Title
The Evolution of a Transcriptional Circuit Governing Fungal Morphology

Author
Cain, Christopher William

Publication Date
2010

Peer reviewed|Thesis/dissertation

The Evolution of a Transcriptional Circuit Governing Fungal Morphology

by

Christopher W. Cain

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

[redacted]

I dedicate this thesis to my mother, Patrice, and to my little sister, Sarah. Sarah's limitless curiosity and excitement served as constant inspiration. Mom's strength and confidence gave me the will to persevere.

Acknowledgements

This thesis would not have been possible without the support of countless friends and colleagues. I am so grateful to all of those who have inspired, encouraged, counseled, and consoled me over the past five years.

I have to start by thanking those who provided the foundation for my graduate work, from my time as an undergraduate at UCSB. Kathy Foltz was an incredible mentor, and took my raw excitement and naïve interest in science and helped develop it into a true passion which I will have for the rest of my life. Peggy Cotter took me on as an undergraduate in her lab and with Corinne Williams began teaching me a practical understanding of laboratory science – something I am still learning more about every day. My experiences in the Cotter lab convinced me to dive headfirst into graduate school, and while there have been some choppy waters, it is a decision I will never regret.

At UCSF I have been fortunate to have an incredibly supportive committee to support and guide me. First and foremost I must thank Anita Sil for the incredible generosity she has shown with her time, and for her constant encouragement. As chair of my qualifying exam committee she gave me the confidence to begin this project, and she has been there during many difficult times. I can honestly say I would not have completed this thesis without her encouragement and concern for my progress. I thank my advisor Sandy Johnson for his generosity in supporting such an open ended project, especially when it was unclear where the results would lead, and for fostering a lab atmosphere where all members are supportive and collaborative. Hiten Madhani has been an endless source of information on filamentous growth, and his advice and suggestions were always helpful

in guiding my project. Cynthia Kenyon was a great source of enthusiasm and her perspective helped ensure my project was of interest to those outside of my fungal niche.

The Johnson lab has been an incredible place to spend the last four years, and this work could not have been accomplished without the support of so many of my labmates. Matthew Lohse, Aaron Hernday, and Becky Zordan were my white-opaque support group. Matt has been a great labmate and a great friend, his constant optimism and support kept me going, and his work ethic is inspirational. I am fortunate to have been able to work with Matt on our publications, and I am grateful he continues to solve the mystery that is Wor1. Aaron and his incredible attention to detail and technique development was essential for the success of my project, it could not have been done had he not laid the groundwork. Becky was a mentor who I could always count on for support and advice, she was my biggest cheerleader in the lab and it was greatly appreciated.

As my project moved towards evolutionary analysis, Oliver Homann provided absolutely critical support, advice, and mentorship; especially in the last year. His generosity with his time is unmatched, and he has taught me a great deal about evolutionary biology, computational analysis, and most of all good science. Lauren Booth provided support and sanity, our coffee breaks and talks about both projects and life really did keep me going. I have been lucky to share a bay at various points with Chris Baker and Brad Green, who taught me much about music and pub trivia, though I still have a lot to learn. I could always count on Quinn Mitrovich to ask the most insightful questions in lab meeting, and his cheery disposition and willingness to help with any question were always appreciated. I'm truly grateful for everyone in the Johnson Lab – it's been a wonderful environment and I could not ask for better teachers of all things science.

The friends I have made here at UCSF have provided an incredible support system, and this thesis would not have been possible without them. Evelyn Chang and Bryant Chhun have kept me sane and taught me much about good music, good food, and good wine. Whether it was game night or a night at the opera, I could always count on them for a good time at the end of a long week. Eric, Arthur, Karin, Ashley, Lisa, and I started playing futsal as the "1st years" many seasons ago, and while the team has evolved and changed, it has always been something to look forward to every week. I've also enjoyed many squash matches with Arthur over the past few years, and perhaps one day I will win one! I really could not ask for a better set of classmates – from our over-the-top first year study sessions, to learning how to ski on trips to Tahoe, to Chilean adventures. I am so grateful I was able to spend time with such a great group of people, and I look forward to many more adventures with them in the future.

Mom, you work so hard every day for me and our family, and I want you to know your dedication and common sense inspires and guides me every day. Dad, Robert, and Sarah, knowing you are always there from me mean so much, I don't tell you that enough. I love you all.

Finally, I have to acknowledge Carlo Torrano. Carlo, you have been my most important source of support as I worked towards the completion of this thesis. You have had an enormous influence on my life in these past few years. Your patience, understanding, advice, and companionship gave me the confidence and strength to persevere through the most difficult times. Thank you.

The Evolution of a Transcriptional Circuit Governing Fungal Morphogenesis

Christopher W. Cain

Fungi have evolved diverse morphological strategies to adapt and thrive in harsh environmental conditions. Although the transcriptional circuits that underlie these cellular changes have been studied in many systems, it is not known how they evolve to regulate specialized morphologies in distant fungal species. I approached this question beginning with the characterization of *Saccharomyces cerevisiae* Mor1, a conserved transcriptional regulator of fungal morphology. I demonstrated that Mor1 is required for haploid invasive growth and diploid pseudohyphal development, and that this is due to the direct regulation of gene expression through a specific and conserved DNA sequence motif. Whole genome ChIP-Chip further identified Mor1 as a novel core component of the filamentous growth regulatory circuit in *S, cerevisiae*. Mor1 is an ortholog of *Candida albicans* Wor1, the master regulator of the white to opaque phenotypic switch. This switch is essential for mating, alters morphology, and contributes to biofilm formation of *C. albicans*. By comparing Mor1 target genes to those of Wor1, I observed that while there is in general a widespread divergence of downstream targets, there is a striking overlap of a subset of targets encompassing many transcriptional regulators involved in morphogenesis. In addition, Mor1, Wor1, and the *Histoplasma capsulatum* ortholog Ryp1, all can function at the same discrete DNA motif within the Flo11 promoter when expressed in *S. cerevisiae*. These results suggest that despite dramatically diverged morphological phenotypes, a conserved regulator governs a conserved set of transcription factors required for basic morphological processes throughout the fungi.

Table of Contents

List of Tables

Chapter 1. Introduction – No tables

Chapter 2. The Evolution of a Transcriptional Circuit Governing Fungal Morphogenesis

Appendix 1. A Distinct Class of DNA-Binding Domains is Exemplified by a Master Regulator of Phenotypic Switching in *Candida albicans*

List of Figures

Chapter 1

Introduction

The fungal kingdom is home to an incredible diversity of species encompassing many distinct unicellular and multi-cellular morphologies. Remarkably, there is often morphological diversity within a single species, as fungi rapidly adapt their shape and size in response to the environment. In recent years it has become clear that cellular remodeling is accomplished through the coordinated regulation of hundreds of genes. These genes code for proteins involved in a wide range of processes, and include cell-cycle regulators, outer membrane proteins, nutritional transporters, and transcriptional regulators. Many of these transcriptional regulators are deeply conserved, and are required for evolutionarily distant fungal species to undergo morphological switching. Here I briefly review the key findings that have enhanced our understanding of the regulation of fungal morphogenesis, and provide evidence that a recently discovered transcriptional regulator acts in a conserved network to regulate morphology in distantly related fungal species.

It was known for decades that a multitude of environmental and growth signals fed into the control of fungal morphogenesis[1], however the mechanistic basis of this control was not well understood. Approximately 20 years ago genetic screens in many model fungi, including *Aspergillus nidulans, Candida albicans,* and *S. cerevisiae*, began to identify genes required for proper fungal development. Many of these genes coded for predicted DNA binding proteins, and these studies gave the first hints of the complexity required to coordinate these cellular transitions.

In *A. nidulans*, a number of genes were identified that are required to trigger the cellular transitions involved in sexual and asexual development[2]. Among these genes were the transcriptional regulators AbaA and StuA, which are required for the proper development

of conidia (asexual spores)[3]. These genes are expressed in a temporal pattern, and epistasis studies determined that they are dependent on each other for proper expression. These studies gave the first hint that an interconnected set of regulators was required to execute a morphological program in fungi.

Soon after these results, a detailed analysis was published of a morphological transition in *S. cerevisiae*. Termed "pseudohyphal growth", this cellular transition takes place in diploid cells starved for nitrogen and is characterized by the elongation of cell shape, a switch to unipolar cell division, and the upregulation of adhesins[4]. Together these cellular changes permit cells to chain together into filaments and invade agar. It was later shown that haploid cells could also adhere to and invade agar, in a process termed "haploid invasive growth"[5]. Collectively these processes are referred to as "filamentous growth". These studies made available the power of yeast genetics to identify and analyze in detail the genes involved in a dimorphic transition.

What followed over the next decade was the identification of numerous transcriptional regulators with roles in this process. A MAPK cascade involved in mating was found to activate filamentous growth, signaling through the Ste12 and Tec1 transcriptional regulators[6,7]. Protein Kinase A was found to signal through another regulator, Flo8, which is also required for filamentous growth[8]. It is a mutation in Flo8 that prevents most standard lab strains of *S. cerevisiae* from filamenting[9]. Other critical regulators identified included Phd1, Sok2, and Mga1[10,11,12]. From these studies came the first hints that orthologous regulators may be playing conserved roles in morphological transitions in distant fungi. Phd1 and Sok2 are orthologs of *A. nidulans* StuA. In addition, Tec1 is the ortholog of *A. nidulans* AbaA, and expression of AbaA can complement for a

Tec1 deletion and restore filamentous growth[7]. This suggested that despite their divergent cellular phenotypes, orthologous transcriptional regulators may play a conserved role in fungal development.

These studies in *S.cerevisiae* laid the foundation for dissecting filamentous growth in *C. albicans*, the most prevalent fungal pathogen of humans. *C. albicans* undergoes a dramatic cellular transition from the yeast form to long hyphal filaments, in response to high temperature or serum. This transition is required for virulence[13]. The first identified regulator of this process was Cph1, the ortholog of *S. cerevisiae* Ste12[14]. Subsequently orthologs of Phd1, Sok2, Tec1, and Flo8 were all demonstrated to play roles in the yeast-hyphal transition[15,16,17]. Moreover expression of these genes in *S. cerevisiae* found that they were functionally interchangeable. This suggested that filamentous growth in *S. cerevisiae* and *C. albicans* were highly related processes, despite major differences in cell shape.

Recent studies have begun to reveal a great complexity in the downstream targets of these transcription factors. In 2006 a study was published that performed whole genome binding analysis on six core *S. cerevisiae* filamentation regulators by coupling chromatin immunoprecipitation with microarray analysis (ChIP-Chip)[18]. This study found that most of these regulators bind to their own promoters as well as each other's, and share hundreds of downstream targets. The complexity of this regulation makes sense in light of the enormity of the task required. Cells must respond sensitively to a wide range of environmental conditions, and integrate these inputs through multiple signaling cascades to regulate the extent of their response. This requires the coordinated remodeling of the cell-cycle, nutritional uptake, and cell wall composition. These outputs are tuned in

species specific ways to give rise to specialized morphological transitions. While it is clear from the above studies that many regulators play conserved roles in fungal development, how their regulatory networks evolve to give rise to these distinct cellular outputs has not been examined in detail.

A recently identified conserved regulator of fungal morphogenesis

In this work we examine in detail the function of a conserved transcriptional regulator present in all fungi. This regulator was initially identified in *Candida albicans* as Wor1, the master regulator of the white-opaque morphological transition[19]. White-opaque switching is a heritable switch between two distinct cell types, white and opaque, which each possess distinct cell surface, shape, and biological properties. Subsequently the *Histoplasma capsulatum* ortholog of Wor1 was identified as Ryp1, the master regulator of the mycelial-yeast transition[20]. Deletion of these proteins significantly affects the expression of hundreds of genes, and Wor1 binds to the promoters of hundreds of genes as demonstrated by *in vivo* ChIP-Chip experiments[21]. However questions remained as to how this family of proteins regulates these developmental processes. First, Wor1 contains no homology to any known functional domains, leaving open the question of how it acts to regulate transcription. In addition, Wor1 associates in large blobs across intergenic regions as long as 10kb, seemingly unusual behavior for a site-specific DNA binding transcription factor. Second, white-opaque switching is a very different developmental transition from mycelial-yeast form growth. While both involve changes to cell shape and size, the extent of the changes and their functional role in each species is dramatically different. I sought to understand how this conserved regulator is able to

regulate critical morphological switches with distinctly different characteristics in fungi separated by 400 million years of evolution[22].

The approach I took was to study the function of the *Saccharomyces cerevisiae* orthologs of Wor1, Yel007w and Yhr177w. Neither protein had been functionally characterized, however there were data suggesting that the proteins might play a role in the filamentous growth pathway – the best studied of fungal morphological transitions[23](Fink lab, unpublished communication). I reasoned that I could gain insight into the role Yel007w plays in filamentous growth by performing genome-wide expression and binding experiments, then integrating the results with the vast ChIP and expression data available in *S. cerevisiae*. This would also allow me to compare the Yel007w regulatory circuit to the Wor1 regulatory circuit and investigate how a morphological transition evolves. I would further be able to take advantage of the genetic tools available to identify how Yel007w was associating with DNA, whether directly or indirectly, thus gaining insight into how this novel and conserved family of regulators functions.

The results of this work are largely contained in Chapter 2 of this thesis. These results can be broken down into two major categories:

1. Functional characterization of Mor1 and orthologous proteins.
2. Evolutionary analysis of the Mor1 regulatory network.

We first identified the Yel007w gene as a master regulator of filamentous growth in *S. cerevisiae*, and named it Mor1 (Morphology Regulator 1). Mor1 is required for filamentous growth, it binds the promoters of nearly all known filamentous regulators, and its promoter is reciprocally bound by those same regulators. Over-expression of

Mor1 is sufficient to drive filamentous growth in non-inducing conditions. Taking advantage of reporter plasmid constructs available in *S. cerevisiae*, we identified the specific site of action for Mor1 at the Flo11 promoter. This contributed to the identification of a discrete DNA motif required for Mor1 function (see Appendix 1). We further showed that Wor1 and Ryp1 act on this same motif, demonstrating that core DNA-recognition and transcriptional activities are deeply conserved in these proteins.

We then compared the regulatory networks of Mor1 in *S. cerevisiae* and Wor1 in *C. albicans*. We found that while there was a greater overlap in their downstream targets than would be expected by chance alone, there was a significant divergence in targets between the two species. The exceptions to this finding were transcriptional regulators, of which 50% were found to be conserved targets. Many of these transcriptional regulators have morphological phenotypes in distant fungal species. This suggests that the interactions between many of the transcriptional regulators targeted in this circuit may be conserved; forming a core network regulating cellular differentiation in fungi; and it is the divergence of their downstream targets that gives rise to specialized cell types.

References

1. Odds FC (1985) Morphogenesis in Candida albicans. Crit Rev Microbiol 12: 45-93.
2. Mirabito PM, Adams TH, Timberlake WE (1989) Interactions of three sequentially expressed genes control temporal and spatial specificity in Aspergillus development. Cell 57: 859-868.
3. Miller KY, Wu J, Miller BL (1992) StuA is required for cell pattern formation in Aspergillus. Genes Dev 6: 1770-1782.
4. Gimeno CJ, Ljungdahl PO, Styles CA, Fink GR (1992) Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. Cell 68: 1077-1090.

5. Roberts RL, Fink GR (1994) Elements of a single MAP kinase cascade in Saccharomyces cerevisiae mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev 8: 2974-2985.
6. Liu H, Styles CA, Fink GR (1993) Elements of the yeast pheromone response pathway required for filamentous growth of diploids. Science 262: 1741-1744.
7. Gavrias V, Andrianopoulos A, Gimeno CJ, Timberlake WE (1996) Saccharomyces cerevisiae TEC1 is required for pseudohyphal growth. Mol Microbiol 19: 1255-1263.
8. 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. The EMBO Journal 18: 1257-1269.
9. Liu H, Styles CA, Fink GR (1996) Saccharomyces cerevisiae S288C has a mutation in FLO8, a gene required for filamentous growth. Genetics 144: 967-978.
10. Gimeno CJ, Fink GR (1994) Induction of pseudohyphal growth by overexpression of PHD1, a Saccharomyces cerevisiae gene related to transcriptional regulators of fungal development. Mol Cell Biol 14: 2100-2112.
11. Ward MP, Gimeno CJ, Fink GR, Garrett S (1995) SOK2 may regulate cyclic AMP-dependent protein kinase-stimulated growth and pseudohyphal development by repressing transcription. Mol Cell Biol 15: 6854-6863.
12. Lorenz MC, Heitman J (1998) Regulators of pseudohyphal differentiation in Saccharomyces cerevisiae identified through multicopy suppressor analysis in ammonium permease mutant strains. Genetics 150: 1443-1457.
13. Lo HJ, Kohler JR, DiDomenico B, Loebenberg D, Cacciapuoti A, et al. (1997) Nonfilamentous C. albicans mutants are avirulent. Cell 90: 939-949.
14. Liu H, Kohler J, Fink GR (1994) Suppression of hyphal formation in Candida albicans by mutation of a STE12 homolog. Science 266: 1723-1726.
15. Schweizer A, Rupp S, Taylor BN, Rollinghoff M, Schroppel K (2000) The TEA/ATTS transcription factor CaTec1p regulates hyphal development and virulence in Candida albicans. Mol Microbiol 38: 435-445.
16. Cao F, Lane S, Raniga PP, Lu Y, Zhou Z, et al. (2006) The Flo8 transcription factor is essential for hyphal development and virulence in Candida albicans. Mol Biol Cell 17: 295-307.
17. Stoldt VR, Sonneborn A, Leuker CE, Ernst JF (1997) Efg1p, an essential regulator of morphogenesis of the human pathogen Candida albicans, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. EMBO J 16: 1982-1991.
18. Borneman AR, Leigh-Bell JA, Yu H, Bertone P, Gerstein M, et al. (2006) Target hub proteins serve as master regulators of development in yeast. Genes Dev 20: 435-448.
19. 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. Proceedings of the National Academy of Sciences of the United States of America 103: 12807-12812.
20. Nguyen VQ, Sil A (2008) Temperature-induced switch to the pathogenic yeast form of *Histoplasma capsulatum* requires Ryp1, a conserved transcriptional regulator.

Proceedings of the National Academy of Sciences of the United States of America 105: 4880-4885.
21. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. PLoS Biology 5: e256.
22. Taylor JW, Berbee ML (2006) Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia 98: 838-849.
23. Li F, Palecek SP (2005) Identification of Candida albicans genes that induce Saccharomyces cerevisiae cell adhesion and morphogenesis. Biotechnol Prog 21: 1601-1609.

Chapter 2

The Evolution of a Transcriptional Circuit Governing Fungal Morphology

The Evolution of a Transcriptional Circuit Governing Fungal Morphology

Christopher W. Cain, Matthew B. Lohse, and Alexander D. Johnson

Introduction

The transcriptional regulatory networks directing cellular differentiation are highly complex[1]. Transcriptional regulators interact with each other and hundreds of downstream targets to effect dramatic changes in cell shape and structure. Fungi, which rapidly undergo diverse cellular adaptations in response to environmental cues, provide a powerful model system in which to study the structure of these networks [2]. The growing collection of sequenced fungal genomes allows for detailed study of the regulatory evolution underlying cellular differentiation[3]. Here, we examine the evolution of a recently described transcriptional regulator involved in diverse cellular transitions in the fungal lineage.

Wor1, a novel DNA-binding transcriptional regulator, was recently identified as the critical component of a cellular differentiation switch in the human fungal pathogen *C. albicans* [4,5,6]. Under low-temperature, anaerobic, or DNA-damaging conditions, *C. albicans* cells can transition between two heritable cellular states, referred to as white and opaque[7]. White and opaque cells have different shapes and cell surfaces, mating competencies, and may play distinct roles in pathogenesis. Wor1 associates with the promoters of 221 genes[8,9], is required for opaque cell formation, and can drive all cells to the opaque state when ectopically expressed.

Although white-opaque switching is limited to *C. albicans* and closely related species, orthologs of Wor1 are found in all sequenced fungi, raising the question of their role outside of this clade. Ryp1, the *H. capsulatum* ortholog, is required for mycelial differentiation to the pathogenic yeast form[10]. This suggests that orthologs of these genes may be important to cellular transitions in diverse fungal species. *S. cerevisiae* has two orthologs of Wor1 as a result of the whole genome duplication, Yel007w and Yhr177w, however little is known about their function. We sought to take advantage of the extensive data on transcriptional networks in this yeast to place these orthologs in the context of known regulatory circuits, and to study the evolution of their regulatory targets across distant fungal species. We hypothesized Yel007w and Yhr177w might also play a role in cellular differentiation in *S. cerevisiae*.

While it was believed for many years that *S. cerevisiae* did not undergo the complex cellular differentiation pathways characteristic of more pathogenic fungi, work over the last 2 decades has clearly demonstrated this was due to laboratory selection [11]. Wild *S. cerevisiae* strains respond to nutrient starvation by triggering a differentiation program which causes cells to elongate and chain together into filaments capable of invading agar[12,13]. While distinct environmental cues trigger modified versions of this program in haploids and diploids, the same signaling pathways and transcriptional regulators are required for both haploid invasive growth and diploid pseudohyphal growth, and here we refer to both processes collectively as “filamentous growth”.

Two clear themes have emerged from previous studies of this process: (1) Regulation of the cell wall flocculin Flo11 is a critical control point for filamentous growth; and (2) A highly interconnected transcriptional network governs *S. cerevisiae* differentiation. Flo11

expression is required for agar invasion in both haploid invasive and diploid pseudohyphal growth [14]. The Flo11 promoter is one of the largest in the genome and is directly regulated by at least 10 transcriptional regulators and both MAPK and PKA signaling pathways [15]. Genome-wide ChIP analysis of many of these regulators has identified a core set of 6 filamentation factors (Mga1, Phd1, Sok2, Ste12, Tec1, and Flo8) which regulate each other and a network of hundreds of downstream targets [2]. Orthologs of these regulators appear in many pathogenic fungi, where they are often associated with a virulence or differentiation defect. This suggests that at least some of their function may be maintained across fungi, but how they evolve to regulate diverse cellular programs is not well understood.

Here, we identify Yel007w as a central transcriptional regulator of the *S. cerevisiae* filamentous growth program, and name it Mor1 (Morphogenesis regulator). Mor1 directly regulates and is regulated by most known filamentous growth regulators, is required for filamentous growth, and its ectopic expression can drive filamentous growth. We find Mor1, Wor1, and Ryp1 activate transcription from the same discrete DNA motif within the Flo11 promoter, arguing that this function has been conserved over 400 million years of evolution[16]. We then compare Mor1 bound gene targets to those bound by Wor1 in *C. albicans,* and find that despite a massive rewiring of downstream genes; orthologs of 10 Mor1 bound *S. cerevisiae* transcriptional regulators are bound by Wor1 in *C. albicans*. Most of these regulators play a role in differentiation processes in *S. cerevisiae, C. albicans,* and more distantly related fungal species. We suggest that regulatory interactions are maintained between this core set of transcriptional regulators over a large evolutionary distance, even as their differentiation phenotypes and

downstream genes diverge. This provides insight into how fungi evolve distinct differentiation programs, and suggests an explanation for why orthologous regulators have a conserved role in regulating these processes.

Results

Mor1 is required for haploid invasive growth and diploid pseudohyphal growth

We first establish the function of Mor1 in *S. cerevisiae*. Despite a wealth of genome-wide analysis available for this yeast, no phenotype had been published for a Mor1 deletion. However two results suggested it was involved in the *S. cerevisiae* filamentous growth program. First, the over-expression of the *C. albicans* homolog Wor1 induced haploid invasive growth in the *S. cerevisiae* S288c lab strain. Second, a deletion of Mor1 in the filamentation-competent Σ1278b strain compromised haploid invasive growth (Summers and Fink, unpublished communication[17,18]). We deleted Mor1 in the related Σ2000 background, which resulted in recapitulation of the invasion defect in haploid cells, and also showed a pseudohyphal growth defect in diploid cells (Fig 1A, B). Deletion of Mor2, a second *S. cerevisiae* ortholog of Wor1 retained after the whole genome duplication, had no effect on either phenotype. Whole genome expression experiments showed that Flo11 expression was at least 40-fold lower in a haploid cell deleted for Mor1 than in the unmodified parent strain (Table S1). Because Flo11 expression is required for filamentous growth, this observation likely accounts for the phenotype of the Mor1 deletion. In strains where the Flo11 coding sequence was replaced with GFP, there was a dramatic reduction in GFP production in the Mor1 deletion compared to the parent strain, providing further evidence that Mor1 is required for Flo11 expression (Fig 1C).

We next determined whether ectopic expression of Mor1 was sufficient to induce pseudohyphal growth under conditions where it does not normally take place. Indeed, ectopic expression of Mor1 from the TEF promoter drives diploid cells to invade agar in rich media conditions (Fig 1D). Thus, Mor1 fulfills the requirements of a master regulator of filamentous growth: its deletion prevents it and its ectopic expression drives it.

Mor1 is a core member of the *S. cerevisiae* filamentous growth circuit

The *C. albicans* Wor1 transcriptional regulator is a sequence-specific DNA-binding protein (Lohse et al. Submitted). To more fully understand the role of Mor1 in regulating filamentous growth, we utilized immunoprecipitation combined with whole genome tiling microarrays (ChIP-Chip), to identify its direct downstream targets. Our initial attempts at immunoprecipitation with a peptide-derived antibody were successful in identifying the highest scoring targets, but low levels of enrichment prevented us from undertaking a comprehensive analysis. To circumvent this problem a GFP tag was introduced at the C-terminus of Mor1 at its normal chromosomal location. The tagged protein was fully competent to promote invasive growth. Immunoprecipitation with a GFP antibody identified 87 discrete binding locations across the genome (Fig 2A). These 87 locations were upstream of 121 genes and were rigorously identified from two highly reproducible high density tiling experiments, minimizing false positives (see materials and methods). All known flocculin genes are bound by Mor1, as well as 25 transcriptional regulators, all sequence specific DNA binding proteins, including many of those previously implicated in filamentous growth. We also attempted to characterize Mor2 binding using a peptide derived antibody, but were unsuccessful at identifying any

enrichment in a wildtype strain. We found that in strains deleted for Mor1, Mor2 expression was upregulated ~8 fold (Table S1) and under these conditions we identified 49 gene targets, 40 of which are also bound by Mor1 (Table Sx). These results suggest that Mor2 recognizes the same DNA sequence as Mor1, but does not play an obvious role in regulating filamentous growth under the conditions examined.

We next compared the Mor1 target genes with those of the previously described transcriptional regulators of filamentous growth. Previous Chip-ChIP analysis of 6 known regulators, Phd1, Sok2, Mga1, Flo8, Tec1, and Ste12, identified a highly interactive regulatory network where each factor regulates, and is regulated by, most of the other transcriptional regulators (Fig 2B)[2]. All of these proteins bind the upstream region of Mor1[19]. Mor1 binds to its own upstream region, as well as those of Phd1, Sok2, Mga1, and Tec1. Mor1 thus appears to be a central regulator of filamentous growth and fits the criteria of a "hub protein" as defined by Bornemann et al. We compared the Mor1 ChIP data with the data for Ste12, Tec1, and Sok2 to determine whether downstream targets were also shared between these regulators. We chose these 3 regulators because of the additional availability of high-density ChIP-Chip data[20]. We found that 77% (67/87) of the Mor1 bound sites were also bound by Sok2, 62% of sites were bound by Tec1, and 56% by Ste12 (Fig 2C). This very strong overlap further supports the conclusion that Mor1 is a central regulator of the filamentous growth program.

Mor1 acts on a specific motif within the Flo11 promoter

To determine whether Mor1 association with DNA, as determined by ChIP, was functionally relevant, we analyzed Mor1's binding to the upstream region of the Flo11 gene in detail. To identify the minimal DNA sequence required for Mor1 function we utilized a plasmid collection containing overlapping 400bp fragments of the FLO11 promoter cloned into a pCyc1-LacZ plasmid construct lacking its native upstream activating sequence. These plasmids were previously used to identify regions of the FLO11 upstream region required for activation by Ste12, Tec1, and Flo8[15]. We transformed this plasmid set into both wildtype and *Δmor1* haploid Σ2000 *S. cerevisiae* strains, and identified two 400bp fragments with Mor1 dependent activity (Fig 3B). These two fragments overlap by 200bp. This 200bp fragment is distinct from the sites of action of Ste12 and Tec1, and it corresponds to the highest enriched site for Mor1 in the ChIP experiment.

To further refine a DNA element required for activation, we generated, from the 200bp overlap, a series of plasmids containing overlapping 50bp pieces of the FLO11 upstream region. Two fragments, from 1025-1075 and 1175-1225 upstream, showed Mor1-dependent transcriptional activation (Fig 3C). Both fragments contained a high-scoring motif recognized by the *C. albicans* Wor1 protein. To test if this putative motif was needed for Mor1-dependent activation, we mutated 3 critical residues in the central core of this motif in the 1175-1225 fragment. When assayed in the reporter system described above, these mutations eliminated Mor1 dependent transcriptional activity (Fig 3D). This experiment revealed that the motif and function of Mor1 and Wor1 must be very similar.

Mor1, Wor1, and Ryp1 recognize the same DNA sequence

As described above, Mor1 and Wor1 appear to recognize the same DNA sequence. To test this idea explicitly, we ectopically expressed *C. albicans* Wor1 in a strain where both Mor1 and Mor2 were deleted. Wor1 activated transcription from the 1175-1225 sequence of the Flo11 promoter to high levels (Fig 4A); the point mutations that inactivated the binding site, as described above, destroyed activation by Wor1. In addition, we verified that ectopic expression of Wor1 drives agar invasion of diploids in rich media conditions (Fig 4C). To further investigate the overlapping DNA binding specificity of Wor1 and Mor1, we immunoprecipitated Wor1 in the *Δmor1Δmor2 S. cerevisiae* strain using a C-terminal peptide derived antibody, and found that it associates not only with the Flo11 promoter, but additional Mor1 target upstream regions(Fig 4D). The levels of enrichment were lower for the Wor1 experiment than for the Mor1 experiment, so we cannot state the precise overlap between the regulators, but we do know it is very high. Of 16 Wor1 peaks, all 16 are also bound by Mor1. To directly test if Wor1 binds the *S. cerevisiae* Mor1 sites, electro mobility shift analysis (EMSA) was performed with *E. coli* purified MBP-Wor1(1-321), containing the first 321 amino acids of the protein, and radio-labeled fragments of the Flo11 promoter. MBP-Wor1(1-321) binds and shifts the Flo11 promoter in the absence of any other factors, and this binding requires the putative DNA binding motif; binding is eliminated when the motif is mutated (Fig 4E). Taken together, these results indicate the DNA-binding specificity of these regulators is conserved between distant fungal species.

To test whether the even more distantly related Ryp1 from *H. capsulatum* recognizes the same DNA sequence, we ectopically expressed Ryp1 in the *Δmor1Δmor2* strain. We found that Ryp1 activated transcription from the 1175-1225 upstream region and that

mutations in the Mor1 binding site destroyed this activation (Fig 4B). Thus, Mor1, Ryp1, and Wor1 proteins all show a similar DNA binding specificity. The ability of Wor1 to bind DNA maps to the 300 AA region of the protein that is broadly conserved across the fungal lineage. Thus it is very likely that its DNA-binding specificity has been preserved over at least 400 million years.

Comparison of Mor1 and Wor1 Target Genes

Given that Mor1 and Wor1 recognize the same DNA sequence, we asked how their direct target genes have changed during the approximately 300 million years they have been diverging from a common ancestor[16]. We began with the Mor1 target genes (as determined by ChIP) and mapped them to their *C. albicans* orthogroups (See materials and methods). Note that it was not possible to carry out this analysis for Ryp1 because the orthology relationships are uncertain due to the large evolutionary distance between these species. Of 87 *S. cerevisiae* Mor1 targets with at least one ortholog in *C. albicans*, we found only 16 targets that were bound in both species, an observation consistent with the high level of rewiring documented for the ascomycete lineage [21,22] (Fig 5A). Although the overlap is small, it is statistically significant, indicating that a small portion of the ancestral circuit has been preserved in *C. albicans* and *S. cerevisiae*. 10 of these 16 conserved targets are transcriptional regulators (Table 1). While transcriptional regulators are enriched overall in the Mor1 bound set, this bias for conserved transcription factors in the overlap of the two data sets is significant ($P<.05$)(Fishers Exact Test). We performed the same analysis in the opposite direction – mapping Wor1 targets to *S. cerevisiae* orthologs, and, as expected, found similar results (Fig 5 A-C). These results show that while most of the downstream targets of Mor1 and Wor1 have diverged from one

another, a core set of conserved transcriptional regulators have been preserved as targets of Wor1/Mor1 from the common ancestor of *S. cerevisiae* and *C. albicans* (Fig5D, Fig S1).

Discussion

Members of the Mor1 family are conserved through all fungal lineages, and Mor1 orthologs have been identified as critical regulators of the white-opaque switch in *C. albicans,* the mycelial-yeast transition in *H. capsulatum*, and conidiation and pathogenesis in the plant pathogen *Fusarium oxysporum*[5,10,23]. We investigated the Mor1 protein in *S. cerevisiae* to better understand how this family of regulators has evolved to regulate these related yet phenotypically distinct cellular differentiation processes.

We first considered whether divergence at the level of protein function could account for the diverse cellular phenotypes governed by these orthologs. While we cannot rule out the contribution of species-specific protein function, we determined that Mor1 and its *C. albicans* ortholog Wor1 act on the same discrete DNA motif upstream of *S. cerevisiae* Flo11 gene, and that Wor1 binds this sequence directly *in vitro*. Moreover, Ryp1 from the distantly related dimorphic pathogen *H. capsulatum* can activate transcription from this same discrete sequence. These results demonstrate that these novel DNA binding proteins act on a highly conserved motif, and that their transcriptional activity is conserved across a great distance of fungal evolution.

We next sought to determine the extent of rewiring that has occurred between the *S.cerevisiae* Mor1 and *C. albicans* Wor1 regulatory circuits. These species diverged from

a common ancestor ~300 million years ago. We performed whole-genome ChIP-Chip experiments and mapped the direct gene targets of Mor1 in *S. cerevisiae*, then compared them to the published targets of Wor1 in *C. albicans*. There was dramatic divergence in the downstream targets of Mor1/Wor1; over 80% of Mor1 targets with a clear ortholog are not bound by Wor1 in *C. albicans.* This result is similar to previous reports, arguing that there is significant divergence in transcription factor binding across fungal species [21,22]. In contrast to the overall trend of target divergence, there was conserved binding upstream of orthologous transcriptional regulators in each species, with 50% of these transcriptional regulators bound by both Mor1 and Wor1. While this bias towards conservation of transcription factor targets has been previously noted between closely related *Saccharomyces* species [22], we were surprised to see such strong conservation here, given the larger evolutionary distance and phenotypic differences between these two species.

To gain insight into the genetic interactions between Mor1 and these conserved transcriptional regulators, we took advantage of the extensive genome-wide binding data available for *S. cerevisiae[2]*. The upstream region of Mor1 is bound by all major filamentous regulators, suggesting it is a central member of this network. Remarkably, 8 of the 10 transcription factors which are conserved targets of Mor1/Wor1 are also targets of all 6 core members of the core filamentous growth network in *S. cerevisiae[19]* (Rox1 is bound by 5 members, Mor2 is only bound by Sok2). In addition, there is a significant overlap of downstream Mor1 gene targets with the known filamentous regulators Sok2, Ste12, and Tec1. This supports the model that these regulators are members of a highly interconnected network which regulates a complex developmental program [2]. In

support of this idea, Mor1 is necessary for filamentous growth, and its ectopic expression can drive filamentation under normally non-inducing conditions.

Taken together, this analysis suggests two distinct tiers of regulatory targets in the Mor1 circuit. One tier consists of a conserved set of transcriptional regulators, which form a highly inter-connected regulatory network. These regulators have a broadly conserved function in governing fungal cellular differentiation, serving critical roles in filamentation, condidiation, and appressorium formation in distant fungal species - all processes which involve dramatic cell wall and cell shape rearrangements (Table S2 collects many of these phenotypes). A second, largely divergent tier of targets involves genes downstream of these transcriptional regulators, such as cell wall and membrane proteins. Rewiring of targets at this level could allow species to specialize these processes to meet their specific needs, without dramatically altering the architecture of the circuit.

This model is demonstrated by the case of Mor1 and Wor1, where species-specific phenotypes may be attributed to major divergence in their downstream gene targets, even as their connections to other regulators are conserved. Cell wall proteins and trans-membrane transporters are highly enriched in the target sets of both Mor1 and Wor1 (GO-Analysis, Table S5 and S6), suggesting that while the specific gene targets may change; a similar core functionality is retained. *C. albicans* white-opaque switching differs from *S. cerevisiae* filamentous growth in many ways, but many of the same mechanistic principles are shared between the two processes, including the elongation of cell shape and restructuring of the cell wall. It is tempting to speculate that the interactions between many of the transcriptional regulators targeted in this circuit may be

conserved; forming a core network regulating cellular differentiation in fungi; and it is the divergence of their downstream targets that gives rise to specialized cell types.

Materials and Methods

Strains and Growth Conditions. Standard laboratory media was used as described in[24]. YEP media supplemented with 2% glucose was used for all microarray and ChIP experiments. B-Galactosidase assays were performed in SCD media lacking either URA, HIS, TRP, URA TRP, or URA HIS to maintain selection of plasmids. Synthetic low ammonium dextrose (SLAD) nitrogen starvation media was prepared as previously described, washing agar three times to remove traces of nitrogen[13]. Strains were constructed in the Σ2000 background, a gift from the lab of Hiten Madhani. Selectable markers were amplified with 50-bp of homology on each side of the gene to be replaced, and transformed using standard methods[25]. Mor1-GFP was generated using a fusion of a S65T GFP cassette in frame at the C-terminus of the protein, and nuclear localization was visualized as had been previously described.

Plasmids. To analyze activity from fragments upstream of the Flo11 gene, plasmids pFlo11 2/1 to pFlo11 15/14 were used, which were previously constructed from the pLG669z backbone[15]. To narrow a functional region of Mor1 activity, 50bp fragments were synthesized which overlapped by 25bp and had XhoI compatible ends, annealed, and ligated into pLG669z digested with XhoI. For the overexpression of Ryp1, its coding region was amplified from *Histoplasma capsulatum* G217B and ligated into digested CEN linked pTEF vector. For overexpression of Wor1, a previously described codon-

optimized version was digested and ligated into CEN linked pTEF. For Mor1, all attempts at PCR amplification of the Mor1 open reading frame were unsuccessful. We suspect this might be due to highly repetitive sequence in the C-terminal half of the protein. The N-terminal half of Mor1 was PCR amplified up until a HindIII site, and cloned into a CEN linked pTEF vector. To complete the open reading frame, plasmid YGPM-25i24, containing a genomic fragment encompassing the entire Mor1 ORF, was obtained as a gift from the lab of Greg Prelich[26], digested with HindIII and SmaI, and cloned in-frame to the pTEF-5'Mor1 plasmid.

Electromobility Gel Shift Assays. All assays were performed identically as described in Lohse et al. (in press).

Invasion Assays. Invasion assays were performed as previously described[14]. Briefly, haploid-invasion was tested by plating cells on YEPD media for 2 days, then washing gently under a steady stream of running water for 20 seconds. Diploid pseudohyphal growth was tested by plating cells at a low density on SLAD agar, allowing individual colonies to form over 7 days, and then photographing colonies under a dissection scope. Induction of the pseudohyphal growth programs was analyzed by patching cells on selectable SD media for 2 days, then washing under a steady stream of running water. Pseudohyphal microcolonies were observed by plating cells at low density under a cover slip on a slide topped with selectable SD media. Microcolonies were allowed to form overnight and then were then photographed at 63X magnification on an inverted Axiovert microscope using Axiovision software.

B-Galactosidase Assays. β-galactosidase assays were performed using a standard protocol [27]. Strains were grown in selectable media to maintain selection for one or both plasmids depending on the experiment. For each strain three colonies were grown overnight, diluted back, and allowed to reach log phase. Cells were harvested, permeablized, and activation assays performed.

Chromatin Immunoprecipitation. Chromatin immunoprecipitation experiments were performed as previously described in detail by our lab[28]. 200mls of log phase culture was crosslinked with 1% formaldehyde, quenched with 125mm glycine, and harvested for experiments. Polyclonal anti-GFP antibody (ab290) was purchased from Abcam. Additional antibodies were generated against N-terminal and C-terminal peptides sequences from Yel007w and Yhr177w, by Bethyl. Peptides were chosen to minimize cross reactivity. C-terminal antibodies were used in initial experiments against Yel007w, and for all experiments against Yhr177w. Yeast Chip on Chip arrays were ordered from Agilent Technologies, and contained ~244k probes with an average resolution of ~50nt. We note that these arrays were designed against the *S. cerevisiae* S288c reference strain, as the genome sequence of Σ had not yet been completed, however the minor sequence variation between these genomes would not impact our experiments. Agilent Chip Analytics software v 1.2 was used to determine peaks of enrichment in 2 independent experiments. Data was visualized using Mochiview, (http://johnsonlab.ucsf.edu)./[29]. Our criteria for calling enriched peaks required that Chip Analytics called a peak in 1 of 2 experiments, that the peak was not significantly enriched over a tRNA or ribosomal protein genes, and that it was visually discernable as a discrete peak with ~> log 2 enrichment in experimental strains and no enrichment in control untagged or deletion

strains. We chose to eliminate tRNA and ribosomal protein genes due to high background signal found in many independent experiments with different antibodies and conditions in our lab. We chose to confirm peaks by eye due to the high level of background noise being called by Chip Analytics – all peaks which we called as bound are presented in supplementary material. Independent analysis of hand calls with a recently developed peak-calling utility within Mochiview found almost complete concordance.

Orthology Mapping. All mapping was done using the published orthogroups generated by the SYNERGY algorithm and available at the MIT website http://www.broadinstitute.org/regev/orthogroups/[30]. All statistical analyses hold true for this purely systematic approach and the hand annotations described below. Three definitive hand annotations were made for the purpose of this analysis. **1.** Mor1 and Mor2 are clearly orthologs of Wor1, as demonstrated by protein sequence alignment across many fungal species. Therefore they are mapped 2:1 to *C. albicans* Wor1. *C. albicans* Pth2 was lost prior to the WGD. While it is an ortholog to *S. pombe* pac2 and many other proteins, it is not present in *S. cerevisiae.* This is reclassified from a 2:2 relationship. **2.** Tec1 has a definitive TEA/ATTS DNA binding domain, and has been identified in *C. albicans* for many years as having one clear ortholog, known as CaTec1[31]. SYNERGY mis-annotates this as a 1:4 annotation, calling 4 possible proteins in *C. albicans* as orthologs. This has been re-classified a 1:1 orthologous relationship. **3.** Rfg1 has been shown by this lab to be an ortholog to the *S. cerevisiae* protein Rox1[32], and looking at a tree of orthologs makes it clear that it is an ortholog of Rox1, and the only ortholog of Rox1. This is reclassified from a 0 to 0 to a 1:1 relationship.

1. Materna SC, Davidson EH (2007) Logic of gene regulatory networks. Curr Opin Biotechnol 18: 351-354.
2. Borneman AR, Leigh-Bell JA, Yu H, Bertone P, Gerstein M, et al. (2006) Target hub proteins serve as master regulators of development in yeast. Genes Dev 20: 435-448.
3. Tuch BB, Li H, Johnson AD (2008) Evolution of eukaryotic transcription circuits. Science 319: 1797-1799.
4. Huang G, Wang H, Chou S, Nie X, Chen J, et al. (2006) Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. Proceedings of the National Academy of Sciences of the United States of America 103: 12813-12818.
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, Borneman AR, Daniels KJ, Pujol C, Wu W, et al. (2006) TOS9 regulates white-opaque switching in Candida albicans. Eukaryot Cell 5: 1674-1687.
7. Lohse MB, Johnson AD (2009) White-opaque switching in Candida albicans. Curr Opin Microbiol 12: 650-654.
8. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. PLoS Biology 5: e256.
9. 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.
10. Nguyen VQ, Sil A (2008) Temperature-induced switch to the pathogenic yeast form of *Histoplasma capsulatum* requires Ryp1, a conserved transcriptional regulator. Proceedings of the National Academy of Sciences of the United States of America 105: 4880-4885.
11. Gancedo JM (2001) Control of pseudohyphae formation in Saccharomyces cerevisiae. FEMS Microbiol Rev 25: 107-123.
12. Gagiano M, Bauer FF, Pretorius IS (2002) The sensing of nutritional status and the relationship to filamentous growth in Saccharomyces cerevisiae. FEMS Yeast Res 2: 433-470.
13. Gimeno CJ, Ljungdahl PO, Styles CA, Fink GR (1992) Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. Cell 68: 1077-1090.
14. Lo WS, Dranginis AM (1998) The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by Saccharomyces cerevisiae. Mol Biol Cell 9: 161-171.
15. Rupp S, Summers E, Lo HJ, 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.
16. Taylor JW, Berbee ML (2006) Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia 98: 838-849.
17. Li F, Palecek SP (2005) Identification of Candida albicans genes that induce Saccharomyces cerevisiae cell adhesion and morphogenesis. Biotechnol Prog 21: 1601-1609.

18. Huang G, Wang H, Chou S, Nie X, Chen J, 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.
19. Monteiro PT, Mendes ND, Teixeira MC, d'Orey S, Tenreiro S, et al. (2008) YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in Saccharomyces cerevisiae. Nucleic Acids Res 36: D132-136.
20. Borneman AR, Zhang ZD, Rozowsky J, Seringhaus MR, Gerstein M, et al. (2007) Transcription factor binding site identification in yeast: a comparison of high-density oligonucleotide and PCR-based microarray platforms. Funct Integr Genomics 7: 335-345.
21. Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD (2008) The evolution of combinatorial gene regulation in fungi. PLoS Biol 6: e38.
22. Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, et al. (2007) Divergence of transcription factor binding sites across related yeast species. Science 317: 815-819.
23. Michielse CB, van Wijk R, Reijnen L, Manders EM, Boas S, et al. (2009) The nuclear protein Sge1 of Fusarium oxysporum is required for parasitic growth. PLoS Pathog 5: e1000637.
24. (1991) Guide to yeast genetics and molecular biology. Methods Enzymol 194: 1-863.
25. Longtine MS, McKenzie A, 3rd, Demarini DJ, Shah NG, Wach A, et al. (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast 14: 953-961.
26. Jones GM, Stalker J, Humphray S, West A, Cox T, et al. (2008) A systematic library for comprehensive overexpression screens in Saccharomyces cerevisiae. Nat Methods 5: 239-241.
27. Rupp S (2002) *LacZ* assays in yeast. Methods in Enzymology 350: 112-131.
28. Nobile CJ, Nett JE, Hernday AD, Homann OR, Deneault JS, et al. (2009) Biofilm matrix regulation by Candida albicans Zap1. PLoS Biol 7: e1000133.
29. Homann OR, Johnson AD MochiView: versatile software for genome browsing and DNA motif analysis. BMC Biol 8: 49.
30. Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. Nature 449: 54-61.
31. Schweizer A, Rupp S, Taylor BN, Rollinghoff M, Schroppel K (2000) The TEA/ATTS transcription factor CaTec1p regulates hyphal development and virulence in Candida albicans. Mol Microbiol 38: 435-445.
32. Kadosh D, Johnson AD (2001) Rfg1, a protein related to the Saccharomyces cerevisiae hypoxic regulator Rox1, controls filamentous growth and virulence in Candida albicans. Mol Cell Biol 21: 2496-2505.

Figure 1: Mor1 is required for haploid invasive growth and pseudohyphal growth in *S. cerevisiae*

A. Haploid invasive growth assay. Haploid "a" cells were plated on YPD and grown for 2 days at 30 degrees, and photographed. Plates were then washed under a gentle stream of water for 15 seconds and photographed again. The genotypes of the strains are shown.

B. Pseudohyphal growth assay. Diploid cells were plated on SLAD media and grown for 7 days, then photographed. A representative colony from each plate is shown. *Δmor1* and *Δmor1Δmor2* plates never displayed pseudohyphal growth.

C. Flo11 expression measurements. The coding sequence of Flo11 was replaced with GFP(S65T). Haploid "a" cells were grown in YPD to log phase, then photographed. All panels have 500ms exposure times. While WT and *Δmor2* cells display bright and variegated expression of GFP, *Δmor1* and *Δmor1Δmor2* cells have undetectable expression under these conditions.

D. Mor1 expressed from the pTEF promoter drives pseudohyphal growth in SD media conditions. Cells grown on SD under coverslip 16 hours.

Figure 2: Mor1 is a central transcriptional regulator of filamentous growth

A. ChIP plots demonstrate binding of Mor1 to key pseudohyphal regulatory targets genome wide. X-axis is a plot of S. cerevisiae chromosomes, centered on peak of enrichment. Y-Axis is log2 enrichment over input. Purple bars indicate Chip-analytics binding calls.

B. Network of interactions between Mor1 and known pseudohyphal regulators. Connections are based on Borneman et al and Mor1 ChIP data. Mor1 driven from a pTEF promoter in a CEN linked plasmid drives invasive growth in SD selective media.

C. Venn diagrams demonstrating overlap between the binding peaks called for Sok2, Ste12, or Tec1 (Borneman et al), with peaks called for Mor1.

Figure 3: Mor1 acts on a discrete motif within the Flo11 Promoter

A. Schematic of UAS-less pCyc1 plasmid
B. 400bp fragments of the Flo11 promoter were inserted into a UAS-less Cyc1 promoter upstream of a LacZ reporter construct (Rupp et al.) These plasmids were transformed into WT Haploid "a" cells, or *Δmor1* haploid "a" cells, and standard β-galactosidase assays were performed in log-phase SD –URA conditions. β-galactosidase activity is plotted on the y-axis.

C. 50bp fragments of the Flo11 promoter were inserted into the UAS-less Cyc1 plasmid. Strains containing these plasmids were measured for β-galactosidase activity under the same conditions as A. Fragment 1175-1225 has the strongest activity, and it is Mor1 dependent.

D. A Wor1 binding motif was identified in *C. albicans*, and predicted to exist within the 1175-1225 Flo11 promoter fragment. The motif is displayed as an inset LOGO. 3 bases of this motif were mutated as shown, and the corresponding plasmid was transformed into yeast and β-galactosidase activity was measured. "1175-1225 MUT" indicates construct with 3 mutated bases.

Figure 4: A highly diverged circuit regulates a conserved set of transcriptional regulators.

(A-C). The number of genes mapped from species A and also found to be in the Mor1/Wor1 bound gene set of species B, as a fraction of the total genes bound in species A that can be mapped to species B.

A. When only 1:1 orthologous relationships are considered
B. When all orthologous relationships are considered
C. When only sequence specific DNA binding proteins are considered

D. A network diagram of conserved transcription factor targets. Each node represents one **orthologous pair** of factors, where at least one member of the pair is bound by either Wor1 or Mor1. For example: ScSok2 and ScPhd1 participate in a 2:2

relationship with CaEfg1 and CaEfh1. Therefore there are 4 possible pairs of interactions – ScSok2:CaEfg1, ScSok2:CaEfh1, ScPhd1:CaEfg1, and ScPhd1:CaEfh1. Two of these pairs are represented as blue nodes bound by both Mor1 and Wor1. Two of these pairs are represented as pink nodes bound only by Mor1, because Efh1 is not a Wor1 target.

This diagram demonstrates that while there is a conserved set of transcription factors bound by both Mor1 and Wor1, there is also dramatic divergence as many targets in one species are not bound in the other.

Figure 5: DNA binding and activation functions of this class of protein are conserved throughout fungi

A. A reporter plasmid carrying either Cyc1 with the 1175-1225 Mutated fragment, or Cyc1 with the 1175-1225 Flo11 promoter fragment, was co-transformed with a plasmid containing the TEF promoter driving either an empty vector or *C. albicans* Wor1. All strains are haploid "a" Δmor1Δmor2. Both plasmids were selected for, and β-galactosidase activity is plotted on the y-axis. Only a strain containing both the 1175-1225 LacZ reporter and pTEF-Wor1 is able to rescue activity in a Δmor1Δmor2 strain background.

B. Same as A, however instead of pTEF-Wor1, a pTEF-Ryp1 plasmid is used.

C. Wor1 and Ryp1 driven by a CEN linked pTEF promoter induce invasive pseudohyphal growth on SD media. Cells were spotted and washed under a stream of water after 2 days growth, pre and post wash are shown. Single cells were plated on solid SD media underneath a coverslip and allowed to grow for 16 hours. Photos of representative micro colonies were taken.

D. ChIP-Chip plot demonstrating that Wor1 associates with the Flo11 promoter *in vivo*. Wor1 was immunoprecipitated from a strain carrying the pTEF-Wor1 plasmid, or a control strain carrying vector plasmid. Fold change is depicted as log2 scale on the y-axis. Wor1 is 16 fold enriched at the Flo11 promoter. Plotted on the same graph is a ChIP where the C-terminal peptide antibody of Mor1 was used in a WT or Δmor1 strain. Enrichment in the two IPs is comparable.

E. MBP-Wor1(1-321) was purified from E.coli and used in an EMSA assay with radio-labeled probe at indicated concentrations. Flo11 1175-1225 is the fragment responsive *in vitro*, Flo11 1175-1225 MUT contains 3 point mutations in the core of the DNA motif as indicated in figure 2.

Table 1: Conserved targets of Mor1 and Wor1

Supplemental Figure 1

A. An alternative way to display the network diagram. All possible orthologous pairings of transcriptional regulators where at least one member of the pair is bound are counted. They are displayed in a Venn diagram as either bound by Mor1, bound by Wor1, or bound by both
B. The same as A, however taking into consideration ALL possible ortholog pairings where at least one member of the pair is bound. This is not restricted to transcriptional regulators, and is not depicted as a network diagram because the number of nodes is too large to display. A comparison of (E) and (F) indicates how much more overlap there is for transcriptional regulators, than for targets as a whole

Supplemental Table 1: Expression Microarray Data

Supplemental Table 2: A collection of phenotypes for conserved regulators bound by Mor1 and Wor1

Supplemental Table 3: Mor1 Bound Genes

Supplemental Table 4: Mor2 Bound Genes

Supplemental Table 5: GO Enrichment for Mor1 Bound Genes

Supplemental Table 6: GO Enrichment for Wor1 Bound Genes

Figure 1.

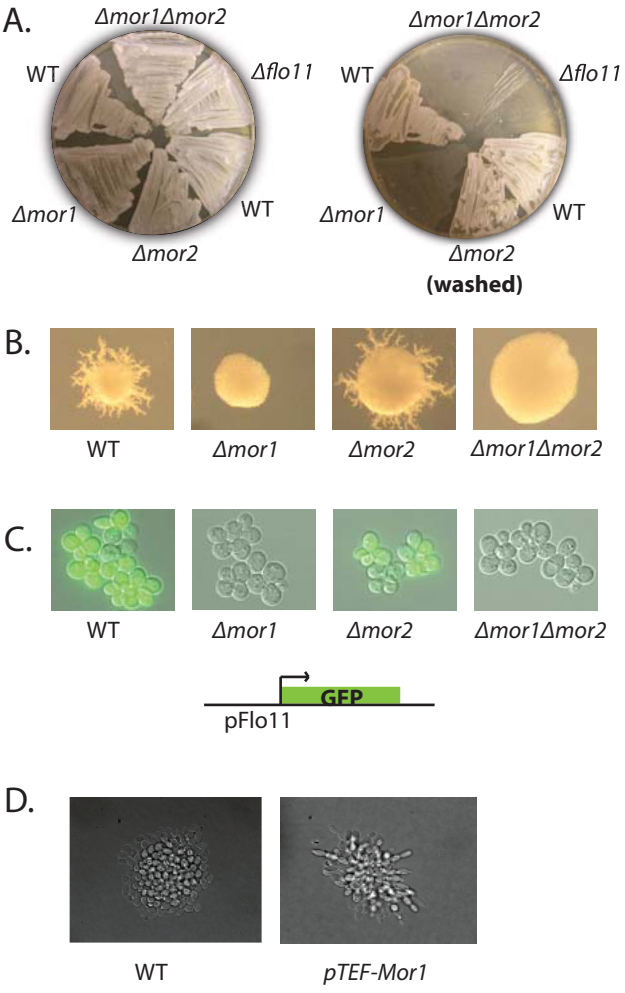
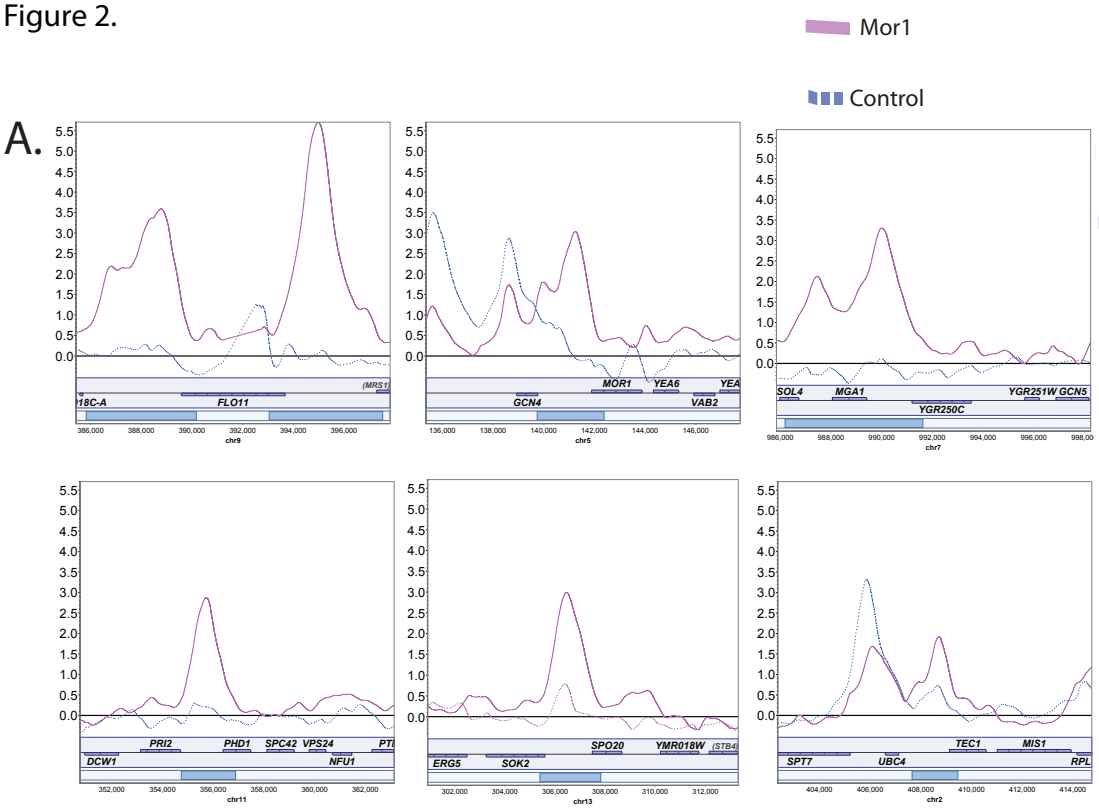
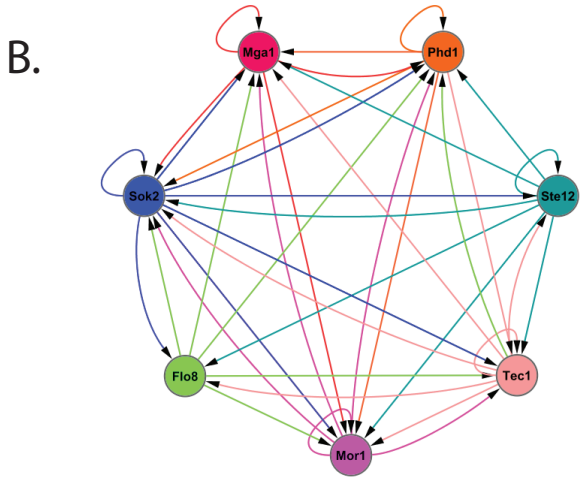

A.
Δmor1Δmor2
WT
Δflo11
Δmor1
WT
Δmor2
Δmor1Δmor2
WT
Δflo11
Δmor1
WT
Δmor2
(washed)
B.
WT
Δmor1
Δmor2
Δmor1Δmor2
C.
WT
Δmor1
Δmor2
Δmor1Δmor2
GFP
pFlo11
D.
WT
pTEF-Mor1

Figure 2.


Mor1
Control
A.
FLO11
MOR1
YEA6
GCN4
VAB2
SOL4
MGA1
YGR251W
GCN5
YGR250C
PRI2
PHD1
SPC42
VPS24
DCW1
NFU1
SPO20
YMR018W
ERG5
SOK2
TEC1
MIS1
SPT7
UBC4
RPL


B.
Mga1
Phd1
Sok2
Ste12
Flo8
Tec1
Mor1

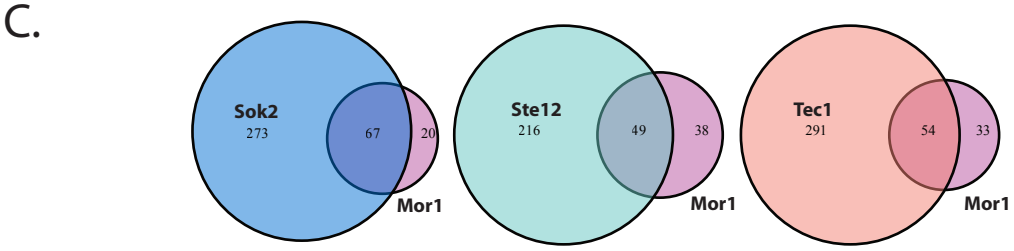

C.
Sok2
273
67
20
Mor1
Ste12
216
49
38
Mor1
Tec1
291
54
33
Mor1

Figure 3.

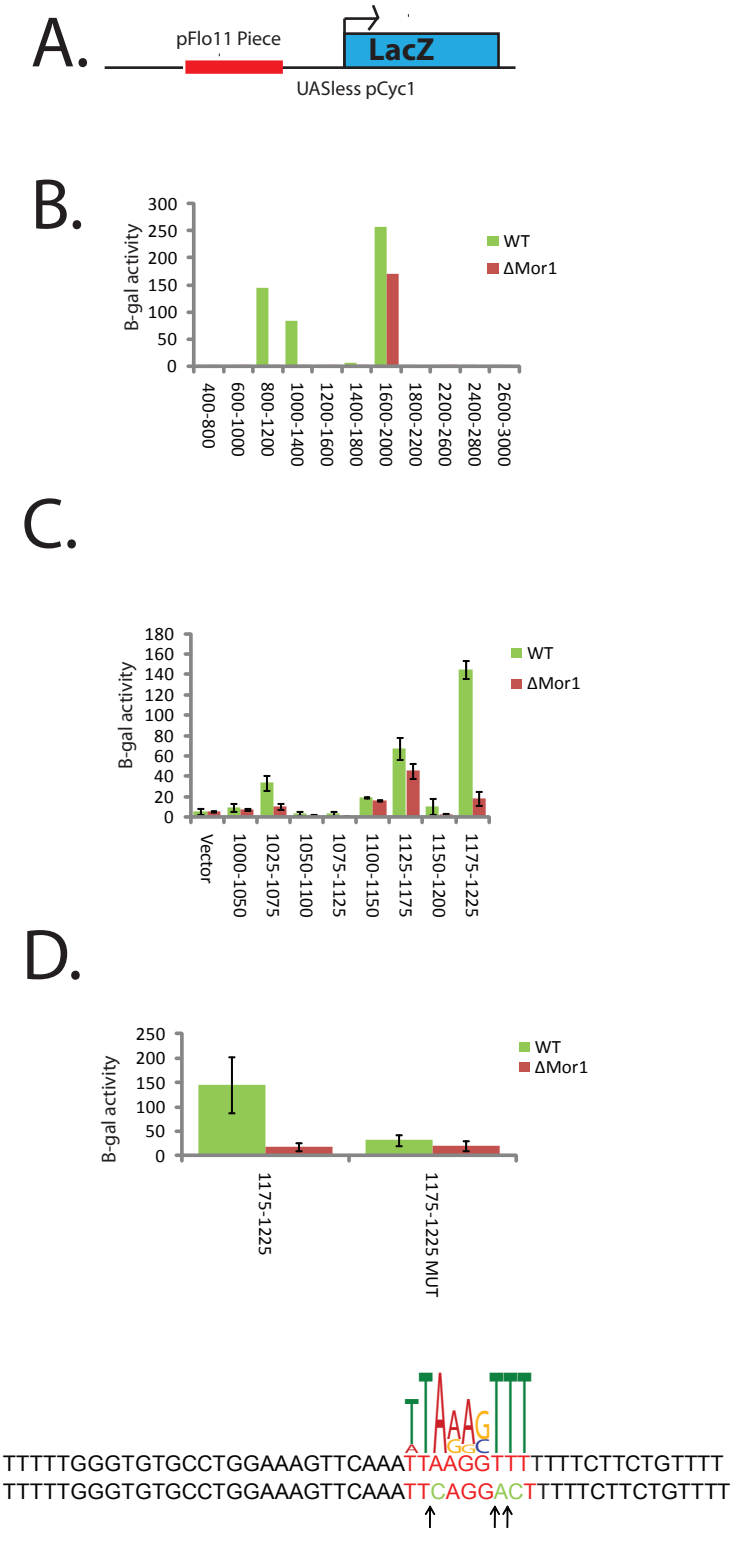

A.
pFlo11 Piece
LacZ
UASless pCyc1
B.
B-gal activity
300
250
200
150
100
50
0
WT
ΔMor1
400-800
600-1000
800-1200
1000-1400
1200-1600
1400-1800
1600-2000
1800-2200
2200-2600
2400-2800
2600-3000
C.
B-gal activity
180
160
140
120
100
80
60
40
20
0
WT
ΔMor1
Vector
1000-1050
1025-1075
1050-1100
1075-1125
1100-1150
1125-1175
1150-1200
1175-1225
D.
B-gal activity
250
200
150
100
50
0
WT
ΔMor1
1175-1225
1175-1225 MUT
TTTTTGGGTGTGCCTGGAAAGTTCAAATTAAGGTTTTTTTCTTCTGTTTT
TTTTTGGGTGTGCCTGGAAAGTTCAAATTCAGGACTTTTTCTTCTGTTTT

Figure 4.

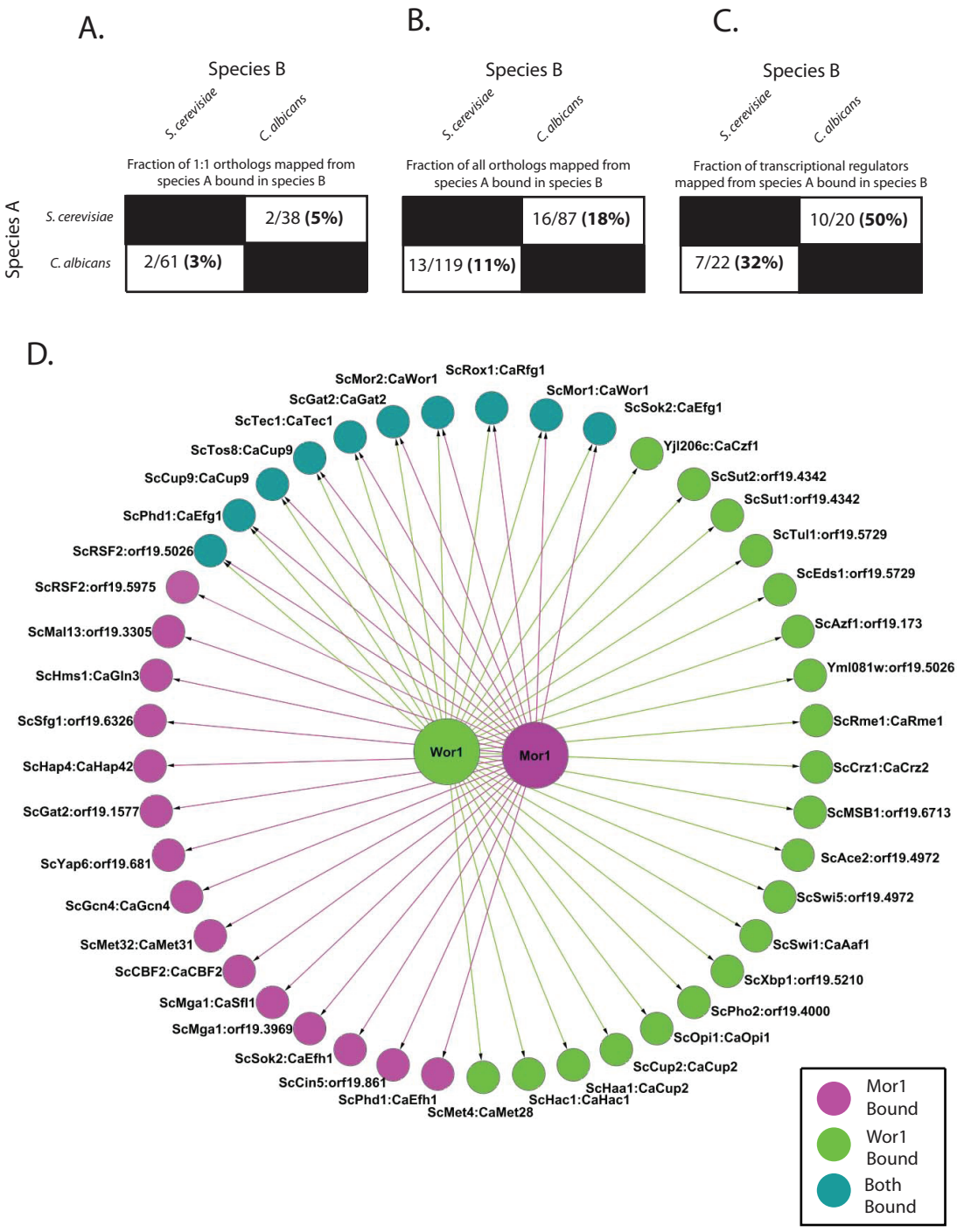

A.
B.
C.
Species B
S. cerevisiae
C. albicans
Fraction of 1:1 orthologs mapped from species A bound in species B
Fraction of all orthologs mapped from species A bound in species B
Fraction of transcriptional regulators mapped from species A bound in species B
Species A
S. cerevisiae
C. albicans
2/38 (5%)
2/61 (3%)
16/87 (18%)
13/119 (11%)
10/20 (50%)
7/22 (32%)
D.
ScRox1:CaRfg1
ScMor2:CaWor1
ScMor1:CaWor1
ScGat2:CaGat2
ScSok2:CaEfg1
ScTec1:CaTec1
Yjl206c:CaCzf1
ScTos8:CaCup9
ScSut2:orf19.4342
ScCup9:CaCup9
ScSut1:orf19.4342
ScPhd1:CaEfg1
ScTul1:orf19.5729
ScRSF2:orf19.5026
ScEds1:orf19.5729
ScRSF2:orf19.5975
ScAzf1:orf19.173
ScMal13:orf19.3305
Yml081w:orf19.5026
ScHms1:CaGln3
ScRme1:CaRme1
ScSfg1:orf19.6326
Wor1
Mor1
ScCrz1:CaCrz2
ScHap4:CaHap42
ScMSB1:orf19.6713
ScGat2:orf19.1577
ScAce2:orf19.4972
ScYap6:orf19.681
ScSwi5:orf19.4972
ScGcn4:CaGcn4
ScSwi1:CaAaf1
ScMet32:CaMet31
ScXbp1:orf19.5210
ScCBF2:CaCBF2
ScPho2:orf19.4000
ScMga1:CaSfl1
ScOpi1:CaOpi1
ScMga1:orf19.3969
ScCup2:CaCup2
ScSok2:CaEfh1
ScHaa1:CaCup2
ScCin5:orf19.861
ScHac1:CaHac1
ScPhd1:CaEfh1
ScMet4:CaMet28
Mor1 Bound
Wor1 Bound
Both Bound

Figure 5.

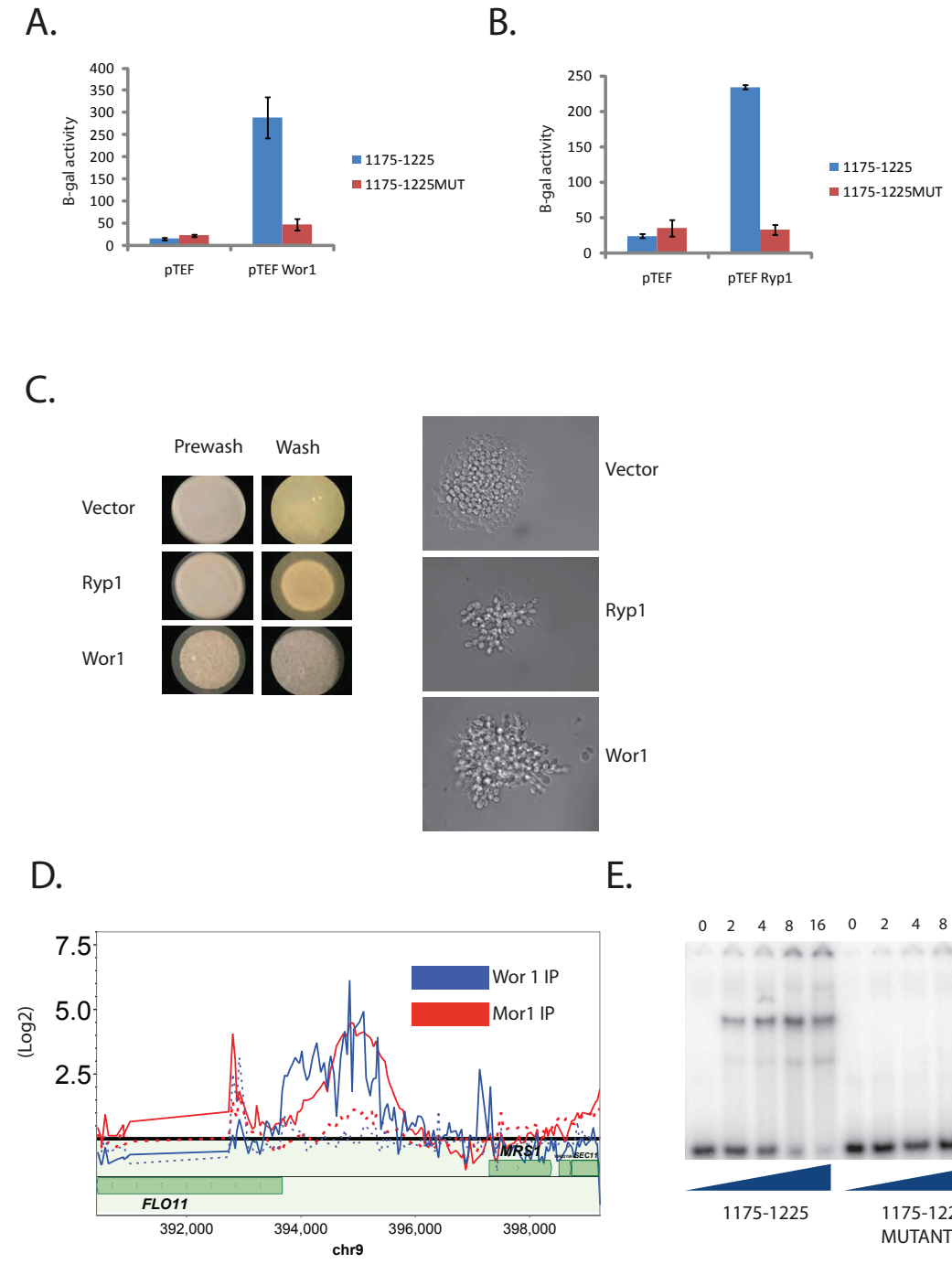

A.
B-gal activity
400
350
300
250
200
150
100
50
0
pTEF
pTEF Wor1
1175-1225
1175-1225MUT
B.
B-gal activity
250
200
150
100
50
0
pTEF
pTEF Ryp1
1175-1225
1175-1225MUT
C.
Prewash
Wash
Vector
Ryp1
Wor1
Vector
Ryp1
Wor1
D.
(Log2)
7.5
5.0
2.5
Wor 1 IP
Mor1 IP
MRS1
SEC11
FLO11
392,000
394,000
396,000
398,000
chr9
E.
0 2 4 8 16 0 2 4 8 16
1175-1225
1175-1225
MUTANT

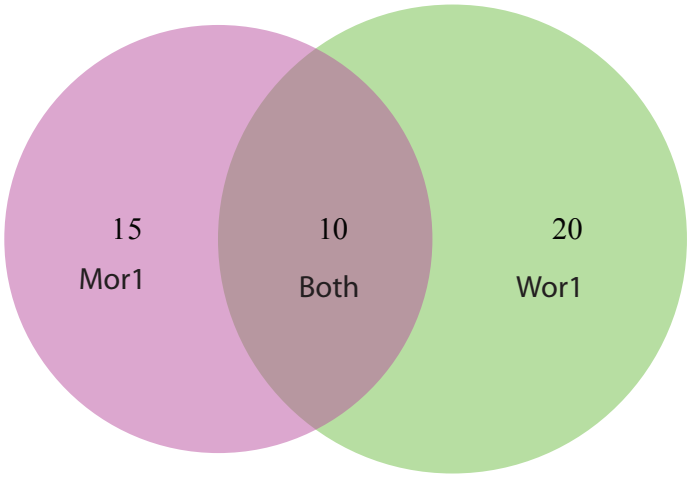
Table 1.

S. cerevisiae Bound Target	*C. albicans* Bound Target	Gene Description *S. cerevisiae*	Gene description *C. albicans*
Mor1	Wor1	Master regulator of pseudohyphal growth	Master regulator of W/O switch
Mor2		Involved in pseudohyphal growth network	Master regulator of W/O switch
Tos8	Cup9	Target of SBF 8 (no known phenotype), bound by all 7 PHD regulators	Hyphal regulated, deletion enhances spider invasion
Cup9		Regulator of peptide transporters	Hyphal regulated, deletion enhances spider invasion
Phd1	Efg1	Overexpression enhances pseudohyphal growth	Regulates W/O switch, ☒lamentation, cell wall
Sok2		Deletion enhances pseudohyphal growth, positively regulates meiosis	Regulates W/O switch, ☒lamentation, cell wall
Rox1	Rfg1	Repressor of hypoxic genes, required for pseudohyphal growth	Repressor of ☒lamentous growth
Gat2	Gat2	Bound by all 7 PHD regulators	Deletion has severe ☒lamentation defects
Rsf2	orf19.5026	Involved in glycerol based growth	No known function
Tec1	Tec1	Required for pseudohyphal growth	Required for hyphal growth
Ena1	Ena21	Plasma membrane Na+-ATPase transporter, regulated by numerous pathways	Regulated in numerous conditions, antifungals
Pma1	Pma1	Plasma membrane H+-ATPase, major ATPase ,essential, glucose regulated	Plasma membrane H+-ATPase, highly expressed
Pmc1	Pmc1	Vacuolar Ca2+ ATPase, calcineurin regulated, mutant is calcium sensitive	Mutant is drug sensitive
Sun4	Sun41	Glucanase involved in cell septation	Cell wall glycosidase, hyphal induced, Efg1 regulated
Cln1	Hgc1	G1 Cyclin, late G1 speci☒c expression promotes S phase	Hypha speci☒c G1 expression, required for hyphal growth
Sps4	orf19.7502	Induced during sporulation	No known function

Transcription Factor

Figure S1.

A.


All pairs of orthologous transcription factors bound in at least one species
15
Mor1
10
Both
20
Wor1

B.

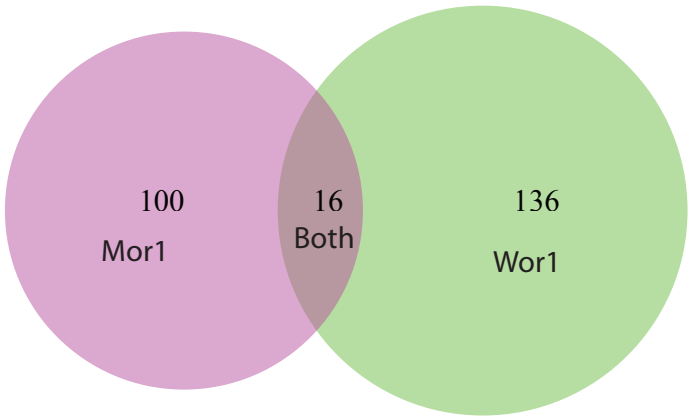

All pairs of orthologs bound in at least one species
100
Mor1
16
Both
136
Wor1

Table S1 (Page 1 of 2)

Top 20 Mor1 bound genes upregulated upon deletion of Mor1 or Mor2

SGD ORF	NAME	ΔMor1	ΔMor1ΔMor2
YHR177W	MOR2	1.8025	-0.67235
YER033C	ZRG8	1.7365	1.86375
YJL078C	PRY3	1.72655	1.74665
YDR259C	YAP6	1.08925	1.4699
YFL051C	YFL051C	0.9626	0.68125
YKL109W	HAP4	0.81665	1.10085
YKR102W	FLO10	0.6835	1.269
YGR247W	CPD1	0.6784	0.2985
YJL105W	SET4	0.66625	0.77275
YGR287C	YGR287C	0.6544	0.80425
YBR083W	TEC1	0.5394	0.4811
YJL214W	HXT8	0.50775	0.63815
YBR238C	YBR238C	0.47985	0.41035
YGL096W	TOS8	0.47835	0.2131
YGR249W	MGA1	0.40235	0.11335
YGR138C	TPO2	0.3911	1.00915
YMR011W	HXT2	0.3905	0.444
YJR127C	RSF2	0.38955	0.4679
YAR050W	FLO1	0.38425	0.6427
YHR210C	YHR210C	0.37275	0.4875

Table S1 (Page 2 of 2)

Top 20 Mor1 bound genes downregulated upon deletion of Mor1 or Mor2

SGD ORF	NAME	ΔMor1	ΔMor1ΔMor2
YIR019C	MUC1	-5.4465	
YEL007W	MOR1	-2.665	-1.16315
YDR077W	SED1	-1.205	0.046
YGR214W	RPS0A	-0.828	-0.367
YKL043W	PHD1	-0.74175	-0.52635
YGR148C	RPL24B	-0.661	-0.49
YMR199W	CLN1	-0.5664	-0.59175
YBL030C	PET9	-0.547	0.118
YLR333C	RPS25B	-0.54665	-0.5241
YDR525W-A	SNA2	-0.51545	-0.43065
YNR060W	FRE4	-0.469	-0.56355
YIL011W	TIR3	-0.445	-0.62835
YGR108W	CLB1	-0.41425	-0.65975
YPL177C	CUP9	-0.3596	-0.3084
YDR040C	ENA1	-0.33465	-0.14965
YIL119C	RPI1	-0.321	-0.48275
YMR016C	SOK2	-0.299	0.04385
YAR035W	YAT1	-0.2789	-0.1284
YMR006C	PLB2	-0.2442	0.01675
YLR213C	CRR1	-0.2321	-0.1213

Table S2.

S. cerevisiae gene	Orthologs with morphological function	Phenotype	References
Mor1/Mor2	*H. capsulatum* Ryp1, *F. oxysporum* Sge1	Required for *H. capsulatum* yeast form and proper conidiation, *F. oxysporum* conidiation and pathogenesis	Proc Natl Acad Sci U S A. 2008 Mar 25;105(12): 4880-5.PLoS Pathog. 2009 Oct;5(10):e1000637.
Tos8/Cup9	*M. grisea* Pth12	Required for appressorium formation	PLoS Genet. 2009 Dec;5(12):e1000757.
Phd1/Sok2	*A. nidulans* , *M. grisea* , *P. marneffei* , *U. maydis* StuA	Required for *A. nidulans* Hulle cell and cleistothecium formation, *A.nidulans* and P. *marneffei* conidiation, *M. grisea* appressorium formation, *U. maydis* dimorphic growth	Mol Microbiol. 2000 Dec;38(5):1034-47. Genes Dev. 1992 6: 1770-1782 Mol Microbiol. 2002 May;44(3):621-31. Mol Plant Microbe Interact. 2010 Feb;23(2):211-22 Mol Plant Microbe Interact. 2007 Sep;20(9):1102-11.
Tec1	*A. nidulans* , *P. marneffei* AbaA	Required for *A. nidulans* and *P. marneffei* conidiation, as well as *P.marneffei* yeast form growth	Cell Volume 57, Issue 5, 2 June 1989, Pages 859-868
Gat2	*A. nidulans* NsdD	Required for Hulle cell and cleistothecium formation, conidiation	Mol Microbiol. 2001 Jul;41(2):299-309.

Table S3 (Page 1 of 3)

Chr.	START	END	STRAND	Gene 1	Name	Gene 2	Name
chr1	27356	30225	+	YAL063C	FLO9		
chr1	47751	49078	+	YAL051W	OAF1		
chr1	189442	190849	+	YAR035W	YAT1		
chr1	195289	199462	+	YAR035C-A			
chr1	202213	203776	+	YAR050W	FLO1		
chr10	25500	27435	+	YJL216C		YJL214W	HXT8
chr10	223083	224891	+	YJL105W	SET4		
chr10	293169	295312	+	YJL078C	PRY3	YJL077W-B	
chr10	422007	423332	+	YJL008C	CCT8	YJL005W	CYR1
chr10	605212	608946	+	YJR094C	IME1	YJR094W-A	RPL43B
chr10	639021	640540	+	YJR115W		YJR116W	
chr10	661898	664545	+	YJR127C	RSF2		
chr11	67601	69271	+	YKL201C	MNN4		
chr11	146470	147905	+	YKL164C	PIR1		
chr11	216073	217789	+	YKL120W	OAC1		
chr11	229348	232264	+	YKL110C	KTI12	YKL109W	HAP4
chr11	354739	356879	+	YKL043W	PHD1		
chr11	641575	644238	+	YKR102W	FLO10	YKR100C	SKG1
chr11	644334	646307	+	YKR102W	FLO10		
chr12	35938	37642	+	YLL053C		YLL052C	AQY2
chr12	369483	371679	+	YLR110C	CCW12	YLR113W	HOG1
chr12	567457	568805	+	YLR213C	CRR1	YLR214W	FRE1
chr12	808770	810202	+	YLR342W	FKS1		
chr13	279235	280982	+	YMR006C	PLB2		
chr13	286782	288339	+	YMR011W	HXT2		
chr13	305396	307824	+	YMR016C	SOK2	YMR017W	SPO20
chr13	387727	388967	+	YMR056C	AAC1	YMR058W	FET3
chr13	539614	541740	+	YMR135C	GID8	YMR136W	GAT2
chr13	556555	557852	+	YMR145C	NDE1		
chr13	660483	663005	+	YMR199W	CLN1		

Table S3 (Page 2 of 3)

chr14	300957	303713	+	YNL178W	RPS3	YNL180C	RHO5
chr14	350876	353236	+	YNL147W	LSM7	YNL146C-A	
chr14	500261	502075	+	YNL066W	SUN4		
chr14	617640	620246	+	YNL008C	ASI3	YNL007C	SIS1
chr14	737954	740212	+	YNR060W	FRE4		
chr15	109062	111733	+	YOL110W	SHR5	YOL109W	ZEO1
chr15	382300	385860	+	YOR028C	CIN5	YOR030W	DFG16
chr15	390983	392859	+	YOR032C	HMS1	YOR032W-A	
chr15	670218	672463	+	YOR178C	GAC1	YOR179C	SYC1
chr15	903338	905671	+	YOR315W	SFG1	YOR313C	SPS4
chr16	114033	115239	+	YPL230W	USV1		
chr16	213393	216354	+	YPL177C	CUP9		
chr16	583658	588834	+	YPR011C		YPR013C	
chr16	590604	592051	+	YPR015C			
chr16	678440	680148	+	YPR065W	ROX1	YPR063C	
chr16	828079	829922	+	YPR148C		YPR149W	NCE102
chr2	163581	166976	+	YBL030C	PET9	YBL029C-A	
chr2	407651	409456	+	YBR083W	TEC1	YBR082C	UBC4
chr2	465596	467204	+	YBR112C	CYC8	YBR114W	RAD16
chr2	696690	698410	+	YBR238C			

chr2	799198	800506	+	YBR296C-A		YBR297W	MAL33
chr4	155694	156878	+	YDL170W	UGA3	YDL171C	GLT1
chr4	384807	388270	+	YDL037C	BSC1	YDL038C	
chr4	464435	466181	+	YDR011W	SNQ2		
chr4	538317	540096	+	YDR040C	ENA1	YDR041W	RSM10
chr4	598787	601321	+	YDR077W	SED1		
chr4	721654	723462	+	YDR132C			
chr4	964533	965699	+	YDR253C	MET32	YDR254W	CHL4
chr4	974631	977787	+	YDR259C	YAP6	YDR260C	SWM1
chr4	1079961	1081576	+	YDR309C	GIC2		
chr4	1344529	1346255	+	YDR441C	APT2		
chr4	1489269	1491243	+	YDR524W-A		YDR524C-B	
chr5	76640	79105	+	YEL040W	UTR2		
chr5	139741	142380	+	YEL009C	GCN4	YEL007W	

Table S3 (Page 3 of 3)

chr5	220792	223112	+	YER033C	ZRG8	YER034W	
chr5	237181	239185	+	YER044C	ERG28	YER043C	SAH1
chr5	490183	493020	+	YER158C		YER159C	BUR6
chr6	18092	20769	+	YFL053W	DAK2	YFL056C	AAD6
chr6	30985	33151	+	YFL051C			
chr7	322738	325555	+	YGL096W	TOS8		
chr7	482976	484346	+	YGL007C-A		YGL008C	PMA1
chr7	558439	559938	+	YGR036C	CAX4	YGR038W	ORM1
chr7	699982	703951	+	YGR108W	CLB1	YGR106C	
chr7	765370	767455	+	YGR138C	TPO2	YGR140W	CBF2
chr7	787352	789282	+	YGR148C	RPL24B	YGR149W	
chr7	918628	921498	+	YGR213C	RTA1	YGR214W	RPS0A
chr7	984257	985458	+	YGR246C	BRF1	YGR247W	CPD1
chr7	986185	991622	+	YGR249W	MGA1		
chr7	1068570	1070075	+	YGR287C		YGR288W	MAL13
chr8	15951	18540	+	YHL039W		YHL046C	
chr8	396249	397714	+	YHR149C	SKG6	YHR150W	PEX28
chr8	454783	457033	+	YHR177W			
chr8	521469	523500	+	YHR210C			
chr9	53266	54768	+	YIL155C	GUT2	YIL153W	RRD1
chr9	137378	143451	+	YIL119C	RPI1	YIL118W	RHO3
chr9	332073	333649	+	YIL013C	PDR11	YIL011W	TIR3
chr9	337956	339764	+	YIL009W	FAA3		
chr9	385826	390166	+	YIR018C-A			
chr9	393045	397523	+	YIR019C	MUC1	YIR021W	MRS1
chr9	426472	428477	+	YIR038C	GTT1		

Table S4

Chr.	START	END	STRAND	Gene 1	Name	Gene 2	Name
chr1	197318	199408	+	YAR050W	FLO1	YAR035C-A	
chr10	223185	224524	+	YJL105W	SET4		
chr10	605619	606978	+	YJR094C	IME1	YJR094W-A	RPL43B
chr10	662069	664646	+	YJR127C	RSF2		
chr11	354739	356523	+	YKL043W	PHD1		
chr11	644448	645941	+	YKR102W	FLO10		
chr13	305552	307423	+	YMR016C	SOK2	YMR017W	SPO20
chr13	387662	389381	+	YMR056C	AAC1	YMR058W	FET3
chr13	661094	662364	+	YMR199W	CLN1		
chr13	914851	915984	+	YMR319C	FET4		
chr14	493985	496173	+	YNL069C	RPL16B		
chr14	763518	764771	+	YNR070W		YNR069C	BSC5
chr15	384002	385860	+	YOR028C	CIN5	YOR030W	DFG16
chr15	391169	392469	+	YOR032W-A		YOR032C	HMS1
chr15	423802	425210	+	YOR049C	RSB1		
chr15	670462	672320	+	YOR178C	GAC1		
chr15	903530	905563	+	YOR315W	SFG1	YOR313C	SPS4
chr16	585441	587525	+	YPR013C			
chr16	828267	829842	+	YPR149W	NCE102	YPR148C	
chr16	920537	921617	+	YPR192W	AQY1		
chr2	164574	166206	+	YBL029C-A		YBL029W	
chr4	385472	387291	+	YDL037C	BSC1		
chr4	538317	539997	+	YDR040C	ENA1	YDR041W	RSM10
chr4	1055649	1056783	+	YDR297W	SUR2		
chr5	139643	143965	+	YEL009C	GCN4	YEL007W	
chr5	490690	491931	+	YER158C			
chr6	31032	33233	+	YFL051C			
chr7	206178	208476	+	YGL158W	RCK1	YGL157W	
chr7	323097	324556	+	YGL096W	TOS8		
chr7	765455	767225	+	YGR138C	TPO2	YGR140W	CBF2
chr7	919088	920511	+	YGR213C	RTA1	YGR214W	RPS0A
chr8	75100	78087	+	YHL015W	RPS20	YHL014C	YLF2
chr9	137824	139712	+	YIL119C	RPI1	YIL118W	RHO3
chr9	141212	142703	+	YIL117C	PRM5	YIL116W	HIS5
chr9	393807	396262	+	YIR019C	MUC1	YIR021W	MRS1
chr9	426596	428651	+	YIR038C	GTT1		

Table S5.

S. cerevisiae GO Terms

Terms from the Process Ontology			
Gene Ontology term	**Cluster frequency**	**Background frequency**	**P-value**
flocculation |AmiGO	5 out of 121 genes, 4.1%	10 out of 7166 background genes, 0.1%	9.67E-05
biological regulation |AmiGO	40 out of 121 genes, 33.1%	1171 out of 7166 background genes, 16.3%	0.0013
cell growth |AmiGO	9 out of 121 genes, 7.4%	79 out of 7166 background genes, 1.1%	0.00205
flocculation via cell wall protein-carbohydrate interaction |AmiGO	4 out of 121 genes, 3.3%	9 out of 7166 background genes, 0.1%	0.00297
regulation of cell size | AmiGO	10 out of 121 genes, 8.3%	109 out of 7166 background genes, 1.5%	0.0044
regulation of cellular component size |AmiGO	10 out of 121 genes, 8.3%	114 out of 7166 background genes, 1.6%	0.00654
regulation of anatomical structure size |AmiGO	10 out of 121 genes, 8.3%	114 out of 7166 background genes, 1.6%	0.00654
regulation of metabolic process| AmiGO	25 out of 121 genes, 20.7%	631 out of 7166 background genes, 8.8%	0.01311
regulation of cellular metabolic process | AmiGO	24 out of 121 genes, 19.8%	613 out of 7166 background genes, 8.6%	0.02332
regulation of transcription |AmiGO	19 out of 121 genes, 15.7%	427 out of 7166 background genes, 6.0%	0.02814
pseudohyphal growth | AmiGO	7 out of 121 genes, 5.8%	65 out of 7166 background genes, 0.9%	0.03307
regulation of biological process| AmiGO	32 out of 121 genes, 26.4%	973 out of 7166 background genes, 13.6%	0.03636
regulation of primary metabolic process | AmiGO	23 out of 121 genes, 19.0%	591 out of 7166 background genes, 8.2%	0.03715
Terms from the Function Ontology			
Gene Ontology term	**Cluster frequency**	**Background frequency**	**P-value**
sequence-specific DNA binding | AmiGO	19 out of 121 genes, 15.7%	203 out of 7166 background genes, 2.8%	8.11E-08
DNA binding | AmiGO	24 out of 121 genes, 19.8%	337 out of 7166 background genes, 4.7%	1.12E-07
transcription regulator activity | AmiGO	22 out of 121 genes, 18.2%	323 out of 7166 background genes, 4.5%	1.38E-06
transcription factor activity |AmiGO	10 out of 121 genes, 8.3%	56 out of 7166 background genes, 0.8%	2.10E-06
mannose binding | AmiGO	4 out of 121 genes, 3.3%	4 out of 7166 background genes, 0.1%	6.80E-06
monosaccharide binding |AmiGO	4 out of 121 genes, 3.3%	7 out of 7166 background genes, 0.1%	0.00023
sugar binding | AmiGO	4 out of 121 genes, 3.3%	8 out of 7166 background genes, 0.1%	0.00045
carbohydrate binding |AmiGO	4 out of 121 genes, 3.3%	9 out of 7166 background genes, 0.1%	0.0008
active transmembrane transporter activity | AmiGO	10 out of 121 genes, 8.3%	124 out of 7166 background genes, 1.7%	0.00368
specific RNA polymerase II transcription factor activity |AmiGO	6 out of 121 genes, 5.0%	55 out of 7166 background genes, 0.8%	0.02657
transmembrane transporter activity | AmiGO	14 out of 121 genes, 11.6%	300 out of 7166 background genes, 4.2%	0.04374
Terms from the Component Ontology			
Gene Ontology term	**Cluster frequency**	**Background frequency**	**P-value**
external encapsulating structure |AmiGO	12 out of 121 genes, 9.9%	100 out of 7166 background genes, 1.4%	9.39E-06
cell wall |AmiGO	12 out of 121 genes, 9.9%	100 out of 7166 background genes, 1.4%	9.39E-06
fungal-type cell wall |AmiGO	12 out of 121 genes, 9.9%	100 out of 7166 background genes, 1.4%	9.39E-06
plasma membrane |AmiGO	17 out of 121 genes, 14.0%	282 out of 7166 background genes, 3.9%	0.00044
membrane fraction |AmiGO	12 out of 121 genes, 9.9%	198 out of 7166 background genes, 2.8%	0.01197
insoluble fraction |AmiGO	12 out of 121 genes, 9.9%	198 out of 7166 background genes, 2.8%	0.01197
cell fraction |AmiGO	13 out of 121 genes, 10.7%	233 out of 7166 background genes, 3.3%	0.01426
membrane |AmiGO	34 out of 121 genes, 28.1%	1139 out of 7166 background genes, 15.9%	0.04312

Table S6

C.albicans GO Terms

Terms from the Process Ontology			
Gene Ontology term	**Cluster frequency**	**Background frequency**	**P-value**
protein amino acid O-linked glycosylation | AmiGO	7 out of 221 genes, 3.2%	18 out of 6435 background genes, 0.3%	0.00053
filamentous growth | AmiGO	31 out of 221 genes, 14.0%	420 out of 6435 background genes, 6.5%	0.01653
amine transport | AmiGO	9 out of 221 genes, 4.1%	51 out of 6435 background genes, 0.8%	0.02205
biofilm formation | AmiGO	9 out of 221 genes, 4.1%	51 out of 6435 background genes, 0.8%	0.02205
growth | AmiGO	31 out of 221 genes, 14.0%	441 out of 6435 background genes, 6.9%	0.04182
Terms from the Function Ontology			
Gene Ontology term	**Cluster frequency**	**Background frequency**	**P-value**
active transmembrane transporter activity| AmiGO	16 out of 221 genes, 7.2%	147 out of 6435 background genes, 2.3%	0.00497
alpha-1,3-mannosyltransferase activity |AmiGO	4 out of 221 genes, 1.8%	7 out of 6435 background genes, 0.1%	0.00555
transcription factor activity | AmiGO	17 out of 221 genes, 7.7%	172 out of 6435 background genes, 2.7%	0.00981
transmembrane transporter activity |AmiGO	24 out of 221 genes, 10.9%	300 out of 6435 background genes, 4.7%	0.01091
inorganic anion transmembrane transporter activity | AmiGO	5 out of 221 genes, 2.3%	18 out of 6435 background genes, 0.3%	0.03452
substrate-specific transmembrane transporter activity | AmiGO	21 out of 221 genes, 9.5%	269 out of 6435 background genes, 4.2%	0.04306
Terms from the Component Ontology			
Gene Ontology term	**Cluster frequency**	**Background frequency**	**P-value**
plasma membrane |AmiGO	37 out of 221 genes, 16.7%	488 out of 6435 background genes, 7.6%	0.00028
intrinsic to plasma membrane | AmiGO	12 out of 221 genes, 5.4%	73 out of 6435 background genes, 1.1%	0.00053
plasma membrane part |AmiGO	14 out of 221 genes, 6.3%	101 out of 6435 background genes, 1.6%	0.00071
integral to plasma membrane | AmiGO	10 out of 221 genes, 4.5%	53 out of 6435 background genes, 0.8%	0.00092
cell surface | AmiGO	17 out of 221 genes, 7.7%	166 out of 6435 background genes, 2.6%	0.00458
extracellular region |AmiGO	9 out of 221 genes, 4.1%	56 out of 6435 background genes, 0.9%	0.00979
external encapsulating structure | AmiGO	14 out of 221 genes, 6.3%	148 out of 6435 background genes, 2.3%	0.04885
cell wall | AmiGO	14 out of 221 genes, 6.3%	148 out of 6435 background genes, 2.3%	0.04885

Appendix 1

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

This appendix was primarily based on the biochemical work of Matthew Lohse. I contributed to the design and construction of experiments for the *in vivo* validation of the Wor1 DNA binding motif. I built the *Δmor1Δmor2* strains where the Wor1 orthologs in *S. cerevisiae* were knocked out, and contributed the pTEF-Wor1 plasmid and protocols for reporter assay construction. I also contributed to the alignment of the Wor1 conserved domain and the delineation of two distinct sub-families with conserved differences within this domain. These results are presented in Figure 4 and Supplementary Figure 1.

At the time this thesis was submitted, this paper was in revision at PNAS.

Classification: Biological Sciences, Biochemistry

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

[*]Corresponding author. Mailing Address: 600 16th Street, University of California San Francisco, San Francisco, CA 94158 Phone: 4154768783 Fax: 4155024315 E-mail: ajohnson@cgl.ucsf.edu

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 new class of motif (likely representing a new 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 several 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 immuneprecipitation 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. 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 1d). 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* (21). 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 (22)).

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 (23) 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 in a sequence dependent manner (Supplemental Figure S2b) although it is not clear if the Wor1 motif represents the optimal sequence for Pth2 binding.

Compared to the motifs recognized by some sequence-specific DNA-binding proteins, the Wor1 position-specific weight matrix is relatively information poor. This explains, at least in part, why it was not possible to computationally determine this motif based solely on the ChIP-Chip data.

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 UAS-less version of the *CYC1* promoter driving expression of LacZ. 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. It is worth noting that overexpressing Wor1 in *S. cerevisiae* induced increased adherence to plates and clumping in solution, consistent with previously reported results (24).

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 S3a). 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 S4a, S4b), 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 S3b, S3c). 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 novel DNA-binding motif 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, based on family members Wor1, Ryp1, and the related family member Pac2.

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 new superfamily of DNA-binding proteins. Although this protein family is deeply 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 (23), using the one site per sequence option with minimum and maximum motif size constraints of 10 and 20bp.

Cloning

Primers used in this study are included in Supplemental Table S3. Plasmids used in this study are included in Supplemental Table S4. Strains used in this study are included in Supplemental Table S5.

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 (EMD4Biosciences, 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 (28) which uses the lacZ system first described by Guarente *et al*. (29) . 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 (30). 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 (31). 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 (32). 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 (33). 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 Madhani lab for plasmids and strains and the Stroud lab, particularly Akram Alian and Sarah Griner, for help with chromatography. The authors thank Oliver Homann and Aaron Hernday for valuable comments on the manuscript.

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. 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.

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

23. 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.

24. Li F, Palecek SP (2005) Identification of Candida albicans genes that induce Saccharomyces cerevisiae cell adhesion and morphogenesis. *Biotechnol Prog* 21:1601-1609.

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. 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.

29. 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.

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

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

32. 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.

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

Figure Legends:

Figure 1: Wor1 is a member of a conserved family of fungal proteins. (a) Alignment of Wor1 homologs across 5 fungal species for 7 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 were chosen to incorporate ones that have been characterized as well as a wide range of fungal species. (b) Location of the two predicted globular domains of Wor1 from 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 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.

Figure 1.

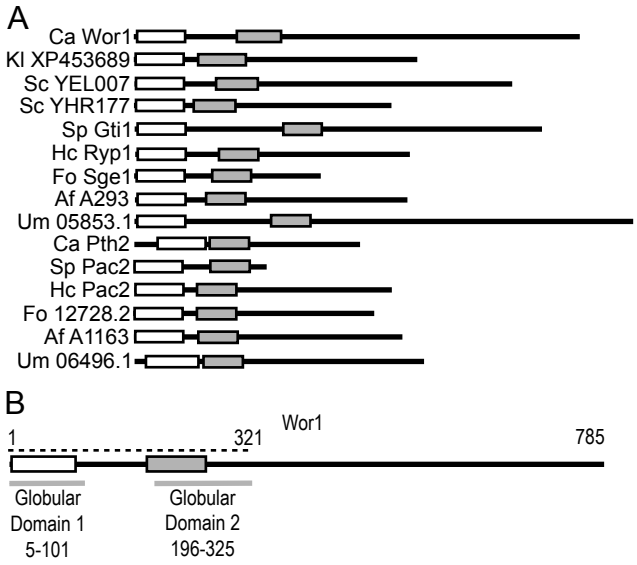

A
Ca Wor1
Kl XP453689
Sc YEL007
Sc YHR177
Sp Gti1
Hc Ryp1
Fo Sge1
Af A293
Um 05853.1
Ca Pth2
Sp Pac2
Hc Pac2
Fo 12728.2
Af A1163
Um 06496.1
B
Wor1
1
321
785
Globular
Domain 1
5-101
Globular
Domain 2
196-325

Figure 2.

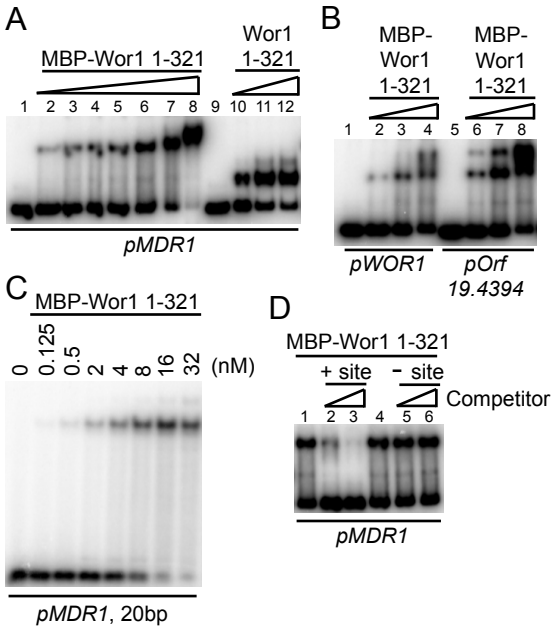

A
MBP-Wor1 1-321
Wor1 1-321
1 2 3 4 5 6 7 8 9 10 11 12
pMDR1
B
MBP-Wor1 1-321
MBP-Wor1 1-321
1 2 3 4 5 6 7 8
pWOR1
pOrf 19.4394
C
MBP-Wor1 1-321
0 0.125 0.5 2 4 8 16 32 (nM)
pMDR1, 20bp
D
MBP-Wor1 1-321
+ site
- site
Competitor
1 2 3 4 5 6
pMDR1

Figure 3.

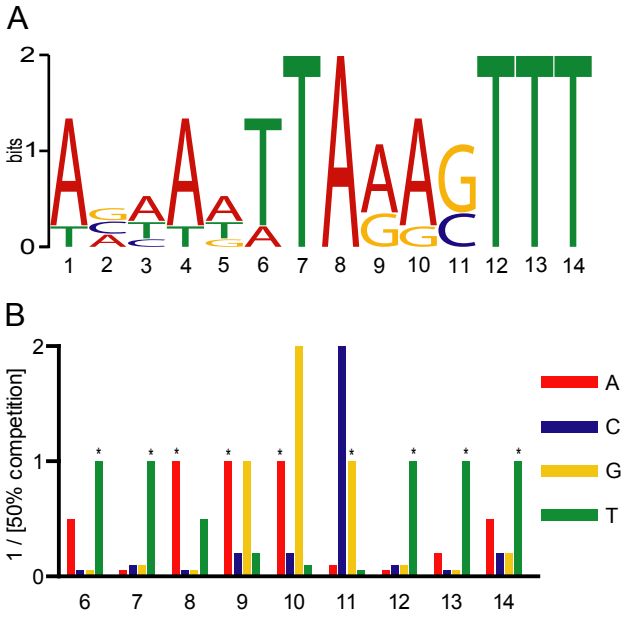

A
bits
2
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
1 / [50% competition]
2
1
0
6 7 8 9 10 11 12 13 14
A
C
G
T

Figure 4.

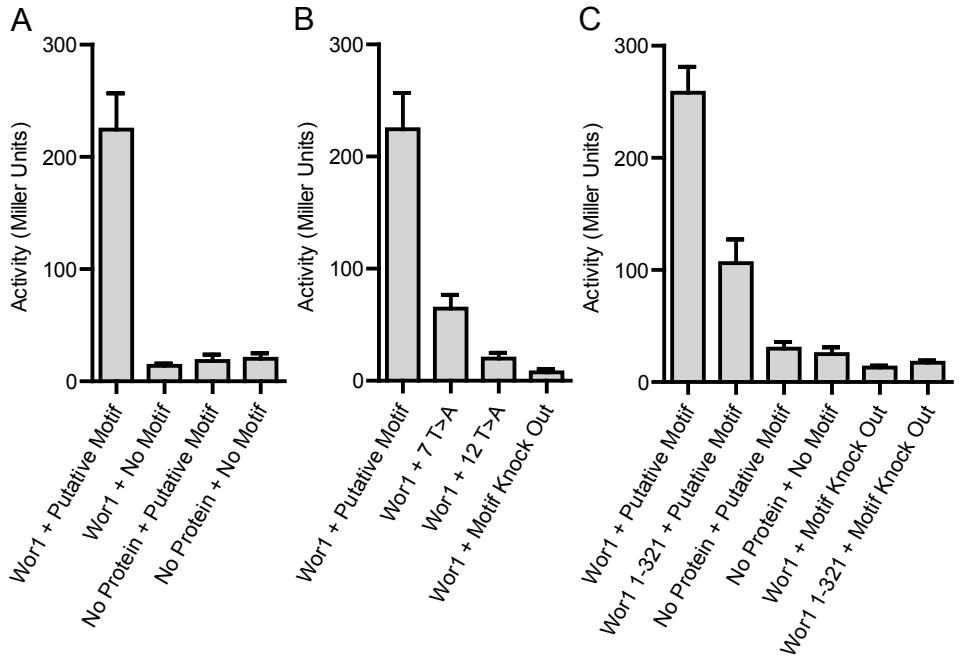

A
B
C
Activity (Miller Units)
300
200
100
0
Wor1 + Putative Motif
Wor1 + No Motif
No Protein + Putative Motif
No Protein + No Motif
Wor1 + 7 T>A
Wor1 + 12 T>A
Wor1 + Motif Knock Out
Wor1 1-321 + Putative Motif
No Protein + No Motif
Wor1 1-321 + Motif Knock Out

Figure 5.

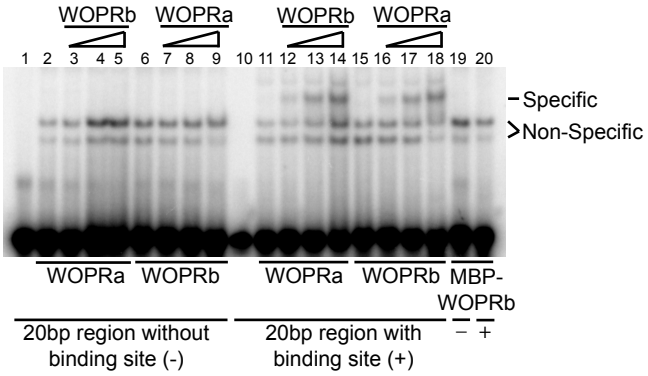
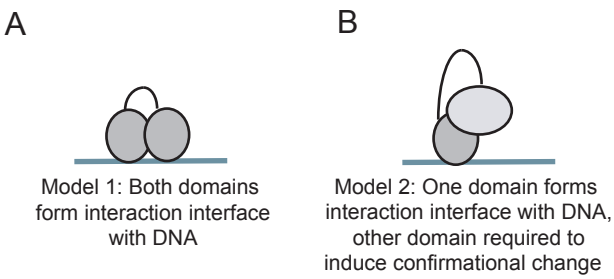

WOPRb
WOPRa
WOPRb
WOPRa
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Specific
Non-Specific
WOPRa
WOPRb
WOPRa
WOPRb
MBP-WOPRb
20bp region without binding site (-)
20bp region with binding site (+)
– +

Figure 6.


A
B
Model 1: Both domains form interaction interface with DNA
Model 2: One domain forms interaction interface with DNA, other domain required to induce confirmational change

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. Genes were chosen to incorporate family members that have been characterized as well as a diverse range of fungal species.

Supplemental Figure S2: Wor1 and Pth2 bind to specific sequences of DNA as monitored by mobility shift assays (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 (1) 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].

Supplemental Figure S3: Wor1 binds DNA as a monomer and sequence-specific DNA binding requires the presence of both conserved domains. (a) 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. (b) 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. (c) Lanes 1-8 as in (b) but with non-specific competitor included in the binding reactions to reduce non-specific binding. As in (b), Wor1 binding is only seen in the presence of both domains and the Wor1 binding site [lane 8].

Supplemental Figure S4: 6xHis Wor1 1-321 is a monomer is 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.

Supplemental Table S1: Wor1 constructs used in this study.

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

Supplemental Table S3: List of primers used in this study.

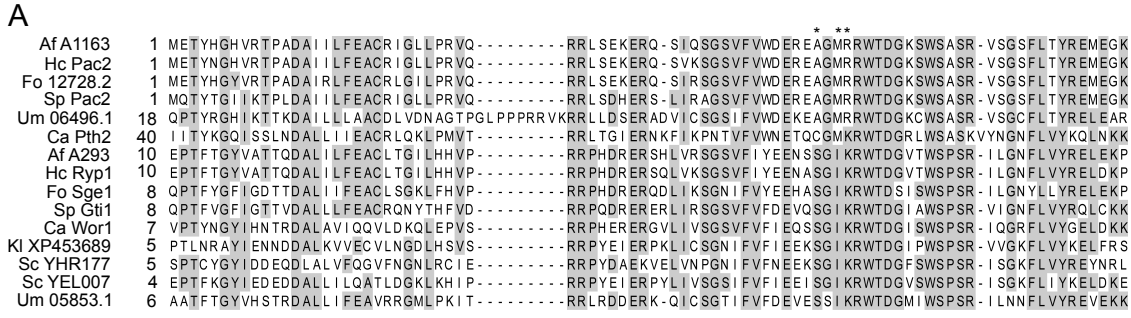
Supplemental Table S4: List of plasmids used in this study.

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

1. 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.

Figure S1.

A


Af A1163 1 METYHGHVRTPADAIILFEACRIGLLPRVQ---------RRLSEKERQ-SIQSGSVFVWDEREAGMRRWTDGKSWSASR-VSGSFLTYREMEGK
Hc Pac2 1 METYNGHVRTPADAIILFEACRIGLLPRVQ---------RRLSEKERQ-SVKSGSVFVWDEREAGMRRWTDGKSWSASR-VSGSFLTYREMEGK
Fo 12728.2 1 METYHGYVRTPADAIRLFEACRLGILPRVQ---------RRLSEKERQ-SIRSGSVFVWDEREAGMRRWTDGKSWSASR-VSGSFLTYREMEGK
Sp Pac2 1 MQTYTGIIKTPLDAIILFEACRIGLLPRVQ---------RRLSDHERS-LIRAGSVFVWDEREAGMRRWTDGKSWSASR-VSGSFLTYREMEGK
Um 06496.1 18 QPTYRGHIKTTKDAILLLAACDLVDNAGTPGLPPPRRVKRRLLDSERADVICSGSIFVWDEKEAGMRRWTDGKCWSASR-VSGCFLTYRELEAR
Ca Pth2 40 IITYKGQISSLNDALLIIEACRLQKLPMVT---------RRLTGIERNKFIKPNTVFVWNETQCGMKRWTDGRLWSASKVYNGNFLVYKQLNKK
Af A293 10 EPTFTGYVATTQDALILFEACLTGILHHVP---------RRPHDRERSHLVRSGSVFIYEENSSGIKRWTDGVTWSPSR-ILGNFLVYRELEKP
Hc Ryp1 10 EPTFTGYVATTQDALILFEACLTGILHHVP---------RRPHDRERSQLVKSGSVFIYEENASGIKRWTDGVTWSPSR-ILGNFLVYRELDKP
Fo Sge1 8 QPTFYGFIGDTTDALIIFEACLSGKLFHVP---------RRPHDRERQDLIKSGNIFVYEEHASGIKRWTDSISWSPSR-ILGNYLLYRELEKP
Sp Gti1 8 QPTFVGFIGTTVDALLLFEACRQNYTHFVD---------RRPQDRERERLIRSGSVFVFDEVQSGIKRWTDGIAWSPSR-VIGNFLVYRQLCKK
Ca Wor1 7 VPTYNGYIHNTRDALAVIQQVLDKQLEPVS---------RRPHERERGVLIVSGSVFVFIEQSSGIKRWTDGISWSPSR-IQGRFLVYGELDKK
Kl XP453689 5 PTLNRAYIENNDDALKVVECVLNGDLHSVS---------RRPYEIERPKLICSGNIFVFIEEKSGIKRWTDGIPWSPSR-VVGKFLVYKELFRS
Sc YHR177 5 SPTCYGYIDDEQDLALVFQGVFNGNLRCIE---------RRPYDAEKVELVNPGNIFVFNEEKSGIKRWTDGFSWSPSR-ISGKFLVYREYNRL
Sc YEL007 4 EPTFKGYIEDEDDALLILQATLDGKLKHIP---------RRPYEIERPYLIVSGSIFVFIEEISGIKRWTDGVSWSPSR-ISGKFLIYKELDKE
Um 05853.1 6 AATFTGYVHSTRDALLIFEAVRRGMLPKIT---------RRLRDDERK-QICSGTIFVFDEVESSIKRWTDGMIWSPSR-ILNNFLVYREVEKK

B

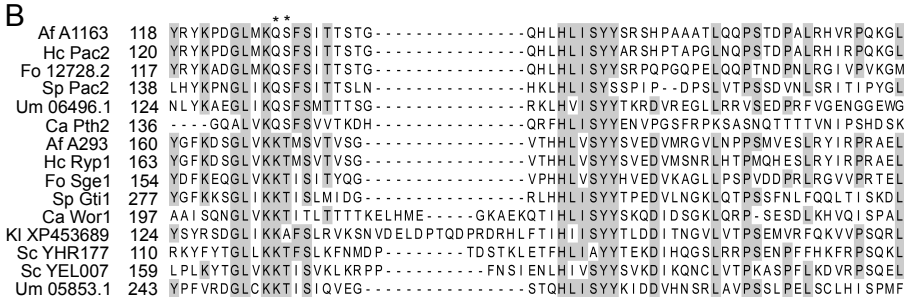

Af A1163 118 YRYKPDGLMKQSFSITTSTG---------------QHLHLISYYSRSHPAAATLQQPSTDPALRHVRPQKGL
Hc Pac2 120 YRYKPDGLMKQSFSITTSTG---------------QHLHLISYYARSHPTAPGLNQPSTDPALRHIRPQKGL
Fo 12728.2 117 YRYKADGLMKQSFSITTSTG---------------QHLHLISYYSRPQPGQPELQQPTNDPNLRGIVPVKGM
Sp Pac2 138 LHYKPNGLIKQSFSITTSLN---------------HKLHLISYSSPIP--DPSLVTPSSDVNLSRITIPYGL
Um 06496.1 124 NLYKAEGLIKQSFSMTTTSG---------------RKLHVISYYTKRDVREGLLRRVSEDPRFVGENGGEWG
Ca Pth2 136 ----GQALVKQSFSVVTKDH---------------QRFHLISYYENVPGSFRPKSASNQTTTTVNIPSHDSK
Af A293 160 YGFKDSGLVKKTMSVTVSG----------------VTHHLVSYYSVEDVMRGVLNPPSMVESLRYIRPRAEL
Hc Ryp1 163 YGFKDSGLVKKTMSVTVSG----------------VTHHLVSYYSVEDVMSNRLHTPMQHESLRYIRPRAEL
Fo Sge1 154 YDFKEQGLVKKTISITYQG----------------VPHHLVSYYHVEDVKAGLLPSPVDDPRLRGVVPRTEL
Sp Gti1 277 YGFKKSGLIKKTISLMIDG----------------RLHHLISYYTPEDVLNGKLQTPSSFNLFQQLTISKDL
Ca Wor1 197 AAISQNGLVKKTITLTTTTKELHME-----GKAEKQTIHLISYYSKQDIDSGKLQRP-SESDLKHVQISPAL
Kl XP453689 124 YSYRSDGLIKKAFSLRVKSNVDELDPTQDPRDRHLFTIHIISYYTLDDITNGVLVTPSEMVRFQKVVPSQRL
Sc YHR177 110 RKYFYTGLLKKTFSLKFNMDP--------TDSTKLETFHLIAYYTEKDIHQGSLRRPSENPFFHKFRPSQKL
Sc YEL007 159 LPLKYTGLVKKTISVKLKRPP----------FNSIENLHIVSYYSVKDIKQNCLVTPKASPFLKDVRPSQEL
Um 05853.1 243 YPFVRDGLCKKTISIQVEG----------------STQHLISYYKIDDVHNSRLAVPSSLPELSCLHISPMF

Figure S2.

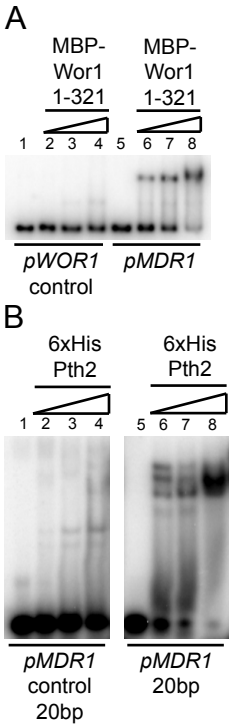

A
MBP-
Wor1
1-321
MBP-
Wor1
1-321
1 2 3 4 5 6 7 8
pWOR1
control
pMDR1
B
6xHis
Pth2
6xHis
Pth2
1 2 3 4
5 6 7 8
pMDR1
control
20bp
pMDR1
20bp

Figure S3.

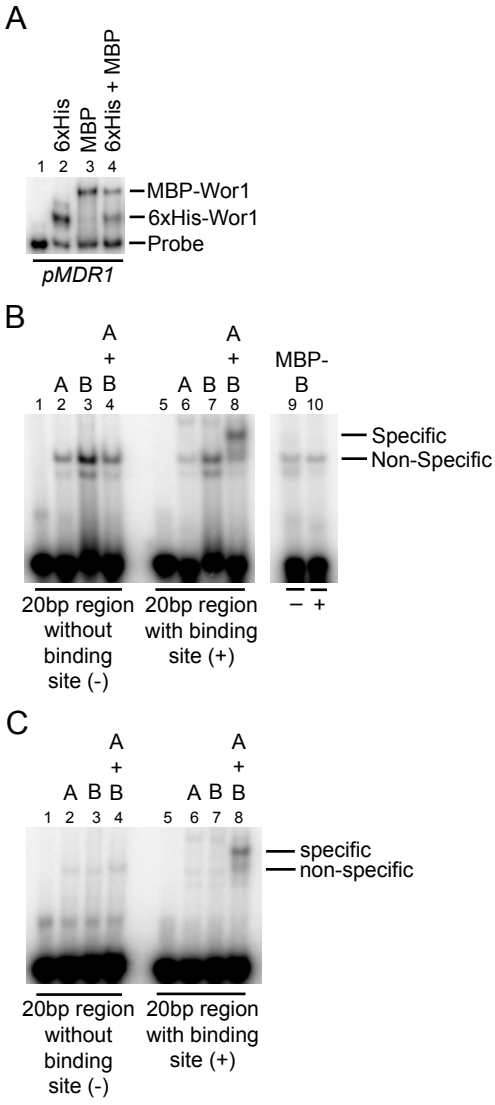

A
6xHis
MBP
6xHis + MBP
1 2 3 4
MBP-Wor1
6xHis-Wor1
Probe
pMDR1
B
A + B
A B B
1 2 3 4
5 6 7 8
MBP-B
9 10
Specific
Non-Specific
20bp region without binding site (-)
20bp region with binding site (+)
– +
C
specific
non-specific

Figure S4.

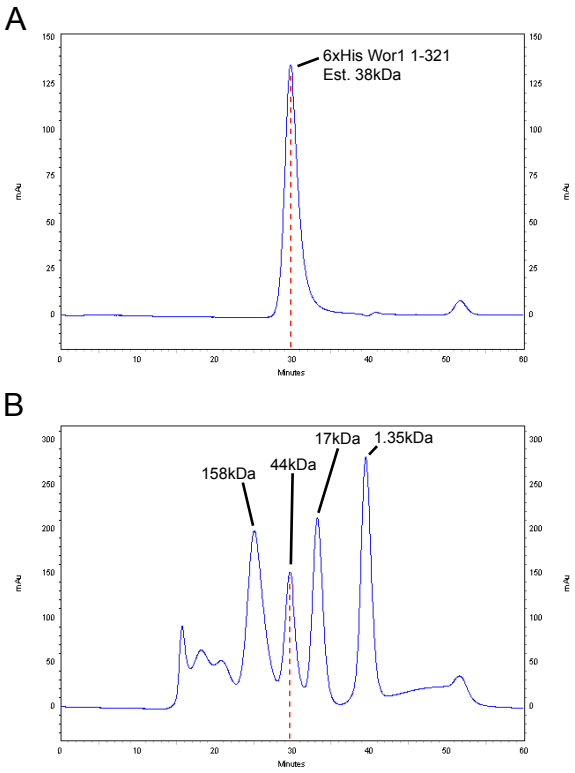

A
6xHis Wor1 1-321
Est. 38kDa
mAu
Minutes
B
158kDa
44kDa
17kDa
1.35kDa
mAu
Minutes

Table S1.

Called	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

Table S2.

Site	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

Table S3 (Page 1 of 4)

Wor1 CUG Codon Changes	
Name	Description
wor1.1 ClaI	Wor1 5' for (ClaI)
wor1.2 for	Wor1 CUG#1 for
wor1.2 rev	Wor1 CUG#1 rev
wor1.3 for	Wor1 CUG#2 for
wor1.3 rev	Wor1 CUG#2 rev
wor1.4 for	Wor1 CUG#3 for
wor1.4 rev	Wor1 CUG#3 rev
wor1.5 for	Wor1 CUG#4 for
wor1.5 rev	Wor1 CUG#4 rev
wor1.6 for	Wor1 CUG#5 for
wor1.6 rev	Wor1 CUG#5 rev
wor1.7 for	Wor1 CUG#6 for
wor1.7 rev	Wor1 CUG#6 rev
wor1.8 XhoI	Wor1 3' rev (XhoI)
Cloning Wor1	
RZO373	Wor1 5' NheI
RZO379	Wor1 (1-101) stop 3' XhoI
RZO377	Wor1 (1-321) stop 3' XhoI
MBL1015	Wor1 (196-321) 5' NheI
MBL1016	Wor1 (196-321) stop 3' XhoI
200bp EMSA Fragments	
MBL721	*pWOR1* 5'
MBL722	*pWOR1* 3'
MBL729	*pWOR1* control 5'
MBL730	*pWOR1* control 3'
MBL735	pOrf19.4394 5'
MBL736	pOrf19.4394 3'
MBL739	*pMDR1* 5'
MBL740	*pMDR1* 3'
20bp competition primers [only forward strand listed]	
pMDR1	
MBL771	120-140bp

Table S3 (Page 2 of 4)

MBL773	130-150bp
MBL775	140-160bp
MBL777	150-170bp
MBL779	160-180bp
pOrf19.4394	
MBL837	0-20bp
MBL839	10-30bp
MBL841	20-40bp
MBL843	30-50bp
MBL845	40-60bp
MBL847	50-70bp
MBL849	60-80bp
MBL851	70-90bp
MBL853	80-100bp
MBL855	120-140bp
MBL857	130-150bp
MBL859	140-160bp
pWOR1	
MBL861	minus 20-0bp
MBL863	minus 10-10bp
MBL865	0-20bp
MBL867	10-30bp
MBL869	20-40bp
MBL871	30-50bp
MBL873	40-60bp
MBL875	50-70bp
MBL877	60-80bp
MBL887	180-200bp
MBL889	190-210bp
MBL891	200-220bp
Motif Mutation Scan	
MBL893	2 T>C
MBL895	2 T>A
MBL897	3 A>G
MBL899	3 A>T
MBL901	6 G>A
MBL903	6 G>C

Table S3 (Page 3 of 4)

MBL905	7 T>C
MBL907	7 T>A
MBL909	8 T>C
MBL911	8 T>A
MBL913	9 T>C
MBL915	9 T>A
MBL917	6 G>T
MBL919	9 T>G
MBL921	WT Control
MBL925	3 A>C
MBL927	4 A>G
MBL929	5 A>G
MBL935	2 T>G
MBL939	4 A>C
MBL941	4 A>T
MBL943	5 A>C
MBL945	5 A>T
MBL947	7 T>G
MBL949	8 T>G
MBL951	1 T>A
MBL953	1 T>C
MBL955	1 T>G
S. cerevisiae cloning	
ΔYEL007 KO 5'	KO Forward
ΔYEL007 KO 3'	KO Reverse
ΔYHR177KO 5'	KO Forward
ΔYHR177 KO 3'	KO Reverse
Motifs for Activation	
MBL1039	Putative Motif Forward
MBL1040	Putative Motif Reverse
MBL1045	Motif KO Forward [7T>A, 8A>C,12T>A,13T>C]

Table S3 (Page 4 of 4)

MBL1046	Motif KO Reverse [7T>A, 8A>C,12T>A,13T>C]
MBL1048	Motif 7 T>A Forward
MBL1049	Motif 7 T>A Reverse
MBL1054	Motif 12 T>A Forward
MBL1055	Motif 12 T>A Reverse
CYC990	External Check Primer
Pth2 Codon Changes	
MBL1072	Pth2 5' [BamHI/NheI]
MBL1075	Pth2 CUG#1 5'
MBL1074	Pth2 CUG#1 3'
MBL1073	Pth2 3' [HindIII/XhoI]

Table S4.

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

Table S5

Description	Figure	Strain	Source
Sigma 2000 Starting Background	not shown	MY1404	Madhani Lab Stocks
ΔYEL007 ΔYHR177	not shown	CC150	This Paper
Wor1 + Putative Motif	2a, 2b, 2c	MLY780	This Paper
Wor1 + No Motif	2a, 2b, 2c	MLY768	This Paper
No Protein + Putative Motif	2a, 2c	MLY773	This Paper
No Protein + No Motif	2a, 2c	MLY766	This Paper
Wor1 + 7 T>A	2b	MLY783	This Paper
Wor1 + 12 T>A	2b	MLY785	This Paper
Wor1 + Motif Knock Out	2b	MLY782	This Paper
No Protein + 7 T>A	not shown	MLY776	This Paper
No Protein + 12 T>A	not shown	MLY778	This Paper
No Protein + Motif Knock Out	not shown	MLY775	This Paper
Wor1 1-321 + Putative Motif	2c	MLY789	This Paper
Wor1 1-321 + No Motif	not shown	MLY791	This Paper
Wor1 1-321 + Motif Knock Out	2c	MLY790	This Paper

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.

Author Signature

6-9-10

Date