformation into *C. albicans.*

Strains

All strains are derivatives of SC5314. Creation of the **a**/**a** selected –Leu –His starting strain RZ47 has been described previously (23). *Candida dublieniensis LEU2* and *Candida maltosa HIS1* were added back to this strain at the *LEU2* locus. The white and opaque strains in this background were used as wild type, independent isolates of this

strain were gifts from Aaron Hernday and Quinn Mitrovich and have been described previously (37). The *EFG1, WOR1, WOR2,* and *CZF1* heterozygous deletion strains and the *efg1/efg1* deletion strain were made using the His or Leu knockout cassettes (38) in the RZ47 background and were a gift from Aaron Hernday and Rebecca Zordan. GFP tagging of Wor1, Wor2, Czf1, and Efg1 was accomplished using the plasmid designs described above. Strains were transformed with linearized plasmids and selected for growth on 200 μg/mL nourseothricin (clonNAT, WERNER BioAgents, Jena, Germany). Insertion was verified with PCR against both flanks and the marker was then flipped out by growth in Yeast Peptone Dextrose (YPD) media supplemented with 2% Maltose for 6-24 hours. Cells were then plated on YPD plus 25 μg/mL nourseothricin to identify small colonies which had lost the nourseothricin resistance marker. Marker excision was then verified by PCR. The second copy of the gene was then tagged by repeating this process, performing additional PCR checks to ensure that untagged copies of the gene did not remain in the cell.

Temperature Shift Assay

Room temperature overnight cultures were started from single colonies with no sectors on SD+aa+Urd plates. Cultures were diluted back to OD_{600}=0.02 in the morning and allowed to grow back to OD_{600}=0.1-0.2 at room temperature. Cells were then diluted (OD_{600}=0.00012 for opaques and OD_{600}=0.00002 for whites) in 500mL SD+aa+Urd in 2.8L flasks and grown at 25°C overnight. When cultures reached OD_{600}=0.1, 0hr samples were removed and the remaining culture was transferred to pre-warmed 37°C flasks and swirled in a 37-40°C water bath for 5 minutes to ensure the temperature quickly reached 37°C. Cultures were then placed in a 37°C shaking incubator.

Modifications of this assay for single cell experiments (Figure 2c-e, Supplemental Figure S2) and some heterozygous strain assays (Figure 5a) are described later. Modifications of this assay for the commitment microarray analysis are described in the Supplemental Text.

Samples were then harvested at 1.5, 3, 4.5, 6. 7.5, and 9 hour time points. Samples for microarray analysis and qPCR were harvested by centrifugation and frozen in liquid nitrogen. Samples for western blotting were centrifuged, resuspended in 1mL PBS, transferred to 2mL tubes, pelleted, decanted, and frozen in liquid nitrogen. Samples for microscopy were centrifuged, washed 2x with PBS, and placed on slides for analysis. At each time point, cells were also plated on room temperature SD+aa+Urd plates which were scored for colony phenotype 5-7 days later to determine percent switching. Between 100 and 500 cells were plated at each time point, generally 100 cells were plated on each of two plates.

Quantitative RT-PCR

Sample preparation, RT-qPCR, and data analysis were performed as previously described (23). Normalization was versus *PAT1* (orf19.3792), as described in that reference.

Microarrays

Total RNA was extracted from samples via buffered phenol extractions. Following cDNA synthesis, samples were dye coupled to both Cy3 and Cy5 in separate reactions. A pooled reference was created from the Cy5 labeled fractions. Equal amounts of Cy3 labeled sample and Cy5 labeled reference were hybridized overnight to custom designed Agilent 8*15k Microarrays (AMADID #020166, Array Express accession number A-MEXP-1805) containing a minimum of 2 independent probes for

each ORF. Following hybridization, slides were processed and then scanned on a Genepix 4000A or 4000B Axon Instrument Scanner (Axon/Molecular Devices, Sunnyvale, CA USA). Arrays were gridded using GenePix Pro version 5.1. Global Lowess normalization analysis was performed for each array using a Goulphar script (39) for R (The R Foundation for Statistical Computing). Normalized data was then transformed before collapsing all points for an ORF to the median value using a custom java script written by Oliver Homann. Temperature shifted opaque samples were transformed versus a pre-shift opaque sample to determine how genes had changed relative to their starting values. Temperature shifted white cells compared to a pre-shifted white population allowed us to control for general temperature responses as well as entry into stationary phase at late time points. Transformations of an opaque sample at a time point versus a white sample at the same time point were also performed. Data shown in Figures 3a and 4 represent the median values from the four experiments. Transformations used for commitment characterization are discussed in the Supplemental Materials.

Microarray data were clustered using Cluster version 3.0 (40) and visualized using Java TreeView Version 1.1.3 (41). The microarray data has been uploaded to Array Express (http://www.ebi.ac.uk/microarray-as/ae/), the accession number for the time course experiments is E-MEXP-2705 and the accession number for the commitment analysis experiments is E-MEXP-2834.

Western Blotting

Extracts were prepared in Urea Lysis buffer (42). 10µL of each sample was run on an SDS-PAGE gel and then analyzed by Western Blotting. Blots were analyzed with rabbit

affinity-purified antibody generated against synthetic peptides unique to each protein from the various genes (Bethyl Laboratories Inc., Montgomery TX). Wor1 antibodies have been previously described (23). Wor2 antibodes were raised against the sequences CSAVINRVSVADLLK and CYSPNSPYSLPTRPSN. Efg1 antibodies were raised against the sequences CQANQSASTVAKEEK and CYGQYNAPGKNQNTPA. Czf1 antibodies were raised against the sequences CKVLRGIVEYRSK and CLPSNVSPPNSRAVPT. To reduce non-specific binding, the Wor2, Efg1, and Czf1 antibodies were additionaly purified as follows. Lysates of *C. albicans* strains deleted for the respective genes were bound to activated Sepharose 4B. A mixture of both antibodies for each protein were then incubated with the resin for 12-18 hours before elution. Elution fractions with high protein concentrations were pooled for further use, creating a mixture of both antibodies against a given protein. Rat α-tubulin (no. ab1616; Abcam, Cambridge, MA) was used as a loading control. Secondary antibodies were goat α-rabbit IrDye800 (611-132-122, Rockland Immunochemicals Inc., Gilbertsville, PA) and goat α-rat AlexaFluor680 (A-21096, Invitrogen, Carlsbad, CA). Blots were scanned on a Oddessey Imaging System scanner (LI-COR Biotechnology, Lincoln, NE) using both the 700 and 800 channels. Protein levels in each lane were first normalized versus the tubulin loading control, then normalized to the maximum value (set to 1) for each protein. Data shown for the four proteins in Figure 2b and Supplemental Figure S1 represent the mean of four (Wor1 and Efg1) or three (Wor2 and Czf1) independent experiments and the error bars in Supplemental Figure S1 reflect the standard error of the mean.

Single Cell Fluorescence and Data Analysis

The temperature shift assay was modified as follows for the single cell experiments. GFP strains were grown overnight at room temperature, diluted to OD_{600}=0.2, and allowed to grow at room temperature. At OD_{600}=0.6, samples were harvested, an equal amount of pre-warmed 37°C SD+aa+Urd was added, and samples were swirled in a 37°C water bath for 5 minutes. At each time point, samples were washed twice with PBS and imaged on a Zeiss Axiovert 200M microscope (Carl Zeiss, Oberkochen, Germany). Constant GFP exposure times were used for each strain during an assay. The Rhodamine channel was used to detect autofluoresence. Images were analyzed using a custom MATLAB (The MathWorks Inc., Natick, MA) script which determined mean GFP fluorescence levels for each cell. Experiments with the heterozygous strains shown in Figure 5a also used this general temperature shift assay approach, although without the microscopy.

Supplemental Materials

Significance Analysis of Microarrays, White and Opaque Enriched Gene Set

As a comparison to the stringent but not statistically based criteria used to develop the enriched gene list, we performed SAM (version 3.09) (43) on the white and opaque microarray data from the T=0hr time point. Data were analyzed using the "Two Class Unpaired" format with a Δ=1.054, which produced an estimated false discovery rate of 0.35%. These criteria produced a list of 260 genes (172 opaque enriched and 88 white enriched). All but 3 of the 209 genes (98.6%) from our enriched gene set were included on this list. Of the 54 genes not on our list, 28 were opaque enriched and 26 were white enriched. A majority of the genes discovered by SAM but not by our criteria appear to have approximately a two-fold median and average change where only two of the four individual changes were at least two fold. Based on this observation and the near total

inclusion of our gene set within the SAM set, we decided to use our more stringent enriched gene criteria in all subsequent analyses.

Regulation of additional groups of opaque-enriched genes

Numerous types of genes (Supplemental Table S3) are represented in early down-regulated opaque-enriched genes (Figure 4a Sets 1 and 2), including those coding for cell wall proteins, various transporters, metabolic genes, transcriptional regulators, as well as a large number of proteins of unknown function. The down-regulation of a third group of 11 opaque-enriched genes, including both *WOR1* and *CZF1*, is significantly delayed with 10 out of 11 reaching 50% of their starting values by the 4.5 hour time point (Figure 4a; Set3; Supplemental Table S4). The remaining 42 opaque enriched genes exhibit a variety of behaviors; levels of *WOR2* change only at the 9 hour time point and several other genes do not change significantly over the time course, suggesting they decrease well after the switching has been completed. The most significant conclusions from these data are that changes in the regulatory proteins lag well behind the bulk of the opaque-to-white expression changes.

Cluster Design for Figure 4

Data presented in Figure 4 was analyzed and clustered as follows. Following transformation, data were collapsed down to 1 point for each ORF, and the ORF list culled down to the 209 opaque and white enriched genes as previously described. Data were clustered using Cluster 3.0 (40), using the "Hierarchical Complete Linkage" setting, with "Genes" set to "Cluster" and "Arrays" set to "Calculate Weights". "Cutoff" was set at 0.1 and "Exponent" at 1. Both similarity metrics were set as "Correlation (uncentered)". Weighting was set to 1 for all the individual transformations in the

"Opaque versus Opaque" and "Opaque versus White" classes, and to 0 for all in the "White versus White" class. Although medians were calculated for each specific time point within a class [e.g. the four Opaque 1.5 hours versus Opaque 0 hour transformations], these data were weighted to 0 for clustering purposes. Gene groups 1 and 3 and groups 4 and 5 were flipped for easier visualization in Java TreeView 1.1.3 (41). The data presented in Figure 4 are median values of the four replicates for each specific time point.

Commitment Profile Microarray Setup

The 50% commitment point microarray analysis was performed as described in materials and methods with the following changes. White and opaque samples were harvested for both the wild type and *WOR1/wor1* strains prior to the temperature shift. Samples were harvested at the 50% commitment point of the wild type (5hr, 10min) and *WOR1/wor1* (2hr, 45min) strains. Only wild type samples were harvested at the 5hr 10min time point. Three independent experiments were performed for the opaque *WOR1/wor1* strain, all other strains had two independent experiments. Samples were processed as described in methods and materials, using the same microarray design (AMADID #020166). Post-temperature shift opaque samples were transformed against the pre-shift samples to determine how the genes had changed relative to their starting values (Supplemental Figure 4b, Transformations $T_{3\text{-}1HET}$, $T_{2\text{-}1WT}$, and $T_{4\text{-}1WT}$). The *WOR1/wor1* 2hr 45min samples were transformed against the wild type 2hr 45min and 5hr 10min samples to identify genes whose levels differed at these points (Supplemental Figure 4b, Transformations $T_{3\text{-}2}$ and $T_{3\text{-}4}$). The *WOR1/wor1* 2hr 45min versus 0hr 00min transformations were transformed against the wild type 2hr 45min versus 0hr 00min

transformations to identify groups of genes whose levels differ between the two strains while accounting for differences in enrichment between the two strains (Supplemental Figure 4b, Transformation $T_{(T_{3\text{-}1HET})\text{-}(T_{2\text{-}1WT})}$).

Commitment Profile Microarray Analysis

We wanted to examine changes in the transcriptome occurring at the 50% commitment points for both the wild type and *WOR1/wor1* strains in order to distinguish between changes related to commitment and experimental artifacts, thus identifying transcriptome changes reflective of commitment. Two different sets of criteria were used to identify the transcriptional commitment signature (genes that are changing at the 50% commitment point for the *WOR1/wor1* strain but not at that time point for the wild type strain). First, we required at least a 1.41-fold change (log2=0.5) in three replicates when transforming the *WOR1/wor1* 2hr 45min samples against the wild type 2hr 45min samples (Transformation $T_{3\text{-}2}$). This requirement focuses on genes whose levels are changing more rapidly in the *WOR1/wor1* strain than in the wild type, the ones most likely to be tracking with the change in time to commitment. Second, we required at least a 1.41-fold difference (log2=0.5) in at least two of three replicates when transforming the *WOR1/wor1* 2hr 45min versus 0hr 00min transformation against the wild type 2hr 45min versus 0hr 00min transformations (Transformation $T_{(T_{3\text{-}1HET})\text{-}(T_{2\text{-}1WT})}$). This requirement accounts for changes in levels of white- and opaque-enrichment in the *WOR1/wor1* strain.

Applying these criteria created a list of 36 genes (18 down-regulated more in the *WOR1/wor1* strain, 18 up-regulated more), of which 27 were on the list of white- and opaque-enriched genes from Supplemental Table S2 (14 opaque-enriched, 13 white-

enriched) (Supplemental Figure S4c). These genes include *WOR1*, *CZF1*, and *EFG1*; the remainder of these genes are listed in Supplemental Table S5 (several of the 9 genes that are not on the enriched gene list are enriched to some extent but did not meet our stringent criteria). The remaining genes include a variety of transporters, GPI-anchored proteins, the *WH11* (orf19.3548.1)-*HSP12* (orf19.3160)-orf19.4216 group of genes, and a large number of genes with no known function. This list of 36 genes includes many members of gene sets 3 (*WOR1*/*CZF1* cluster, 9/11), 5 ("middle" white-enriched genes, 5/11), and 6 ("late" white-enriched genes, 7/10) from Figures 4a and 4b. Transforming the *WOR1/wor1* 2hr 45min sample versus the wild type 5hr 10min sample shows that the levels of a majority of these genes are similar at the 50% commitment points in the two strains.

References

1. Slutsky B, et al. (1987) "White-opaque transition": a second high-frequency switching system in *Candida albicans*. *J Bacteriol* 169:189-197.
2. Soll DR, Morrow B, Srikantha T (1993) High-frequency phenotypic switching in *Candida albicans*. *Trends Genet* 9:61-65.
3. Bennett RJ, Uhl MA, Miller MG, Johnson AD (2003) Identification and characterization of a *Candida albicans* mating pheromone. *Mol Cell Biol* 23:8189-8201.
4. Johnson A (2003) The biology of mating in *Candida albicans*. *Nat Rev Microbiol* 1:106-116.
5. Lockhart SR, Daniels KJ, Zhao R, Wessels D, Soll DR (2003) Cell biology of mating in *Candida albicans*. *Eukaryot Cell* 2:49-61.
6. Lohse MB, Johnson AD (2009) White-opaque switching in *Candida albicans*. *Curr Opin Microbiol* 12:650-654.
7. Morschhäuser J (2010) Regulation of white-opaque switching in Candida albicans. *Med Microbiol Immunol* 199:165-172.
8. Soll DR (2009) Why does *Candida albicans* switch? *FEMS Yeast Res* 9:973-989.
9. Lan CY, et al. (2002) Metabolic specialization associated with phenotypic switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 99:14907-14912.
10. Tsong AE, Miller MG, Raisner RM, Johnson AD (2003) Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115:389-399.

11. Miller MG, Johnson AD (2002) White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* 110:293-302.

12. Kvaal CA, Srikantha T, Soll DR (1997) Misexpression of the white-phase-specific gene WH11 in the opaque phase of *Candida albicans* affects switching and virulence. *Infect Immun* 65:4468-4475.

13. Kvaal C, et al. (1999) Misexpression of the opaque-phase-specific gene PEP1 (SAP1) in the white phase of *Candida albicans* confers increased virulence in a mouse model of cutaneous infection. *Infect Immun* 67:6652-6662.

14. Geiger J, Wessels D, Lockhart SR, Soll DR (2004) Release of a potent polymorphonuclear leukocyte chemoattractant is regulated by white-opaque switching in *Candida albicans*. *Infect Immun* 72:667-677.

15. Lohse MB, Johnson AD (2008) Differential phagocytosis of white versus opaque *Candida albicans* by *Drosophila* and mouse phagocytes. *PLoS One* 3:e1473.

16. Rikkerink EH, Magee BB, Magee PT (1988) Opaque-white phenotype transition: a programmed morphological transition in *Candida albicans*. *J Bacteriol* 170:895-899.

17. Kolotila MP, Diamond RD (1990) Effects of neutrophils and in vitro oxidants on survival and phenotypic switching of *Candida albicans* WO-1. *Infect Immun* 58:1174-1179.

18. Dumitru R, et al. (2007) In vivo and in vitro anaerobic mating in *Candida albicans*. *Eukaryot Cell* 6:465-472.

19. Ramírez-Zavala B, Reuss O, Park YN, Ohlsen K, Morschhäuser J (2008) Environmental induction of white-opaque switching in *Candida albicans*. *PLoS Pathog* 4:e1000089.

20. Huang G, Srikantha T, Sahni N, Yi S, Soll DR (2009) CO(2) regulates white-to-opaque switching in *Candida albicans*. *Curr Biol* 19:330-334.

21. Huang G, et al. (2006) Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 103:12813-12818.

22. Srikantha T, et al. (2006) TOS9 regulates white-opaque switching in *Candida albicans*. *Eukaryot Cell* 5:1674-1687.

23. Zordan RE, Galgoczy DJ, Johnson AD (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci U S A* 103:12807-12812.

24. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* 5:e256.

25. Vinces MD, Kumamoto CA (2007) The morphogenetic regulator Czf1p is a DNA-binding protein that regulates white opaque switching in *Candida albicans*. *Microbiology* 153:2877-2884.

26. Czerny T, et al. (1999) twin of eyeless, a second Pax-6 gene of *Drosophila*, acts upstream of eyeless in the control of eye development. *Mol Cell* 3:297-307.

27. Silver SJ, Rebay I (2005) Signaling circuitries in development: Insights from the retinal determination gene network. *Development* 132:3-13.

28. Molkentin JD, Olson EN (1996) Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. *Proc Natl Acad Sci U S A* 93:9366-9373.

29. Tapscott SJ (2005) The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* 132:2685-2695.

30. Sudbery P, Gow N, Berman J (2004) The distinct morphogenic states of *Candida albicans*. *Trends Microbiol* 12:317-324.

31. Hnisz D, Schwarzmüller T, Kuchler K (2009) Transcriptional loops meet chromatin: a dual-layer network controls white-opaque switching in *Candida albicans*. *Mol Microbiol* 74:1-15.

32. Bergen MS, Voss E, Soll DR (1990) Switching at the cellular level in the white-opaque transition of *Candida albicans*. *J Gen Microbiol* 136:1925-1936.

33. Srikantha T, Tsai LK, Daniels K, Soll DR (2000) EFG1 null mutants of *Candida albicans* switch but cannot express the complete phenotype of white-phase budding cells. *J Bacteriol* 182:1580-1591.

34. Scheper W, Copray S (2009) The Molecular Mechanism of Induced Pluripotency: A Two-Stage Switch. *Stem Cell Rev* 5:204-223.

35. Hochedlinger K, Plath K (2009) Epigenetic reprogramming and induced pluripotency. *Development* 136:509-523.

36. Reuss O, Vik A, Kolter R, Morschhäuser J (2004) The SAT1 flipper, an optimized tool for gene disruption in *Candida albicans*. *Gene* 341:119-127.

37. Mitrovich QM, Tuch BB, Guthrie C, Johnson AD (2007) Computational and experimental approaches double the number of known introns in the pathogenic yeast *Candida albicans*. *Genome Res* 17:492-502.

38. Noble SM, Johnson AD (2005) Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. *Eukaryot Cell* 4:298-309.

39. Lemoine S, Combes F, Servant N, Le Crom S (2006) Goulphar: rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics* 7:467.

40. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20:1453-1454.

41. Saldanha AJ (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics* 20:3246-3248.

42. Ubersax JA, et al. (2003) Targets of the cyclin-dependent kinase Cdk1. *Nature* 425:859-864.

43. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response *Proc Natl Acad Sci U S A* 98:5115-5121.

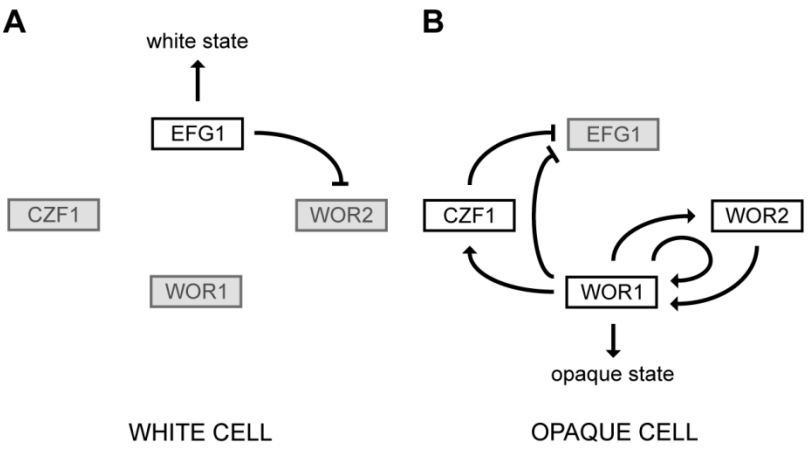


Figure 1: Working model of the white-opaque regulatory circuit and its activity in white and opaque cells. (a) In white cells, *EFG1* represses *WOR1* indirectly through *WOR2* to maintain white cell identity. (b) In opaque cells, *WOR1*, *WOR2*, and *CZF1* establish a series of positive feedback loops, maintaining opaque cell identity and repressing *EFG1*. Up-regulated genes and active relationships are indicated in black. Down-regulated genes are indicated in gray. Arrows and bars represent activation and repression respectively. Figure adapted from Zordan *et al*., 2007 (24).

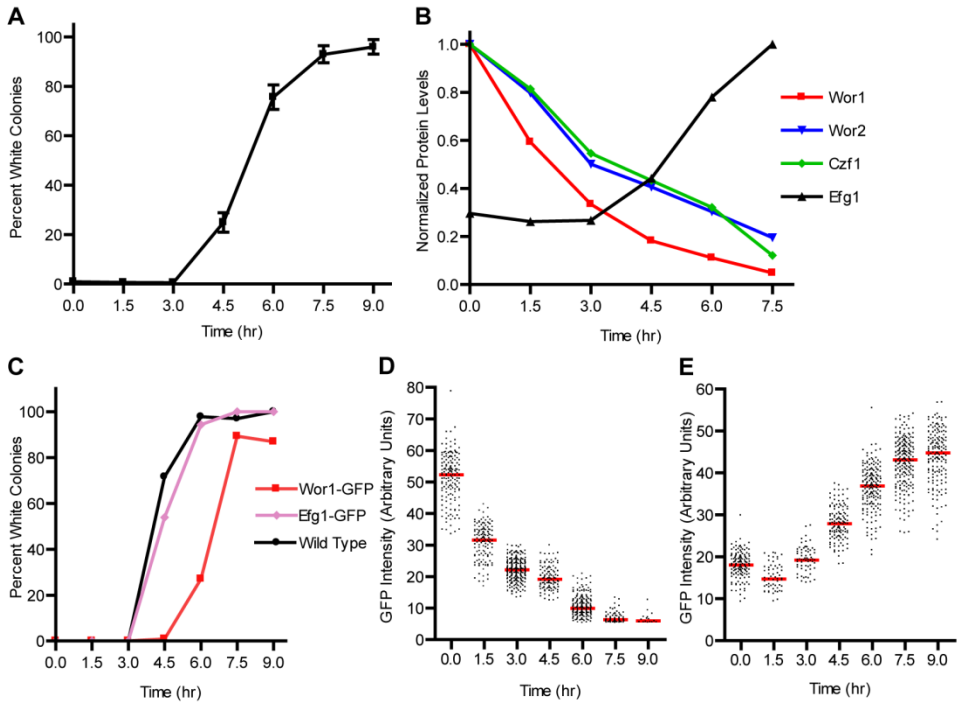


Figure 2: Changes in protein levels of Wor1, Wor2, Czf1, and Efg1 during the opaque to white switch. (a) Commitment point. Opaque cells in suspended culture were shifted from 25°C to 37°C at T=0 hr. At each time point, an aliquot of cells was collected and plated at room temperature to determine the percentage of cells that had switched to the white cell type. Data points reflect the mean of six experiments (except for the 9 hour time point which reflects 5) and error bars represent the standard error of the mean. (b) Levels of the four regulatory proteins as determined by western blotting. As described in the methods, protein levels were normalized to the maximum levels for each regulator. For Wor1, Wor2, and Czf1 the maximum level occurs in opaque cells, while Efg1 is maximally expressed in white cells. Data points reflect the mean of four (Wor1 and Efg1) or three (Wor2 and Czf1) experiments; error bars for each protein level are shown in Supplemental Figure S1. (c) Commitment point for the experiments in (d) and (e); the percentage of colonies with the white colony phenotype is plotted for wild type, Wor1-GFP, and Efg1-GFP containing strains. (d) Single cell measurements of Wor1. Wor1-

GFP levels in individual cells were quantitated as described in the methods. The red line represents the median value at each time point and each dot represents a single cell. (e) Single cell measurements of Efg1; symbols are as described in (d).

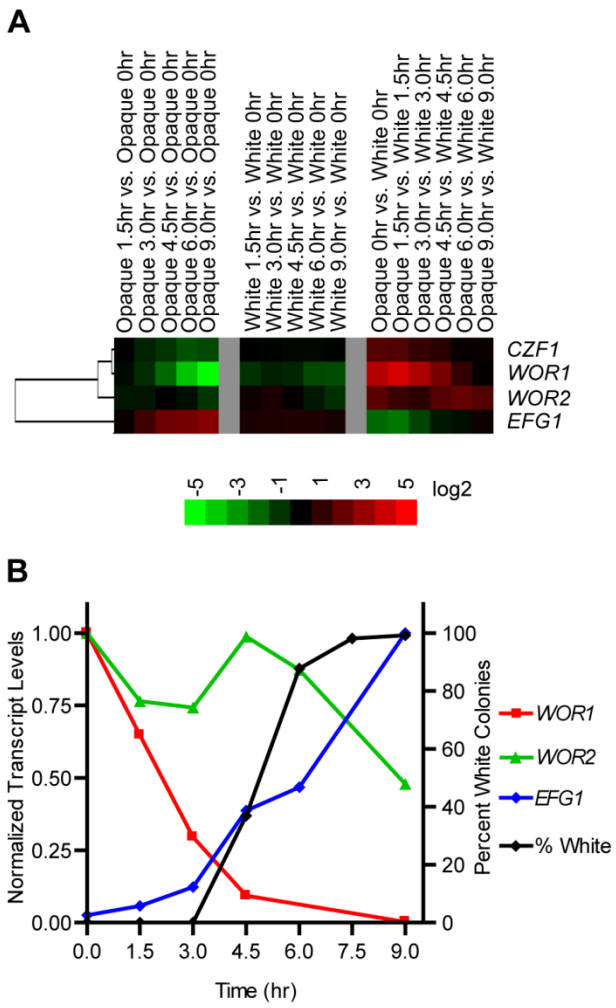


Figure 3: Transcript levels of *WOR1*, *WOR2*, *CZF1*, and *EFG1*. (a) Microarray transcript data, with the median of four independent experiments shown. The first block represents changes in opaque cells during the time course, normalized to the starting (T=0 hr) opaque cells. The second block highlights the changes due to temperature alone by comparing a time course performed on white cells (which do not switch) to the starting (T=0 hr) culture of white cells. The last block shows the profile of the opaque cell cultures at each time point, compared to the white cell cultures at the same time point. The color key is based on a log2 scale. (b) Levels of *WOR1*, *WOR2*, and *EFG1* transcripts as determined by RT-qPCR (left axis). The levels of each transcript are

normalized to their maximum levels in opaque (*WOR1*, *WOR2*) or white (*EFG1*) cells. The right axis displays the percentage of colonies with the white colony phenotype for the same experiment.

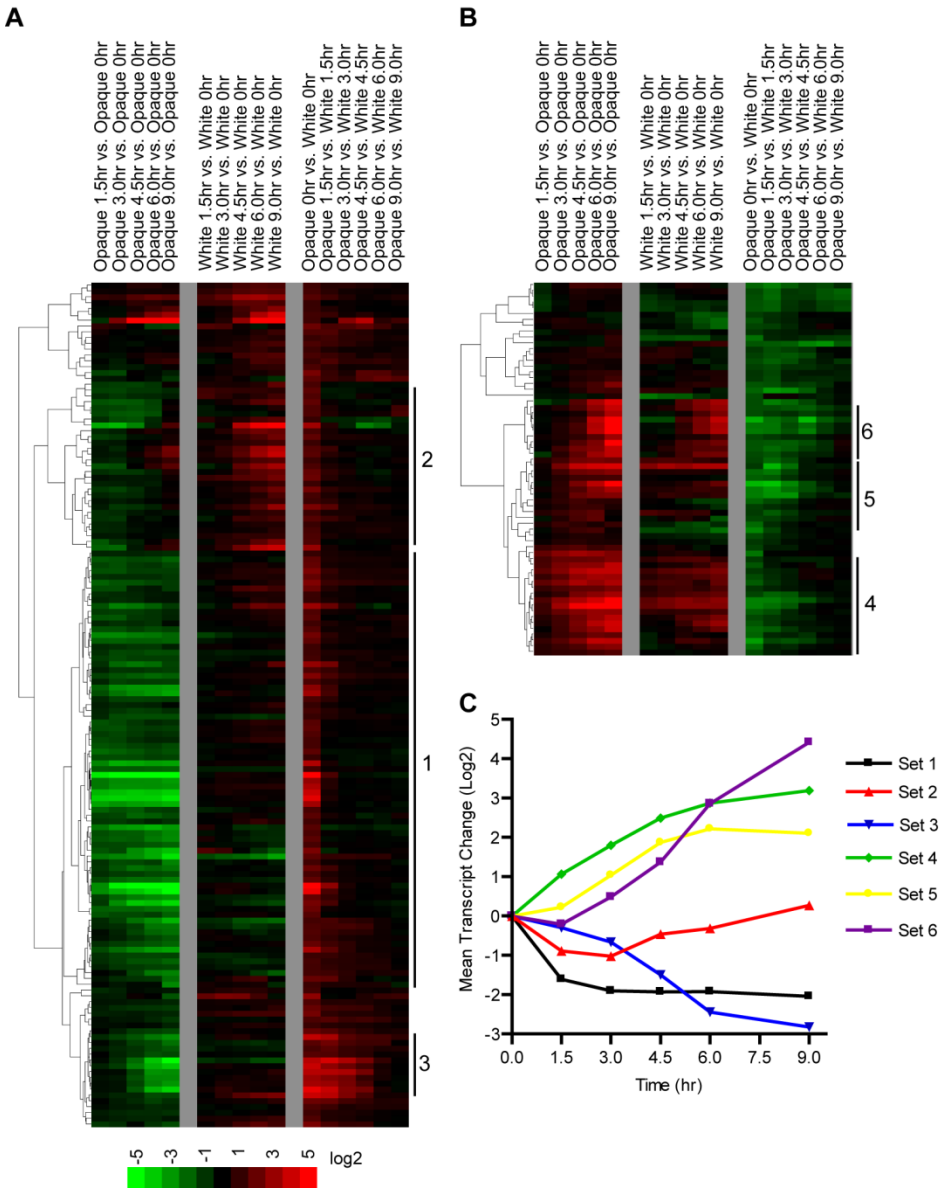


Figure 4: Behavior of 209 white- and opaque-enriched genes over the time course of switching. The three column blocks are described in Figure 3a, with the median value from the four experiments shown for each gene. (a) Changes in the 145 opaque-enriched genes during the time course. (b) Changes in the 64 white-enriched genes during the time course. The color key for (a) and (b) is based on a log2 scale. (c) The six gene sets were binned and the mean transcriptional change at each time point for each set was determined in opaque cells. Sets of genes (1, 2, 3, etc.) are described in more detail in the

text. Sets 1 and 2 are the opaque-enriched genes that are rapidly down-regulated upon the temperature shift. Set 3 is the 11 gene *WOR1*/*CZF1* cluster of opaque-enriched genes. Set 4 contains the white-enriched "early" genes. Set 5 is white-enriched "middle" genes, and Set 6 is the white-enriched "late" gene group. Data are from the same set of experiments shown Figure 3a.

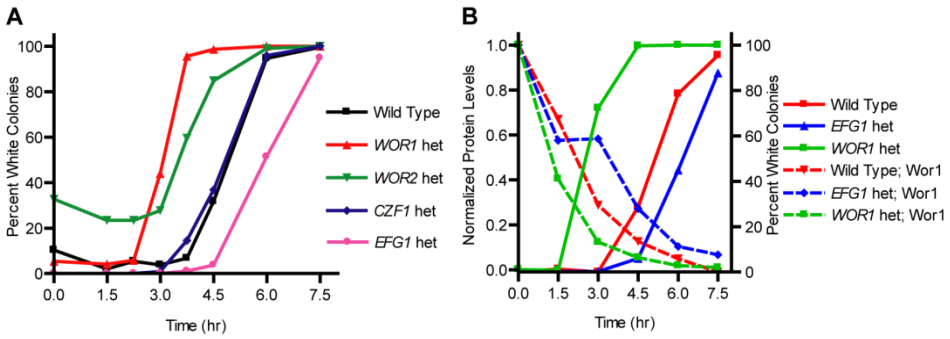


Figure 5: Effects of deleting one copy of *WOR1*, *WOR2*, *CZF1*, or *EFG1* on the commitment point. (a) Temperature-induced switching from opaque to white cell types for the wild type strain and for heterozygous deletions of the four known regulators. The percentage of colonies with the white colony phenotype is plotted for each strain. The 50% commitment points for these strains are 5 hours (wild type and *CZF1* het), 3 hours (*WOR1* het), and 6 hours (*EFG1* het). As the *WOR2* heterozygous strain had a consistent background of 20-30% white cells, we have defined the commitment point for this strain as the point at which half the opaque colonies, or 60-65% of the total colonies, are white (approximately 4 hours). All temperature shift experiments in (a) were performed in parallel on the same day. (b) Wor1 protein levels in wild type, *WOR1* heterozygous, and *EFG1* heterozygous strains. Western blotting data (left axis, dashed lines) and percent commitment (right axis, solid lines) are plotted for Wor1 in wild type (red), *WOR1* heterozygous (green), and *EFG1* heterozygous (blue) strains. Data points represent the mean of two experiments.

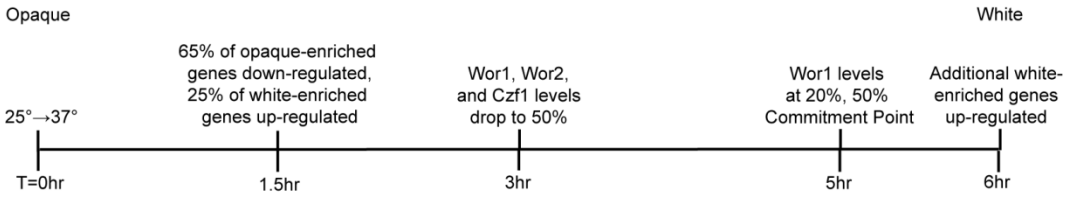
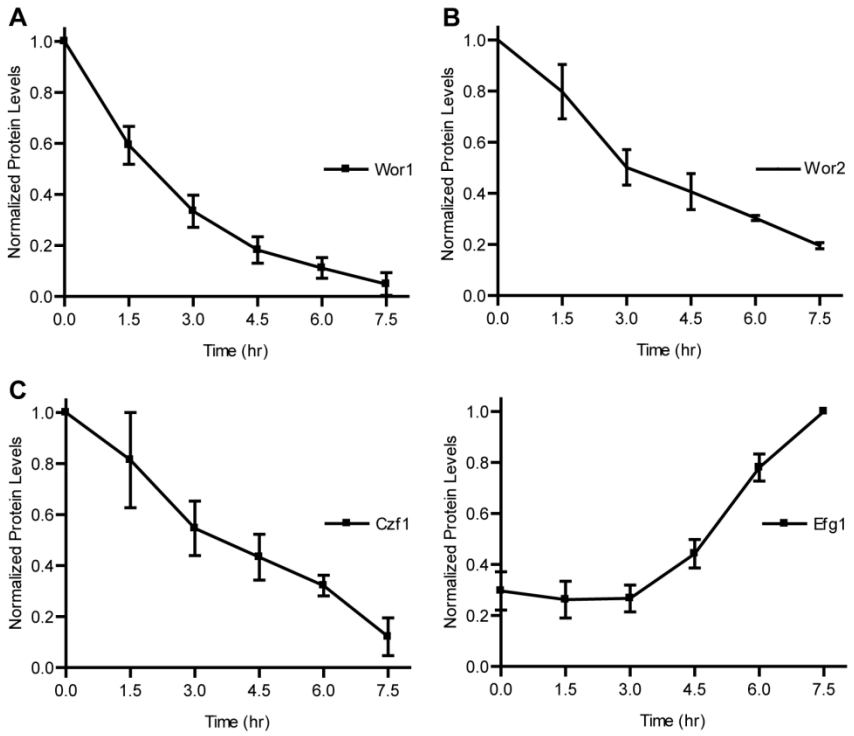
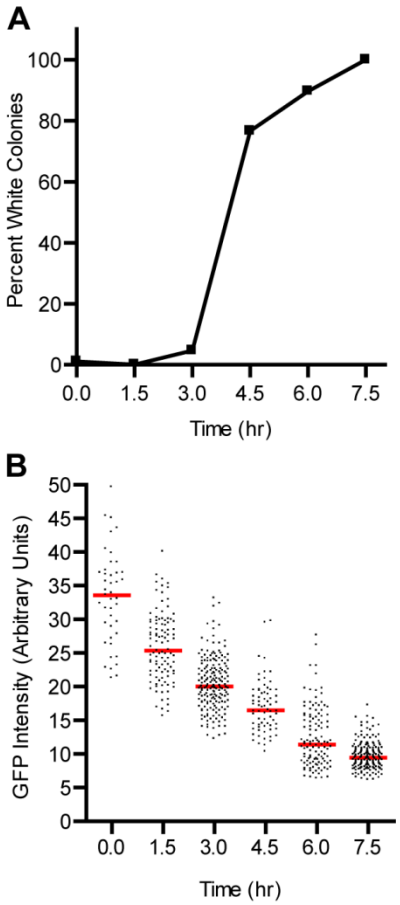


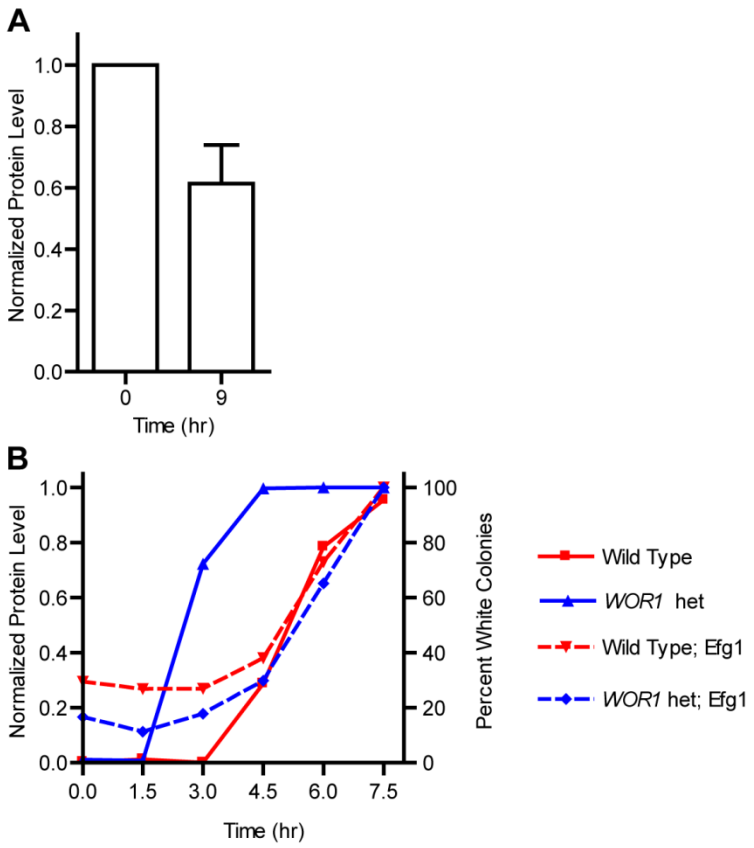
Figure 6: Order and timing of events following temperature shift of an opaque population.



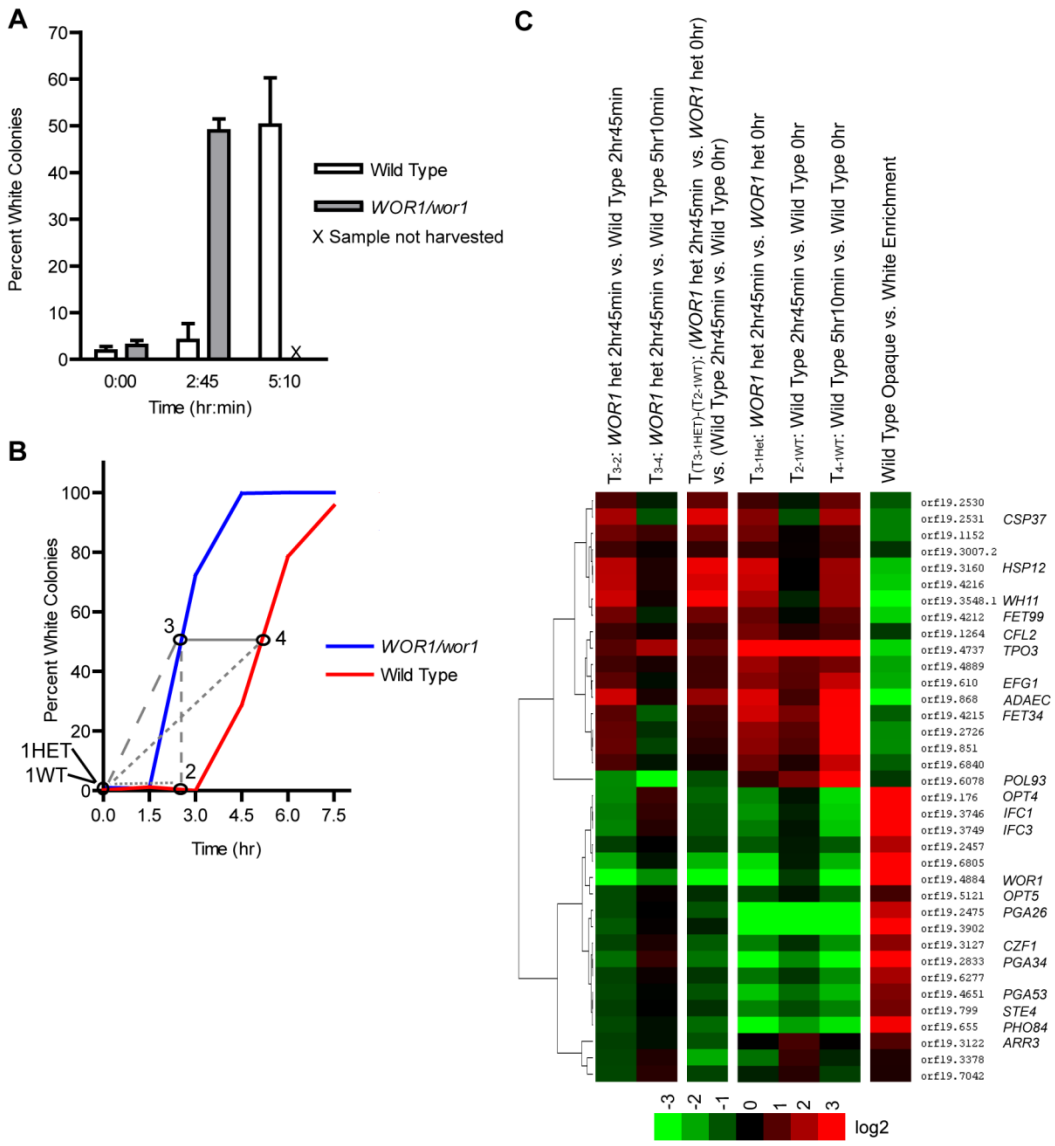
Supplemental Figure S1: Changes in Wor1, Wor2, Czf1, and Efg1 protein levels during the opaque to white switch. (a) Wor1. (b) Wor2. (c) Czf1. (d) Efg1. Western blotting data for the four regulatory proteins. As described in the methods, protein levels were normalized to the maximum levels for each regulator. Data points reflect the average of four (Wor1 and Efg1) or three (Wor2 and Czf1) experiments and are the same as in Figure 2b, error bars represent the standard error of the mean.



Supplemental Figure S2: Single cell measurements of Wor2 during the opaque-to-white shift. (a) Commitment point for the experiment in (b); the percentage of colonies with the white phenotype is plotted for the Wor2-GFP strain. (b) Wor2-GFP levels in individual cells were quantitated as described in the methods. The red line represents the median value at each time point. Each dot represents a single cell.



Supplemental Figure S3: Effects of perturbations of *WOR1* and *EFG1* levels on other members of the white-opaque regulatory circuit. (a) Changes in Wor1 levels following a temperature shift in an opaque *efg1/efg1* strain as determined by western blotting. As described in the methods, Wor1 levels were normalized to their levels prior to the temperature shift. Data points reflect the average of two experiments and error bars represent the standard error of the mean. (b) Efg1 protein levels in wild type and *WOR1* heterozygous strains. Western blotting data (left axis, dashed lines) and percent commitment (right axis, solid lines) for Efg1 in wild type (red) and *WOR1* heterozygous (green) strains. As described in the methods, Efg1 levels were normalized to the levels at the 7.5hr timepoint. Data points reflect the mean of two experiments.



Supplemental Figure S4: Identification of 36 genes closely associated with the commitment point. (a) Temperature-induced switching from opaque to white cell types for wild type (clear) and *WOR1/wor1* (grey) strains. Opaque cells in suspended culture were shifted from 25° to 37° at T=0 hr. At each time point, an aliquot of cells was collected and plated at room temperature to determine the percentage of cells that had switched cell types. The percentage of colonies with the white colony phenotype is plotted. Data points reflect the mean of three experiments for the *WOR1/wor1* strain and two experiments for the wild type strain. Error bars represent the standard error of the

mean. Samples were not collected for the *WOR1/wor1* strain at the 5hr 10min time point. (b) Schematic of the microarray sample transformations used to identify the commitment-associated group of genes based on analysis of temperature shifted wild type (red) and *WOR1/wor1* (blue) opaque populations. Grey lines illustrate several common transformations used. For example, transformation of the *WOR1/wor1* 2hr 45min sample (3) against the wild type 2hr 45min sample (2) identifies genes with different levels in the two strains at the *WOR1/wor1* strain commitment point; the short hand representation of this transformation is $T_{3\text{-}2}$. Transformations are described in more detail in the Supplemental Text. Commitment point data in this panel were taken from representative temperature shift assays for the two strains. (c) Microarray transcript data, with the median of two ($T_{2\text{-}1WT}$, and $T_{4\text{-}1WT}$) or three ($T_{3\text{-}1HET}$, $T_{3\text{-}2}$, $T_{3\text{-}4}$, and $T_{(T3\text{-}1HET)}$-($T_{2\text{-}1WT}$)) transformations shown. The first block represents the difference between *WOR1/wor1* samples and wild type samples. The second block represents differences in transcript levels at the *WOR1/wor1* commitment point while accounting for differing enrichment levels between the strains. The third block represents changes in opaque cells during the time course, normalized to the starting (T=0 hr) opaque cells for the relevant strain. The final block shows the profile of wild type opaque cell cultures at the starting point, compared to the equivalent white cell cultures, illustrating the level of white- or opaque-enrichment of these genes. Data in this final block were taken from the time course temperature shift assay shown in Figures 3a and 4, the median of the four independent experiments is shown. The color key uses a log2 scale.

Supplemental Table S1: List of genes changing in both white and opaque cells at 1.5 or 3 hours.

ORF Number	Name	Median Opaque 1.5hr vs. 0hr, log2	Median White 1.5hr vs. 0hr, log2	Present In 3 Hour Set?
1.5 Hour Set				
orf19.4477	*CSH1*	2.13875	1.9185	yes
orf19.7654	*CPR6*	1.84275	1.74975	
orf19.1847	*ARO10*	1.7315	1.30375	yes
orf19.5811	*MET1*	1.2745	1.85	yes
orf19.473	*TPO4*	1.2245	1.45	yes
orf19.1925	*TLO5*	1.1975	1.27825	
orf19.5645	*MET15*	1.191	1.43275	yes
orf19.6191	*TLO8*	1.1905	1.16125	
orf19.6515	*HSP90*	1.165	1.288	
orf19.424	*TRP99*	-1.19225	-1.2685	yes
orf19.7238	*NPL3*	-1.22475	-1.528	yes
orf19.169	*CHO2*	-1.4575	-1.20875	yes
orf19.3895	*CHT2*	-1.50025	-1.75575	yes
orf19.756	*SAP7*	-1.57075	-1.27075	yes
orf19.2669	-	-2.9355	-3.13	yes
orf19.2668	*RHD2*	-3.23475	-3.3405	yes
orf19.7585	*INO1*	-4.60625	-5.14425	
3 Hour Set				
orf19.3646	*CTR1*	2.4575	1.22825	
orf19.473	*TPO4*	2.4345	2.455	
orf19.4477	*CSH1*	2.42425	1.162	
orf19.5811	*MET1*	2.11425	2.614	
orf19.3433	*OYE23*	1.90075	1.20475	
orf19.5842	-	1.8525	2.43425	
orf19.909	*STP4*	1.7935	1.9875	
orf19.4941	*TYE7*	1.785	1.7195	
orf19.1847	*ARO10*	1.72	1.09925	
orf19.6079	-	1.69575	1.47025	
orf19.3794	*CSR1*	1.69	1.2245	
orf19.409	-	1.679	2.393	
orf19.689	*PLB1*	1.575	1.4315	
orf19.4076	*MET10*	1.5575	1.3445	
orf19.7313	*SSU1*	1.5415	1.90175	
orf19.7278	-	1.43375	2.79075	

ORF Number	Name	Median Opaque 1.5hr vs. 0hr, log2	Median White 1.5hr vs. 0hr, log2	Present In 3 Hour Set?
orf19.1868	*RNR22*	1.42775	1.831	
orf19.4099	*ECM17*	1.422	1.15675	
orf19.5180	*PRX1*	1.387	1.178	
orf19.5645	*MET15*	1.334	1.438	
orf19.431	*ZCF2*	1.3085	1.278	
orf19.1113	-	1.29325	1.1955	
orf19.822	-	1.238	1.65275	
orf19.3139	-	1.20075	1.396	
orf19.5785	-	1.14825	1.39175	
orf19.5610	*ARG3*	1.14425	1.523	
orf19.5753	*HGT10*	-1.03675	-2.47675	
orf19.7238	*NPL3*	-1.06475	-1.269	
orf19.424	*TRP99*	-1.234	-1.224	
orf19.588	-	-1.29825	-1.3305	
orf19.6996	-	-1.33525	-1.756	
orf19.169	*CHO2*	-1.43825	-1.068	
orf19.756	*SAP7*	-1.56875	-1.2055	
orf19.3895	*CHT2*	-2.186	-1.84025	
orf19.2669	-	-2.695	-2.878	
orf19.2668	*RHD2*	-3.53325	-3.4745	

Supplemental Table S1: List of genes changing in both white and opaque cells at 1.5 or 3 hours. Median change (log2 scale) from four replicates listed for each gene. To be included in this set, transcript levels must exhibit at least a 2 fold change in 3 of 4 opaque and 3 of 4 white arrays, have changed in the same direction in both cases, and NOT be white or opaque enriched. Transformations are 1.5 or 3 hours versus 0 hour, so positive values represent an increase in transcript levels

Supplemental Table S2: List of 209 white- and opaque-enriched genes.

ORF number	Name	Median Wild Type Enrichment, Opaque Versus White, log2	Median *WOR1/wor1* Enrichment, Opaque Versus White, log2	*WOR1/wor1* versus Wild Type Enrichment, log2
orf19.4934	*OP4*	6.0225	4.899	-1.1235
orf19.3902	-	5.9725	5.106	-0.8665
orf19.2460	-	4.6505	4.067	-0.5835
orf19.176	*OPT4*	4.595	4.654	0.059
orf19.3668	*HGT2*	4.484	3.479	-1.005
orf19.4279	*MNN1*	3.952	5.094	1.142
orf19.4527	*HGT1*	3.9485	2.749	-1.1995
orf19.2833	*PGA34*	3.929	4.939	1.01
orf19.4972	-	3.7935	2.885	-0.9085
orf19.6805	-	3.588	3.798	0.21
orf19.4884	*WOR1*	3.5645	4.929	1.3645
orf19.893	*PGA17*	3.4495	3.155	-0.2945
orf19.4072	*IFF6*	3.373	3.9765	0.6035
orf19.2253	-	3.2515	2.128	-1.1235
orf19.3749	*IFC3*	3.2275	3.575	0.3475
orf19.4082	*DDR48*	3.169	2.9055	-0.2635
orf19.1370	-	3.158	2.752	-0.406
orf19.1541	-	3.141	3.569	0.428
orf19.3282	*BMT3*	3.05725	3.3055	0.24825
orf19.3746	*IFC1*	3.03175	3.269	0.23725
orf19.7053	*GAC1*	2.9245	2.361	-0.5635
orf19.655	*PHO84*	2.901	3.318	0.417
orf19.5144	*PGA28*	2.897	1.198	-1.699
orf19.6983	-	2.7885	1.9175	-0.871
orf19.5307	*JEN2*	2.697	1.522	-1.175
orf19.467	-	2.525	2.999	0.474
orf19.1709	-	2.51725	1.712	-0.80525
orf19.742	*ALD6*	2.426	2.321	-0.105
orf19.7434	*GLG2*	2.4065	1.48	-0.9265
orf19.3707	*YHB1*	2.40425	1.915	-0.48925
orf19.2062	*SOD4*	2.386	2.542	0.156
orf19.4505	*ADH3*	2.3465	1.476	-0.8705
orf19.2475	*PGA26*	2.319	2.095	-0.224
orf19.4066	-	2.26875	1.37	-0.89875
orf19.3612	*PST2*	2.2485	2.225	-0.0235

ORF number	Name	Median Wild Type Enrichment, Opaque Versus White, log2	Median *WOR1/wor1* Enrichment, Opaque Versus White, log2	*WOR1/wor1* versus Wild Type Enrichment, log2
orf19.899	-	2.227	1.942	-0.285
orf19.2457	-	2.109	2.482	0.373
orf19.4459	-	2.093	2.08	-0.013
orf19.4654	-	2.0805	2.628	0.5475
orf19.2886	*CEK1*	2.08025	2.0765	-0.00375
orf19.3926	-	2.072	2.416	0.344
orf19.3278	*GSY1*	2.0555	1.195	-0.8605
orf19.4255	*ECM331*	2.026	2.158	0.132
orf19.3363	*VTC4*	1.995	1.8395	-0.1555
orf19.6277	-	1.974	2.721	0.747
orf19.4381	*VTC3*	1.9685	1.4885	-0.48
orf19.2030	-	1.963	1.5635	-0.3995
orf19.2247	-	1.93975	1.4675	-0.47225
orf19.4157	*SPS20*	1.93925	1.5025	-0.43675
orf19.6254	*ANT1*	1.9235	1.493	-0.4305
orf19.6993	*GAP2*	1.8885	1.447	-0.4415
orf19.1344	-	1.88525	1.047	-0.83825
orf19.532	*RBR2*	1.8675	1.911	0.0435
orf19.1288	*FOX2*	1.85775	1.203	-0.65475
orf19.5723	*POX1*	1.8165	1.174	-0.6425
orf19.6445	*ECI1*	1.81475	1.539	-0.27575
orf19.7094	*HGT12*	1.809	0.25	-1.559
orf19.5992	*WOR2*	1.8065	1.613	-0.1935
orf19.5305	*RHD3*	1.79525	3.031	1.23575
orf19.2192	*GDH2*	1.7945	2.173	0.3785
orf19.6229	*CAT1*	1.7625	0.89	-0.8725
orf19.5760	*IHD1*	1.76	1.357	-0.403
orf19.4979	*KNS1*	1.742	1.966	0.224
orf19.4749	-	1.7295	2.079	0.3495
orf19.7514	*PCK1*	1.706	0.964	-0.742
orf19.6143	-	1.7055	1.4025	-0.303
orf19.5615	*AYR2*	1.702	1.3855	-0.3165
orf19.3127	*CZF1*	1.67525	2.2445	0.56925
orf19.2125	-	1.661	1.548	-0.113
orf19.1239	-	1.6555	1.599	-0.0565
orf19.4591	*CAT2*	1.64975	0.941	-0.70875
orf19.2724	-	1.6485	1.046	-0.6025

ORF number	Name	Median Wild Type Enrichment, Opaque Versus White, log2	Median *WOR1/wor1* Enrichment, Opaque Versus White, log2	*WOR1/wor1* versus Wild Type Enrichment, log2
orf19.7513	-	1.626	1.275	-0.351
orf19.4666	-	1.59225	1.2505	-0.34175
orf19.5921	-	1.5695	1.1615	-0.408
orf19.3663	*PHO91*	1.5665	1.4695	-0.097
orf19.535	*RBR1*	1.5645	2.005	0.4405
orf19.2156	*NAG1*	1.555	1.9035	0.3485
orf19.3419	*MAE1*	1.5495	1.259	-0.2905
orf19.7521	*REP1*	1.54775	1.5965	0.04875
orf19.1830	-	1.5445	0.83	-0.7145
orf19.3338	-	1.5445	1.4615	-0.083
orf19.7512	-	1.5445	1.059	-0.4855
orf19.3337	-	1.5255	1.8035	0.278
orf19.1562	-	1.51	2.025	0.515
orf19.4651	*PGA53*	1.50825	1.934	0.42575
orf19.413	-	1.4935	1.223	-0.2705
orf19.1539	-	1.485	1.2185	-0.2665
orf19.4607	-	1.484	1.401	-0.083
orf19.6830	-	1.484	0.852	-0.632
orf19.6784	*PGA32*	1.476	1.174	-0.302
orf19.1171	-	1.475	0.496	-0.979
orf19.5140	-	1.475	1.142	-0.333
orf19.5170	*ENA21*	1.46325	1.5225	0.05925
orf19.4833	*MLS1*	1.45775	1.193	-0.26475
orf19.4539	-	1.4195	1.305	-0.1145
orf19.3924	-	1.40475	1.4815	0.07675
orf19.1187	*CPH2*	1.4035	0.925	-0.4785
orf19.3325	-	1.4015	0.987	-0.4145
orf19.2124	-	1.399	1.2185	-0.1805
orf19.2941	*SCW4*	1.39275	1.2605	-0.13225
orf19.799	*STE4*	1.373	1.95	0.577
orf19.5455	-	1.36925	1.0985	-0.27075
orf19.6066	-	1.36675	0.5935	-0.77325
orf19.3733	*IDP2*	1.35	0.8665	-0.4835
orf19.5138	*IFA21*	1.35	1.646	0.296
orf19.575	*HYR3*	1.3485	1.078	-0.2705
orf19.3330.3	*POX18*	1.3475	0.953	-0.3945
orf19.5157	-	1.34025	1.5675	0.22725

ORF number	Name	Median Wild Type Enrichment, Opaque Versus White, log2	Median *WOR1/wor1* Enrichment, Opaque Versus White, log2	*WOR1/wor1* versus Wild Type Enrichment, log2
orf19.3984	-	1.339	1.772	0.433
orf19.6570	*NUP*	1.3385	1.659	0.3205
orf19.3121	*GST1*	1.33475	1.839	0.50425
orf19.1313	*CDR3*	1.3095	1.352	0.0425
orf19.1655	*PXP2*	1.3075	0.779	-0.5285
orf19.7314	*CDG1*	1.30475	0.834	-0.47075
orf19.2701	-	1.29625	0.8045	-0.49175
orf19.5267	-	1.289	1.334	0.045
orf19.4530.1	-	1.2815	0.955	-0.3265
orf19.7111.1	*SOD3*	1.238	0.235	-1.003
orf19.5994	*RHB1*	1.23675	1.278	0.04125
orf19.164	-	1.2305	1.072	-0.1585
orf19.460	*CEK2*	1.2205	1.063	-0.1575
orf19.4590	*RFX2*	1.2185	0.528	-0.6905
orf19.3415	*PTK2*	1.2	0.8875	-0.3125
orf19.1652	*POX1-3*	1.18725	0.741	-0.44625
orf19.1600	-	1.1825	0.887	-0.2955
orf19.2758	*PGA38*	1.1765	1.3025	0.126
orf19.5563	*RNH1*	1.16525	0.9075	-0.25775
orf19.5614	-	1.1495	0.694	-0.4555
orf19.918	*CDR11*	1.148	1.039	-0.109
orf19.2841	*PGM2*	1.1445	0.751	-0.3935
orf19.7500	*PXA1*	1.141	0.757	-0.384
orf19.3931	*SFC1*	1.13375	0.48225	-0.6515
orf19.577	-	1.13275	0.669	-0.46375
orf19.1704	*FOX3*	1.1185	0.552	-0.5665
orf19.1277	-	1.1145	1.28	0.1655
orf19.4476	-	1.1025	1.2	0.0975
orf19.4090	-	1.10225	0.9805	-0.12175
orf19.7437	-	1.09025	1.1155	0.02525
orf19.3897	-	1.087	1.145	0.058
orf19.4570	-	1.087	0.889	-0.198
orf19.2018.1	-	1.08225	0.7645	-0.31775
orf19.4056	*GAT2*	1.074	0.961	-0.113
orf19.5711	-	1.0465	0.7125	-0.334
orf19.5031	*SSK1*	1.0425	0.4935	-0.549
orf19.3713	-	-1.04275	-0.6825	0.36025

ORF number	Name	Median Wild Type Enrichment, Opaque Versus White, log2	Median *WOR1/wor1* Enrichment, Opaque Versus White, log2	*WOR1/wor1* versus Wild Type Enrichment, log2
orf19.2028	*MXR1*	-1.04525	-0.6895	0.35575
orf19.2583.1	-	-1.048	-1.283	-0.235
orf19.3461	-	-1.05425	-1.087	-0.03275
orf19.7219	*FTR1*	-1.073	-1.2575	-0.1845
orf19.6192	-	-1.07675	-0.46	0.61675
orf19.3868	-	-1.093	-0.6155	0.4775
orf19.4215	*FET34*	-1.116	-1.005	0.111
orf19.3040	*EHT1*	-1.1425	-1.202	-0.0595
orf19.1442	*PLB4.5*	-1.144	-1.377	-0.233
orf19.3159	*UTP20*	-1.19025	-0.816	0.37425
orf19.3869	-	-1.236	-1.0025	0.2335
orf19.84	*CAN3*	-1.25275	-1.399	-0.14625
orf19.2583.2	-	-1.253	-1.2105	0.0425
orf19.946	*MET14*	-1.289	-1.119	0.17
orf19.6937	*PTR2*	-1.33	-1.172	0.158
orf19.2989	-	-1.3395	-1.076	0.2635
orf19.6874	-	-1.3525	-1.471	-0.1185
orf19.6770	-	-1.3555	-0.972	0.3835
orf19.4907	-	-1.4055	-0.734	0.6715
orf19.2244	-	-1.41375	-1.4165	-0.00275
CaalfMp03	*NAD1*	-1.42275	-0.569	0.85375
orf19.6311	-	-1.451	-1.0225	0.4285
orf19.2898	-	-1.45725	-0.866	0.59125
orf19.5063	-	-1.464	-1.147	0.317
orf19.5673	*OPT7*	-1.475	-1.341	0.134
orf19.1906	-	-1.48575	-1.4	0.08575
orf19.1152	-	-1.507	-1.463	0.044
orf19.2531	*CSP37*	-1.50825	-1.3965	0.11175
orf19.2881	*MNN4*	-1.516	-1.861	-0.345
orf19.4653	-	-1.53875	-1.5125	0.02625
CaalfMp13	*NAD5*	-1.539	-0.3435	1.1955
CaalfMp09	*NAD2*	-1.60975	-0.162	1.44775
orf19.4195.1	*FCA1*	-1.6215	-1.226	0.3955
CaalfMp10	*NAD3*	-1.6395	-0.0635	1.576
orf19.2726	-	-1.6535	-1.722	-0.0685
orf19.851	-	-1.6545	-1.771	-0.1165
orf19.24	*RTA2*	-1.672	-1.548	0.124

ORF number	Name	Median Wild Type Enrichment, Opaque Versus White, log2	Median *WOR1/wor1* Enrichment, Opaque Versus White, log2	*WOR1/wor1* versus Wild Type Enrichment, log2
orf19.54	*RHD1*	-1.67625	-1.6915	-0.01525
orf19.1415	*FRE10*	-1.681	-1.791	-0.11
orf19.1153	*GAD1*	-1.6995	-0.486	1.2135
orf19.5025	*MET3*	-1.8065	-1.526	0.2805
orf19.3448	-	-1.858	-1.208	0.65
orf19.609	-	-1.862	-1.633	0.229
orf19.2948	*SNO1*	-1.9385	-1.774	0.1645
CaalfMp06	*ATP6*	-1.95125	-0.413	1.53825
orf19.85	*GPX2*	-1.95675	-1.569	0.38775
orf19.4889	-	-1.958	-2.734	-0.776
orf19.610	*EFG1*	-2.05025	-1.948	0.10225
orf19.4769	*IPT1*	-2.068	-1.68	0.388
orf19.23	*RTA3*	-2.085	-1.997	0.088
orf19.2946	*HNM4*	-2.12325	-2.41975	-0.2965
orf19.4679	*AGP2*	-2.21325	-2.082	0.13125
orf19.3160	*HSP12*	-2.285	-2.574	-0.289
orf19.6837	-	-2.302	-1.2545	1.0475
orf19.4216	-	-2.4105	-2.696	-0.2855
orf19.1189	-	-2.4435	-2.283	0.1605
orf19.4212	*FET99*	-2.4745	-2.3405	0.134
orf19.4737	*TPO3*	-2.5145	-0.87	1.6445
orf19.1979	*GIT1*	-2.5225	-2.614	-0.0915
orf19.6202	*RBT4*	-2.64725	-2.3005	0.34675
orf19.2947	*SNZ1*	-3.0945	-3.237	-0.1425
orf19.3548.1	*WH11*	-3.483	-3.125	0.358
orf19.868	*ADAEC*	-3.619	-2.968	0.651

Supplemental Table S2: List of 209 white- and opaque-enriched genes and the median difference (log2 scale) in transcript levels between opaque and white cells observed for each. Values greater than zero indicate enrichment in opaque cells.

Supplemental Table S3: List of 92 opaque enriched genes down-regulated within 90 minutes of the shift to 37°C, median change plotted (log2 scale).

Group	ORF Number	Name	Median change, 1.5hr vs 0hr, opaque cells, log2	Description
Transporters	orf19.5307	*JEN2*	-1.816	Protein described as carboxylic acid transporter; lactate transport
	orf19.655	*PHO84*	-0.936	Protein similar to high-affinity phosphate transporters
	orf19.3663	*PHO91*	-1.18625	Putative low-affinity phosphate transporter
	orf19.3668	*HGT2*	-4.0705	Putative glucose transporter of the major facilitator superfamily
	orf19.4527	*HGT1*	-3.584	High-affinity glucose transporter
	orf19.6993	*GAP2*	-1.3645	Possible amino acid permease
	orf19.7094	*HGT12*	-2.097	Glucose Transporter
	orf19.4090	-	-0.74725	Predicted membrane transporter
	orf19.7500	*PXA1*	-0.913	Putative peroxisomal ABC Transporter
	orf19.918	*CDR11*	-0.648	Putative transporter
Metabolic	orf19.6445	*ECI1*	-1.68025	Protein similar to S. cerevisiae Eci1p, which is involved in fatty acid oxidation
	orf19.742	*ALD6*	-1.5585	aldehyde dehydrogenase (NAD) activity, ethanol metabolic
	orf19.3330.3	*POX18*	-1.6435	predicted preoxisomal protein
	orf19.1655	*PXP2*	-1.1495	Putative acyl-CoA oxidase; enzyme of fatty acid beta-oxidation
	orf19.5723	*POX1*	-1.511	Predicted acyl-CoA oxidase, fatty acid beta-oxidation

Group	ORF Number	Name	Median change, 1.5hr vs 0hr, opaque cells, log2	Description
	orf19.7514	*PCK1*	-2.283	Phosphoenolpyruvate carboxykinase; role in gluconeogenesis
	orf19.2192	*GDH2*	-1.243	Glutamate dehydrogenase
	orf19.3419	*MAE1*	-0.8125	Malic enzyme, gluconeogenesis
	orf19.3325	-	-0.801	Glycogen synthesis initiator
	orf19.3931	*SFC1*	-0.56	Acyltransferase
	orf19.4833	*MLS1*	-1.59125	malate synthase
	orf19.1704	*FOX3*	-0.728	Fatty acid oxidation
	orf19.7434	*GLG2*	-0.508	self-glucosylating initiator of glycogen synthesis
	orf19.1288	*FOX2*	-1.18525	3-hydroxyacyl-CoA epimerase, fatty acid beta-oxidation
	orf19.7053	*GAC1*	-1.9015	Regulatory subuit of phosphoprotein phosphatase, glycogen metabolic process
	orf19.1652	*POX1-3*	-0.91975	Predicted acyl-CoA oxidase, fatty acid beta-oxidation
	orf19.3278	*GSY1*	-1.4615	glycogen synthase
	orf19.4591	*CAT2*	-0.93	carnitine acetyl transferase, fatty acid beta-oxidation
	orf19.3282	*BMT3*	-1.57675	beta-mannosyltransferase
Cell Wall/Surface	orf19.2475	*PGA26*	-1.098	Putative GPI-anchored protein of unknown function
	orf19.5267	-	-0.967	Putative cell wall protein;
	orf19.5305	*RHD3*	-2.3305	Putative GPI-anchored protein
	orf19.575	*HYR3*	-1.4325	Putative GPI-anchored protein of unknown function
	orf19.532	*RBR2*	-1.8495	Cell wall protein
	orf19.4072	*IFF6*	-2.344	Putative GPI-anchored protein of unknown function

Group	ORF Number	Name	Median change, 1.5hr vs 0hr, opaque cells, log2	Description
	orf19.5144	*PGA28*	-2.35	Predicted Kex2p substrate; putative GPI-anchored protein
	orf19.2941	*SCW4*	-1.3675	Cell Wall Protein
	orf19.5760	*IHD1*	-1.662	Putative GPI-anchored protein of unknown function
	orf19.6784	*PGA32*	-1.51775	Putative GPI-anchored protein of unknown function
	orf19.893	*PGA17*	-2.631	Putative GPI-anchored protein of unknown function
	orf19.535	*RBR1*	-1.727	GPI anchored cell wall protein
Signaling	orf19.7521	*REP1*	-0.6265	Regulator of MDR1
	orf19.4590	*RFX2*	-0.913	Transcriptional Repressor, filamentation, DNA damage, heat response
	orf19.3415	*PTK2*	-0.786	Protein kinase, polyamine import
	orf19.1187	*CPH2*	-1.08025	Transcriptional activator of hyphal growth
	orf19.4056	*GAT2*	-0.8885	Transcription Factor
	orf19.2886	*CEK1*	-0.3055	ERK Family kinase, mating
	orf19.5031	*SSK1*	-0.58925	Regulator of two-component system
	orf19.5994	*RHB1*	-0.76675	small G protein, Ras family, cell well integrity
Stress	orf19.3707	*YHB1*	-1.63725	Nitric oxide dioxygenase
	orf19.6229	*CAT1*	-1.2685	Catalase, resistance to oxidative stress
	orf19.4082	*DDR48*	-3.3075	Stress response
	orf19.3612	*PST2*	-1.064	Oxidoreductase
Unknown	orf19.2018.1	-	-0.76375	Predicted ORF
	orf19.3121	-	-0.69175	Predicted ORF
	orf19.2247	-	-1.23575	Predicted ORF
	orf19.2460	-	-2.812	Predicted ORF

Group	ORF Number	Name	Median change, 1.5hr vs 0hr, opaque cells, log2	Description
	orf19.3902	-	-2.774	Predicted ORF
	orf19.3924	-	-0.903	Predicted ORF
	orf19.3926	-	-1.047	Predicted ORF
	orf19.1171	-	-0.7585	Predicted ORF
	orf19.1541	-	-1.5255	Predicted ORF
	orf19.1562	-	-0.7	Predicted ORF
	orf19.467	-	-1.259	Predicted ORF
	orf19.4157	*SPS20*	-1.31625	Predicted ORF
	orf19.5711	*PBP2*	-0.69075	Predicted ORF
	orf19.6254	-	-0.951	Predicted ORF
	orf19.1239	-	-1.6485	Predicted ORF
	orf19.899	-	-2.6625	Predicted ORF
	orf19.7513	-	-1.5445	Predicted ORF
	orf19.5615	*AYR2*	-1.59925	Predicted ORF
	orf19.4539	-	-1.9475	Predicted ORF
	orf19.2701	-	-1.08525	Predicted ORF
	orf19.1370	-	-3.221	Predicted ORF
	orf19.4934	*OP4*	-4.645	Ala- Leu- and Ser-rich protein; N-terminal hydrophobic region
	orf19.5157	-	-1.80125	Predicted ORF
	orf19.4654	-	-1.8695	Predicted ORF
	orf19.3897	-	-1.46175	Predicted ORF
	orf19.1344	-	-1.602	Predicted ORF
	orf19.2030	-	-0.8945	Predicted ORF
	orf19.3337	-	-1.2665	Predicted ORF
	orf19.3338	-	-1.35575	Predicted ORF
	orf19.3984	-	-0.70825	Predicted ORF
	orf19.6066	-	-0.7555	Predicted ORF
	orf19.1600	-	-0.3475	Predicted ORF
	orf19.577	-	-0.5195	Predicted ORF
	orf19.1709	-	-1.90225	Predicted ORF
	orf19.6143	-	-1.6225	Predicted ORF
	orf19.7512	-	-1.18025	Predicted ORF
	orf19.2724	-	-0.36275	Predicted ORF
	orf19.6830	-	-0.4655	Putative enoyl-CoA hydratase

Group	ORF Number	Name	Median change, 1.5hr vs 0hr, opaque cells, log2	Description
	orf19.4979	*KNS1*	-0.391	Protein kinase

Supplemental Table S3: List of 92 opaque enriched genes down-regulated within 90 minutes of the shift to 37°C, median change plotted (log2 scale). Genes are grouped based on hand annotation as many *C. albicans* genes lack Gene Ontology terms due to a lack of a recognizable *S. cerevisiae* homolog. These genes correspond to Sets 1 and 2 in Figure 4a.

Group	ORF Number	Name	Description
Signaling	orf19.4884	*WOR1*	Transcriptional regulator
	orf19.3127	*CZF1*	Transcriptional regulator
	orf19.799	*STE4*	Beta subunit of heterotrimeric G protein of mating signal transduction pathway
	orf19.4972	-	Putatuve Zinc finger Transcription factor
Transporters	orf19.176	*OPT4*	Oligopeptide transporter
	orf19.3749	*IFC3*	Oligopeptide transporter
	orf19.3746	*IFC1*	Oligopeptide transporter
Cell Wall/Surface	orf19.2833	*PGA34*	Putative GPI-anchored protein
Unknown	orf19.6805	-	Uncharacterized
	orf19.6277	-	Uncharacterized
	orf19.4749	-	Uncharacterized

Supplemental Table S4: List of 11 opaque enriched genes down-regulated at 3 or 4.5 hours that cluster with *WOR1* and *CZF1*. Genes are grouped based on hand annotation as many *C. albicans* genes lack Gene Ontology terms due to a lack of a recognizable *S. cerevisiae* homolog. These genes correspond to Set 3 in Figure 4a.

Supplemental Table S5: Genes associated with commitment.

ORF number	Name	Median *WOR1/wor1* 2h45min vs. Wild Type 2h45min, log2	Median (*WOR1/wor1* 2h45min vs. 0h) vs. (Wild Type 2h45min vs 0h), log2	In Wild Type Enriched Gene Set?
orf19.4884	WOR1	-4.013	-3.946	yes
orf19.3378	-	-0.819	-2.053	no
orf19.6805	-	-1.952	-2.164	yes
orf19.3122	ARR3	-0.7665	-0.9825	no
orf19.2475	PGA26	-0.879	-0.998	yes
orf19.2833	PGA34	-1.306	-1.383	yes
orf19.176	OPT4	-1.615	-1.197	yes
orf19.3127	CZF1	-0.808	-1.0935	yes
orf19.4651	PGA53	-0.849	-0.9855	yes
orf19.7042	-	-0.839	-0.807	no
orf19.3746	IFC1	-1.483	-1.0365	yes
orf19.6078	POL93	-1.641	-0.9575	no
orf19.3749	IFC3	-1.563	-1.016	yes
orf19.655	PHO84	-0.869	-1.264	yes
orf19.5121	OPT5	-0.895	-0.549	no
orf19.1264	CFL2	0.723	0.728	no
orf19.4889	-	0.7745	0.766	yes
orf19.851	-	1.207	0.51	yes
orf19.4212	FET99	1.334	1.3125	yes
orf19.2530	-	0.922	1.166	no
orf19.1152	-	1.321	1.327	yes
orf19.868	ADAEC	2.456	1.773	yes
orf19.4216	-	2.211	2.56	yes
orf19.3160	HSP12	2.195	2.813	yes
orf19.2531	CSP37	1.9675	2.652	yes
orf19.3548.1	WH11	2.433	2.967	yes
orf19.2457	-	-0.746	-0.772	yes
orf19.6277	-	-0.788	-0.644	yes
orf19.799	STE4	-0.748	-0.588	yes
orf19.4737	TPO3	1.004	1.142	yes
orf19.4215	FET34	1.052	0.768	yes
orf19.610	EFG1	1.085	0.773	yes
orf19.3902	-	-1.062	-0.369	yes
orf19.3007.2	-	0.7795	0.636	no

ORF number	Name	Median *WOR1/wor1* 2h45min vs. Wild Type 2h45min, log2	Median (*WOR1/wor1* 2h45min vs. 0h) vs. (Wild Type 2h45min vs 0h), log2	In Wild Type Enriched Gene Set?
orf19.2726	-	1.173	0.678	yes
orf19.6840		0.9145	0.253	no

Supplemental Table S5: Genes associated with commitment. List of 36 genes exhibiting at least a 1.41-fold change (log2=0.5) in three replicates when transforming the *WOR1/wor1* 2hr 45min samples against the wild type 2hr 45min samples ($T_{3\text{-}2}$) and at least a 1.41-fold difference (log2=0.5) in at least two of three replicates when transforming the *WOR1/wor1* 2hr 45min versus 0hr 00min transformation against the wild type 2hr 45min versus 0hr 00min transformations ($T_{(T_{3\text{-}1HET})\text{-}(T_{2\text{-}1WT})}$). All values in log2 scale. Genes that are on the enriched gene list have been indicated.

Supplemental Table S6: List of *C. albicans* strains used in this study.

Description	Strain	Genotype	Reference
Wild Type (white)	MLY537	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	This Study
Wild Type (white), independent isolate	MLY14	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	Reference 1
Wild Type (opaque)	MLY589	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	This Study
Wild Type (opaque), independent isolate	MLY594	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	This Study
Wor1-GFP (opaque)	MLY625	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR1-GFP/WOR1-GFP	This Study
Wor2-GFP (opaque)	MLY634	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR2-GFP/WOR2-GFP	This Study
Czf1-GFP (opaque)	MLY637	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} CZF1-GFP/CZF1-GFP	This Study
Efg1-GFP (opaque)	MLY627	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} EFG1-GFP/EFG1-GFP	This Study
WOR1 het (opaque)	MLY482	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a WOR1/wor1Δ::C.m.LEU2	This Study
WOR2 het (opaque)	MLY591	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a WOR2/wor2Δ::C.d.HIS1	This Study

Description	Strain	Genotype	Reference
CZF1 het (opaque)	MLY592	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a CZF1/czf1Δ::C.d.HIS1	This Study
EFG1 het (opaque)	MLY593	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a EFG1/efg12Δ::C.d.HIS1	This Study
WOR1 het (white)	RZY187	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a WOR1/wor1Δ::C.m.LEU2	This Study
efg1/efg1 (opaque)	MLY643	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a Δefg1:HIS/Δefg1::LEU2	This Study

References

1. Mitrovich QM, Tuch BB, Guthrie C, Johnson AD. *Genome Res* 2007, 17:492-502

Supplemental Table S6: List of *C. albicans* strains used in this study.

Supplemental Table S7: List of primers used in this study.

Description	Primer sequence 5'-3'
Primers for Creation of Wild Type Strain Independent Isolate	Reference 1
Deletion Strains	
WOR1 deletion related primers	Reference 2
General deletion system related primers	Reference 3
WOR2 5' flank (forward)	TTTAACCTGTAAGACTCATCCTTC
WOR2 5' flank (reverse)	CACGGCGCGCCTAGCAGCGGTAGCTTCACACTTGATTTTG
WOR2 3' flank (forward)	GTCAGCGGCCGCATCCCTGCTAATAAATCCAATATATTCATACTTTTG
WOR2 3' flank (reverse)	TTAACAATAGTCAATATATGTGTTCTC
WOR2 5' flank check (forward)	TTATACTATGATCTCTCGATTTCCG
WOR2 5' flank check (reverse)	AAGAATTTTGAGTTTGTGGG
CZF1 5' flank (forward)	TATAGCAAAATTCAAAGGGC
CZF1 5' flank (reverse)	CACGGCGCGCCTAGCAGCGGCCAGATAGTTTTCGTTTGAATG
CZF1 3' flank (forward)	GTCAGCGGCCGCATCCCTGCTAAGCTTCTCTGTGTTGGAGG
CZF1 3' flank (reverse)	CAAGTAATATGGCCAACAAAC
CZF1 5' flank check (forward)	CCTCAACATATTCTATATACCCAAC
CZF1 5' flank check (reverse)	CTTTACACACGACACCAATTAC
EFG1 5' flank (forward)	TTGATTTAGTGTATTACATCCAGCC
EFG1 5' flank (reverse)	CACGGCGCGCCTAGCAGCGGAATGGGTTAAGGGTTGGTTG
EFG1 3' flank (forward)	GTCAGCGGCCGCATCCCTGCAGGTTCAGTTCACCCTTCAC
EFG1 3' flank (reverse)	CCATCGAGTAAAATATACTTGTTCG
EFG1 5' flank check (forward)	CTGACACAGTCAAAAAGTTAGCAGAG
EFG1 5' flank check (reverse)	ACGCCACAAACTATCATCTC
CZF1 5' CDS Loss Check (forward)	ACCCAATATCAATTGGAATGACCCTAAC
CZF1 3' CDS Loss Check (reverse)	ACATCATGGCATTTTGCTCG
EFG1 5' CDS Loss Check (forward)	CTATACCCTATTACAATCAAATGAACGG
EFG1 3' CDS Loss Check (reverse)	TGCATTGTCGATACATGTGG
WOR2 5' CDS Loss Check (forward)	ATGACACAATTACCTTCTGTTTCAG
WOR2 3' CDS Loss Check (reverse)	TGTACTGGCAATTGTGACTC
RT-qPCR Primers	

Description	Primer sequence 5'-3'
PAT1 (forward)	TTATCGGAATGGTCCTCGTG
PAT1 (reverse)	CCAGAAGAACCATCATCAAC
EFG1 (forward)	CATCACAACCAGGTTCTACAACCAAT
EFG1 (reverse)	CTACTATTAGCAGCACCACCC
WOR1 (forward)	AGTGGCTGGTGTAGGAGCAC
WOR1 (reverse)	CTTGATCCACCAGGTGTTGA
WOR2 (forward)	CCA CCA AGG AAA AGA AAT CGT GAT AC
WOR2 (reverse)	AAC TAT TGA TTG TTG ATT CTG TTG TTG TTG
GFP Tagging and Verification	
GFP 5' (forward)	GGT GGT GGT TCT AAA GGT GAA GAA TTA TTC ACT
GFP 3' (reverse)	GCG GCC GCC TCG AG TTA TTT GTA CAA TTC ATC CAT ACC AT
SphI-*WOR1* 5' flank (forward)	gttacagcatgcCTA GTA GCA ACA TAA CTA CAA ATT CCA
WOR1 5' flank (reverse)	TCA CCT TTA GAA CCA CCA CC AGT ACC GGT GTA ATA CGA CCC AGA AGA ATT TCC AAC TGC G
WOR1 3' flank (forward)	CTC GAG GCG GCC GC TAG TTG AAT TAA TAC GGT GAT TCT GTT ATT ATT TT
WOR1 3' flank-AatII (reverse)	ctcggagacgtc CTA GCC AAT TCG TTC AGA TAT TCA TAC
WOR1 flank check (forward)	CAA CAA CAA CAG TAG CTC ACG AAC AAC AGG
WOR1 flank check (reverse)	GAT ATG AGA ATT AGG GTT ATG GTA TGA TGA TTT TCT GGA TTT CCG
WOR1 junction check (reverse)	GCC CCA AAA TAA TAA CAG AAT CAC CGT ATT AAT TCA ACT AAG TAC
SphI-*EFG1* 5' flank (forward)	gttacagcatgcTCC GGT AAA TAC CAA GGC TG
EFG1 5' flank (reverse)	TCA CCT TTA GAA CCA CCA CC CTT TTC TTC TTT GGC AAC AGT GCT A
EFG1 3' flank (forward)	CTC GAG GCG GCC GC TAA TAA ATA TCA TTC GTG TAC ATC ACC TTC
EFG1 3' flank-AatII (reverse)	ctcggagacgtc CTT GCT CCT GCT TGG TAC TTC
EFG1 flank check (forward)	AAT GCA GCT GCT GCA ACG GC
EFG1 flank check (reverse)	GGC ATA CTT ACC ACA ATG CAC AGT TGT TGT ATT TC
EFG1 junction check (reverse)	CAG AAA GCA GAA GGT GAT GTA CAC GAA TGA TAT TTA TTA CTT TT
SphI-*CZF1* 5' flank (forward)	gttacagcatgcCAG TCA CAA CAA TGG GAC G
CZF1 5' flank (reverse)	TCA CCT TTA GAA CCA CCA CC TTT ACT TCT GTA TTC AAC AAT ACC TCT C

Description	Primer sequence 5'-3'
CZF1 3' flank (forward)	CTC GAG GCG GCC GC TAA GCT TCT CTG TGT TGG AGG GAT
CZF1 3' flank-AatII (reverse)	ctcggagacgtc ACC TTT ACA CAC GAC ACC AAT TAC
CZF1 flank check (forward)	TAT GTG CAA ATG AAT CAA TTG CCC AAT CAG CAT TAC
CZF1 flank check (reverse)	CTG TGC TTT TGT TTG AAT GGT TTG TCT ACG C
CZF1 junction check (reverse)	GTT TTT TTT AAA CGA TAT CCC TCC AAC ACA GAG AAG CTT ATT TAC
SphI-*WOR2* 5' flank (forward)	gttacagcatgcTTA CCA AGA CAT GGC CAG AG
WOR2 5' flank (reverse)	TCA CCT TTA GAA CCA CCA CC TTT AAG TAA ATC AGC CAC TGA AAC TCT ATT AAT TA
WOR2 3' flank (forward)	CTC GAG GCG GCC GC TAA GAG TAT AGT AGA ACA AAA TAA TGT GTA TAT AGA T
WOR2 3' flank-AatII (reverse)	ctcggagacgtc CAA CAC CGT ACA ATA ATT TGA CAG G
WOR2 flank check (forward)	CAG CCA GTG ATG ACA CCA CCA C
WOR2 flank check (reverse)	TGC TGC TGT TGT TCA CAA TTG GTG GAC
WOR2 junction check (reverse)	CTA TAT ACA CAT TAT TTT GTT CTA CTA TAC TCT TAT TTA A
GFP 5' check	GGT TGG CCA TGG AAC TGG CA
GFP 3' Check	GGT GAT GGT CCA GTC TTG TTA CCA GAC
SAT1 flipper 5' check	CGA TGC ATA CGA CTA CAT CAA TGA AAT CCA GAC A
SAT1 flipper 3' check	GGT TCT CGG GAG CAC AGG ATG AC
References	1. Mitrovich QM, Tuch BB, Guthrie C, Johnson AD. *Genome Res* 2007, 17:492-502.
	2. Zordan RE, Galgoczy DJ, Johnson AD. *Proc Natl Acad Sci U S A* 2006, 103:12807-12812.
	3. Noble SM, Johnson AD. *Eukaryot Cell* 2005, 4:298-309.

Supplemental Table S7: List of primers used in this study.

Chapter 5:

Conclusions and Future Directions

The work described in this thesis has increased our understanding of both specific components of the transcriptional network regulating white-opaque switching in *Candida albicans* and of more general behaviors exhibited by both cell types. This thesis represents an effort to examine core components of the regulation of switching and to situate them in the broader context of the innate immune system rather than focusing on one particular aspect of the system. By studying both the regulation of the white-opaque switch and its role in *C. albicans'* interactions with the host immune system, we have been able to develop new hypotheses linking the two and propose benefits that might arise in the host (1).

From a minimalist standpoint, understanding how the master regulator Wor1 functions is central to comprehending the regulation of the white-opaque switch. Although Wor1 belongs to a protein family conserved across the fungal lineage, we knew little about it or its homologs when this project began. We have since shown that the conserved region of Wor1 exhibits sequence-specific DNA binding (2). Building on this discovery, we identified the nucleotide sequence recognized by Wor1 and demonstrated that this sequence was sufficient for Wor1-dependent transcriptional activation *in vivo*. We also conducted an initial characterization of *C. albicans* Pth2, a second member of this family (2). These studies demonstrate that Wor1 and its homologs represent a new superfamily of DNA-binding proteins. Looking beyond the regulatory circuit, we have also examined the role of Wor1 in the expression of the opaque characteristics of a cell, a topic discussed in Appendix 3.

Moving beyond a minimalistic model, understanding the white-opaque switch requires a consideration of the entire regulatory circuit as well as changes in the

transcriptome. To better understand the function and dynamics of the regulatory circuit, we investigated the order of regulatory changes during the temperature induced opaque-to-white switch (1). We found that changes in key transcriptional regulators occur gradually during this transition, extending over several cell divisions with little cell-to-cell variation in regulator levels. Our transcriptome analysis revealed that opaque cells begin to globally resemble white cells well before they irreversibly commit to switching. We have proposed that this characteristic of the switching process, a rapid transcriptome change preceding regulatory changes, permits *C. albicans* to "test the waters" before making an all-or-none decision. Temporarily separating the bulk of the transcriptome from its regulatory circuit may allow opaque cells to sample a new environment, switching cell types only if a favorable match occurs between the new gene expression profile and the environment.

Given the dramatic differences between white and opaque cells (3-6), we wondered how the ability to switch between the two cell types benefits *C. albicans* in its host. What advantages merit such a dramatic reorganization of the cell or making this reorganization sensitive to such a wide range of environmental cues? We demonstrated that differences between the white and opaque cell types directly affect interactions with the host innate immune system; both *Drosophila* hemocyte-derived S2 cells and mouse macrophage-derived RAW264.7 cells preferentially phagocytose white cells over opaque cells (7). This difference is seen for both the overall percentage of cultured cells that phagocytose white versus opaque *C. albicans* and the average number of *C. albicans* cells taken up by each phagocytic cell. This differing susceptibility to phagocytic innate immune cells is an important distinction between white and opaque *C. albicans*, one that

may allow *C. albicans* to escape phagocytosis by switching from the prevalent white form into the rarer opaque form. Assuming this difference is relevant *in vivo*, it may explain the apparent inertia that must be overcome for the opaque-to-white transition to happen. Given the increased risked of phagocytosis by the innate immune system in the white state, the white-opaque regulatory circuit may have faced selective pressure to make sure that elevated temperatures would persists before switching to the white cell type.

Advances in our understanding of white-opaque switching over the past five years have raised new questions about this phenomena and suggested several potentially fruitful lines of inquiry. The observation that white and opaque cells are phagoctosed at different levels by a variety of phagocytic cells makes the molecular basis for the differential recognition by the innate immune system a topic of interest. As increasing numbers of white and opaque-enriched genes are deleted, these strains can be screened to identify the gene or genes that contribute to differential regulation by the innate immune system. Of broader interest, our ever increasing understanding of the white-opaque regulatory circuit should allow for the creation of white and opaque locked strains that can be used to look for host niches preferred by each cell type and identify host barriers that either cell type cannot cross. Such *in vivo* assays will move us beyond an *in vitro* observation to examine the implications of switching on commensalism and infection.

Our work with Wor1 has established that it binds DNA directly and identified the DNA sequence it recognizes, but many unanswered questions remain about its function. We do not understand the basis for Wor1's unique binding pattern, with binding frequently spreading across several kilobases of intergenic DNA (8). Possibly related to

this issue, we do not know if Wor1 binds DNA in a cooperative manner with one or more co-factors. Identification of the motifs recognized by Wor2, Efg1, and Czf1 and a bioinformatic analysis of the co-occurrence of these with the Wor1 motif might address this issue. Assuming a transcription factor's motif co-occurs with Wor1 in a statistically significant manner, expressing and purifying that factor for *in vitro* cooperativity studies might be informative. Using the plasmids developed for Wor1 *in vivo* activation assays, existing *S. cerevisiae* transcription factor deletion libraries (9) could be screened for co-factors by looking for loss of Wor1 [or its homolog, *S. cerevisiae* YEL007] dependent activation. Wor1 binds at both white and opaque enriched genes in *C. albicans*, thus we expect that it functions as a repressor of some genes. Combining the newly identified Wor1 motif, published Wor1 binding data, and our knowledge of genes up-regulated in a *WOR1/wor1* opaque (Appendix 3), we possess the basis for developing plasmids to examine Wor1-dependent repression *in vivo*. These plasmids could then be used in a co-factor screen similar to the one described above. Finally, crystallization of Wor1 (or a Wor1 homolog) bound to DNA would greatly increase our understanding of this protein family, a topic discussed in Appendix 2.

Given the experimental infrastructure in place in *C. albicans*, it would possible to expand on the initial work from Chapter 3 to better characterize Pth2. Such work would be the first detailed characterization of the second, less studied, member of this protein family. The motif recognized by Pth2 could be characterized using the techniques from Chapter 3 or one of many available high-throughput approaches. RNA-sequencing indicated that Pth2 is up-regulated in biofilms (10), it might be worth performing ChIP-chip on it under that condition. A phenotypic analysis across a range of conditions,

similar to that in Homann et al. (11), might suggest other conditions where Pth2 plays a role.

Studies on the temperature-induced opaque-to-white switch also suggest areas of interest. The set of 36 genes closely associated with commitment to the white cell type (1), discussed in Chapter 4, make an excellent candidate list for deletion. Given the apparent importance of these genes for establishment or maintenance of the white or opaque cell types, this deletion library could be screened for phenotypic differences, changes in switching frequencies, and differences in time to 50% commitment in the temperature shift assay.

References

1. Lohse MB, Johnson AD (2010) Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans. *Mol Microbiol* 78:331-343.
2. Lohse MB, Zordan RE, Cain CW, Johnson AD (2010) Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in Candida albicans. *Proc Natl Acad Sci U S A* 107:14105-14110.
3. Lan CY, et al. (2002) Metabolic specialization associated with phenotypic switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 99:14907-14912.
4. Tsong AE, Miller MG, Raisner RM, Johnson AD (2003) Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115:389-399.
5. Lohse MB, Johnson AD (2009) White-opaque switching in *Candida albicans*. *Curr Opin Microbiol* 12:650-654.
6. Geiger J, Wessels D, Lockhart SR, Soll DR (2004) Release of a potent polymorphonuclear leukocyte chemoattractant is regulated by white-opaque switching in *Candida albicans*. *Infect Immun* 72:667-677.
7. Lohse MB, Johnson AD (2008) Differential phagocytosis of white versus opaque *Candida albicans* by *Drosophila* and mouse phagocytes. *PLoS One* 3:e1473.
8. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* 5:e256.
9. Akache B, Wu K, Turcotte B (2001) Phenotypic analysis of genes encoding yeast zinc cluster proteins. *Nucleic Acids Res* 29:2181-2190.

10. Tuch BB, et al. (2010) The Transcriptomes of Two Heritable Cell Types Illuminate the Circuit Governing Their Differentiation. *PLoS Genetics* 6:e1001070.

11. Homann OR, Dea J, Noble SM, Johnson AD (2009) A phenotypic profile of the Candida albicans regulatory network. *PLoS Genetics* 5:e1000783.

Appendix 1

Influence of Two Glucose Sensing Pathways on White-Opaque Switching

Introduction

C. albicans has been characterized as a Crabtree-negative facultative aerobe, it prefers to metabolize carbon through respiration rather than fermentation if oxygen is present (1, 2). White and opaque cell types of *C. albicans* differ in their metabolic preferences; transcriptional analysis of the two cell types reveals differential regulation iron, phosphate, and glucose related genes (3-5). Opaque cells appear starved for glucose and phosphate, even if both substances are abundant in the growth medium. As a result, opaque cells take *C. albicans'* preference for respiration beyond that seen in white cells, further down-regulating genes involved in glycolysis while up-regulating genes associated with the Citric Acid Cycle and usage of alternative carbon sources (3-5). Given the range of metabolic conditions *C. albicans* would be expected to encounter in different host environments, it is tempting to speculate that the metabolic specialization of the two cell types represents specialization for specific host niches.

The discovery of *C. albicans* mating in mice suggested a paradox (6), opaque cells did not appear stable at temperatures encountered in the host (7) but mating only occurs between opaque cells (8). Recent discoveries have resolved this apparent paradox; opaque cells are stabilized by the presence of carbon dioxide (9) or absence of oxygen (10, 11) at elevated temperatures. Interestingly, opaque-cells down regulate many genes associated with the Citric Acid Cycle and alternative carbon source usage when exposed to elevated temperature, becoming more like white cells even though they have not yet committed to switching (5). Unpublished data from Aaron Hernday in the Johnson lab has suggested another environmental cue capable of stabilizing opaque cells

at 37°C, growth on media lacking glucose. Currently, it is not known how glucose or oxygen and carbon dioxide levels work to stabilize the opaque cell type.

Given the links between opaque cell stability, glucose, and temperature, we wanted to determine the influence of glucose sensing on white-opaque switching. *C. albicans* is known to use three different pathways to synthesize a response to internal and external glucose levels (Figure 1, see (12) for review). The first pathway (Figure 1a) detects extracellular glucose through the membrane bound sensor Hgt4 (orf19.5962); Hgt4 derepresses glucose transporters, fermentation, and alternative respiration genes when sufficient extracellular glucose is present by triggering the degradation of Std1 (orf19.6173) and the inactivation of the transcriptional factor Rgt1 (orf19.2747) (13-15). The second pathway measures intracellular glucose levels (Figure 1b), higher intracellular glucose concnetrations result in increased ATP levels which causes the Reg1 (orf19.2005) /Glc7 (orf19.6285) phosphatase complex to deactivate the kinase Snf1 (orf19.1936) (S. cerevisiae effect discussed in (16)), ultimately allowing the transcription factor Mig1 (orf19.4318) to repress the utilization of non-fermentable carbon sources and gluconeogenesis (17, 18). The third pathway, referred to as the adenylate cyclase pathway, incorporates signals from intra- and extracellular sources and affects PKA activity through the inactivation of its Bcy1 (orf19.2014) inhibitory subunit (for review see (12)). The activated PKA then can repress Rgt1 from the first pathway as well as inducing a range of hyphal related genes via phosphorylation of the transcription factor Efg1 (orf19.610) (19). As the third pathway was already under investigation in the Johnson lab, this study focused on the influence of the first two pathways on white-opaque switching.

Methods and Materials

Genes were deleted using the His and Leu cassettes first described in (20). Deletions were conducted in the previously described RZ47 **a/a** switching competent background (21). Primers used in this study are listed in Table 1, strains are listed in Table 2. White-opaque switching assays were performed based on previously described techniques (21), only one isolate per strain was used for opaque-to-white switching assays. Control strains with *HIS1* and *LEU2* added back were gifts from Aaron Hernday and Quinn Mitrovich and have been described previously (5). The stability of opaque cells at 37°C in the presence or absence of glucose was examined using an unpublished assay developed by Aaron Hernday. In brief, opaque cells were stuck out on Spider medium plates that had been coated with either glucose solution or sterile water. Glucose solutions were designed to mimic the amount of glucose present in a SD+aa+Ura plate. Plates were then incubated at 37°C for 1 to 3 days, and representative colonies restruck onto SD+aa+Ura plates. These plates were grown at room temperature to determine colony phenotype.

Results

Selection of Candidate Genes for Deletion

Based on existing annotations, three genes from the extracellular pathway (*SNF3* (orf19.5962), *STD1* (orf19.6173), and *RGT1* (orf19.2747)) and 7 from the intracellular sensing pathway (*REG1* (orf19.2005), *SNF1* (orf19.1936), *SNF4* (orf19.5768), *MIG1* (orf19.4318), *MIG2* (orf19.5326), *HXK2* (orf19.542), and orf19.3840 (S.cer *SAK1* homolog) were targeted for deletion. *SNF3*, *STD1*, and *RGT1* were chosen for deletion to give complete coverage of the extracellular sensing pathway. *REG1*, orf19.3840, and

SNF4 were chosen to include up-stream signals in the intracellular sensing pathway in addition to *SNF1* and *MIG1*. *HXK2* was chosen to affect levels to phosphorylated glucose. *MIG2* (orf19.5326) was included as it was not clear if it played a role in this pathway. Two additional genes, *GAC1* (orf19.7053) and *SHA3* (orf19.3669), we also included as both were opaque-enriched and possibly played a role in the regulation of glucose or glycogen related processes. Other genes like *GLC7* (orf19.6285) were not included as annotation at the Candida Genome Database (http://www.candidagenome.org/) suggested they might be essential. It did not prove possible to delete *REG1* and *SNF1*, which are both likely essential genes. It was also not possible to delete *RGT1*, although this gene has been deleted in other backgrounds where we have since determined that it did not have a significant impact on white-to-opaque switching (data not shown).

Determination of Switching Rates

Using these 10 deletion strains, white-to-opaque switching frequencies were determined using a standard assay (21). We observed that deletion of *HXK2*, orf19.3840, *MIG1*, and *SNF4* reduced white-to-opaque switching frequencies but did not block them entierely (Table 3). Deletion of *MIG2, GAC1, SHA3,* or *STD1* did not have a measurable effect on switching rates. Deletion of *SNF3* appears to significantly reduce white-to-opaque switching (data not shown), however, the one opaque sector we found gave rise to extremely filamentous colonies so we do not know whether we missed other switching events. Looking at opaque-to-white switching frequencies at room temperature (Table 4), we found slight effects on opaque-cell stability but none were of a sufficient magnitude to merit further study.

Opaque Cell Stablity at 37°C

We also examined the stability of opaque isolates of these strains at 37°C in the presence and absence of glucose. Using the protocol developed by Aaron Hernday, cells were stuck out on Spider medium plates coated with glucose solution or sterile water and incubated at 37°C. Representative colonies were restruck to SD+aa+Ura plates and grown at room temperature to determine colony phenotype. In these experiments, summarized in Table 5, no strain was entirely locked in the opaque cell type in the absence of glucose. The *MIG1* and *MIG2* deletions resulted in some stable opaque colonies, but not to a greater extent than the wild type cells. The *STD1* deletion gave rise to only white cells, this deletion may cause cells to mistakenly believe glucose is present. Looking at cells grown on glucose coated plates, the *MIG1* delete did give rise to a small fraction of opaque cells suggesting that these cells might not recognize that glucose is present. The degree of interconnectivity between the various sugar sensing pathways (12) may contribute to the lack of major changes observed in this assay.

Future Directions

None of the glucose-sensing related genes deleted had a large influence on white-opaque switching frequencies or the stability of opaque cells at 37°C. It may be worth repeating the 37°C stability assays with a more rigorous scoring of the phenotypes of the various strains to look for more subtle differences in these strains. An examination of the effects of these deletions on the time to 50% commitment when opaque cells are exposed to elevated temperatures could be of interest, although it may be difficult to control for changes in growth rates in these strains. In the future, an examination into the *snf3/snf3* opaque cell hyperfilimentation, a phenotype not observed in *snf3/snf3* white

cells, may present an opportunity to probe the links between glucose, opaque cells, and filimentation.

References

1. Franzblau SG, Sinclair NA (1983) Induction of fermentation in Crabtree-negative yeasts. *Mycopathologia* 82:185-190.
2. De Deken RH (1966) The Crabtree effect: a regulatory system in yeast. *J Gen Microbiol* 44:149-156.
3. Lan CY, et al. (2002) Metabolic specialization associated with phenotypic switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 99:14907-14912.
4. Tsong AE, Miller MG, Raisner RM, Johnson AD (2003) Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115:389-399.
5. Lohse MB, Johnson AD (2010) Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans. *Mol Microbiol* 78:331-343.
6. Hull CM, Raisner RM, Johnson AD (2000) Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. *Science* 289:307-310.
7. Rikkerink EH, Magee BB, Magee PT (1988) Opaque-white phenotype transition: a programmed morphological transition in *Candida albicans*. *J Bacteriol* 170:895-899.
8. Miller MG, Johnson AD (2002) White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* 110:293-302.
9. Huang G, Srikantha T, Sahni N, Yi S, Soll DR (2009) CO(2) regulates white-to-opaque switching in *Candida albicans*. *Curr Biol* 19:330-334.

10. Dumitru R, et al. (2007) In vivo and in vitro anaerobic mating in *Candida albicans*. *Eukaryot Cell* 6:465-472.

11. Ramírez-Zavala B, Reuss O, Park YN, Ohlsen K, Morschhäuser J (2008) Environmental induction of white-opaque switching in *Candida albicans*. *PLoS Pathog* 4:e1000089.

12. Sabina J, Brown V (2009) Glucose sensing network in Candida albicans: a sweet spot for fungal morphogenesis. *Eukaryot Cell* 8:1314-1320.

13. Brown V, Sabina J, Johnston M (2009) Specialized sugar sensing in diverse fungi. *Curr Biol* 19:436-441.

14. Brown V, Sexton JA, Johnston M (2006) A glucose sensor in Candida albicans. *Eukaryot Cell* 5:1726-1737.

15. Sexton JA, Brown V, Johnston M (2007) Regulation of sugar transport and metabolism by the Candida albicans Rgt1 transcriptional repressor. *Yeast* 24:847-860.

16. Wilson WA, Hawley SA, Hardie DG (1996) Glucose repression/derepression in budding yeast: SNF1 protein kinase is activated by phosphorylation under derepressing conditions, and this correlates with a high AMP:ATP ratio. *Curr Biol* 6:1426-1434.

17. Murad AM, et al. (2001) Transcript profiling in Candida albicans reveals new cellular functions for the transcriptional repressors CaTup1, CaMig1 and CaNrg1. *Mol Microbiol* 42:981-993.

18. Zaragoza O, Rodriguez C, Gancedo C (2000) Isolation of the MIG1 gene from Candida albicans and effects of its disruption on catabolite repression. *J Bacteriol* 182:320-326.

19. Harcus D, Nantel A, Marcil A, Rigby T, Whiteway M (2004) Transcription profiling of cyclic AMP signaling in Candida albicans. *Mol Biol Cell* 15:4490-4490.

20. Noble SM, Johnson AD (2005) Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. *Eukaryot Cell* 4:298-309.

21. Zordan RE, Galgoczy DJ, Johnson AD (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci U S A* 103:12807-12812.

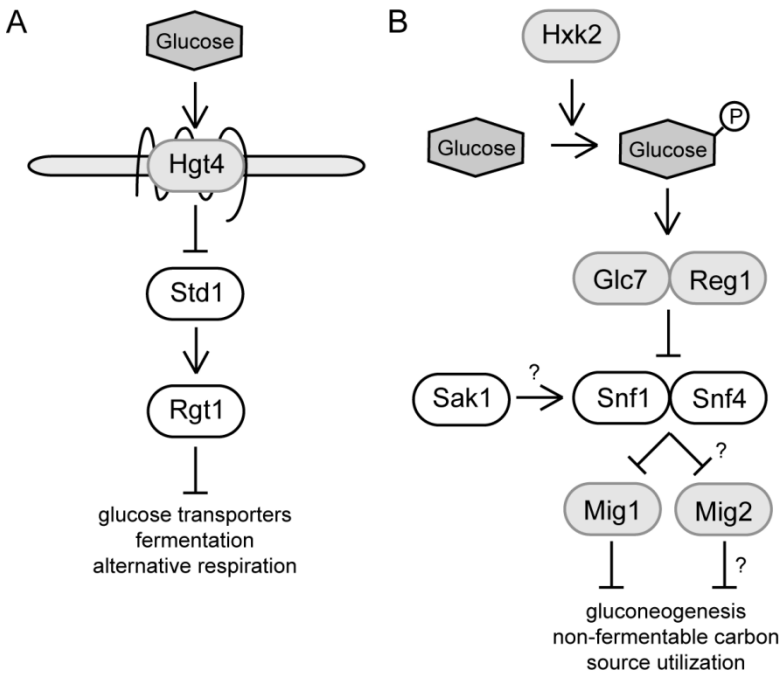


Figure 1: An overview of two of the glucose-sensing pathways in *C. albicans*. [A] *HGT4* extracellular glucose-sensing pathway. [B] *MIG1* intracellular glucose-sensing pathway. Figure adapted from Sabina and Brown 2009 (12), arrows and bars represent activation and repression respectively. Genes in grey are upregulated in the presence of glucose.

Table 1: List of primers used in this study.

Name	Description	Primer sequence 5'-3'	Reference
	General deletion system related primers		Reference 1
MBL618	*SNF3 5' flank (forward)*	ATT TAC ACT AAT CAG ATC AAA TCC CAC	This Study
MBL616	*SNF3 5' flank (reverse)*	cacggcgcgcctagcagcgg CCT AAT CAC TCA AAA CAA ATA GTT ATT CG	This Study
MBL617	*SNF3 3' flank (forward)*	Gtcagcggccgcatccctgc GTG TCA TAT AGA GGT TGA TAT TTT TTT TTT TTT TC	This Study
MBL619	*SNF3 3' flank (reverse)*	TTG GTT AGC ATT ACT CAT TGT TGT TAG	This Study
MBL620	*SNF3 5' flank check (forward)*	TGT CGT CGT TTT GGT GCG	This Study
MBL621	*SNF3 3' flank check (reverse)*	CTG CCC AAT GTA ATG GAG TTC	This Study
MBL622	*SNF3 5' CDS Loss Check (forward)*	GAT CCT TTC TGC TAT TGT TCC GTT ATA TC	This Study
MBL623	*SNF3 3' CDS Loss Check (reverse)*	CCC AAC ACC TCC CCA AAA C	This Study
MBL634	*STD1 5' flank (forward)*	TTT ACA ACA ATA GAC TTC GGA TAA ACT G	This Study
MBL632	*STD1 5' flank (reverse)*	cacggcgcgcctagcagcgg CGA TTC AAC AAC TCT GAT GAT GG	This Study
MBL633	*STD1 3' flank (forward)*	Gtcagcggccgcatccctgc GCC TGA CAA GAA AGT CAA GTT G	This Study
MBL635	*STD1 3' flank (reverse)*	GAT GGT TCA ATA TGA AAC AGC CAG	This Study
MBL636	*STD1 5' flank check (forward)*	ATG TTG GGA GGC TAC AAG AAG	This Study
MBL637	*STD1 3' flank check (reverse)*	ATG ACC CAG TAA TAC CAC CG	This Study
MBL638	*STD1 5' CDS Loss Check (forward)*	ACG CTC GGT GAG ATC AG	This Study
MBL639	*STD1 3' CDS Loss Check (reverse)*	TCG GGG TTT ATT GAC GAT GAG	This Study

Name	Description	Primer sequence 5'-3'	Reference
MBL586	*REG1 5' flank (forward)*	TTA ATT TAC AAC AAC CAA ATA ACT GCT C	This Study
MBL584	*REG1 5' flank (reverse)*	cacggcgcgcctagcagcgg GGA ATA ACC GAT ACC AAC GAC	This Study
MBL585	*REG1 3' flank (forward)*	Gtcagcggccgcatccctgc ACC GTC AGT AAT GAG TTT TGT TG	This Study
MBL587	*REG1 3' flank (reverse)*	CTT TTA TCT CTT ACA ATC GTG ATT TCC	This Study
MBL656	*REG1 5' flank check (forward)*	CTA CAC TTA TTC ACT CAT TCA AAC TCA TAC	This Study
MBL589	*REG1 3' flank check (reverse)*	CAG ATA GTA GTT CTC CAC AAC CC	This Study
MBL590	*REG1 5' CDS Loss Check (forward)*	GTG TAT TCA GAT GAT GAG CAA CG	This Study
MBL591	*REG1 3' CDS Loss Check (reverse)*	TGA CAT TAT GAG AAA GAC TCG ACG	This Study
MBL519	*SNF1 5' flank (forward)*	GGT CTA GAA CTG ATA GAG ATA GAG	This Study
MBL517	*SNF1 5' flank (reverse)*	cacggcgcgcctagcagcgg CTT TAA CTC TAT GTT ATT ATA TAA TGT GTA TTA C	This Study
MBL518	*SNF1 3' flank (forward)*	Gtcagcggccgcatccctgc AAA GTA AGT AAG TAC TAG CTT AGA TTG	This Study
MBL520	*SNF1 3' flank (reverse)*	TGT GTA GTT CCT GTA TAT AGG C	This Study
MBL521	*SNF1 5' flank check (forward)*	GGA AAA ACC AAC CAT TGT CCA GAA GAA TG	This Study
MBL522	*SNF1 3' flank check (reverse)*	GTC TTG AAT AAT CTT TCT TTA ATT GTT TCT TGA AAG TTG	This Study
MBL533	*SNF1 5' CDS Loss Check (forward)*	GTG ATT TAA AGC CAG AAA ACT TAT TGT TAG AC	This Study
MBL534	*SNF1 3' CDS Loss Check (reverse)*	CAG CAA CTT GTT TGT TTG CCT TTT C	This Study
MBL578	*SNF4 5' flank (forward)*	CAT ATT CAT ATT CTT GGT CTC CTT CC	This Study
MBL576	*SNF4 5' flank (reverse)*	cacggcgcgcctagcagcgg GTC AAA CCA GAA CAC TGT AAG TAG	This Study
MBL577	*SNF4 3' flank (forward)*	Gtcagcggccgcatccctgc TTG TAC AAT AAA GTC CAT GAT GGT G	This Study

Name	Description	Primer sequence 5'-3'	Reference
MBL578	*SNF4 3' flank (reverse)*	GTG TGT GAG TGC TTT CAC TG	This Study
MBL654	*SNF4 5' flank check (forward)*	CAA TTC TTC CTT CAA TTT ACT CAT ATC TTG	This Study
MBL581	*SNF4 3' flank check (reverse)*	GAA AAT TCG ATA TAA GGA CCA ATG AAC C	This Study
MBL582	*SNF4 5' CDS Loss Check (forward)*	CAA TTC CCA GAA AAG TTT GAA CTA G	This Study
MBL583	*SNF4 3' CDS Loss Check (reverse)*	CTC AAC AAA GCA TCT CCA ACA G	This Study
MBL570	*MIG1 5' flank (forward)*	CCG CCA ACA GTT ATT TTT TTC C	This Study
MBL568	*MIG1 5' flank (reverse)*	cacggcgcgcctagcagcgg GCG ACG TTG ATT AAC TTG ATT AAT AAG	This Study
MBL569	*MIG1 3' flank (forward)*	Gtcagcggccgcatccctgc GGT GTT TTA ATT TAT AGA TAG ATT TAC AAT ACG G	This Study
MBL571	*MIG1 3' flank (reverse)*	TTC ACA TTC GGA TAG AAC TAA CAA	This Study
MBL572	*MIG1 5' flank check (forward)*	GTA TAT GAT GAT GTC AAC TGG ATC TG	This Study
MBL648	*MIG1 3' flank check (reverse)*	AAA AAA AGA AAA ACA AAC TCA AAG GAA GG	This Study
MBL574	*MIG1 5' CDS Loss Check (forward)*	GGA GTT ACA GTT GTT CCT AAT ACT G	This Study
MBL575	*MIG1 3' CDS Loss Check (reverse)*	GAT GAT GAA CTA GCA TTA AAT AAT CGA G	This Study
MBL626	*MIG2 5' flank (forward)*	TCC TCA ATT CAG ATA TAA TGT GGT TTC	This Study
MBL624	*MIG2 5' flank (reverse)*	cacggcgcgcctagcagcgg GAG TGG TAT GGT GAT GGT TTT G	This Study
MBL625	*MIG2 3' flank (forward)*	Gtcagcggccgcatccctgc CGT CTC TTC GTC TCC ACC	This Study
MBL627	*MIG2 3' flank (reverse)*	CAA TGC TGG TGT GGC TG	This Study
MBL628	*MIG2 5' flank check (forward)*	CGC CTT CGC AAT ACT CCA	This Study
MBL629	*MIG2 3' flank check (reverse)*	TAC TCG CCA GAC CAC G	This Study
MBL630	*MIG2 5' CDS Loss Check (forward)*	GAG AAA AAC CCC ATC AGT GTA C	This Study

Name	Description	Primer sequence 5'-3'	Reference
MBL631	*MIG2 3' CDS Loss Check (reverse)*	GTA CTG CTA TTG CTG GCA G	This Study
MBL594	*HXK2 5' flank (forward)*	CAC TAT AAC TCC TCC CTT CCC	This Study
MBL592	*HXK2 5' flank (reverse)*	cacggcgcgcctagcagcgg GCA ACC AAA AGT TAA AGT GAT TGA C	This Study
MBL593	*HXK2 3' flank (forward)*	Gtcagcggccgcatccctgc GTA ATG TGT ATC ATG ACA GTT TTG AC	This Study
MBL595	*HXK2 3' flank (reverse)*	GGA TCT ACT TTC TTC AAT CAA GCG	This Study
MBL658	*HXK2 5' flank check (forward)*	TGA GAA CTG CCA CCC TGG	This Study
MBL597	*HXK2 3' flank check (reverse)*	CAC CAA AAC CCA ACA TCA TAC C	This Study
MBL598	*HXK2 5' CDS Loss Check (forward)*	GAA GGT AAA TGT CCA TCA GAT ATT CC	This Study
MBL599	*HXK2 3' CDS Loss Check (reverse)*	AGC ACA ATG AGC GGT TTT G	This Study
MBL503	*GAC1 5' flank (forward)*	CTA CCA CAG TAA CAA CCA AAG	This Study
MBL501	*GAC1 5' flank (reverse)*	cacggcgcgcctagcagcgg TGA TCG ATT GAT TGA TTG ATT GAT TG	This Study
MBL502	*GAC1 3' flank (forward)*	Gtcagcggccgcatccctgc GAA TTA ATA CAA CAA AAA ACA AGG AC	This Study
MBL504	*GAC1 3' flank (reverse)*	CTC AAA GAT TTC CAA ATA TAG TTG G	This Study
MBL505	*GAC1 5' flank check (forward)*	CCT TCT CTT ATT GTC ATT CCC TTT CTG TTC	This Study
MBL506	*GAC1 3' flank check (reverse)*	TGG AGC TTG GGT AGA CGA GAA AGA AAC	This Study
MBL531	*GAC1 5' CDS Loss Check (forward)*	ATC GGT TCG TTT TGC ATC AAG ATT AG	This Study
MBL532	*GAC1 3' CDS Loss Check (reverse)*	TTA CCA TCT GTC TTT AAC ACA TAT GCT GAA	This Study
MBL511	*SHA3 5' flank (forward)*	AAG TCC TCC TCT GCT TCC CT	This Study
MBL509	*SHA3 5' flank (reverse)*	cacggcgcgcctagcagcgg TGT AGG TAA TTG TTA TAG GTA AGT TAA TAA	This Study

Name	Description	Primer sequence 5'-3'	Reference
MBL510	*SHA3 3' flank (forward)*	Gtcagcggccgcatccctgc GTA CAC ATA AAT AAT TAA TGA CAA TTA ATA ATA AC	This Study
MBL512	*SHA3 3' flank (reverse)*	TTT CAA AGT CCC TCT TCC AT	This Study
MBL513	*SHA3 5' flank check (forward)*	GTG GGT GGT AGT AAT TGA ATG ATA GTA TCA GAT G	This Study
MBL514	*SHA3 3' flank check (reverse)*	TCA CCA ACA AAC AAC CCA ACA GCA AC	This Study
MBL535	*SHA3 5' CDS Loss Check (forward)*	CAC CCA AAC ATA GCG ACT ATT CAT C	This Study
MBL536	*SHA3 3' CDS Loss Check (reverse)*	GGT GTT GTA ATT GGT GGT TCT TTC TG	This Study
MBL642	*RGT1 5' flank (forward)*	GTC ACT GGT TAG GTT GGG TAG	This Study
MBL640	*RGT1 5' flank (reverse)*	cacggcgcgcctagcagcgg GTA GTA GTT AGT AGT GTG TGT GTG G	This Study
MBL641	*RGT1 3' flank (forward)*	Gtcagcggccgcatccctgc CCC CCA TGT ATG TAT GTA TGA ATG	This Study
MBL643	*RGT1 3' flank (reverse)*	CAA CAA CAA CAA CAT CAC CAA GG	This Study
MBL644	*RGT1 5' flank check (forward)*	AAC CAA AAA AAA TAA AAT CTT CGC ACC	This Study
MBL678	*RGT1 3' flank check (reverse)*	GTG ATT TCA TTA TTA GTT TTC CTT TGT TGT GG	This Study
MBL646	*RGT1 5' CDS Loss Check (forward)*	GAT TAT ACC AGT GAT GCT GAA TCA G	This Study
MBL647	*RGT1 3' CDS Loss Check (reverse)*	GTA GTT TCA TTG GGA AAT GGA TAT GAT	This Study
MBL602	*SAK1 5' flank (forward)*	TTA GTT TTG GAA GTA ATG CCT TTA TGT AC	This Study
MBL600	*SAK1 5' flank (reverse)*	cacggcgcgcctagcagcgg TTG TCT GTT TCA AGC TAA ACA ACC	This Study
MBL601	*SAK1 3' flank (forward)*	Gtcagcggccgcatccctgc GGT CGA TTT TCT AGA ACG CAT TAC	This Study
MBL603	*SAK1 3' flank (reverse)*	ACA ACT TAT TTT CTG GAG GCG	This Study
MBL604	*SAK1 5' flank check (forward)*	CAA AAT ACC AAC TTC CCT TGA TCG	This Study

Name	Description	Primer sequence 5'-3'	Reference
MBL605	*SAK1 3' flank check (reverse)*	GGA GAG CGT GGA GTT AAG AC	This Study
MBL606	*SAK1 5' CDS Loss Check (forward)*	GGT GGG TTG AAG GAT GGA G	This Study
MBL607	*SAK1 3' CDS Loss Check (reverse)*	CAA TTT GAG GAG CCT CCG AAA C	This Study

References 1. Noble SM, Johnson AD. (2005) *Eukaryot Cell* 4:298-309.

Table 1: List of primers used in this study.

Table 2: List of strains used in this study.

Description	Strain	Genotype	Reference
Wild Type (white)	MLY537	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	Ch4, Reference 1
Wild Type (opaque)	MLY589	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	Ch4, Reference 1
snf3/snf3 (white)	MLY745	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δsnf3:LEU2/Δsnf3:HIS1	This study
snf3/snf3 (opaque)	MLY753	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δsnf3:LEU2/Δsnf3:HIS1	This study
std1/std1 (white)	MLY749	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δstd1:LEU2/Δstd1:HIS1	This study
std1/std1 (opaque)	MLY752	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δstd1:LEU2/Δstd1:HIS1	This study
snf4/snf4 (white)	MLY743	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δsnf4:LEU2/Δsnf4:HIS1	This study
snf4/snf4 (opaque)	MLY747	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δsnf4:LEU2/Δsnf4:HIS1	This study
mig1/mig1 (white)	MLY742	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δmig1:LEU2/Δmig1:HIS1	This study
mig1/mig1 (opaque)	MLY746	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δmig1:LEU2/Δmig1:HIS1	This study
mig2/mig2 (white)	MLY744	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δmig2:LEU2/Δmig2:HIS1	This study
mig2/mig2 (opaque)	MLY751	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δmig2:LEU2/Δmig2:HIS1	This study

Description	Strain	Genotype	Reference
hxk2/hxk2 (white)	MLy734	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δhxk2:LEU2/Δhxk2:HIS1	This study
hxk2/hxk2 (opaque)	MLY750	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δhxk2:LEU2/Δhxk2:HIS1	This study
sha3/sha3 (white)	MLY713	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δsha3:LEU2/Δsha3:HIS1	This study
sha3/sha3 (opaque)	MLY723	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δsha3:LEU2/Δsha3:HIS1	This study
gac1/gac1 (white)	MLY714	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δgac1:LEU2/Δgac1:HIS1	This study
gac1/gac1 (opaque)	MLY724	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δgac1:LEU2/Δgac1:HIS1	This study
sak1/sak1 (white)	MLY733	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δsak1:LEU2/Δsak1:HIS1	This study
sak1/sak1 (opaque)	MLY748	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} Δsak1:LEU2/Δsak1:HIS1	This study

References

1. Lohse MB, Johnson AD. (2010) *Mol Microbiol* 78:331-343.

Table 2: List of strains used in this study.

Strain	Percent Colonies with Opaque sectors
Control	3.11
ΔΔ*gac1*	3.17
ΔΔsha3	4.36
Strain	Percent Colonies with Opaque sectors
Control	2.09
ΔΔsak1	0.49
ΔΔhxk2	0.20
ΔΔmig1	0.07
ΔΔsnf4	0.14
Strain	Percent Colonies with Opaque sectors
Control	1.77
ΔΔmig1	0.06
ΔΔsnf4	≤0.14
ΔΔmig2	1.88
ΔΔsnf3	≤0.07
ΔΔstd1	2.00

Table 3: Frequency of white-to-opaque switching, expressed as a percentage of colonies with one or more opaque sectors.

Strain	Percent Colonies with White sectors
Control	9.71
ΔΔmig1	9.88
ΔΔsnf4	9.21
Strain	Percent Colonies with White sectors
Control	5.36
ΔΔmig2	2.49
ΔΔhxk2	11.54
ΔΔstd1	3.00

Table 4: Frequency of opaque-to-white switching. expressed as a percentage of colonies with one or more white sectors or that were entirely white.

Gene	Colony Phenotypes No Glucose, 37°C	Colony Phenotypes Glucose, 37°C
Control	white-opaque mix	white
ΔΔ*mig1*	white-opaque mix	white, few opaque colonies
ΔΔhxk2	no growth	white
ΔΔsak1	no growth	white
ΔΔsnf4	no growth	white
ΔΔmig2	white-opaque mix	white
ΔΔstd1	white	white
ΔΔsnf3	white-opaque mix	white

Table 5: Opaque stability in the presence and absence of glucose at 37°C.

Appendix 2:

Crystallization of the novel DNA-binding transcription factor Wor1

Introduction

White-opaque switching is an epigenetic phenomenon in the human fungal pathogen *Candida albicans* whereby two distinct and heritable cell types (named white and opaque) arise from the same genome (1-3). The transcription factor Wor1 (orf19.4884) functions as the master regulator of this phenomenon and is necessary for both the establishment and the maintenance of the opaque cell type (4-7). Although white-opaque switching appears limited to *C. albicans* and its very close pathogenic relatives, proteins similar to Wor1 are found across the fungal kingdom. In other words, Wor1-like proteins are found in species encompassing a billion years of evolutionary history.

The conserved portion of this protein family consists of about 300 amino acids at the N-terminus of the protein (Figure 1). This region, which we have termed the WOPR box, consists of two conserved blocks of sequence (in Wor1 roughly positions 1 to 100 (WOPRa) and 185 to 265 (WOPRb)) separated by a less conserved region of variable length. The length and sequence of the gap between these conserved regions and the unconserved c-terminal region vary widely among this family. Based on conserved differences at several residues in each conserved regions, it is possible to identify two distinct family members in most fungal species (8).

Although nearly every sequenced fungal genome has Wor1 homologs, this family remains largely uncharacterized. Of those that have been characterized, Ryp1 from *Histoplasma capsulatum* is a key regulator of the yeast-to-mycelial transition (9), Gti1 from *Schizosaccharomyces pombe* plays a role in regulating gluconate uptake in response to starvation conditions (10), and Sge1 from the plant pathogen *Fusarium oxysporum* is

required for parasitic growth (11). The second family member is even less characterized, with Pac2 from *S. pombe* playing a role in the regulation of mating (12). Despite the limited characterization of this family, it appears that it commonly is involved in the regulation of major morphological changes. The Johnson lab has recently shown that Wor1 (and thus presumably the entire Wor1 family) recognizes and binds to DNA in a sequence specific manner and that Wor1 binding is sufficient to activate transcription *in vivo*. Given these observations, this protein family represents a previously uncharacterized DNA-binding domain (8).

Although we know that Wor1 binds DNA directly, we do not know how it performs this function. Homology detection programs such as PSI-BLAST (13) only show a similarity to other members of this family, there is no a clear relationship to any known DNA-binding protein family. A variety of structural prediction programs suggest that the two conserved blocks each code for globular domains, as discussed above, but structure homology search tools like Phyre (14) fail to find a match between either region and any published structure. As a result, we do not know how this family binds DNA or which residues are most important for this interaction. Similarly, we cannot easily determine whether there is any significance to differences between the two distinct members of this family.

To address these questions, we started a collaboration with the lab of Robert Stroud at UCSF to crystallize Wor1 in the presence of DNA to determine the first three-dimensional structure for a member of this family. This project is ongoing as this thesis is in preparation, so this appendix represents a summary of progress to that point.

Methods and Materials and Results

Selection of Wor1 constructs for Crystallization

Several different fragments of Wor1 could conceivably be chosen for crystallization. Although the WOPRa (1-101aa) or WOPRb (196-321aa) boxes could be screened, their lack of independent binding to DNA renders them unattractive (8). Taking a maximalist approach, one could attempt to crystallize full length Wor1. There are several reasons we decided against such an approach. First, expression levels for the full length Wor1 are much lower than seen for Wor1 1-321 constructs (data not shown), a major concern given the milligram quantities needed for crystal screens. The predicted unstructured nature of the C-terminal 400aa of Wor1 also is of concern, its disorder might hinder crystallization or make any diffraction pattern impossible to interpret. Since Wor1 1-321 is sufficient for both DNA binding and activation in addition to expressing well (8), we decided to focus on it as the target of crystallization efforts. Efforts to crystallize Wor1 have centered on fusing three affinity tags fused to Wor1 1-321aa. Two of these, 6xHis Wor1 1-321aa and 6xHis-Maltose Binding Protein (MBP) Wor1 1-321aa, have been discussed in Chapter 3 (8) and will be covered in more detail below. The third affinity tagged version, using Glutathione S-transferase (GST), has not proved successful.

Creation of GST Tagged Wor1 constructs

Three independent GST Wor1 1-321 have fusions have been created using the commercial pGEX-6P2 vector (GE Healthcare Life Sciences, Piscataway, NJ), the custom pDAL649 vector from David Low's lab at UC Santa Barbara (cloning performed by Rebecca Zordan), and a custom pET28 vector modified to contain GST that was a gift from Robert Stroud's lab at UC San Francisco. Expression levels ranged from very low to non-existent for these constructs, the reason for GST related expression problems is

not immediately clear. No further purification efforts were made with these constructs as the low expression makes purifying sufficient protein for crystallization difficult at best. As work with these plasmids was abandoned, I have not included maps or details of their construction.

Modification of MBP Wor1 1-321 for crystallization

Construction of the 6xHis and 6xHis-MBP Wor1 1-321aa fusions has been previously discussed and will not be repeated here (8). The 6xHis-MBP Wor1 1-321 construct used in Chapter 3 is not optimal for crystallization as the sequence linking MBP and Wor1 allows for a great deal of flexibility and thus may hinder the formation of crystals or the interpretation of diffraction patterns. To correct this, the plasmid was modified to remove the flexible regions based on consultations with the Stroud lab. In brief, a 63bp/21aa sequence between MBP and Wor1 (including the PreScission Protease Site) was removed. This was achieved by amplifying a ~500bp region of bRZ94 [the starting 6xHis-MBP Wor1 1-321 construct] that contained SacI and EcoRI restriction sites. The primers used for amplification were designed to omit the undesired 63bp region located near the SacI site. This fragment was then ligated into EcoRI/SacI digested bRZ94 to create the desired construct; the deletion of the 63bp was verified via sequencing. Maps of this plasmid (pMBL221) and the template plasmid (bRZ94) have been included in Figure 2. Plasmids are listed in Table 1 and primers used for this process are listed in Table 2.

Ni-NTA purification of 6xHis tagged Wor1 constructs

Although the initial purification steps for 6xHis Wor1 1-321aa and 6xHis-MBP Wor1 1-321aa using Ni-NTA agarose have been discussed in Chapter 3 (8), the protocol

has since been modified to increase yield and purity. Expression was performed in 2xYT media supplemented with 1mM $MgSO_4$ and 0.15% glucose. 5mL cultures were inoculated from frozen stocks, grown at 37°C overnight, diluted back 200-fold in fresh media, and grown to an approximate OD of 0.6 where they were moved to 25°C and induced with 0.1mM IPTG. Induced cultures were grown for 4 hours, pelleted, and frozen in liquid nitrogen before storage at -80°C. 1 liter cultures were grown for test purifications, roughly 4 liters of culture were grown for crystallization once purification conditions had been optimized.

For purification of the 6xHis tagged version as well as Ni-NTA agarose based purification of the 6xHis-MBP tagged version (illustrated in Figure 3a and 3b), cells were resuspended in lysis buffer (50mM NaH_2PO_4, 600mM NaCl, 10mM imidazole, pH 8.0) supplemented with 10mM β-mercaptoethanol (βME), 1mg/mL lysozyme, and 1 Complete Mini Protease inhibitor cocktail tablet, EDTA-free (Roche, Basel, Switzerland) per 10mL volume. Resuspended cells were then lysed using the Agard Lab C3 Emulisifier, with a 10-15 minute run at ~15,000 psi for each 1L of culture. Following this step, the lysate was sonicated (5 sets of 15 pulses, power setting ~4, duty cycle 40%) to ensure shearing of the bacterial DNA. This lysis method proved far more efficient than sonication alone as the insoluble pellet size was greatly reduced. The revised method also eliminated the DNAse step, resulting in improved purity. The lysate was centrifuged and the soluble fraction passed over a .45μm SFCA filter (Nalge Company, Rochester, NY). The filtered fraction was poured on a Ni-NTA agarose column (QIagen, Valencia, CA) at 4°C. After allowing the lysate to flow through, the column was washed with at least 5 column volumes of wash media (50mM NaH_2PO_4, 600mM NaCl, 20mM

imidazole, pH 8.0) followed by elution with roughly 10 column volumes of elution media (50mM NaH_2PO_4, 600mM NaCl, 250mM imidazole, pH 8.0). The elutate was concentrated on an Amicon-Ultra ultracel 10k centrifugal filter (Millipore, Billerica, MA) and passed over a Illustra NAP-5 or a NAP-25 column (GE Healthcare) into 20mM Tris pH 7.4, 150mM NaCl, 2mM DTT before further processing (for the MBP version use 20mM Tris pH 8, 50mM NaCl, 2mM DTT). During this and subsequent steps of the process, care should be taken when concentrating as 6xHis Wor1 1-321aa begins to crash out of solution at concentrations greater than 7-8mg/mL. It should be noted that although it is possible to get an untagged version of Wor1 1-321 following PreScission Protease cleavage of the 6xHis-MBP tagged version (bRZ94/pMBL206), the yield from this approach is low as much of the cleaved Wor1 crashes out of solution or sticks to amylase resin beads used to remove the cleaved MBP tag. Purity of the various constructs was verified on SDS-PAGE gels. Protein concentrations were determined based on comparison to a BSA dilution series on SDS-PAGE gels as well as a BCA Assay versus BSA standards or by automated UV-spectrometers.

Further Purification of 6xHis Wor1 1-321aa constructs

Given the high pI (~8.8) of 6xHis Wor1 1-321, it was purified over a SP cation exchange column (GE Healthcare Life Sciences, Piscataway, NJ) using 20mM Tris pH7.0 (0M and 1M NaCl solutions, 2mM DTT), loading with 50mM NaCl (14 minutes) followed by a 26 minute gradient from 250-350mM or 270-330mM and then a 10 minute 1M NaCl wash step. The Wor1 construct eluted during the 26-minute gradient step. Following concentration, the construct was then further purified using a S200 size

exclusion column (20mM Tris pH7.4, 150mM NaCl, 2mM DTT) run at ~0.3-0.5mL/minute. This protocol is illustrated in Figure 3a.

Amylose Resin Purification of 6xHis-MBP Wor1 1-321aa constructs

The 6xHis-MBP tagged Wor1 construct can also be purified using MBP-binding Amylose Resin (New England Biolabs Inc., Ipswich, MA). This approach has the advantage of allowing for the inclusion of EDTA for protease inhibition and results in higher purity after the initial purificiation step. The final amount of purified protein is not as high for Amylose Resin as for an equivalent amount of Ni-NTA, significantly larger amounts of chromatography medium are required for an equivalent yield. For this process (Figure 3c), cells are resuspended in 20mM Tris pH7.4, 600mM NaCl, 10mM EDTA, 10mM βME, 1mg/mL lysozyme, and 1 Complete Mini Protease inhibitor cocktail tablet (Roche, Basel, Switzerland) per 10mL volume. Resuspended cells were then lysed using the Agard Lab C3 Emulisifier, with a 10-15 minute run at ~15,000 psi for each 1L of culture. Following this step, the lysate was sonicated (5 sets of 15 pulses, power setting ~4, duty cycle 40%) to ensure shearing of the bacterial DNA. As discussed above, this lysis method proved far more efficient that sonication alone, the insoluble pellet size was significantly reduced. The lysate was centrifuged and the soluble fraction passed over a .45μm SFCA filter (Nalge Company, Rochester, NY). The lysate was then poured onto an Amylose Resin column at 4°C. The column was then washed with at least 15 column volumes of Buffer B (20mM Tris pH7.4, 600mM NaCl, 1mM EDTA, 10mM βME), followed by elution in at least 5 column volumes of the Buffer B supplemented with 20mM Maltose. The elutate was concentrated on an Amicon-Ultra ultracel 10k centrifugal filter (Millipore, Billerica, MA) and passed over a Illustra NAP-5

or a NAP-25 column (GE Healthcare) into 20mM Tris pH 8, 50mM NaCl, 2mM DTT before further processing.

Following Ni-NTA or Amylose Resin purification, 6xHis-MBP tagged Wor1 constructs were then purified over a 5mL Q Sepharose HP anion-exchange column (GE Healthcare Life Sciences, Piscataway, NJ) with Tris pH8.0 (0M and 1M NaCl solutions, 2mM DTT), loading with 50mM NaCl (14 minutes). This was followed by a two-step gradient, from 70 to 120mM in 21 minutes then from 120 to 200mM in 13 minutes, finishing with a 12 minute 1M NaCl wash step. We do note, however, that some 6xHis-MBP Wor1 1-321aa elutes during both 1M NaCl wash steps, it is not currently clear why this happens. Purified 6xHis-MBP-Wor1 appears soluble at 4°C with concentrations of at least 5mg/mL, higher concentrations are currently being tested (20mg/mL appears fine for at least a couple of hours). A S200 size exclusion chromatography step was performed as described above, although optional for protein alone it should be performed following incubation with DNA for crystal screens. This protocol is illustrated in Figures 3b and 3c. Concentrations of at least 12mg/mL are tolerated for extended periods of time, making 6xHis-MBP-Wor1 much easier to work with in that regards than 6xHis-Wor1.

Final Preparations for Crystal Screens

Initial screening for crystals incorporated several approaches. In short, 6xHis Wor1 1-321 was screened by itself and with two different DNA fragments. Although a structure of Wor1 bound to DNA would be preferred, Wor1 is being screened by itself as a backup. The first DNA fragment used for co-crystallization screens was a double-stranded 20bp oligo centered on the ~10bp Wor1 motif [forward strand sequence gcgcg

TTAAA GTTTAgcgcg] with 5bp flanks on each side designed to not bind Wor1. The second DNA fragment for co-crystallization uses the same 20bp as a core but contains 1bp sticky ends [Gcgcg TTAAA GTTTAgcgcg G/ CGC GCT AAA CTT TAA CGC GC C] to allow multiple DNA fragments to anneal to each other in the hopes that the DNA will form the foundation for crystallization. The DNA sequences were designed to present only one binding site so that a population of Wor1-DNA complexes does not include a mixture of different species that could complicate diffraction patterns. Depending on the results of further screens, it may prove necessary to try different DNA sequences [varying the motif, flanking DNA, fragment length, or overhang length for example]. Oligos were ordered from Integrated DNA Technologies, Inc. (Coralville, IA) using the 100nmol reaction size and the default purification. Oligos were resuspended in water at 500µM.

Prior to mixing with Wor1, complementary oligos were mixed, heated for 5 minutes at 95°C, and allowed to slowly cool back to room temperature. An excess concentration of DNA was mixed with 6xHis Wor1 1-321 on ice, then ran over a S200 size exclusion column (20mM Tris pH7.4, 150mM NaCl, 2mM DTT; run at ~0.3-0.5mL/ minute) to separate Wor1+DNA from unbound Wor1 or DNA. Samples were then concentrated prior to crystal screens.

Initial screens have been conducted using the Mosquito located in the first floor of Genentech Hall. Five QIAGEN 96 condition crystallization screens were used: Classics 1, PEGS 1, Nucleix, Procomplex, and PACT. When screening, it should be noted that the use of high concentrations of 6xHis Wor1 [~>7mg/mL, when screening without

DNA] often result in a majority of the wells crashing out within 24 hours. This is less of a problem when DNA is present.

For 6xHis-MBP-Wor1, samples were screened alone or with an excess of DNA with S200 purification and concentration performed as previously described. Initial screens have been conducted using the Mosquito located in the first floor of Genentech Hall with five QIAGEN 96 condition crystallization screens: Classics 1, PEGS 1, Nucleix, Procomplex, and PACT. We did not observe precipitation in more than a third of wells for any condition, suggesting that higher protein concentrations that initially used may be needed. Additionally, we noticed a large number of wells with phase separation; as a result a shift to 20mM Tris pH7.4, 50mM NaCl, 1mM DTT might be helpful in further screens.

References

1. Lohse MB, Johnson AD (2009) White-opaque switching in *Candida albicans*. *Curr Opin Microbiol* 12:650-654.
2. Morschhäuser J (2010) Regulation of white-opaque switching in Candida albicans. *Med Microbiol Immunol* 199:165-172.
3. Soll DR (2009) Why does *Candida albicans* switch? *FEMS Yeast Res* 9:973-989.
4. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* 5:e256.
5. Zordan RE, Galgoczy DJ, Johnson AD (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci U S A* 103:12807-12812.
6. Srikantha T, et al. (2006) TOS9 regulates white-opaque switching in *Candida albicans*. *Eukaryot Cell* 5:1674-1687.
7. Huang G, et al. (2006) Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 103:12813-12818.
8. Lohse MB, Zordan RE, Cain CW, Johnson AD (2010) Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in Candida albicans. *Proc Natl Acad Sci U S A* 107:14105-14110.
9. Nguyen VQ, Sil A (2008) Temperature-induced switch to the pathogenic yeast form of *Histoplasma capsulatum* requires Ryp1, a conserved transcriptional regulator. *Proc Natl Acad Sci U S A* 105:4880-4885.

10. Caspari T (1997) Onset of gluconate-H+ symport in *Schizosaccharomyces pombe* is regulated by the kinases Wis1 and Pka1, and requires the gti1+ gene product. *J Cell Sci* 110:2599-2608.

11. Michielse CM, et al. (2009) The Nuclear Protein Sge1 of Fusarium oxysporum Is Required for Parasitic Growth. *PLoS Pathog* 5:e1000637.

12. Kunitomo H, Sugimoto A, Wilkinson CR, Yamamoto M (1995) *Schizosaccharomyces pombe* pac2+ controls the onset of sexual development via a pathway independent of the cAMP cascade. *Curr Genet* 28:32-38.

13. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.

14. Kelley LA, Sternberge MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc* 4:363-371.

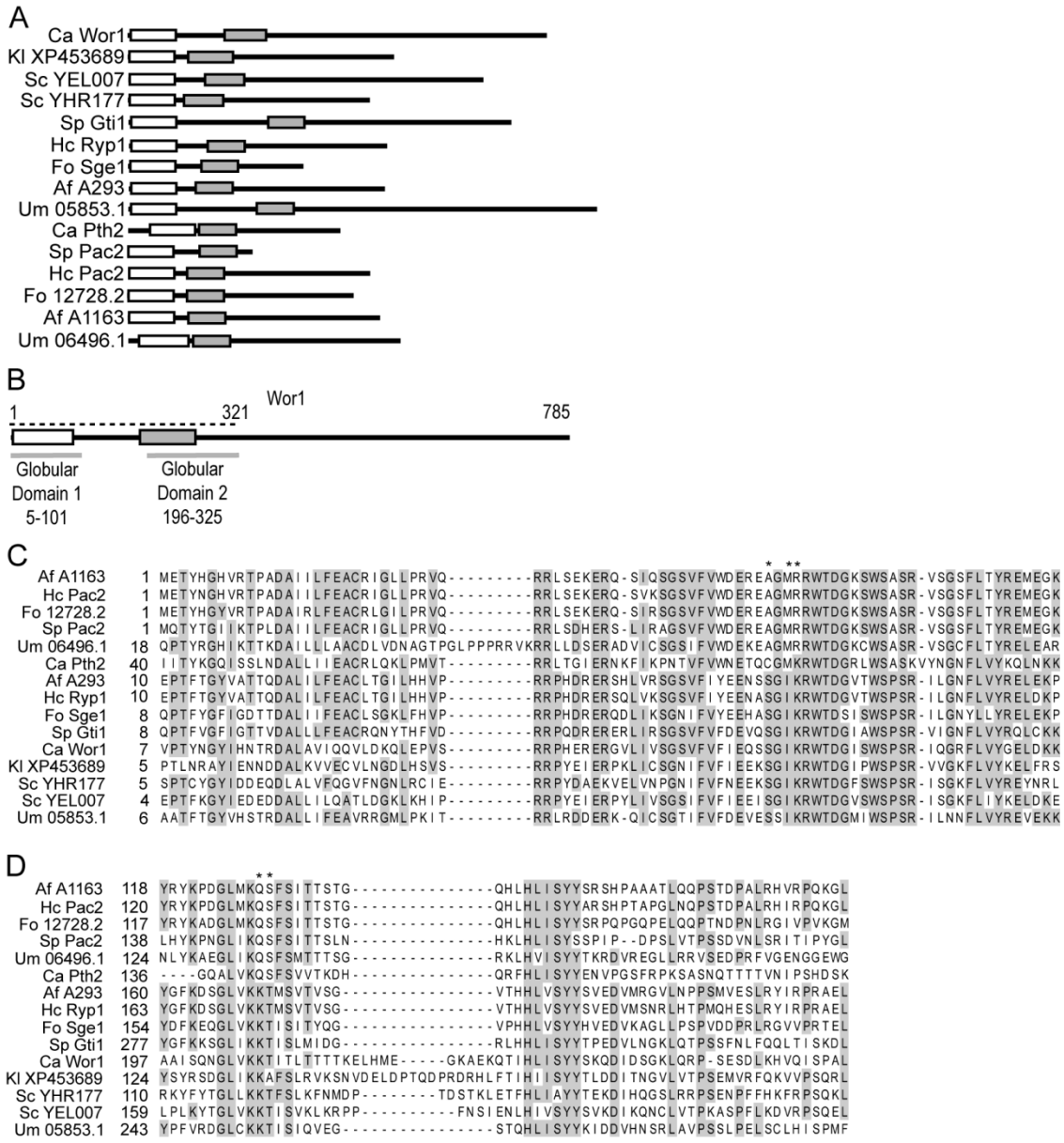


Figure 1: Wor1 is a member of a conserved family of fungal proteins. (a) Alignment of Wor1 homologs across 8 fungal species for 15 representative members of this protein family, *C. albicans* Wor1 and Pth2, *S. cerevisiae* YEL007 and YHR177, *F. oxysporum* Sge1 and 12728.2, *H. capsulatum* Ryp1 and Pac2, *S. pombe* Gti1 and Pac2, *Kluyveromyces lactis* XP-453689, *Ustilago mayis* UM05853.1 and UM06496.1, and *Aspergillus fumigatus* A293 and A1163. Wor1, Gti1, YEL007, YHR177, Ryp1, A293, UM05853.1, XP_453689, and Sge1 all represent one distinct set of this protein family

whereas Pth2, Pac2, HcPac2, A1163, UM06496.1, and 12728.2 represent the second set. Wor1 is 785aa long, the other proteins are drawn to scale. The two conserved domains in this family of proteins are indicated by clear (WOPRa) and gray (WOPRb) filled boxes respectively. Family members are representative of a wide range of fungal species and include those that have been experimentally characterized. (b) Location of the two globular domains of Wor1 predicted by ELM and their positions relative to the WOPRa (clear) and WOPRb (gray) domains conserved in this family of proteins. The region encompassed by the Wor1 1-321aa construct is also illustrated (dashed line). (c) Alignment of WOPRa from the same 15 family members. This region corresponds to residues 7-90 of Wor1. (d) Alignment of WOPRb from the same members of this family. This region corresponds to residues 197-262 of Wor1. In both panels, residues that differ between the two distinct family members are indicated with an asterisk. This figure originally appeared in Lohse *et al*. 2010 (8).

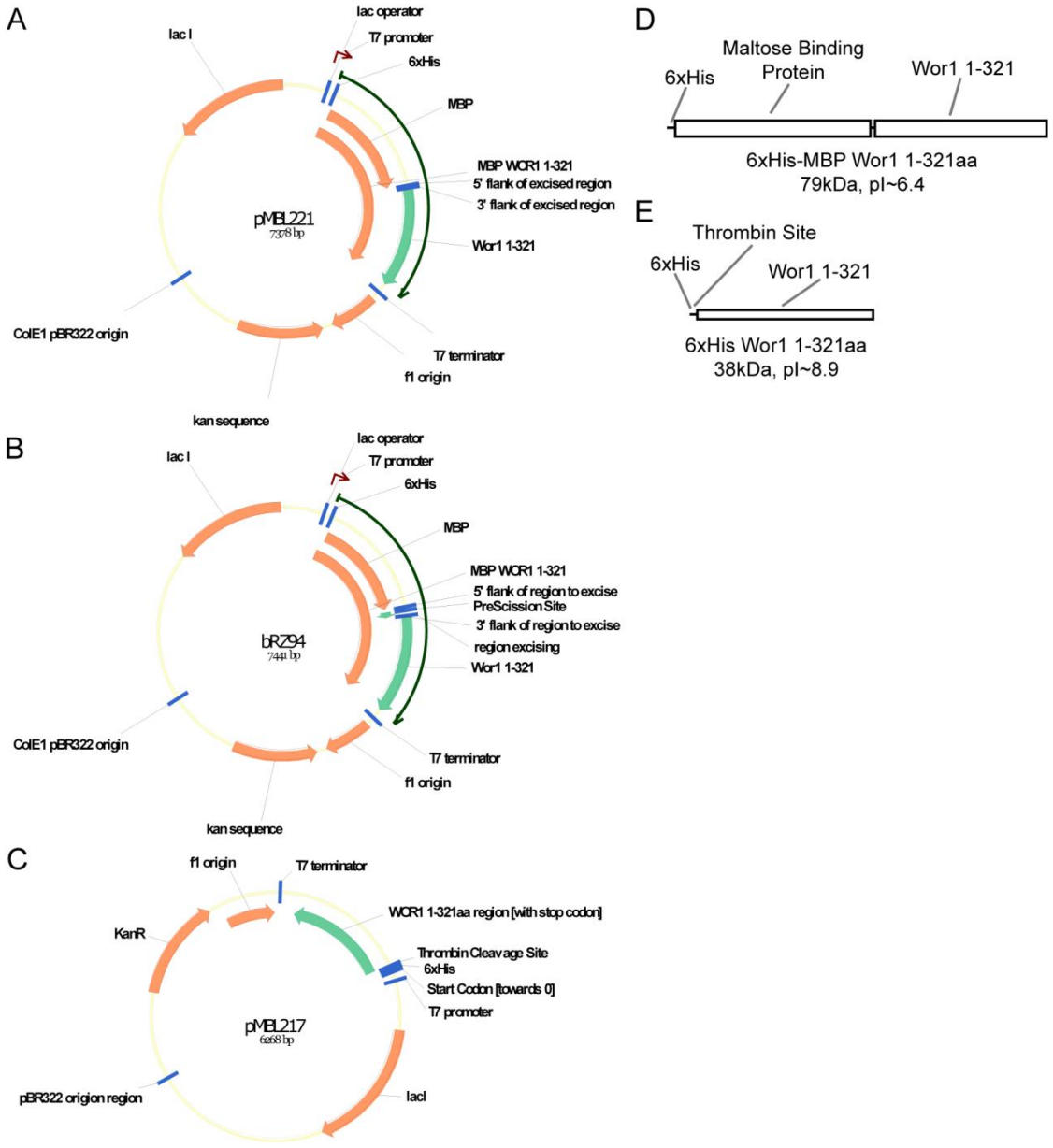


Figure 2: Plasmid and protein maps of affinity tagged Wor1 constructs. [a] Plasmid map of pMBL221, 6xHis-MBP Wor1 1-321aa for crystalization. [b] Plasmid map of bRZ94, the starting 6xHis-MBP Wor1 1-321aa template. [c] Plasmid map of pMBL217, 6xHis Wor1 1-321aa. Plasmids in [a-c] were all based in some way on pET28. [d] Protein map of 6xHis-MBP Wor1 1-321aa, as expressed from pMBL221. [e] Protein map of 6xHis Wor1 1-321aa, as expressed from pMBL217.

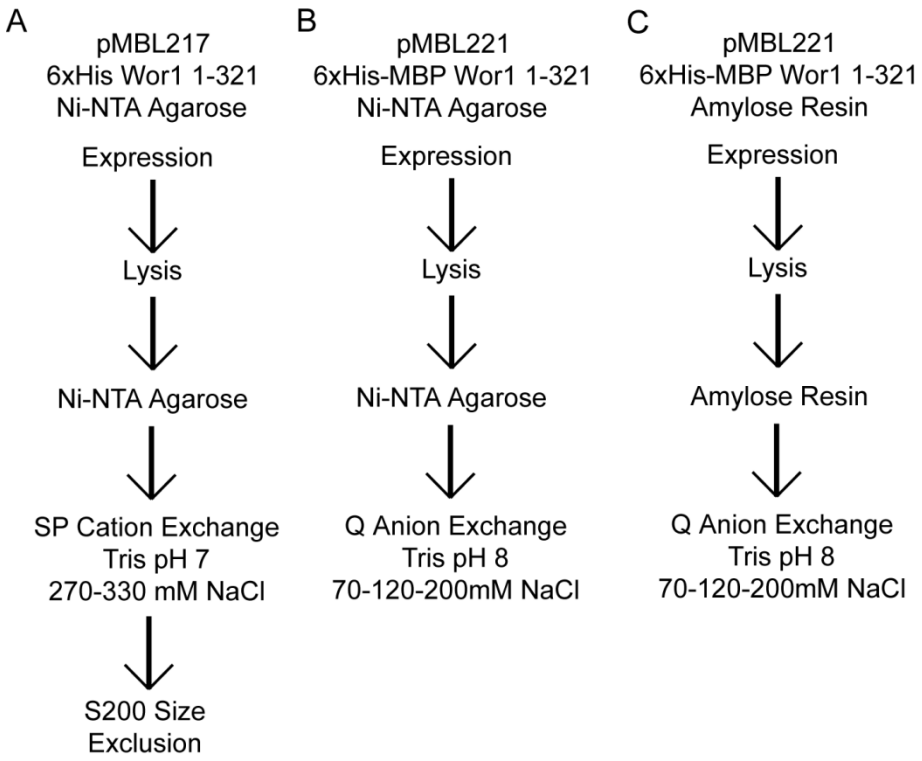


Figure 3: Purification schemes for various Wor1 constructs. [a] Scheme for 6xHis Wor1 1-321aa. [b] Ni-NTA based scheme for 6xHis-MBP Wor1 1-321aa. [c] Amylose resin based scheme for 6xHis-MBP Wor1 1-321aa. Plasmids pMBL217 and pMBL221 have been transformed in to BL21 cells for expression as pMBL219 and pMBL224, respectively.

Description	Name	BL21 stock	Reference
6xHis-MBP Wor1 1-321aa	pMBL221	pMBL224	This Study
6xHis-MBP-PreScission Wor1 1-321aa	bRZ94	pMBL206	Ch3, Reference 1
pET28 6xHis-Wor1 1-321aa	pMBL217	pMBL219	Ch3, Reference 1

References

1. Lohse MB, Zordan RE, Cain CW, Johnson AD. *Proc Natl Acad Sci USA* (2010) 107:14105-10

Table 1: List of plasmids used in this study.

Name	Description	Sequence	Reference
MBP Linker Removal			
MBL1018	SacI-5' end of MBP-Wor1 link	taattcgagctcgaacaacaacaacaataacaataacaacaacTCT AAT TCA AGT ATA GTC CCT ACA TA	This study
MBL1019	3' end of fragment-EcoRI	GGA CCT GAT CGA GAA TTC T	This study
Crystalization			
Name	Description	Sequence	Reference
MBL1062	Motif, no overlap (forward)	Gcgcg TTAAA GTTTAgcgcg	This study
MBL1063	Motif, no overlap (reverse)	CGC GCT AAA CTT TAA CGC GC	This study
MBL1066	Motif, 1bp overlap (forward)	Gcgcg TTAAA GTTTAgcgcg G	This study
MBL1067	Motif, 1bp overlap (reverse)	CGC GCT AAA CTT TAA CGC GC C	This study

Table 2: List of primers used in this study.

Appendix 3:

Role of Wor1 in Expression of Opaque Nature of a Cell:

Characterization of a *WOR1/wor1* strain

Introduction

The human fungal pathogen *C. albicans* differs from the great majority of fungal species by possessing the ability to switch between two distinct types of cells, called white and opaque (1). These two cell types are both stable and heritable, white cells will tend to give rise to white cells and opaque cells to opaque cells (2). The transcription factor Wor1 is the master regulator of this phenomenon and is necessary and sufficient for the establishment of the opaque cell type (Figure 1) (3-6). As discussed in Chapter 3, we know that Wor1 binds DNA directly in a sequence-specific manner and that it can activate transcription *in vivo* (7). Based on several studies, we have a good idea of the role of Wor1 in the transcriptional circuit regulating white-opaque switching. Genome wide ChIP-chip indicates that Wor1's role is not limited to white-opaque switching, it binds hundreds of promoters in opaque cells (3). These promoters are for both opaque-enriched and white-enriched genes, as well as promoters of genes not enriched in either cell type. Given Wor1 binding at white and opaque-enriched genes, it appears that it functions both as a activator [opaque-enriched genes] and a repressor [white-enriched genes] although the latter has yet to be formally proven in an *in vivo* assay. Recently published RNA-Sequencing (RNA-Seq) analysis indicates that approximietely 65 percent of promoters bound by Wor1 are for white or opaque-enriched genes (8), suggesting that Wor1 plays an important role of the expression of the characteristics of an opaque cell and the repression of white characteristics.

To better understand Wor1's function and identify targets for future repression studies, we wanted to examine the role of Wor1 in the establishment of the "opaque" character of a cell. A study like this would normally be conducted by looking at changes

in transcript levels in a strain with the gene of interest, in this case *WOR1*, deleted. In the case of *WOR1*, however, the absolute requirement of *WOR1* for the formation of opaque cells limits the information that could be gathered by such an approach (4-6). Studies with a *WOR1/wor1* heterozygous deletion strain, discussed in Chapter 4, indicated an approach that avoids the problems inherent in the *WOR1* homozygous deletion. The *WOR1/wor1* strain is capable of forming opaque cells, but temperature shift assays indicate that these *WOR1/wor1* opaques are less stable than a wild type opaque (9). This raises the possibility that expression of a number of genes might be affected by the deletion of one copy of *WOR1*. Given this possibility, we used microarrays to determine the genome-wide effects of deleting one copy of *WOR1* in both white and opaque cells.

Methods and Materials

Sample Growth and Treatment

Growth conditions are similar to that reported in Chapter 4 (9). Both wild type and *WOR1/wor1* opaque and white cells were grown in synthetic complete media supplemented with 2% glucose and 100μg/mL uridine (SD+aa+Urd). Room temperature overnight cultures were started from single colonies with no sectors on SD+aa+Urd plates. Cultures were diluted back to OD_{600}=0.02 in the morning and allowed to grow back to OD_{600}=0.1-0.2 at room temperature. Cells were then diluted [OD_{600}=0.00012 for opaques and OD_{600}=0.00002 for whites] in 500mL SD+aa+Urd in 2.8L flasks and grown at 25°C overnight. When cultures reached OD_{600}=0.1, samples were harvested. Two independent samples were harvested for each condition, except for the *WOR1/wor1* opaque which had three. Samples were harvested at the same time as those for the commitment point analysis reported in Chapter 4.

Microarrays

This protocol is based on the one presented in Chapter 4 (9). Total RNA was extracted from samples via buffered phenol extractions. Following cDNA synthesis, samples were dye coupled to both Cy3 and Cy5 in separate reactions. A pooled reference was created from the Cy5 labeled fractions. Equal amounts of Cy3 labeled sample and Cy5 labeled reference were hybridized overnight to custom designed Agilent 8*15k Microarrays (AMADID #020166, ArrayExpress accession number A-MEXP-1805) containing a minimum of 2 independent probes for each ORF. Following hybridization, slides were processed and then scanned on a Genepix 4000A or 4000B Axon Instrument Scanner (Axon/Molecular Devices, Sunnyvale, CA USA). Arrays were gridded using GenePix Pro version 5.1. Global Lowess normalization analysis was performed for each array using a Goulphar script (10) for R (The R Foundation for Statistical Computing). Normalized data was then transformed before collapsing all points for an ORF to the median value using a custom java script written by Oliver Homann. The data for this experiment have been uploaded to ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/) as part of accession number E-MEXP-2834.

Results

Comparison of Microarray and RNA-Sequencing Enrichment Data

We first compare the microarray derivied white and opaque-enriched gene set presented in Chapter 4 (209 genes, 145 opaque and 64 white-enriched) to the RNA-Seq developed enriched gene set from (8). The RNA-Seq set contains 647 opaque and 663 white enriched items, limiting this set to annotated open reading frames results in 416

opaque and 400 white-enriched genes. A total of 165 genes (79%) from the microarray enriched list are on this list (Table 1). The larger number of enriched genes identified by RNA-Seq reflects the greater sensitivity of that technique; genes from the microarray set rank among the most white or opaque enriched in the RNA-Seq set. Looking at the 165 genes on both lists, there is a good correlation of enrichment levels (R^2=0.784, Figure 2) between the two data sets. Given the greater sensitivity of the deep sequencing data set, we will use it to define enriched genes for the purpose of this analysis.

Role of Wor1 in White cells

In order to determine the effect of deleting one copy of *WOR1* in white cells, we transformed white *WOR1/wor1* samples against white wild type samples (two samples each). A comparison of a *WOR1/wor1* white strain to the wild type white reveals that only 13 genes changed by at least 1.41-fold between these strains (minimum difference of log2=0.5 in both transformations, Table 2). Of these 13 genes, 10 are up-regulated and three are down-regulated. One of the down-regulated genes, *HIS1*, reflects the metabolic difference between the two strains (*his1/his1* versus *his1/his1/HIS1*) and is thus an artifact. Both down-regulated genes (*WOR1*, orf19.3924) are on the list of opaque-enriched genes. *WOR1* levels have decreased by approximately 2-fold, as would be expected when one copy is deleted.

Looking at the 10 up-regulated genes, 6 are on this list of enriched genes (4 white-enriched, 2 opaque-enriched). It is not clear why two opaque-enriched genes (*ARO10* (orf19.1847), orf19.341) would be up-regulated in this strain, perhaps this is an artifact of the *his1* auxotrophy. Given the small number of genes in this set and the subtle changes

observed, it is hard to draw conclusions from this set and it may prove more informative to use a *wor1/wor1* strain to look for effects in white cells.

Role of Wor1 in Opaque cells

In order to determine the effect of deleting one copy of *WOR1* in opaque cells, we transformed opaque *WOR1/wor1* samples against opaque wild type samples. These two populations had an average of 3 and 1.7 percent white cells respectively, thus the minority white cell population should not skew this analysis.

A comparison of the *WOR1/wor1* opaque strain to the wild type opaque reveals that 121 genes changed by at least 1.41-fold between these strains (Minimum Difference of log2=0.5 in all three transformations, Table 3). We found 71 genes up-regulated and 50 genes down-regulated. As in white cells, the inclusion of *HIS1* reflects the metabolic difference between the two strains. Of the 71 up-regulated genes, 25 are white-enriched genes while 8 are opaque-enriched (Figure 3, combined 33 of 71, 46%). As before, it is not clear why any opaque enriched genes would be up-regulated in a *WOR1/wor1* background. The set of 71 up-regulated genes contains 8 of the 10 up-regulated in white cells (including opaque-enriched *ARO10* and orf19.341), suggesting that these genes may change as part of be a general response to the loss of *WOR1* or to the *his1* auxotrophy.

Looking at the 50 down-regulated genes, 35 are on the enriched gene list and all of these are opaque-enriched (Figure 3b, 70%). Genes on the enriched list trend towards the most down-regulated genes in the *WOR1/wor1* background, suggesting a link between *WOR1* and the establishment of a fully "opaque" cell.

Transcript levels of *WOR1* in the *WOR1/wor1* opaque ranged from 70 to 100 percent of wild type levels, rather than the 50 percent that might be expected when one

copy was deleted. This result suggests some sort of feedback mechanism or dosage compensation in the regulation of *WOR1* in opaque cells. This observation corresponds with our observations of Wor1 levels in the *WOR1/wor1* opaque strain, which were around 70 to 80 percent of the wild type values.

Comparison of Expression Level Change and Wor1 binding

In an effort to create a list of genes of interest for future studies of Wor1 dependent activation or repression, we compared the list of genes changing in the *WOR1/wor1* opaque to the Wor1 binding site list presented in Chapter 3 (7) that Aaron Hernday developed re-analyzing previously published data (3). We found that 16 of the 71 up-regulated genes (22%) and 11 of the 50 down-regulated genes (22%) had Wor1 binding sites (Table 3). Promoters of these genes, identified in Table 3, are potential targets for future *in vitro* or *in vivo* studies with Wor1. Given the low percentage of genes bound by Wor1, much of the transcriptional change in *WOR1/wor1* cells may depend on one or more intermediate transcription factors.

Transcription Factors Changing in the *WOR1/wor1* background

We then examined the list of genes changing in the *WOR1/wor1* opaque for known transcription factors, in order to identify factors that might contribute to changes in the genes not bound by Wor1. Two transcription factors were down-regulated: orf19.4972 and *ZCF25* (orf19.4568). Both are opaque-enriched and Wor1 bound, suggesting they may play a role in the expression of opaque-enriched genes and/or contribute to opaque cell stability. Five transcription factors were up-regulated: *STP4* (orf19.909), *FCR1* (orf19.6817), *TYE7* (orf19.4941), *CRZ2* (orf19.2356), and orf19.5026. Of these, *STP4*, *FCR1*, and *TYE7* are white-enriched and perhaps can explain the up-

regulation of genes not bound by Wor1. A more definitive answer to this question may not be possible until these transcription factors are better characterized.

Conclusions/Future Directions

Not surprisingly given Wor1's association with hundreds of promoters, we found that Wor1 contributes to the full expression of a cell's "opaqueness". Interestingly, we found that more than one-hundred genes changed in a reproducible manner in opaque cells with one copy of *WOR1* deleted. Many of these genes (>75%) are not bound by Wor1, suggesting that these changes are an indirect effect caused by one or more undetermined intermediates. An analysis of changes in a white *wor1/wor1* strain may also be of interest at some point in the future, especially if performed using a technique that provides greater sensitivity than this microarray-based analysis. It may also be of interest to further examine the set of transcription factors changing in the *WOR1/wor1* background for the gain or loss of the opaque cell characteristics. Finally, this analysis suggests a number of candidate promoters for future attempts to develop an *in vivo* transcriptional repression assay.

References

1. Lohse MB, Johnson AD (2009) White-opaque switching in *Candida albicans*. *Curr Opin Microbiol* 12:650-654.
2. Slutsky B, et al. (1987) "White-opaque transition": a second high-frequency switching system in *Candida albicans*. *J Bacteriol* 169:189-197.
3. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* 5:e256.
4. Zordan RE, Galgoczy DJ, Johnson AD (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci U S A* 103:12807-12812.
5. Huang G, et al. (2006) Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 103:12813-12818.
6. Srikantha T, et al. (2006) TOS9 regulates white-opaque switching in *Candida albicans*. *Eukaryot Cell* 5:1674-1687.
7. Lohse MB, Zordan RE, Cain CW, Johnson AD (2010) Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in Candida albicans. *Proc Natl Acad Sci U S A* 107:14105-14110.
8. Tuch BB, et al. (2010) The Transcriptomes of Two Heritable Cell Types Illuminate the Circuit Governing Their Differentiation. *PLoS Genetics* 6:e1001070.

9. Lohse MB, Johnson AD (2010) Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans. *Mol Microbiol* 78:331-343.

10. Lemoine S, Combes F, Servant N, Le Crom S (2006) Goulphar: rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics* 7:467.

11. Homann OR, Johnson AD (2010) MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol* 8:49.

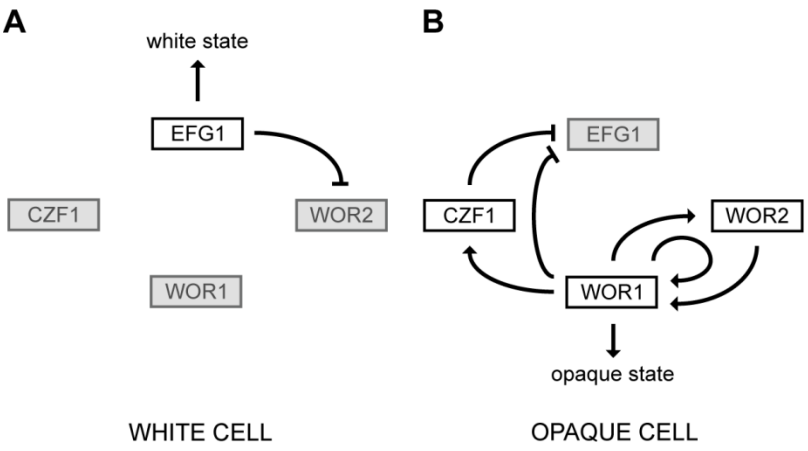


Figure 1: Working model of the white-opaque regulatory circuit and its activity in white and opaque cells. (a) In white cells, *EFG1* represses *WOR1* indirectly through *WOR2* to maintain white cell identity. (b) In opaque cells, *WOR1*, *WOR2*, and *CZF1* establish a series of positive feedback loops, maintaining opaque cell identity and repressing *EFG1*. Up-regulated genes and active relationships are indicated in black. Down-regulated genes are indicated in gray. Arrows and bars represent activation and repression respectively. Figure adapted from Zordan et al. 2007, initially published in Lohse et al. 2010.

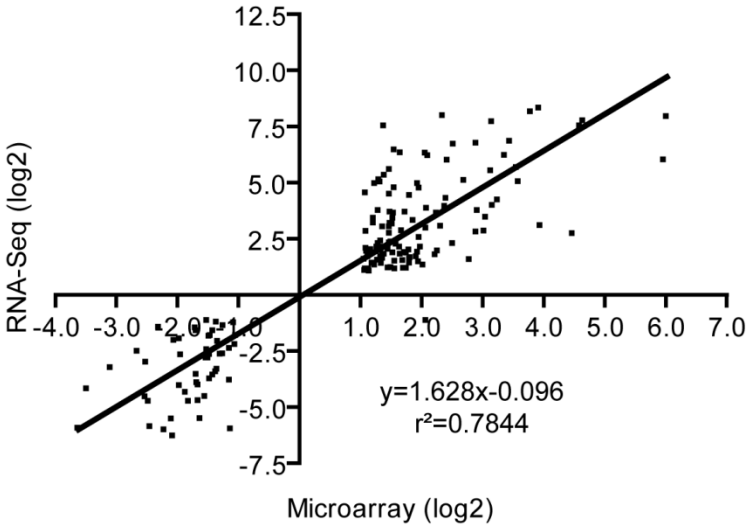


Figure 2: White or opaque-enrichment value as determined by microarray or deep sequencing. Enrichment values (log 2 scale) are plotted for a set of 165-genes in both the microarray (x-axis) and RNA-Sequencing (y-axis) derived enriched gene sets. Values greater than zero represent opaque-enrichment and less than zero represent white-enrichment. The trendline represents the linear regression of the two data sets.

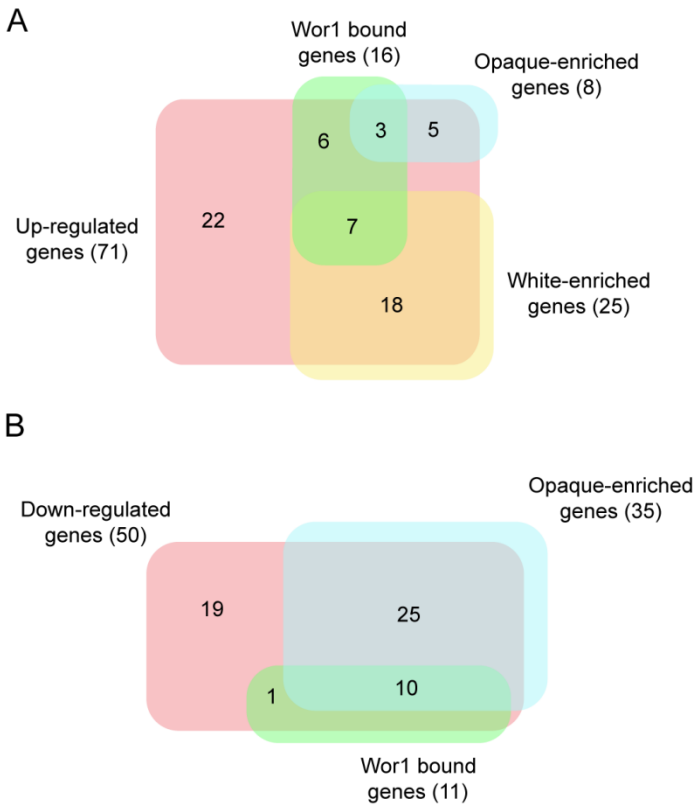


Figure 3: Relationships between expression changes in an opaque *WOR1/wor1* strain and white or opaque-enrichment in wild type cells. [a] Overlap between genes up-regulated in an opaque *WOR1/wor1* strain (pink), binding by Wor1 (green), and white- (yellow) or opaque-enrichment (blue) in wild type cells. [b] Overlap between genes down-regulated in an opaque *WOR1/wor1* strain (pink), binding by Wor1 (green), and opaque-enrichment (blue) in wild type cells. Boxes are not drawn to scale.

Table 1: List of 165 genes on both the microarray and RNA-Sequencing derived enriched gene sets.

Gene	Name	Opaque:White Enrichment, Wild Type, Microarray (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)
orf19.2833	*PGA34*	3.93	8.29
orf19.4972		3.79	8.14
orf19.4505	*ADH3*	2.35	7.97
orf19.4934	*OP4*	6.02	7.92
orf19.2460		4.65	7.74
orf19.1370		3.16	7.69
orf19.176	*OPT4*	4.60	7.50
orf19.2941	*SCW4*	1.39	7.50
orf19.893	*PGA17*	3.45	6.82
orf19.5144	*PGA28*	2.90	6.73
orf19.467		2.53	6.70
orf19.535	*RBR1*	1.56	6.43
orf19.1239		1.66	6.31
orf19.3926		2.07	6.29
orf19.4072	*IFF6*	3.37	6.19
orf19.2457		2.11	6.18
orf19.3902		5.97	5.99
orf19.742	*ALD6*	2.43	5.98
orf19.4884	*WOR1*	3.56	5.65
orf19.1171		1.48	5.56
orf19.1541		3.14	5.51
orf19.3924		1.40	5.31
orf19.1655	*PXP2*	1.31	5.10
orf19.5307	*JEN2*	2.70	5.08
orf19.6805		3.59	5.02
orf19.3121	*GST1*	1.33	5.00
orf19.7111.1	*SOD3*	1.24	4.94
orf19.2247		1.94	4.93
orf19.2156	*NAG1*	1.56	4.75
orf19.6277		1.97	4.74
orf19.3897		1.09	4.52
orf19.4607		1.48	4.46
orf19.7094	*HGT12*	1.81	4.41
orf19.7434	*GLG2*	2.41	4.28

Gene	Name	Opaque:White Enrichment, Wild Type, Microarray (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)
orf19.2253		3.25	4.21
orf19.4082	*DDR48*	3.17	3.96
orf19.2062	*SOD4*	2.39	3.92
orf19.4459		2.09	3.84
orf19.7053	*GAC1*	2.92	3.74
orf19.2701		1.30	3.74
orf19.4979	*KNS1*	1.74	3.66
orf19.1539		1.49	3.66
orf19.3612	*PST2*	2.25	3.62
orf19.7521	*REP1*	1.55	3.62
orf19.3282	*BMT3*	3.06	3.43
orf19.3338		1.54	3.39
orf19.460	*CEK2*	1.22	3.39
orf19.532	*RBR2*	1.87	3.29
orf19.4590	*RFX2*	1.22	3.18
orf19.413		1.49	3.16
orf19.3337		1.53	3.13
orf19.4279	*MNN1*	3.95	3.07
orf19.2475	*PGA26*	2.32	3.03
orf19.799	*STE4*	1.37	3.01
orf19.6784	*PGA32*	1.48	3.00
orf19.2886	*CEK1*	2.08	2.95
orf19.3127	*CZF1*	1.68	2.83
orf19.3746	*IFC1*	3.03	2.82
orf19.4476		1.10	2.80
orf19.655	*PHO84*	2.90	2.78
orf19.5140		1.48	2.72
orf19.3668	*HGT2*	4.48	2.71
orf19.4381	*VTC3*	1.97	2.54
orf19.3984		1.34	2.38
orf19.4651	*PGA53*	1.51	2.33
orf19.6830		1.48	2.29
orf19.1709		2.52	2.27
orf19.7314	*CDG1*	1.30	2.27
orf19.4666		1.59	2.24
orf19.2724		1.65	2.11
orf19.5138	*IFA21*	1.35	2.11

Gene	Name	Opaque:White Enrichment, Wild Type, Microarray (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)
orf19.4530.1		1.28	2.11
orf19.3363	*VTC4*	1.99	2.10
orf19.4090		1.10	2.05
orf19.5170	*ENA21*	1.46	2.00
orf19.1600		1.18	1.99
orf19.5305	*RHD3*	1.80	1.98
orf19.1313	*CDR3*	1.31	1.96
orf19.4066		2.27	1.94
orf19.1277		1.11	1.92
orf19.4157	*SPS20*	1.94	1.88
orf19.4749		1.73	1.86
orf19.1830		1.54	1.85
orf19.5760	*IHD1*	1.76	1.84
orf19.2125		1.66	1.79
orf19.2124		1.40	1.79
orf19.899		2.23	1.76
orf19.5455		1.37	1.75
orf19.6254	*ANT1*	1.92	1.68
orf19.6570	*NUP*	1.34	1.68
orf19.3931	*SFC1*	1.13	1.68
orf19.5723	*POX1*	1.82	1.64
orf19.6066		1.37	1.61
orf19.5992	*WOR2*	1.81	1.56
orf19.4056	*GAT2*	1.07	1.56
orf19.6983		2.79	1.55
orf19.5267		1.29	1.52
orf19.5615	*AYR2*	1.70	1.51
orf19.5921		1.57	1.48
orf19.2030		1.96	1.45
orf19.3330.3	*POX18*	1.35	1.43
orf19.3415	*PTK2*	1.20	1.40
orf19.6445	*ECI1*	1.81	1.39
orf19.3733	*IDP2*	1.35	1.38
orf19.5994	*RHB1*	1.24	1.31
orf19.4255	*ECM331*	2.03	1.30
orf19.1562		1.51	1.19
orf19.4833	*MLS1*	1.46	1.18

Gene	Name	Opaque:White Enrichment, Wild Type, Microarray (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)
orf19.6229	*CAT1*	1.76	1.17
orf19.6143		1.71	1.17
orf19.3663	*PHO91*	1.57	1.14
orf19.4570		1.09	1.14
orf19.2841	*PGM2*	1.14	1.13
orf19.2018.1		1.08	1.06
orf19.5614		1.15	1.04
orf19.4654		2.08	-1.16
orf19.2531	*CSP37*	-1.51	-1.16
orf19.6874		-1.35	-1.23
orf19.7219	*FTR1*	-1.07	-1.27
orf19.3869		-1.24	-1.39
orf19.6837		-2.30	-1.47
orf19.1153	*GAD1*	-1.70	-1.50
orf19.2244		-1.41	-1.62
orf19.1415	*FRE10*	-1.68	-1.62
orf19.3713		-1.04	-1.83
orf19.6311		-1.45	-1.83
orf19.5673	*OPT7*	-1.47	-1.86
orf19.4889		-1.96	-1.98
orf19.610	*EFG1*	-2.05	-2.05
orf19.6937	*PTR2*	-1.33	-2.09
orf19.2028	*MXR1*	-1.05	-2.24
orf19.3040	*EHT1*	-1.14	-2.40
orf19.1152		-1.51	-2.47
orf19.6202	*RBT4*	-2.65	-2.54
orf19.84	*CAN3*	-1.25	-2.65
orf19.946	*MET14*	-1.29	-2.66
orf19.5063		-1.46	-2.68
orf19.2948	*SNO1*	-1.94	-2.69
orf19.1906		-1.49	-2.83
orf19.2881	*MNN4*	-1.52	-2.83
orf19.4737	*TPO3*	-2.51	-3.02
orf19.2947	*SNZ1*	-3.09	-3.27
orf19.2989		-1.34	-3.33
orf19.6770		-1.36	-3.45
orf19.54	*RHD1*	-1.68	-3.57

Gene	Name	Opaque:White Enrichment, Wild Type, Microarray (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)
orf19.4907		-1.41	-3.60
orf19.2898		-1.46	-3.77
orf19.1442	*PLB4.5*	-1.14	-3.82
orf19.24	*RTA2*	-1.67	-3.94
orf19.851		-1.65	-4.03
orf19.85	*GPX2*	-1.96	-4.07
orf19.3548.1	*WH11*	-3.48	-4.21
orf19.3448		-1.86	-4.36
orf19.4653		-1.54	-4.55
orf19.1979	*GIT1*	-2.52	-4.57
orf19.2726		-1.65	-4.73
orf19.5025	*MET3*	-1.81	-4.77
orf19.4212	*FET99*	-2.47	-4.77
orf19.4195.1	*FCA1*	-1.62	-5.53
orf19.23	*RTA3*	-2.09	-5.55
orf19.1189		-2.44	-5.89
orf19.868	*ADAEC*	-3.62	-5.96
orf19.4215	*FET34*	-1.12	-6.00
orf19.4679	*AGP2*	-2.21	-6.03
orf19.4769	*IPT1*	-2.07	-6.31

Table 1: List of 165 genes on both the microarray and RNA-Sequencing derived enriched gene sets. Enrichment values for both are on a log2 scale and genes are ordered based on enrichment in the RNA-Seq data set.

Gene	Name	*WOR1/wor1*: Wild Type (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)	Wor1 Peak LOD Score [maximum 4.05]
orf19.4335	TNA1	1.65375	-1.67	
orf19.1354	UCF1	1.0205	-2.49	
orf19.1375	LEU42	0.8075		
orf19.4773	AOX2	0.80325	-2.2	4.038290209
orf19.1847	ARO10	0.79	1.11	
orf19.341	-	0.802	3.67	
orf19.6994	BAT22	0.738		2.060792023
orf19.6899	-	0.68025	-2.16	
orf19.7498	LEU1	0.67		
orf19.6659	GAP6	0.6105		
orf19.3924	-	-0.046	5.31	
orf19.4884	WOR1	-1.085	5.65	4.047563757
orf19.4026	HIS1	-5.2675		

Table 2: List of 13 genes changing at least 1.41 fold (log2=0.5) in a white *WOR1/wor1* strain. Transcript change values for both are on a log2 scale, white or opaque-enrichment values are taken from RNA-Sequencing results [ref] and are in a log2 scale. Values of enrichment less that 2 (log2=1) are not shown. Wor1 motif LOD scores are out of a maximum value of 4.05, a score of 2.86 explains 63% of Wor1 binding events while only occurring in 18% of an equivalently sized set of control regions. Motif LOD scores evaluated in MochiView (11).

Table 3: List of 121 genes changing at least 1.41 fold (log2=0.5) in an opaque *WOR1/wor1* strain.

Gene	Name	*WOR1/wor1*: Wild Type (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)	Wor1 Peak LOD Score [maximum 4.05]
orf19.6899	-	1.9065	-2.16	
orf19.909	STP4	1.71	-3.75	
CaalfMp06	ATP6	1.6525		
orf19.341	-	1.445	3.67	
CaalfMp13	NAD5	1.3985		
orf19.4769	IPT1	1.39	-6.31	3.74
orf19.5735.3	-	1.3685	4.3	
orf19.6249	HAK1	1.328	-2.27	
orf19.6817	FCR1	1.319	-3.88	
orf19.3378	-	1.317	2.62	
orf19.6770	-	1.291	-3.45	4.04
orf19.5025	MET3	1.2425	-4.77	
orf19.1847	ARO10	1.2145	1.11	
CaalfMp05	ATP9	1.212		
orf19.3448	-	1.198	-4.36	
orf19.4445	-	1.171		
orf19.5069	-	1.13	2.29	2.36
orf19.4949	-	1.087		
orf19.4773	AOX2	1.076	-2.2	4.04
CaalfMp03	NAD1	1.066		
CaalfMp07	ATP8	1.065		
CaalfMp14	NAD4	1.038		
orf19.6660	-	1.035	-2.26	
orf19.5282	-	1.0315	-1.46	
orf19.1944	GPR1	1.021		
orf19.6192	-	0.9885		
orf19.4335	TNA1	0.9795	-1.67	
orf19.2898	-	0.9735	-3.77	
orf19.1117	-	0.952	1.55	
orf19.4907	-	0.937	-3.6	
orf19.938	-	0.932	3.2	3.73
orf19.2962	-	0.917	-1.48	4.05

Gene	Name	*WOR1/wor1*: Wild Type (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)	Wor1 Peak LOD Score [maximum 4.05]
orf19.6311	-	0.8985	-1.83	2.36
orf19.1153	GAD1	0.889	-1.5	
orf19.2800	-	0.8655		
orf19.3159	UTP20	0.862		4.04
orf19.4941	TYE7	0.841	-4.47	
orf19.6659	GAP6	0.838		
orf19.2028	MXR1	0.8315	-2.24	
orf19.1774	-	0.82	-3.42	
orf19.1375	LEU42	0.798		
orf19.3427	-	0.782		
orf19.4784	CRP1	0.74		
CaalfMp01	COX2	0.7365		
orf19.3868	-	0.7175		
orf19.24	RTA2	0.716	-3.94	4.05
CaalfMp12	NAD4L	0.706		
CaalfMp02	NAD6	0.703		
orf19.3209	FGR42	0.702		4.05
orf19.21	-	0.6885		
orf19.7411	OAC1	0.686		
orf19.23	RTA3	0.68	-5.55	
orf19.1486	-	0.6745		4.05
orf19.5382	-	0.668		
orf19.2197	-	0.656		
orf19.6937	PTR2	0.656	-2.09	
orf19.6514	CUP9	0.652	-1.81	
orf19.6996	-	0.6505	2.03	2.06
orf19.2356	CRZ2	0.6495		2.53
orf19.6840	-	0.635	-3.02	2.12
orf19.6000	CDR1	0.6275		2.36
orf19.4123	-	0.6265		
orf19.2167	-	0.6245		
orf19.6994	BAT22	0.624		2.06
orf19.5475	-	0.6235		
orf19.6086	LEU4	0.611		
orf19.323	-	0.597		
orf19.3712	-	0.587		
orf19.5026	-	0.582		

Gene	Name	*WOR1/wor1*: Wild Type (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)	Wor1 Peak LOD Score [maximum 4.05]
orf19.149	-	0.562		
orf19.3431	-	0.5405		
orf19.6269	-	-0.512		
orf19.535	RBR1	-0.533	6.43	
orf19.781	-	-0.544	1.25	
orf19.6919	-	-0.555	1.99	2.06
orf19.4590	RFX2	-0.604	3.18	
orf19.3110	-	-0.6045		
orf19.3121	GST1	-0.61	5	
orf19.6784	PGA32	-0.6425	3	
orf19.6608	-	-0.6475		
orf19.575	HYR3	-0.655		3.02
orf19.6869	-	-0.656		
orf19.2941	SCW4	-0.6605	7.5	
orf19.6078	POL93	-0.662		
orf19.3612	PST2	-0.669	3.62	
orf19.4569	-	-0.7375		
orf19.780	DUR1,2	-0.738		
orf19.2247	-	-0.7405	4.93	4.05
orf19.6993	GAP2	-0.7565		
orf19.3444	-	-0.761	3.41	
orf19.4791	-	-0.768		
orf19.2701	-	-0.77	3.74	
orf19.6594	PLB3	-0.8105		
orf19.4972	-	-0.818	8.14	4.05
orf19.4568	ZCF25	-0.827	2.56	
orf19.4066	-	-0.855	1.94	
orf19.5169	-	-0.857	1.73	
orf19.3931	SFC1	-0.861	1.68	
orf19.2253	-	-0.873	4.21	
orf19.2062	SOD4	-0.877	3.92	
orf19.4121	-	-0.9	1.92	
orf19.356	GTT13	-0.914	4.76	
orf19.893	PGA17	-0.922	6.82	
orf19.4530.1	-	-0.9405	2.11	
orf19.3902	-	-0.941	5.99	3.74
orf19.2460	-	-0.974	7.74	

Gene	Name	*WOR1/wor1*: Wild Type (log2)	Opaque:White Enrichment, Wild Type, RNA-Seq (log2)	Wor1 Peak LOD Score [maximum 4.05]
orf19.4072	IFF6	-0.9935	6.19	3.60
orf19.7514	PCK1	-1.03		
orf19.4450.1	-	-1.042		
orf19.929	-	-1.043	3.77	
orf19.5760	IHD1	-1.09	1.84	
orf19.4527	HGT1	-1.101		
orf19.3668	HGT2	-1.16	2.71	3.32
orf19.6983	-	-1.2795	1.55	3.74
orf19.7111.1	SOD3	-1.4025	4.94	
orf19.4934	OP4	-1.564	7.92	
orf19.7434	GLG2	-1.586	4.28	
orf19.1370	-	-1.678	7.69	2.06
orf19.5144	PGA28	-2.151	6.73	4.05
orf19.5307	JEN2	-2.547	5.08	4.05
orf19.4026	HIS1	-5.188		

Table 3: List of 121 genes changing at least 1.41 fold (log2=0.5) in an opaque *WOR1/wor1* strain. Transcript change values for both are on a log2 scale, white or opaque-enrichment values are taken from RNA-Sequencing results [ref] and are in a log2 scale. Values of enrichment less that 2 (log2=1) are not shown. Wor1 motif LOD scores are out of a maximum value of 4.05, a score of 2.86 explains 63% of Wor1 binding events while only occurring in 18% of an equivalently sized set of control regions. Motif LOD scores evaluated in MochiView (11).

Appendix 4:

Experimental measurements taken for modeling studies of the white-opaque regulatory circuit.

Introduction

The human fungal pathogen *C. albicans* differs from the majority of fungal species by possessing the ability to switch between two distinctive types of cells, called white and opaque (1). These two cell types are both stable and heritable, white cells will tend to give rise to white cells and opaque cells to opaque cells. The past decade has seen advances in our understanding of the regulation of white-opaque switching in *C. albicans*, a core circuit (illustrated in Figure 1) consisting of four transcriptional factors that controls the regulation of switching between the two cell types has recently been identified (2-6). Given this increase in information about this circuit, the Johnson lab began a collaboration with Cihan Oguz from Hana El-Samad's lab at UCSF to develop a mathematical model of the regulatory circuit. Prior to this effort, there was only one paper dealing with modeling of this switch (7). A mathematically based model would further our understanding of the regulatory circuit in several ways. It would determine whether the current conceptual model of the circuit could explain circuit behaviors observed *in vivo*. If a mathematical model could not explain certain behaviors, it would suggest that we have not identified critical components of the system. This model might also elucidate which interactions are the most important for specific behaviors. In the long term, information produced by a mathematical model of the regulatory circuit could also serve as the basis for developing testable predictions.

When modeling began, we already possessed a great deal of information that could be used to inform and test different models. Specifically, we knew which transcription factors were up-regulated in each cell type, which factors could drive switching when ectopically expressed, and which factors could be deleted without

affecting the stability of the two cell types (2). Concurrent with the early stages of the modeling process, we developed data about the transition between the opaque and white cell types (see Chapter 4) that was also of use (8). Many possibly informative experimental measurement, however, had not been taken. Foremost among these, we were not sure of the relative levels of the different transcription factors or how these compared to each other. Although estimates could be made based on transcript levels (9, 10), we did not know if these directly related to protein levels.

In order to inform the model, we endeavored to measure relative levels of Wor1, Wor2, Czf1, and Efg1 in both white and opaque cells. We used a combination of western blotting, flow cytometry, and fluorescent microscopy to measure levels in both cell types in a combination of strains previously reported (see Chapter 4)(8) and strains specific to this study (Figure 2, Table 1). We also examined the white-to-opaque switching frequency in strains with one copy of the various regulators deleted.

Methods and Materials

Creation of fluorescently tagged strains.

The development of plasmids for GFP tagging Wor1, Wor2, Czf1, and Efg1 has been discussed in Chapter 4 (8). These plasmids can quickly be digested with SphI and AatII to produce large amounts of previously sequenced DNA to use in transformations, eliminating the need for using PCR to create a fragment for tagging,. This technique results in a larger number of colonies following transformation, increasing the efficiency of strain creation. As these plasmids use the SAT1 flipper system (11), they only leave a 48bp insertion (XhoI-Flp Site-NotI) between the end of GFP and the start of the region 3' of the original gene. With these plasmids, we created strains where one or both copies of

each regulator were GFP tagged (see Table 1 for strains, Figure 3 for plasmid maps). In addition to GFP tagging, we also have equivalent plasmids for YFP tagging Wor1 and Wor2. Any of the GFP or YFP plasmids can be used to transcriptionally or translationally tag genes other than the four mentioned above using custom designed oligonucleotides.

Although not used for parameter estimation, we have also developed a series of plasmids for mCherry tagging members of the circuit. In brief, we modified the sequence for mCherry (sequence in table 2, primers in table 3) reported in Shaner *et al*. (12) to account for alternative usage of the CUG codon in *C. albicans*. During this process, we also compensated for *C. albicans* specific codon usage bias based on the example of Cormack et al 1997 (13). Once the sequence was finalized, *C. albicans* optimized mCherry was then synthesized by DNA2.0 [Menlo Park, CA]. After verifying that *C. albicans* mCherry was fluorescent, we built plasmids for gene tagging based on the approach described above (illustrated in Figure 3, see Table 4 for all plasmids). We note that fluorescence was only observed for mCherry-tagged Wor1and Efg1, and that only Efg1-mCherry appears functional. Of additional note, *C. albicans* mCherry photobleaches quickly, so care must be taken when working with it.

Measurement of Protein Levels

Normalized western blotting was performed as described in Chapter 4 (8). Samples of white and opaque *C. albicans* populations growing in log-phase were harvested at OD=0.7-0.8 and processed as previously described. In addition to the standard wildtype background (MLY537/589), we also used strains provided by Aaron Hernday where one copy of Czf1 or Wor2 had been Myc-tagged. For these strains, we

used a commercially available antibody to Myc for primary blotting. Protein levels determined by these methods are included in Table 5.

Flow cytometry was performed with strains where one or both copies of each regulator were tagged. Overnight cultures for each strain were diluted back to OD~0.2 and allowed to regrow for several hours. Samples were centrifuged, washed twice with cold PBS, and sonicated briefly prior to injection on a LSRII [BD Biosciences, San Jose, CA]. Data were analyzed using FlowJo version 7 (Tree Star Inc., Ashland, OR). Matched strains lacking GFP were used for determination of background fluorescence levels. Protein levels determined by this method are included in Table 5.

We also determined protein levels using a modification of the fluorescence microscopy technique described in Chapter 4 (8). Overnight cultures for each strain were diluted back to OD~0.2 and allowed to regrow for several hours. Samples were centrifuged, washed twice with cold PBS, and then incubated at 4°C for an hour in the presence of DAPI. Cells were imaged on an imaged on a Zeiss Axiovert 200M microscope (Carl Zeiss, Oberkochen, Germany). Constant GFP exposure times were used for each strain during an assay. Images were analyzed using a custom MATLAB (The MathWorks Inc., Natick, MA) script which identified cells based on DAPI levels and then determined the mean GFP fluorescence levels for each cell identified. Matched strains lacking GFP were used for determination of background fluorescence levels. Protein levels determined by this method are included in Table 5.

Switching Frequency Determinations

White-opaque switching assays were performed based on previously described techniques (3). Control strains with *HIS1* and *LEU2* added back were gifts from Aaron

Hernday and Quinn Mitrovich and have been described previously (8), the *his1 leu2* strain has been previously reported. Heterozygous deletion strains are the white equivalent of strains previously reported (8).

Results

Looking at this data, the changes in each protein's level appear similar to the magnitude of changes seen in studies of transcript levels. Wor1 changed the most between the two cell types, being up-regulated by more than 50-fold in opaque cells. Given the scarcity of Wor1 in white cells, changes in Wor1 should be considered as representative rather than an absolute measure of change. Both Wor2 and Czf1 changed on the order of 5-fold between the two cell types. Levels of Efg1 underwent a reciprocal change, having been down-regulated 4 to 5-fold in opaque cells.

Relative Protein Levels

Since the flow cytometry and fluorescence microscopy protein level determinations relied on GFP fusions, it is also possible to make comparisons between levels of the different transcription factors. We used this information to estimate the relative amounts of the four transcription factors in both cell types (Table 6). It should be noted, however, that this comparison relies on the assumption that the different GFP fusions have the same quantum efficiency, which we have not verified. It also assumes that the GFP tagging has not dramatically changed how the different transcription factors are regulated or their degradation rates. Although these assumptions seem reasonable given the behavior of the tagged strains, they have not been proven.

Looking at this data, it appears that levels of Wor1 in opaque cells are similar to those of Efg1 in white cells. Wor1 levels in white cells are much lower than those of

Czf1 or Wor2 in white cells, which are themselves lower than Efg1's levels in opaque cells.

Percent Switching in Regulator Heterozygous Deletion Strains

To provide another data set for parameter estimation, we determined the frequency of white-to-opaque switching in strains with one copy of *WOR1*, *WOR2*, *CZF1*, or *EFG1* deleted (Table 7). We observed that deletion of one copy of any of the three opaque-enriched regulators reduced the frequency of switching, while deletion of a copy of white-enriched *EFG1* increased the frequency of switching.

Conclusions

Working with other members of the Johnson and El-Samad lab groups at UCSF, we were able to quantify relative abundances of the 4 transcription factors controlling the white-opaque switch. Using a combination of fluorescence microscopy, flow cytometry, and western blotting, we determined the levels of key regulatory proteins in both cell types and estimated the relative abundance of these proteins in relation to each other. These data, coupled with information gathered by Rebecca Zordan (2, 3) and the examination of temperature-induced opaque-to-white switching (8), have proven critical in the estimation of parameter sets (e.g Hill coefficients, protein degradation rates) needed for the qualitative evaluation of the current working model of the white-opaque regulatory circuit. As such, they formed part of the foundation for future efforts to evaluate this circuit and to test our assumptions in a mathematically rigorous fashion.

References

1. Lohse MB, Johnson AD (2009) White-opaque switching in *Candida albicans*. *Curr Opin Microbiol* 12:650-654.
2. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* 5:e256.
3. Zordan RE, Galgoczy DJ, Johnson AD (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci U S A* 103:12807-12812.
4. Srikantha T, et al. (2006) TOS9 regulates white-opaque switching in *Candida albicans*. *Eukaryot Cell* 5:1674-1687.
5. Huang G, et al. (2006) Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 103:12813-12818.
6. Vinces MD, Kumamoto CA (2007) The morphogenetic regulator Czf1p is a DNA-binding protein that regulates white opaque switching in *Candida albicans*. *Microbiology* 153:2877-2884.
7. Sriram K, Soliman S, Fages F (2009) Dynamics of the interlocked positive feedback loops explaining the robust epigenetic switching in *Candida albicans*. *Journal of theoretical biology* 258:71-88.
8. Lohse MB, Johnson AD (2010) Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans. *Mol Microbiol* 78:331-343.

9. Lan CY, et al. (2002) Metabolic specialization associated with phenotypic switching in *Candida albicans*. *Proc Natl Acad Sci U S A* 99:14907-14912.

10. Tsong AE, Miller MG, Raisner RM, Johnson AD (2003) Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115:389-399.

11. Reuss O, Vik A, Kolter R, Morschhäuser J (2004) The SAT1 flipper, an optimized tool for gene disruption in *Candida albicans*. *Gene* 341:119-127.

12. Shaner NC, et al. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22:1567-1572.

13. Cormack BP, et al. (1997) Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in Candida albicans. *Microbiology* 143:303-311.

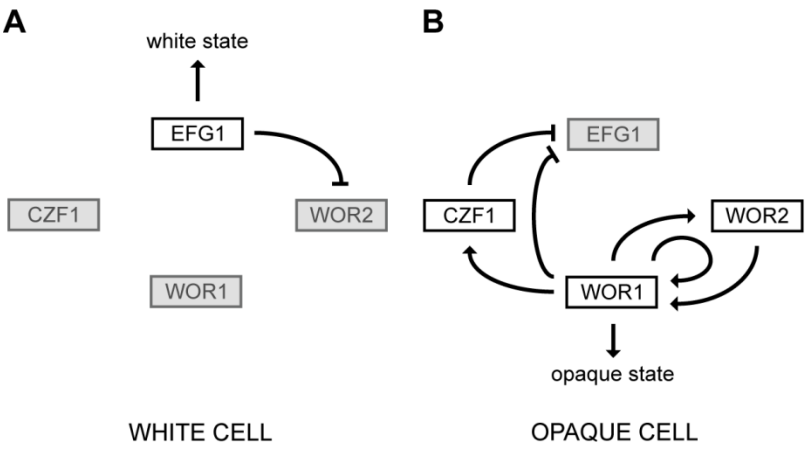


Figure 1: Working model of the white-opaque regulatory circuit and its activity in white and opaque cells. (a) In white cells, *EFG1* represses *WOR1* indirectly through *WOR2* to maintain white cell identity. (b) In opaque cells, *WOR1*, *WOR2*, and *CZF1* establish a series of positive feedback loops, maintaining opaque cell identity and repressing *EFG1*. Up-regulated genes and active relationships are indicated in black. Down-regulated genes are indicated in gray. Arrows and bars represent activation and repression respectively. Figure adapted from Zordan et al. 2007, initially published in Lohse et al. 2010.

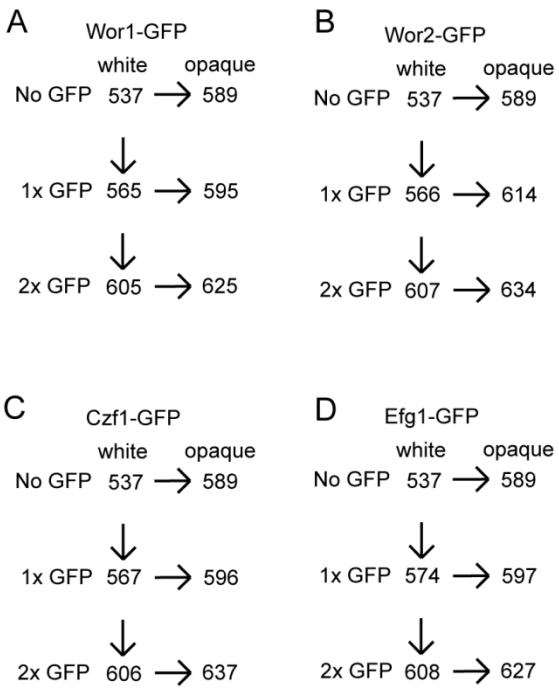


Figure 2: Relationships between the GFP-tagged strains used in this study. Strain numbers (MLY#) shown.

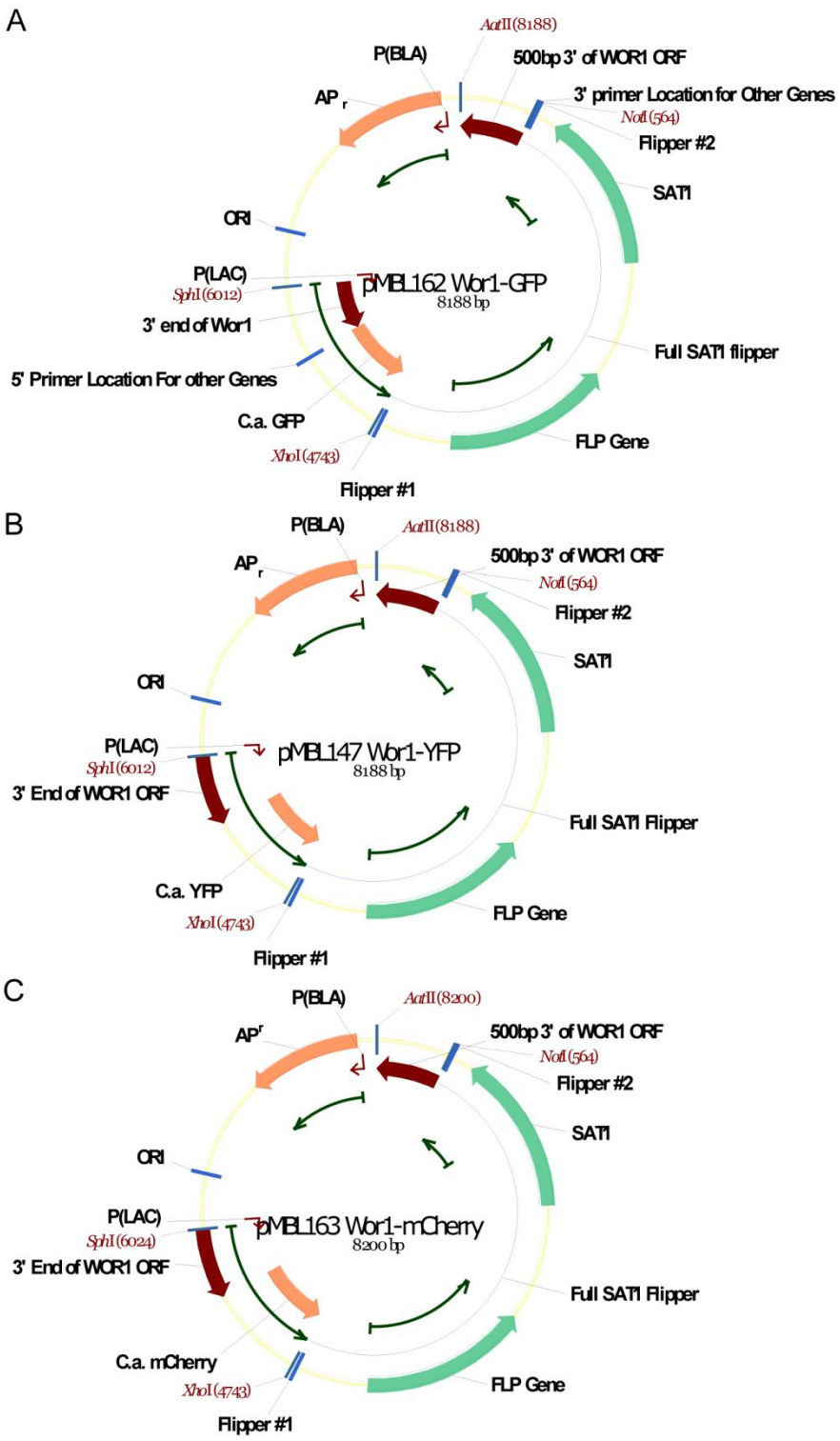


Figure 3: Maps of plasmids for fluorescent tagging of Wor1. [a] Map of pMBL162, for GFP tagging Wor1. [b] Map of pMBL147, for YFP tagging Wor1. [c] Map of pMBL163, for mCherry tagging Wor1. Plasmids for the other three regulators differ

only in the respective ~500bp ORF and 3’ homology regions used for each protein tagged, the overall plasmid arrangement is the same.

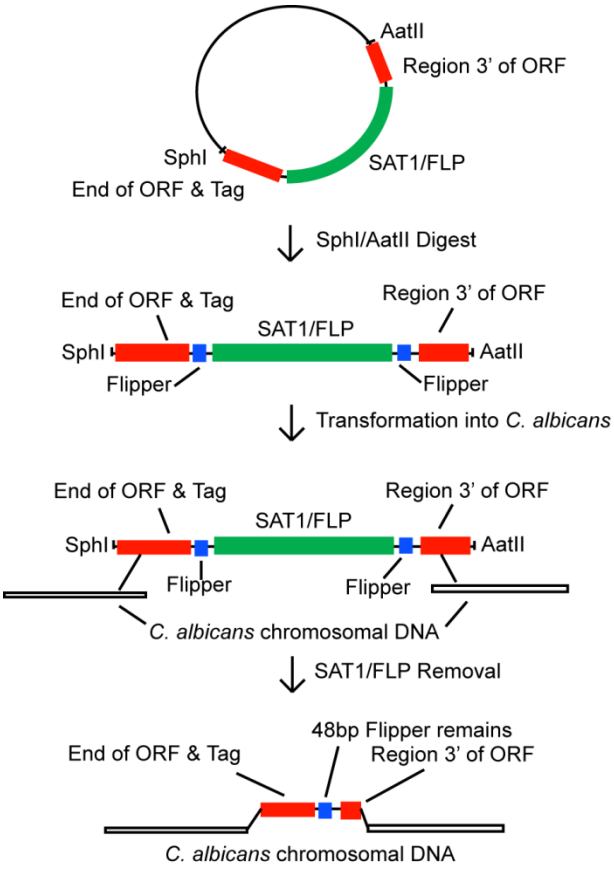


Figure 4: Schematic of the various steps in the transformation process. Following digestion with SphI and AatII, the linearized plasmid is transformed into *C. albicans* where it is integrated at the end of the gene target. The SAT1 and FLP genes are then removed leaving only a 48bp fragment at the end of the gene.

Table 1: List of strains used in this study.

Description	Strain	Genotype	Reference
Wild Type (white)	MLY537	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	Ch4, Reference 1
Wild Type (opaque)	MLY589	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	Ch4, Reference 1
Wor1-GFP, dual (white)	MLY605	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR1-GFP/WOR1-GFP	This Study
Wor1-GFP, dual (opaque)	MLY625	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR1-GFP/WOR1-GFP	Ch4, Reference 1
Wor2-GFP, dual (white)	MLY607	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR2-GFP/WOR2-GFP	This Study
Wor2-GFP, dual (opaque)	MLY634	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR2-GFP/WOR2-GFP	Ch4, Reference 1
Czf1-GFP, dual (white)	MLY606	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} CZF1-GFP/CZF1-GFP	This Study
Czf1-GFP, dual (opaque)	MLY637	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} CZF1-GFP/CZF1-GFP	Ch4, Reference 1
Efg1-GFP, dual (white)	MLY608	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} EFG1-GFP/EFG1-GFP	This Study
Efg1-GFP, dual (opaque)	MLY627	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} EFG1-GFP/EFG1-GFP	Ch4, Reference 1
Wor1-GFP, single (white)	MLY565	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR1-GFP/WOR1	This Study

Description	Strain	Genotype	Reference
Wor1-GFP, single (opaque)	MLY595	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR1-GFP/WOR1	This Study
Wor2-GFP, single (white)	MLY566	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR2-GFP/WOR2	This Study
Wor2-GFP, single (opaque)	MLY614	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR2-GFP/WOR2	This Study
Czf1-GFP, single (white)	MLY567	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} CZF1-GFP/CZF1	This Study
Czf1-GFP, single (opaque)	MLY596	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} CZF1-GFP/CZF1	This Study
Efg1-GFP, single (white)	MLY574	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} EFG1-GFP/EFG1	This Study
Efg1-GFP, single (opaque)	MLY597	a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} EFG1-GFP/EFG1	This Study
Wild Type, *his1/his1, leu2/leu2* (white)	RZY47	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}	Reference 2
WOR1/wor1 (white)	RZY187	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a WOR1/wor1Δ::C.m.LEU2	Ch4, Reference 1
WOR2/wor2 (white)	AHY117a1	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a WOR2/wor2Δ::C.d.HIS1	This Study
CZF1/czf1 (white)	AHY118a1	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a CZF1/czf1Δ::C.d.HIS1	This Study

Description	Strain	Genotype	Reference
EFG1/efg1 (white)	AHY119a	a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a EFG1/efg1Δ::C.d.HIS1	This Study

References

1. Lohse MB, Johnson AD. (2010) *Mol Microbiol* 78:331-343.

2. Zordan RE, Galgoczy DJ, and Johnson AD. (2006) *Proc Natl Adad Sci USA* 103(34): 12807-12.

Table 1: List of strains used in this study.

C. albicans optimized mCherry Sequence
atggtttctaagggtgaagaagataacatggctattatcaaagaatttatgagattcaaagtccacatggaaggttctgttaacggtc acgaattcgaaattgaaggtgaaggtgaaggtagaccatacgaaggtactcaaaccgctaagttgaaagttactaagggtggtcca ttgccattcgcctgggatattttgtctccacaattcatgtacggttctaaggcttacgttaaacatccagctgatatcccagactacttga aattatctttcccagaaggttttaaatgggaaagagttatgaacttcgaagatggtggtgttgtcaccgttactcaagactcttccttgc aagacggtgaattcatttataaagttaagttgagaggtactaatttcccatctgatggtccagttatgcaaaagaaaactatgggttgg gaagcctcttccgaaagaatgtacccagaagacggtgctttgaaaggtgaaattaagcaaagattgaaattaaaggatggtggtca ctacgacgctgaagtcaaaactacctacaaggctaagaaaccagttcaattaccaggtgcctataacgtcaatatcaaattggacat tacctcccataacgaagattacactattgttgaacaatacgaaagagctgaaggtagacactctactggtggtatggatgaattgtac aagtaa

Table 2: Sequence of *C. albicans* optimized mCherry.

Table 3: List of Primers used in this study

Name	Description	Primer sequence 5'-3'	Reference
mCherry Tagging plasmids			
MBL313	WOR1-mCherry 5' (reverse)	GTT AAT TAA ACC TGG GAT TCT TCT ACC AGT ACC GGT GTA ATA CGA CCC AGA AGA	This Study
MBL250	WOR1-mCherry 3' (forward)	CTC GAG GCG GCC GC TAG TTG AAT TAA TAC GGT GAT TCT GTT ATT ATT TT	Ch4, Reference 1
MBL314	WOR2-mCherry 5' (reverse)	GTT AAT TAA ACC TGG GAT TCT TCT ACC TTT AAG TAA ATC AGC CAC TGA AAC TCT ATT AAT	This Study
MBL268	WOR2-mCherry 3' (forward)	CTC GAG GCG GCC GC TAA GAG TAT AGT AGA ACA AAA TAA TGT GTA TAT AGA T	Ch4, Reference 1
MBL325	CZF1-mCherry 5' (reverse)	GTT AAT TAA ACC TGG GAT TCT TCT ACC TTT ACT TCT GTA TTC AAC AAT ACC TCT C	This Study
MBL326	CZF1-mCherry 3' (forward)	CTC GAG GCG GCC GC TAA GCT TCT CTG TGT TGG AGG GAT	Ch4, Reference 1
MBL318	EFG1-mCherry 5' (reverse)	GTT AAT TAA ACC TGG GAT TCT TCT ACC CTT TTC TTC TTT GGC AAC AGT GCT A	This Study
MBL319	EFG1-mCherry 3' (forward)	CTC GAG GCG GCC GC TAA TAA ATA TCA TTC GTG TAC ATC ACC TTC	Ch4, Reference 1
MBL312	linker-mCherry 5' (forward)	ggtAGAAGAatcCCAGGTttaattaacGTT TCT AAG GGT GAA GAA GAT	This Study
MBL271	NotI-XhoI mCherry 3' (reverse)	GCG GCC GCC TCG AG TTA CTT GTA CAA TTC ATC CAT ACC ACC	This Study
MBL383	mCherry 5' check	TTG TGG AGA CAA AAT ATC CCA GGC GAA TG	This Study
MBL384	mCherry 3' check	GTC AAA ACT ACC TAC AAG GCT AAG AAA CCA GTT CA	This Study
YFP Tagging Plasmids, same as for GFP tagging plasmids			
MBL251	WOR1-YFP 5' (reverse)	GCG GCC GCC TCG AG TTA TTT GTA CAA TTC ATC CAT ACC AT	Ch4, Reference 1
MBL250	WOR1-YFP 3' (forward)	CTC GAG GCG GCC GC TAG TTG AAT TAA TAC GGT GAT TCT GTT ATT ATT TT	Ch4, Reference 1

Name	Description	Primer sequence 5'-3'	Reference
MBL267	WOR2-YFP 5' (reverse)	TCA CCT TTA GAA CCA CCA CC TTT AAG TAA ATC AGC CAC TGA AAC TCT ATT AAT TA	Ch4, Reference 1
MBL268	WOR2-YFP 3' (forward)	CTC GAG GCG GCC GC TAA GAG TAT AGT AGA ACA AAA TAA TGT GTA TAT AGA T	Ch4, Reference 1
			Ch4, Reference 1
MBL197	linker-YFP 5' (forward)	GGT GGT GGT TCT AAA GGT GAA GAA TTA TTC ACT	Ch4, Reference 1
MBL251	NotI-XhoI-YFP 3' (reverse)	GCG GCC GCC TCG AG TTA TTT GTA CAA TTC ATC CAT ACC AT	Ch4, Reference 1
MBL209	YFP 5' check	GGT TGG CCA TGG AAC TGG CA	Ch4, Reference 1
MBL382	YFP 3' check	GGT GAT GGT CCA GTC TTG TTA CCA GAC	Ch4, Reference 1
GFP Tagging Plasmids			Ch4, Reference 1
5' SphI primers			Ch4, Reference 1
3' AatII Primers			Ch4, Reference 1
Integration Check Primers			Ch4, Reference 1

References 1. Lohse MB, Johnson AD. (2010) *Mol Microbiol* 78:331-343.

Table 3: List of Primers used in this study

Description	Name	Reference	Comments
Wor1-GFP	pMBL162	Ch4, Reference 1	Visible/Functional
Wor2-GFP	pMBL164	Ch4, Reference 1	Visible/Functional
Czf1-GFP	pMBL171	Ch4, Reference 1	Visible/Functional
Efg1-GFP	pMBL165	Ch4, Reference 1	Visible/Functional
Wor1-YFP	pMBL147	This study	Visible/Functional
Wor2-YFP	pMBL140	This study	Visible/Functional
Wor1-mCherry	pMBL163	This study	Visible/Nuclear Localized but not Functional
Wor2-mCherry	pMBL167	This study	Not Visible
Czf1-mCherry	pMBL166	This study	Not Visible
Efg1-mCherry	pMBL168	This study	Visible/Functional

References 1. Lohse MB, Johnson AD. (2010) *Mol Microbiol* 78:331-343.

Table 4: List of Plasmids used in this study.

Protein	Western Blotting, Peptide Antibodies (Opaque/ White)	Western Blotting, αMyc Antibodies (Opaque/ White)	Flow Cytometry (Opaque/ White)	Fluorescent Microscopy (Opaque/ White)	Combined Average (Opaque/White)
Wor1	116.05	N/A	70.88	69.96	85.63
Wor2	6.41	2.86	5.71	5.79	5.19
Czf1	6.48	3.06	7.02	3.78	5.08
Efg1	0.09	N/A	0.24	0.33	0.22

Table 5: Levels of Wor1, Wor2, Czf1, and Efg1 in white and opaque *C. albicans*, as determined by fluorescence microscopy, flow cytometry, and western blotting. Values are plotted as the ratio of levels in opaque cells relative to white cells, values of greater than one represent opaque-enrichment. Values were averaged within the replicates of each experimental technique, and then the overall averages and medians were calculated.

Protein	Cell Type	Realtive Levels, Flow Cytometry	Relative Levels, Fluorescent Microscopy	Relative Levels, Combined Average
Wor1	white	1.24	0.74	0.99
Wor1	opaque	87.49	131.12	109.31
Wor2	white	11.78	12.09	11.94
Wor2	opaque	63.22	60.81	62.02
Czf1	white	5.36	10.54	7.95
Czf1	opaque	29.03	31.94	30.49
Efg1	white	100.00	100.00	100.00
Efg1	opaque	24.11	27.34	25.73

Table 6: Relative levels of Wor1, Wor2, Czf1, and Efg1 in white and opaque *C. albicans*, as determined by fluorescence microscopy and flow cytometry. Levels are normalized against Efg1 in white cells, which has arbitrarily been set to 100. Values were averaged within the replicates of the two experimental technique, then overall averages were also calculated.

Strain	Percent Colonies with Opaque sectors	n
Control	1.48	1616
Control *his1/his1 leu2/leu2*	1.25	1362
WOR1/wor1	0.056	3523
WOR2/wor2	0.31	1281
CZF1/czf1	0.15	1320
EFG1/efg1	6.39	1799

Table 7: Frequency of white-to-opaque switching, expressed as a percentage of colonies with one or more opaque sectors.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Matthew B. Schroe — 11/17/2010

Author Signature — Date