Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains

Matthew B. Lohse[a,1], Aaron D. Hernday[a,1,2], Polly M. Fordyce[b,c], Liron Noiman[a,d], Trevor R. Sorrells[a,d], Victor Hanson-Smith[a], Clarissa J. Nobile[a], Joseph L. DeRisi[b,c], and Alexander D. Johnson[a,b,3]

Departments of [a]Microbiology and Immunology and [b]Biochemistry and Biophysics, University of California, San Francisco, CA 94158; [c]Howard Hughes Medical Institute, Chevy Chase, MD 20815; and [d]Tetrad Program, Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158

Edited by Robert T. Sauer, Massachusetts Institute of Technology, Cambridge, MA, and approved March 27, 2013 (received for review December 13, 2012)

Sequence-specific DNA-binding proteins are among the most important classes of gene regulatory proteins, controlling changes in transcription that underlie many aspects of biology. In this work, we identify a transcriptional regulator from the human fungal pathogen *Candida albicans* that binds DNA specifically but has no detectable homology with any previously described DNA- or RNA-binding protein. This protein, named White–Opaque Regulator 3 (Wor3), regulates white–opaque switching, the ability of *C. albicans* to switch between two heritable cell types. We demonstrate that ectopic overexpression of *WOR3* results in mass conversion of white cells to opaque cells and that deletion of *WOR3* affects the stability of opaque cells at physiological temperatures. Genome-wide chromatin immunoprecipitation of Wor3 and gene expression profiling of a *wor3* deletion mutant strain indicate that Wor3 is highly integrated into the previously described circuit regulating white–opaque switching and that it controls a subset of the opaque transcriptional program. We show by biochemical, genetic, and microfluidic experiments that Wor3 binds directly to DNA in a sequence-specific manner, and we identify the set of *cis*-regulatory sequences recognized by Wor3. Bioinformatic analyses indicate that the Wor3 family arose more recently in evolutionary time than most previously described DNA-binding domains; it is restricted to a small number of fungi that include the major fungal pathogens of humans. These observations show that new families of sequence-specific DNA-binding proteins may be restricted to small clades and suggest that current annotations—which rely on deep conservation—underestimate the fraction of genes coding for transcriptional regulators.

transcriptional regulation | transcription factor | transcription networks | epigenetic switch

Regulation of gene expression by sequence-specific DNA-binding proteins underlies many biological processes, from environmental responses in single-celled organisms to the development of multicellular structures in animals and plants. Between 5% and 10% of the coding capacity of most genomes is dedicated to these proteins, and they can be arranged into numerous families and superfamilies based on their amino acid sequences and the structural motifs through which DNA is recognized (1). In this paper, we identify a previously uncharacterized family of sequence-specific DNA-binding proteins that appeared recently in the lineage giving rise to *Candida albicans*, the most common fungal pathogen of humans.

C. albicans is a part of the normal human gut microbiome, but it also may cause disease in humans. In immunocompromised individuals, it may lead to a wide range of medical problems, including disseminated bloodstream infections with mortality rates upward of 40%, as well as superficial mucosal infections such as thrush (2–4). *C. albicans* undergoes a process known as white–opaque switching, in which it switches between two genetically identical but phenotypically distinct cell types termed "white" and "opaque" (5–11). These two states are heritable, with white cells giving rise to white cells and opaque cells giving rise to opaque cells. Switching between these two cell types is rare, occurring approximately once every 10,000 generations, in a seemingly stochastic manner under standard laboratory conditions (12). The white–opaque switch is intimately connected with mating in *C. albicans*, as opaque cells are the mating-competent cell type, whereas white cells do not mate (13). Overall, roughly one-sixth of the *C. albicans* genome is differentially regulated between the two cell types (14–16), resulting in different cell and colony morphologies (9), different interactions with the host immune system (17–20), and different metabolic preferences (14).

Previous work has identified five key transcriptional regulators—Wor1, Wor2, Czf1, Efg1, and Ahr1—that control white–opaque switching in *C. albicans* through a series of nested positive-feedback loops (21–24) (Fig. 1). In this paper, we report a sixth regulator of white–opaque switching in *C. albicans* that was identified based on an examination of transcripts up-regulated in opaque cells compared with white cells and on genome-wide binding data for Wor1, the "master regulator" of white–opaque switching. We describe how this regulator, which we have named Wor3, is integrated into the circuitry defined by the previously identified regulators, and we show that an 84-amino acid region of Wor3 can bind to DNA in a sequence-specific manner. Using a variety of strategies, including a microfluidics-based approach in which Wor3 is presented with all possible 8-mer DNA sequences, we identify the *cis*-regulatory sequence recognized by this DNA-binding domain. Finally, we show by numerous criteria that Wor3 exemplifies a distinct family of DNA-binding proteins.

Results

Identification of Wor3 (Orf19.467). Although five regulators of white–opaque switching have been identified, there is no compelling reason to assume these represent the complete set. To identify additional regulators of white–opaque switching, we reexamined the previously published RNA-seq transcriptional

Author contributions: M.B.L., A.D.H., and A.D.J. designed research; M.B.L., A.D.H., P.M.F., and L.N. performed research; M.B.L., A.D.H., P.M.F., and J.L.D. contributed new reagents/analytic tools; M.B.L., A.D.H., P.M.F., T.R.S., V.H.-S., C.J.N., and A.D.J. analyzed data; and M.B.L., A.D.H., P.M.F., L.N., T.R.S., V.H.-S., C.J.N., J.L.D., and A.D.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The genome-wide datasets reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo [accession nos. GSE42134 (microarray data) and GSE42837 (ChIP-chip data).

[1]M.B.L. and A.D.H. contributed equally to this work.

[2]Present address: Amyris, Emeryville, CA 94608.

[3]To whom correspondence should be addressed. E-mail: ajohnson@cgl.ucsf.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221734110/-/DCSupplemental.

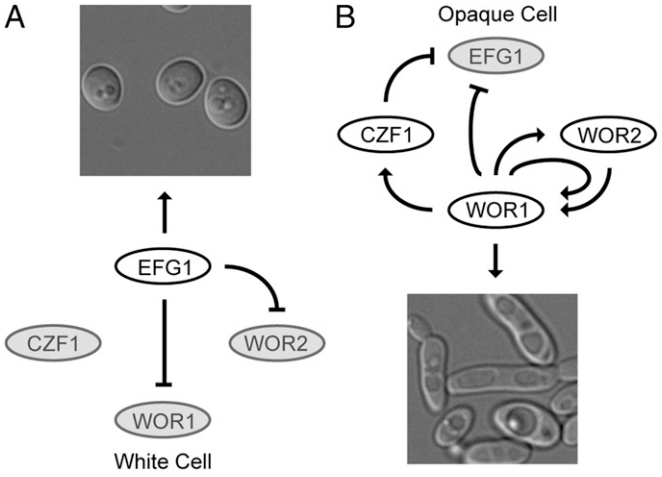

Fig. 1. Working model for the white–opaque regulatory circuit and its activity in white and opaque cell types. (*A*) In white cells, *EFG1* represses *WOR1* directly and indirectly through *WOR2* to maintain white cell identity. (*B*) In opaque cells, *WOR1*, *WOR2*, and *CZF1* establish a series of positive-feedback loops, maintaining opaque cell identity and repressing *EFG1*. Up-regulated genes and active relationships are indicated in black. Down-regulated genes are indicated in gray. Arrows and bars represent activation and repression, respectively. Figure adapted from Zordan et al. (22) and incorporates data from Lassak et al. (47) and Sriram et al. (48).

profiling data of white and opaque cells (16) as well as the genome-wide chromatin immunoprecipitation (ChIP-chip) binding data for Wor1 (22). We also performed microarray-based transcriptional profiling of the two cell types to provide an independent set of data. To identify additional regulators, we searched for genes that were highly up-regulated in opaque cells (in both the RNA-seq and microarray-based profiling datasets) and whose promoters were highly enriched for the presence of Wor1 by ChIP-chip in opaque cells. Only one gene, *ORF19.467*, met both criteria (Dataset S1). This gene is up-regulated 20- to 100-fold in opaque cells and is bound by Wor1 in its relatively long (8-kb) upstream region. When tagged with GFP, Orf19.467 is observed readily in the nucleus in opaque cells but is undetectable in white cells (Fig. 2*A*).

Overexpression of *WOR3* Drives White-to-Opaque Switching. To determine whether *ORF19.467* has a functional role in white–opaque switching, we ectopically overexpressed it in white cells. This resulted in mass conversion of the colonies to the opaque cell type, whereas the control strain remained white (Fig. 2*B* and *SI Appendix*, Table S1). This ectopic expression phenotype is very similar to that previously observed for other switch regulators [e.g., *WOR1* and *CZF1* (21, 22, 24)], and based on this ability we named Orf19.467 "White–Opaque Regulator 3" (Wor3).

Deletion of *WOR3* Affects the Stability of Opaque Cells at Elevated Temperatures. We next examined the effects of deleting *WOR3* on white–opaque switching. Under a standard set of laboratory conditions (25 °C, synthetic dextrose + amino acids + uridine), deletion of *WOR3* had no measurable effects on the white-to-opaque switching rate or on the stability of the opaque state (Fig. 2 *C*–*E* and *SI Appendix*, Table S1). However, at physiological human body temperature (37 °C), the *wor3* deletion strain showed a strong effect on opaque cell stability. Normally, *C. albicans* opaque cells switch *en masse* back to white cells when the temperature is raised from 25 °C to 37 °C in Spider medium supplemented with glucose. However, the *wor3* deletion strain remained opaque at 37 °C under these conditions (*SI Appendix*, Fig. S1). Thus, both *WOR3* overexpression and deletion promote the opaque state. None of the other five white–opaque regulators shows this pattern of behavior, suggesting that Wor3 has a distinctive role in the circuit.

Wor3 Regulates a Subset of the Opaque Cell Transcriptional Program. We examined the effect of *WOR3* overexpression in strains deleted for other members of the circuit. Ectopic expression of *WOR3* could not drive switching to the opaque cell type when *WOR1*, *WOR2*, or *CZF1* was deleted. When *EFG1* was deleted, however, *WOR3* ectopic overexpression strongly promoted switching to the opaque cell type (*SI Appendix*, Table S1). Based on these results, Wor3 appears to be a "modulator" of switching, somewhat reminiscent of Czf1 (22, 24); that is, Wor3 is not absolutely required for switching, but its expression has a large effect on the rate of switching.

We next performed ChIP-chip on a myc-tagged Wor3 protein expressed in opaque cells and found that Wor3 was localized to 87 intergenic regions upstream of 119 genes (Datasets S1 and S2). Wor3 bound to its own upstream region and to the upstream regions of *WOR1*, *WOR2*, *EFG1*, *AHR1*, and *CZF1* (Fig. 3*A*), indicating that it is centrally involved in the white–opaque switch circuit.

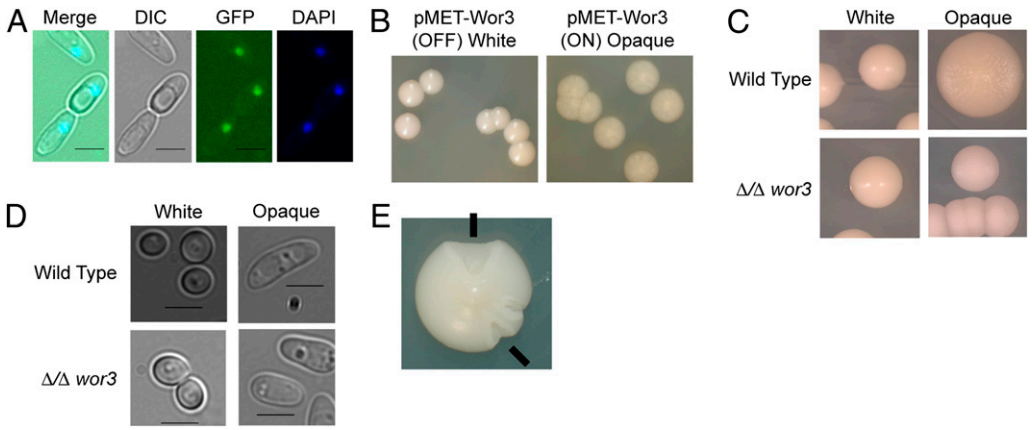

Fig. 2. Ectopic overexpression of *WOR3* drives switching to the opaque cell type. (*A*) Visualization of Wor3–GFP fusion protein in opaque cells. From left to right, merged image, differential interference contrast microscopy (DIC), fluorescence of GFP, staining by DAPI. (*B*) Ectopic overexpression of *WOR3* in white cells at 25 °C results in mass switching to the opaque cell type. (*C*) Wild-type and *wor3* deletion white and opaque colonies. (*D*) Wild-type and *wor3* deletion white and opaque cells. (*E*) Opaque sectors, indicated with lines, are formed by a white colony derived from a *wor3* deletion strain.

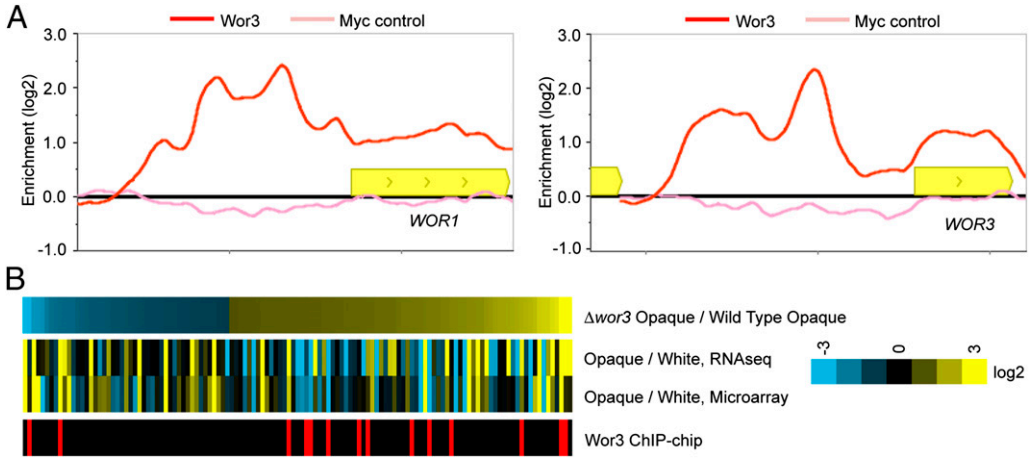

Fig. 3. Chromatin immunoprecipitation and microarray analysis of Wor3. (*A*) Wor3 binds to the upstream regions of *WOR1* (*Left*) and itself (*Right*). ChIP-chip binding data shown for Wor3-myc (red) and myc control (pink); ORFs are represented as yellow boxes. Data were mapped and plotted using MochiView. Binding enrichment ($\log_2$) is plotted on the *y* axis. The full ChIP-chip dataset for Wor3 is described in Dataset S2. (*B*) Transcriptional changes in a *wor3* deletion strain relative to the parent strain (top lane). All genes differentially regulated at least twofold upon deletion of *WOR3* are shown. Opaque or white enrichment of the same genes in a wild-type background (middle lanes). Wor3 binding in vivo as determined by ChIP-chip is indicated in red in the bottom lane. RNA-seq enrichment values are taken from Tuch et al. (16); all other data are from this study.

We next examined the transcriptional changes resulting from the deletion of *WOR3*. Deletion of *WOR3* had minimal transcriptional effects in white cells, exhibiting no changes in transcription greater than twofold. In opaque cells, however, deletion of *WOR3* resulted in 47 genes down-regulated and 78 genes up-regulated at least twofold (Fig. 3*B*). Despite being dispensable for the stability of the opaque cell type under these conditions, Wor3 appears to play a role in the expression of a significant portion of the opaque cell transcriptional program.

Wor3 Is a Sequence-Specific DNA-Binding Protein. The enrichment of Wor3 at specific locations across the genome in the ChIP-chip binding data suggested the possibility that Wor3 binds directly to DNA in a sequence-specific manner. To test this hypothesis, we performed a microfluidics-based DNA experiment based on mechanically induced trapping of molecular interactions (MITOMI 2.0) (25, 26). This technique examines the quantitative binding of an in vitro transcribed and translated protein to a library containing all possible 8-mer DNA sequences (*SI Appendix*, Fig. S2 and Dataset S3). Full-length and two truncated versions of Wor3 exhibited clear sequence-specific DNA binding, with a strong preference for a 5′-ATAACC-3′ sequence (Fig. 4*A* and *SI Appendix*, Figs. S3 and S4). To better characterize the binding of Wor3 to DNA and to examine the effects of flanking sequence, we constructed a Wor3-specific library of oligonucleotides containing systematic substitutions of all possible nucleotides at each position within this target site and directly and quantitatively measured concentration-dependent binding by MITOMI 2.0 (26). These experiments confirm that the core sequence 5′-ATAACC-3′ is critical for Wor3 binding, and the experiment also revealed preferences beyond the core sequence (Fig. 4 *B* and *C*; *SI Appendix*, Fig. S5; and Dataset S4). We further verified that Wor3 specifically recognizes this motif through a series of electrophoretic mobility shift assays (EMSAs) using purified, bacterially

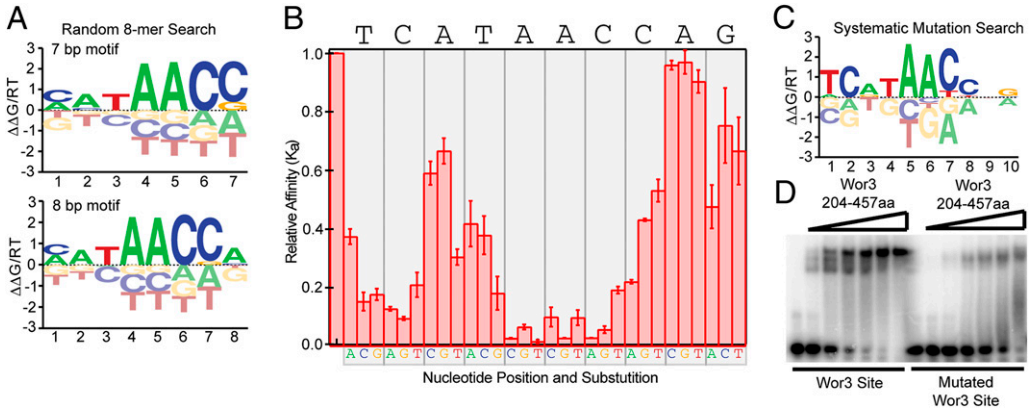

Fig. 4. Wor3 DNA-binding preferences determined via microfluidic affinity analysis (MITOMI 2.0). (*A*) Highest scoring 7- and 8-bp PSAMs from MITOMI 2.0 analysis of a truncated Wor3 construct (amino acids 204–457) binding to a pseudorandom 8-mer library. Each motif is represented as an AffinityLogo, with the relative height of each letter denoting the contribution to overall binding affinity. (*B*) Measured binding affinities (K_a) relative to the "consensus" site affinity (5′-TCATAACCAG) for systematic substitutions of alternate nucleotides at each position. Relative affinities were determined via global fits of measured concentration-dependent binding to a single-site binding model. Values shown are the average of affinities measured in two independent experiments; error bars represent the SEM. (*C*) AffinityLogo representation of the PSAM derived from the relative affinities shown in *B*. (*D*) EMSAs using DNA fragments containing the Wor3 motif or a mutated version of the motif were performed with the Wor3 204–457-aa truncation. From left to right, protein concentrations are 0, 0.5, 1, 2, 4, 8, and 16 nM.

produced Wor3 and DNA sequences containing either this motif or a mutated version of the motif. Wor3 binding to a DNA sequence with the preferred motif occurred with a dissociation constant (K_d) of ~1–2 nM (Fig. 4*D*), consistent with its affinity for DNA being physiologically relevant. Furthermore, we observed that when expressed in *Saccharomyces cerevisiae*, *C. albicans* Wor3 can activate transcription in vivo from a reporter construct that contains its *C. albicans cis*-regulatory sequence (*SI Appendix*, Fig. S6 *A* and *B*).

To directly test the relevance of this motif in *C. albicans* in vivo, we further processed the Wor3 ChIP-chip binding data using MochiView (27) to identify 500-bp regions corresponding to areas of maximum peak enrichment, as previously described (28, 29). We then examined the ability of the MITOMI 2.0-generated Wor3 motif to explain the set of 174 regions of peak enrichment identified by ChIP-chip. Although the Wor3 motif alone did a poor job of predicting this full set of Wor3 binding regions (*SI Appendix*, Fig. S6*C*), there was a strong correlation between Wor3 occupancy and a Wor3 motif plus bound Wor1 (*SI Appendix*, Fig. S6*D*). These results suggest that Wor3 binds cooperatively to DNA with Wor1. Consistent with this idea, the Wor1 and Wor3 ChIP profiles show strong overlap, with 68 of 87 (78%) of the Wor3 intergenic bound regions also bound by Wor1 (*SI Appendix*, Fig. S6*E*).

On the Origins of Wor3. We analyzed DNA binding further by a series of bacterially produced deletion derivatives of Wor3 and identified an 84-aa sequence (amino acids 243–326) that was sufficient for sequence-specific binding to DNA in vitro (*SI Appendix*, Fig. S7). The Wor3 family of proteins is defined by a single conserved region, ~200 amino acids in size, which contains this 84-aa sequence. Perhaps the most striking feature of this region is the presence of 16 conserved cysteine residues, grouped in eight "CxxC" motifs, where x is a variable residue. Three of these eight CxxC motifs are within the 84-aa region sufficient for DNA binding. Clear homologs of Wor3, identifiable by this 200-aa conserved region, appear throughout the CTG clade as well as in *Cyberlindnera jadinii* and *Wickerhamomyces anomalus* (Fig. 5). (The CTG clade includes *C. albicans* as well as species such as *Candida tropicalis*, *Lodderomyces elongisporus*, *Debaryomyces hansenii*, and *Clavispora lusitaniae* and is so named because, in all these species, the CTG codon is translated as serine instead of leucine, as in the conventional genetic code.) We could not identify Wor3 homologs in the *Kluyveromyces lactis* or *S. cerevisiae* clades or in more distantly related species, such as *Yarrowia lipolytica*. The most parsimonious explanation for this arrangement is the emergence of Wor3 in the common ancestor of *C. albicans* and *S. cerevisiae*, after the divergence from *Y. lipolytica*, followed by its loss in the common ancestor of *S. cerevisiae* and *K. lactis*, at a point after the divergence of *C. jadinii* and *W. anomalus* (Fig. 5).

Extensive searches of the known protein databases indicate that the Wor3 family has no detectable homology with any previously studied protein or protein family (*SI Appendix*, Fig. S8 and *SI Materials and Methods*). Searches of the protein data banks (30), using the program HHpred (31), revealed only trivial matches between the *C. albicans* Wor3 sequence and other protein families (*SI Appendix*, Table S2). Although the top search hits found by HHpred were statistically significant—in that their *P* values were less than 1e-4—these matches were based on the shared presence of the amino acid motif CxxC (*SI Appendix*, Table S2). Further searches of protein databases, using randomized sequences containing CxxC, revealed multiple instances of this motif in disparate protein families that have different structures and are generally accepted to be nonhomologous. This strongly suggests that CxxC sequences arose convergently on multiple occasions (Dataset S5). Although Wor3 shares CxxC motifs with other protein families, the presence of the CxxC motif is not sufficient evidence for common ancestry. Taken together, these results indicate that Wor3 represents a distinct protein family, one that either arose de novo or diverged from another protein family to such an extent that vestiges of its ancestry have vanished.

Discussion

White–opaque switching in *C. albicans* is orchestrated by a highly interconnected transcriptional network. In this paper, we identify an additional member of this regulatory network, which we have named Wor3. Its ectopic expression induces the white-to-opaque transition *en masse*, and its deletion affects the stability of the opaque state at physiological temperatures.

In our view, the most significant aspect of this work is the finding that Wor3 represents a distinct family of sequence-specific DNA-binding proteins. From our analysis, we infer that the Wor3 family of transcriptional regulators first appeared ~300 Mya (32), before the divergence of *C. albicans* and *S. cerevisiae*, but after *Y. lipolytica* branched from other Ascomycete species. Two competing hypotheses may explain its origins. According to the first, Wor3 evolved from an existing fungal domain or from a horizontally transferred gene. Traces of such a relationship, however, are not detectable above statistical noise, at least in the current genome sequences. In contrast, other sequence-specific DNA-binding protein families easily may be traced much further back in evolutionary time. The second hypothesis is that Wor3 evolved de novo, perhaps from a previously untranslated DNA sequence (33). This hypothesis is difficult to test rigorously

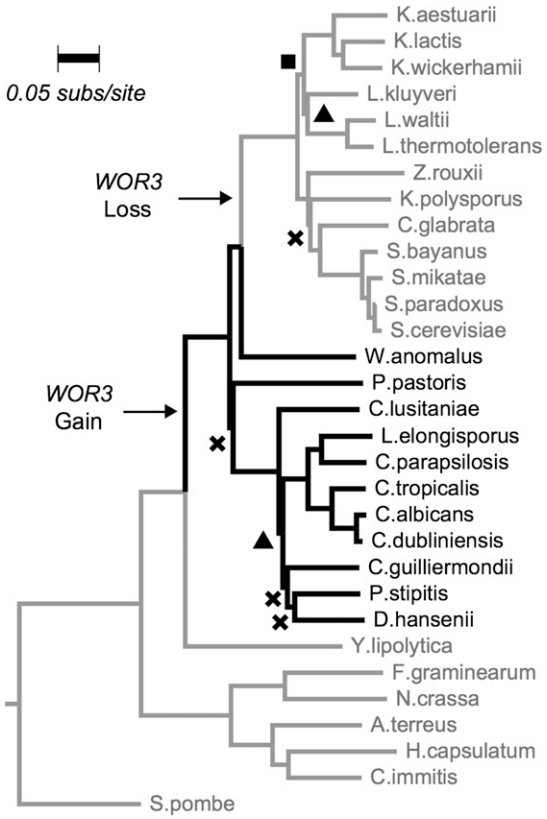

Fig. 5. Phylogenetic tree of 31 fungal species inferred from protein sequences of 79 highly conserved genes. Species containing a Wor3 homolog are indicated in black. Species lacking a Wor3 homolog are gray. The most parsimonious evolutionary explanation for the distribution of *WOR3* is indicated by the "*WOR3* gain" and "*WOR3* loss" labels. Glyphs indicate branch support values as SH-like approximate-likelihood ratios: ×, <0.8; ■, 0.80–0.89; ▲, 0.90–0.99. Branches lacking glyphs have maximum support (= 1.0). Branch lengths express substitutions per site.

because there is no reasonable expectation that noncoding ancestral DNA would be preserved in modern species. Indeed, we did not find evidence of noncoding DNA resembling Wor3 in any genome sequences available at the National Center for Biotechnology Information (34).

Regardless of its evolutionary origins, Wor3 exemplifies a distinctive family of sequence-specific DNA-binding proteins. Although the term "family" is used in many different ways in biology, we use it here, in accordance with the Structural Classification of Proteins definition, to mean a group of proteins with significant amino acid sequence similarity (and, consequently, a strong inference of homology) that does not show significant amino acid similarities to proteins outside the family (1). Does the Wor3 family represent a 3D structure distinct from any of the known structures of sequence-specific DNA binding proteins? Without a Wor3 structure, we cannot answer this question definitively. However, we have analyzed the Wor3 sequences using a wide range of folding and modeling algorithms, and they have not revealed any meaningful matches with known structures. As described above, our analysis also failed to reveal any ancestral relationship between Wor3 and any other protein family. Thus, Wor3 likely exemplifies a distinct protein family that binds DNA through a structure distinct from those previously described for sequence-specific DNA-binding proteins.

Finally, the appearance of a distinctive family of sequence-specific proteins in relatively recent evolutionary history suggests that other evolutionary lineages likely contain newly formed, unannotated transcriptional regulators. We propose that the reliance on deep homology in enumerating and analyzing transcriptional regulators may have inadvertently missed those regulators most relevant to the emergence of new clades.

Materials and Methods

All methods are described briefly below. For additional experimental details, please see *SI Appendix, SI Materials and Methods*.

Growth Conditions. Unless otherwise noted, cells were grown at room temperature (25 °C) in synthetic complete media supplemented with 2% (vol/vol) glucose and 100 μg/mL uridine (SD+aa+uri). To confirm that homogenous cell populations were present before microarrays and ChIP-chip experiments were performed, white and opaque cell populations were assessed by microscopy.

Strain and Plasmid Construction. A list of strains, plasmids, and oligonucleotide sequences used in this study may be found in *SI Appendix*, Tables S3–S5. Details of strain and plasmid construction may be found in *SI Appendix, SI Materials and Methods*.

Switching Assays. Plate-based quantitative white–opaque switching assays and ectopic expression assays using the p*MET3* ectopic expression system were performed as previously described (13, 22).

Temperature and Carbon Source Stability Assay. Strains were grown at room temperature for 7 d on SD+aa+uri plates. White or opaque colonies were streaked onto Spider media plates pretreated with 200 μL of 40% (mass/vol) glucose or water. The Spider plates then were incubated for 2 d at 37 °C; during this time, plates were kept in a cardboard box to minimize drying. After 2 d, individual colonies were restreaked onto SD+aa+uri plates and allowed to grow for 5–7 d, at which point colony morphology was scored. When examining the stability of opaque strains, we sometimes observed isolated white colonies in an otherwise opaque population. We considered a strain "stable" if more than half the colonies that grew up on the restreaked plate were opaque. Two independent Wor3 deletion strains were used for this experiment.

Microarrays. Samples for gene expression microarray analysis were harvested from mid-log phase cultures by centrifugation. Total RNA was extracted using the RiboPure-Yeast RNA kit (Ambion). Reverse transcription and dye coupling to Cy3 and Cy5 dyes were performed as previously described (35). White vs. opaque (AHY135 vs. AHY136) and wild-type vs. mutant (AHY135 vs. AHY207 or AHY136 vs. AHY212) cDNA was competitively hybridized against a mixed reference to custom Agilent 8 × 15K microarrays containing at least two probes per ORF (AMADID #020166). Arrays were scanned using a GenePix 4000B scanner (Axon/Molecular Devices), and the data were extracted using GenePix Pro version 5.1. The Cy3 and Cy5 values were normalized by global Lowess normalization using the Goulphar script (36) for R (The R Foundation for Statistical Computing), and transformations (i.e., white vs. opaque or wild-type vs. deletion strain) were performed before the extraction of median differential expression values for each ORF. Differentially expressed genes were identified using a twofold cutoff. Two biological replicates were performed for each condition. Raw gene expression array data have been deposited in the Gene Expression Omnibus (GEO), www.ncbi.nlm.nih.gov/geo (accession no. GSE42134).

ChIP-chip. Chromatin immunoprecipitation was performed as previously described (37). Briefly, cultures were harvested during mid-log phase by centrifugation and cell pellets were lysed by physical disruption with glass beads. C-terminally myc-tagged Wor3 was immunoprecipitated with a monoclonal anti-myc antibody, and the enriched DNA was amplified, dye coupled, and hybridized against a genomic DNA input control on custom 1 × 244K Agilent tiling microarrays (AMADID 016350). At least two biological replicates were performed for each condition. Array scanning was performed using a GenePix 4000B scanner (Axon/Molecular Devices). The data were extracted and processed as described by Nobile et al. (29), with the following exception: minimum enrichment cutoffs for MochiView peak detection were set to 0.58 for the tagged arrays and 0.27 for the untagged control arrays. Peak sizes were set to 500 bp. The identical peak-detection settings were used to reanalyze the Wor1 ChIP-chip data previously published (22). Called peaks were filtered by subtraction of likely artifactual peaks, based on the fact that these loci showed variable but substantial enrichment in many deletion (control) ChIP-chip experiments that were performed with antibodies against a deleted target (*SI Appendix*, Table S6). The list of bound target genes, with their associated Wor3 enrichment values, was generated by assigning the highest Wor3 ChIP enrichment value from each bound intergenic region to the 5′ intergenic region of each ORF using MochiView. Raw ChIP-chip data have been deposited in GEO, www.ncbi.nlm.nih.gov/geo (accession no. GSE42837).

MITOMI 2.0 Random Library Experiments. MITOMI 2.0 experiments for de novo identification of transcription factor-binding sites using a pseudorandom library of DNA sequences were performed as described previously (25), with the following modifications. First, we used an improved pseudorandom 8-mer DNA library based on a previously published algorithm (38) that included all possible 8-mer DNA sequences within 740 oligonucleotides (Dataset S3). Second, we designed a smaller version of the microfluidic devices with 1,568 chambers arrayed in 28 channels with 56 chambers per channel. Third, we made several changes to the protocol for both mold and device fabrication (*SI Appendix, SI Materials and Methods*). Finally, we printed two arrays per 2 × 3-inch SuperChip epoxysilane glass slide (Thermo Fisher Scientific). Raw data from the three MITOMI 2.0 experiments are presented in *SI Appendix*, Fig. S4. Position-specific affinity matrices (PSAMs) for selected versions of the Wor3 motif from different MITOMI 2.0 experiments are included in Dataset S6.

MITOMI 2.0 Binding-Curve Experiments. Experiments assessing concentration-dependent binding to oligonucleotides containing systematic mutations of candidate "consensus" transcription factor target sites were performed largely as described previously (39), with final concentrations of DNA (before printing) set to be 10 μM, 6.7 μM, 4 μM, 3 μM, 2 μM, 1.3 μM, 0.9 μM, and 0.4 μM. Oligonucleotide sequences used for binding curves are listed in *SI Appendix*, Table S5. Single-site binding model fits shown are from globally fitting all binding curves simultaneously. Binding curves for both repeats of this experiment are included in Dataset S4. The PSAM for the Wor3 motif shown in Fig. 4*C* is included in Dataset S6.

Protein Purification and EMSAs. Purification of 6-His–tagged Wor3 truncation constructs was performed using a previously reported protocol (28). Protein expression was conducted in the BL21 background, and induction was with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 4 h at 25 °C. In brief, bacterial pellets were lysed and protein purified using Ni-NTA agarose (Qiagen). Protein concentrations were determined by comparison with known amounts of BSA on a Coomassie blue-stained SDS/PAGE gel.

EMSAs were performed as previously described (40) using 21-bp oligonucleotides containing the Wor3 motif or a mutated version of the motif. The K_d determination buffer conditions [no poly(deoxyinosinic-deoxycytidylic) acid, 50 mM NaCl] were used.

Intergenic Region Overlap Comparison. Wor1 and Wor3 binding sets were compared using MochiView v1.45 (27). Full details of this process may be found in *SI Appendix, SI Materials and Methods*.

Motif Comparisons. The ability of the Wor3 motif to explain binding sites relative to the genome as a whole was determined using previously reported methods (28, 29, 40). Full details of this process may be found in *SI Appendix, SI Materials and Methods*.

Fungal Phylogeny Development. To construct the phylogenetic tree of 31 yeast species, we chose 79 orthologs present in a single copy in each species according to two previous ortholog maps (41, 42). Sequences of *W. anomalus* were obtained from the US Department of Energy Joint Genome Institute, www.jgi.doe.gov (43). Protein sequences from each species were concatenated and aligned using Clustal (44), and the tree was constructed using PhyML with the BLOSUM62 model (45). Supports for branches were estimated using the Shimodaira–Hasegawa-like (SH-like) approximate likelihood ratio test (46) as implemented in PhyML. Full details on searches for Wor3 homologs may be found in *SI Appendix, SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Oliver Homann for developing MochiView; and Sudarsi Desta, Jeanselle Dea, and Jorge Mendoza for technical assistance. We also thank Michael Winter and Michael Chimenti for advice on protein chemistry and structural analysis of Wor3; Christopher Baker for ideas and suggestions for the manuscript; and Jessica Walter, Wendell Lim, Oren Rosenberg, and Jeff Cox for plasmids. This study was supported by National Institutes of Health Grants R01AI049187 (to A.D.J.) and F32AI071433 (to A.D.H.) and the Howard Hughes Medical Institute (J.L.D. and P.M.F.).

1. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247(4):536–540.
2. Eggimann P, Garbino J, Pittet D (2003) Epidemiology of Candida species infections in critically ill non-immunosuppressed patients. *Lancet Infect Dis* 3(11):685–702.
3. Gudlaugsson O, et al. (2003) Attributable mortality of nosocomial candidemia, revisited. *Clin Infect Dis* 37(9):1172–1177.
4. Wey SB, Mori M, Pfaller MA, Woolson RF, Wenzel RP (1988) Hospital-acquired candidemia. The attributable mortality and excess length of stay. *Arch Intern Med* 148(12):2642–2645.
5. Bennett RJ, Uhl MA, Miller MG, Johnson AD (2003) Identification and characterization of a *Candida albicans* mating pheromone. *Mol Cell Biol* 23(22):8189–8201.
6. Johnson A (2003) The biology of mating in *Candida albicans*. *Nat Rev Microbiol* 1(2): 106–116.
7. Lohse MB, Johnson AD (2009) White-opaque switching in *Candida albicans*. *Curr Opin Microbiol* 12(6):650–654.
8. Morschhäuser J (2010) Regulation of white-opaque switching in Candida albicans. *Med Microbiol Immunol (Berl)* 199(3):165–172.
9. Slutsky B, et al. (1987) "White-opaque transition": A second high-frequency switching system in *Candida albicans*. *J Bacteriol* 169(1):189–197.
10. Soll DR (2009) Why does *Candida albicans* switch? *FEMS Yeast Res* 9(7):973–989.
11. Soll DR, Morrow B, Srikantha T (1993) High-frequency phenotypic switching in *Candida albicans*. *Trends Genet* 9(2):61–65.
12. Rikkerink EH, Magee BB, Magee PT (1988) Opaque-white phenotype transition: A programmed morphological transition in *Candida albicans*. *J Bacteriol* 170(2): 895–899.
13. 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(3):293–302.
14. Lan CY, et al. (2002) Metabolic specialization associated with phenotypic switching in *Candidaalbicans*. *Proc Natl Acad Sci USA* 99(23):14907–14912.
15. Tsong AE, Miller MG, Raisner RM, Johnson AD (2003) Evolution of a combinatorial transcriptional circuit: A case study in yeasts. *Cell* 115(4):389–399.
16. Tuch BB, et al. (2010) The transcriptomes of two heritable cell types illuminate the circuit governing their differentiation. *PLoS Genet* 6(8):e1001070.
17. 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(2):667–677.
18. 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(12):6652–6662.
19. 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(11):4468–4475.
20. Lohse MB, Johnson AD (2008) Differential phagocytosis of white versus opaque *Candida albicans* by *Drosophila* and mouse phagocytes. *PLoS One* 3(1):e1473.
21. Zordan RE, Galgoczy DJ, Johnson AD (2006) Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. *Proc Natl Acad Sci USA* 103(34):12807–12812.
22. 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(10):e256.
23. Wang H, et al. (2011) Candida albicans Zcf37, a zinc finger protein, is required for stabilization of the white state. *FEBS Lett* 585(5):797–802.
24. Vinces MD, Kumamoto CA (2007) The morphogenetic regulator Czf1p is a DNA-binding protein that regulates white opaque switching in *Candida albicans*. *Microbiology* 153(Pt 9):2877–2884.
25. Fordyce PM, et al. (2010) De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nat Biotechnol* 28(9):970–975.
26. Maerkl SJ, Quake SR (2007) A systems approach to measuring the binding energy landscapes of transcription factors. *Science* 315(5809):233–237.
27. Homann OR, Johnson AD (2010) MochiView: Versatile software for genome browsing and DNA motif analysis. *BMC Biol* 8:49.
28. Cain CW, Lohse MB, Homann OR, Sil A, Johnson AD (2012) A conserved transcriptional regulator governs fungal morphology in widely diverged species. *Genetics* 190(2): 511–521.
29. Nobile CJ, et al. (2012) A recently evolved transcriptional network controls biofilm development in Candida albicans. *Cell* 148(1-2):126–138.
30. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242.
31. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server issue): W244–W248.
32. Taylor JW, Berbee ML (2006) Dating divergences in the Fungal Tree of Life: Review and new analyses. *Mycologia* 98(6):838–849.
33. Carvunis AR, et al. (2012) Proto-genes and de novo gene birth. *Nature* 487(7407): 370–374.
34. Cummings L, et al. (2002) Genomic BLAST: Custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiol Lett* 216(2):133–138.
35. Homann OR, Dea J, Noble SM, Johnson AD (2009) A phenotypic profile of the Candida albicans regulatory network. *PLoS Genet* 5(12):e1000783.
36. Lemoine S, Combes F, Servant N, Le Crom S (2006) Goulphar: Rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics* 7:467.
37. Hernday AD, Noble SM, Mitrovich QM, Johnson AD (2010) Genetics and molecular biology in Candida albicans. *Methods Enzymol* 470:737–758.
38. Mintseris J, Eisen MB (2006) Design of a combinatorial DNA microarray for protein-DNA interaction studies. *BMC Bioinformatics* 7:429.
39. Fordyce PM, et al. (2012) Basic leucine zipper transcription factor Hac1 binds DNA in two distinct modes as revealed by microfluidic analyses. *Proc Natl Acad Sci USA* 109(45):E3084–E3093.
40. Lohse MB, Zordan RE, Cain CW, Johnson AD (2010) Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in Candida albicans. *Proc Natl Acad Sci USA* 107(32):14105–14110.
41. Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD (2008) The evolution of combinatorial gene regulation in fungi. *PLoS Biol* 6(2):e38.
42. Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449(7158):54–61.
43. Schneider J, et al. (2012) Genome sequence of Wickerhamomyces anomalus DSM 6766 reveals genetic basis of biotechnologically important antimicrobial activities. *FEMS Yeast Res* 12(3):382–386.
44. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680.
45. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
46. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* 60(5):685–699.
47. Lassak T, et al. (2011) Target specificity of the Candida albicans Efg1 regulator. *Mol Microbiol* 82(3):602–618.
48. Sriram K, Soliman S, Fages F (2009) Dynamics of the interlocked positive feedback loops explaining the robust epigenetic switching in *Candida albicans*. *J Theor Biol* 258(1):71–88.