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

Supplemental Materials Table of Contents

Supplemental Figures

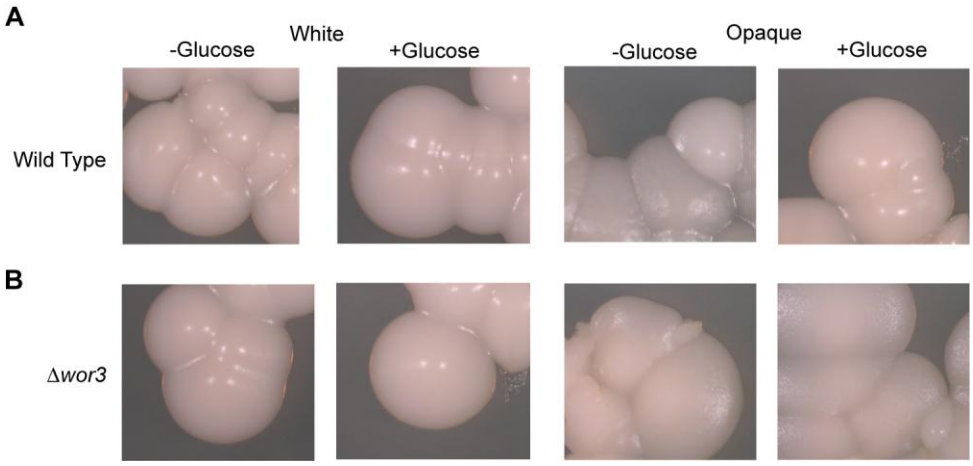
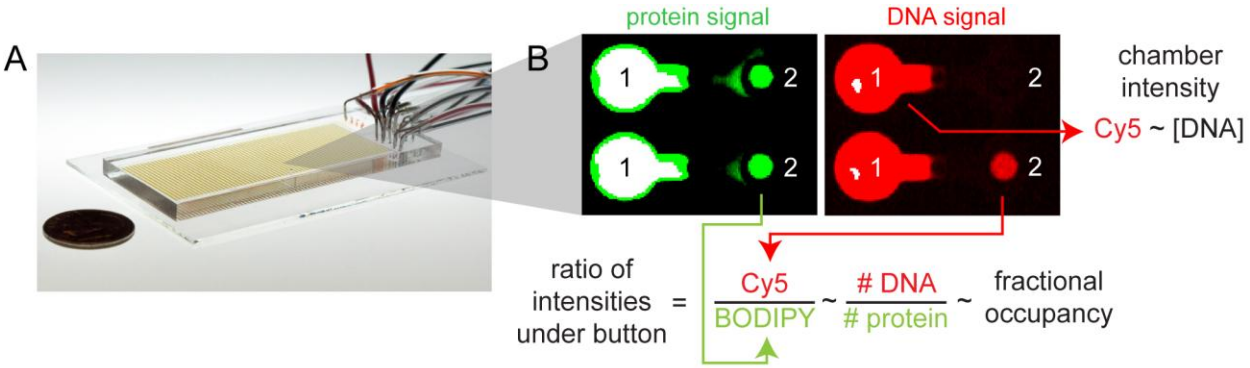
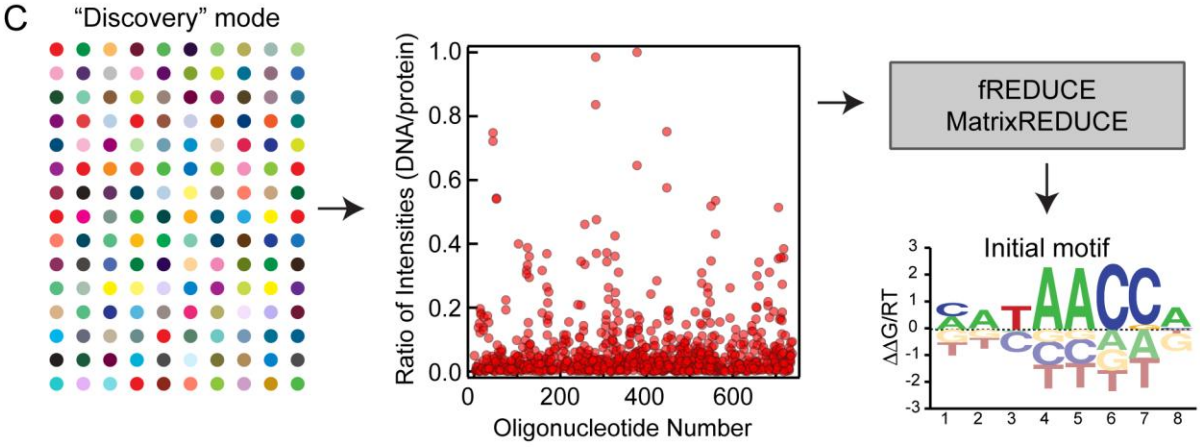


Figure S1: Deletion of Wor3 results in increased opaque cell stability at 37°C in the presence of glucose. (A) Wild-type white and opaque colonies were grown for two days on Spider media with or without glucose at 37°C. Colonies were then re-streaked from the spider plates and grown for one week at room temperature on SD+aa+Uri plates to determine colony phenotypes; typical examples are shown for each permutation. (B) As in panel A, but with *wor3* deletion strains. Typical examples of colony morphologies for the re-streaked colonies of the *wor3* deletion strains are shown for each permutation.


A
B
protein signal
DNA signal
1
2
chamber intensity
Cy5 ~ [DNA]
ratio of intensities under button
=
Cy5 / BODIPY
~
DNA / # protein
~
fractional occupancy


C
"Discovery" mode
Ratio of Intensities (DNA/protein)
Oligonucleotide Number
fREDUCE
MatrixREDUCE
Initial motif
ΔΔG/RT


D
Overlapping 8mers
Oligo #1 CGCGCAATTGCAATGCAAAAAACGTAGCGCAGGTGTGAGTACCCTCGTACGGTTACTCCGGCGGTATGAC
Oligo #2 CGCACGGTTAAGACTAGCGCCAGCATAGACCGGAGGCGTAAACACTGCATCCGCCCTCCGGCGGTATGAC
Oligo #3 CGCATCCGCCCGCCTTCAATGCGGCCTGAACGAGGGGCCAGATGTGGGTGGTCGTCTCCGGCGGTATGAC

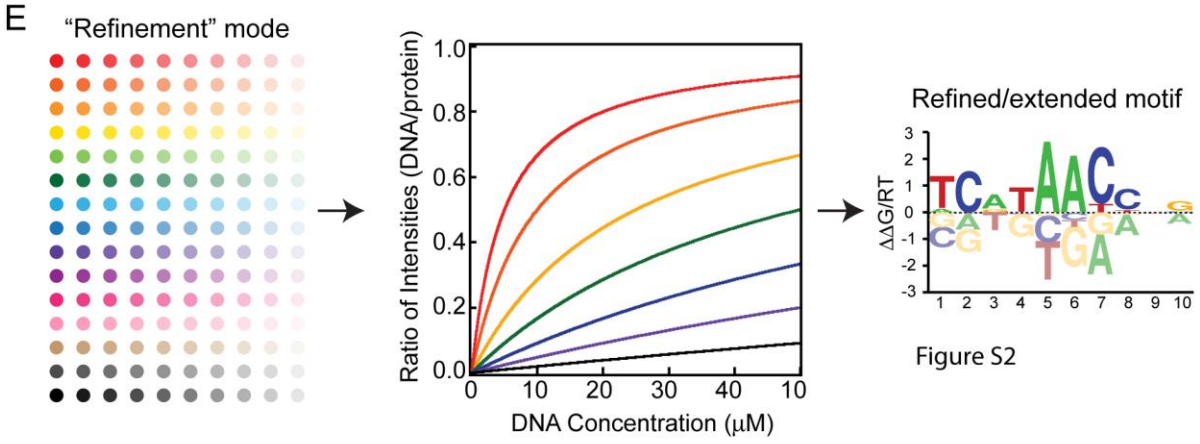

E
"Refinement" mode
Ratio of Intensities (DNA/protein)
DNA Concentration (μM)
Refined/extended motif
ΔΔG/RT

Figure S2


F
Oligo #1 CGCAATTGCGAGTACACTCATGCATGGATCGACCTTCCTCTCCGGCGGTATGAC
Oligo #2 CGCAATTGCGAGTACACACATGCATGGATCGACCTTCCTCTCCGGCGGTATGAC
Oligo #3 CGCAATTGCGAGTACACCCATGCATGGATCGACCTTCCTCTCCGGCGGTATGAC
Oligo #4 CGCAATTGCGAGTACACGCATGCATGGATCGACCTTCCTCTCCGGCGGTATGAC

Figure S2: MITOMI 2.0 experimental setup and analysis pathways, illustrated with simulated data. (A) Photograph of microfluidic device used in MITOMI 2.0 experiments. Each device contains 1,568 individual unit cells that can be used to measure binding affinities for a single transcription factor interacting with multiple DNA oligonucleotides. Unit cells are programmed with individual DNA sequences via alignment to a spotted DNA microarray. (B) Fluorescence images showing recorded intensities from labeled protein molecules (left) and DNA molecules (right) in two individual unit cells, each composed of two chambers. DNA signal intensity in chamber #1 reflects the soluble oligonucleotide concentration. The ratio of the DNA signal to the protein signal in chamber #2 reflects the number of DNA molecules bound by each protein molecule, providing a measurement of binding affinity. (C) Experimental procedure for *de novo* discovery of transcription factor binding sites. A library of oligonucleotides containing all possible 8mer DNA sequences is spotted at a single concentration. Intensity ratios for all oligonucleotides are processed using fREDUCE and MatrixREDUCE algorithms to identify the target site responsible for producing a given pattern of binding. (D) Example sequences from 8mer oligonucleotide library used for target site discovery. Each oligonucleotide sequence is composed of a constant 5' segment (grey), a variable segment (black) containing multiple overlapping 8mer candidate target sites (blue lines), and a constant 3' segment (grey and red). (E) Experimental procedure for refinement and extension of candidate target sites. Each oligonucleotide from a transcription factor-specific library is printed at multiple concentrations to allow measurement of concentration-dependent binding and determination of binding affinities via global fits to a single-site binding model. (F) Example oligonucleotide sequences used for target site

refinement and extension. Each oligonucleotide contains either the candidate target site (orange) or systematic mutations at each position within the candidate target site (green) embedded within constant flanking sequences.

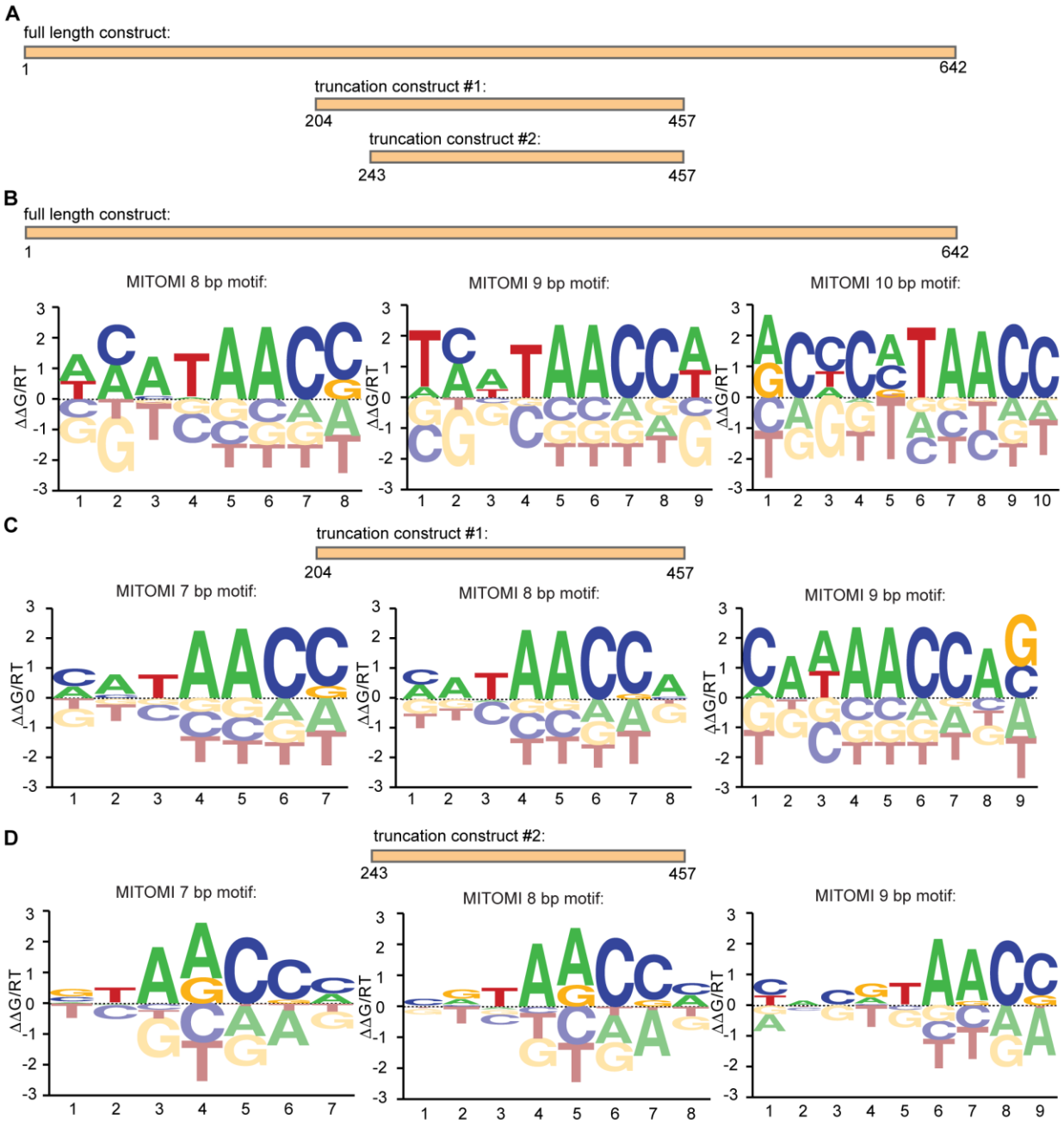


Figure S3: Preferred Wor3 target motifs identified from three MITOMI 2.0 experiments assessing binding of different C-terminally 6xHis-tagged Wor3 protein constructs to a pseudorandom 8mer oligonucleotide library. (A) Diagram showing three different Wor3 constructs used in experiments. (B) AffinityLogo representations of the top-scoring 8-, 9-

and 10-bp PSAMs from MITOMI 2.0 experiments measuring binding of full-length Wor3. (C) Top-scoring MITOMI 2.0 PSAMs for a Wor3 truncation construct spanning amino acids 204-457. (D) Top-scoring MITOMI 2.0 PSAMs for a Wor3 truncation construct spanning amino acids 243-457.

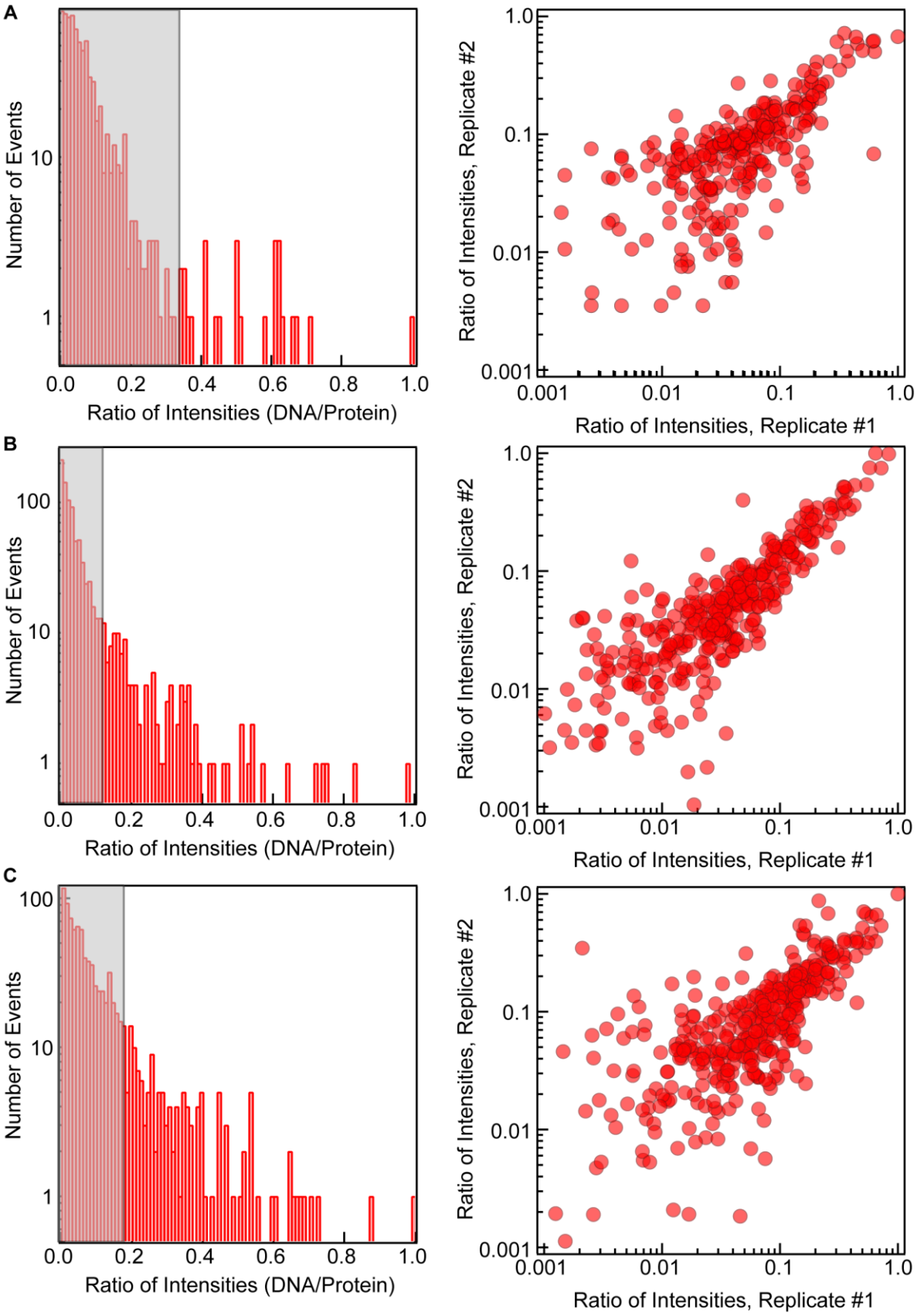

A
Number of Events
10
1
0.0
0.2
0.4
0.6
0.8
1.0
Ratio of Intensities (DNA/Protein)
1.0
0.1
0.01
0.001
Ratio of Intensities, Replicate #2
0.001
0.01
0.1
1.0
Ratio of Intensities, Replicate #1
B
100
Number of Events
C

Figure S4: Raw data from three MITOMI 2.0 experiments assessing binding of different Wor3 protein constructs to a pseudorandom 8mer oligonucleotide library. (A) Data for full-length C-terminally 6xHis-tagged Wor3. Left: distribution of measured fluorescence intensity ratios (DNA/Protein); grey bar shows four standard deviations from the mean as determined by a Gaussian fit to the distribution centered on zero. Right: Scatter plot showing measured intensity ratios for two printed replicates of each oligonucleotide (r^2 = 0.34). (B) Same data as in (A) for a truncated Wor3 construct with a C-terminal 6xHis tag (amino acids 204-457; r^2 = 0.90). (C) Same data as in (A) for a second truncated Wor3 construct with a C-terminal 6xHis tag (amino acids 243-457; r^2 = 0.71).

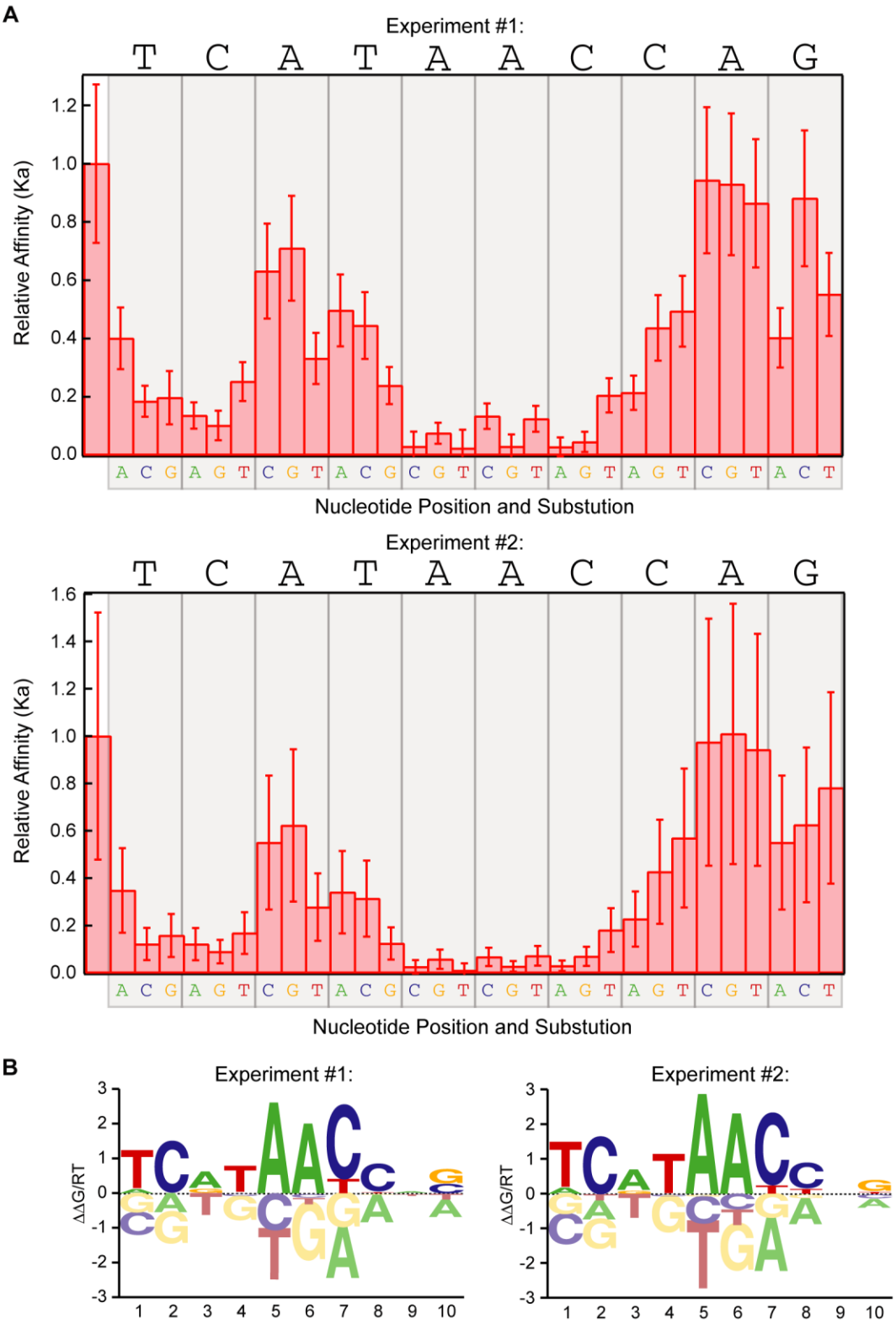


Figure S5: Data from two experimental replicates measuring Wor3 binding to an oligonucleotide library containing either a “consensus” Wor3 target site or single

nucleotide mutations of the Wor3 site. Experiments use the Wor3 204-457aa truncation. (A) Measured binding affinities (K_a) relative to the “consensus” site affinity for systematic nucleotide substitutions at each position within the motif for two experimental replicates. Affinity values and error bars are derived from global fits of binding curves to a single-site binding model. (B) AffinityLogo representations of PSAMs calculated from relative binding affinities shown in (A).

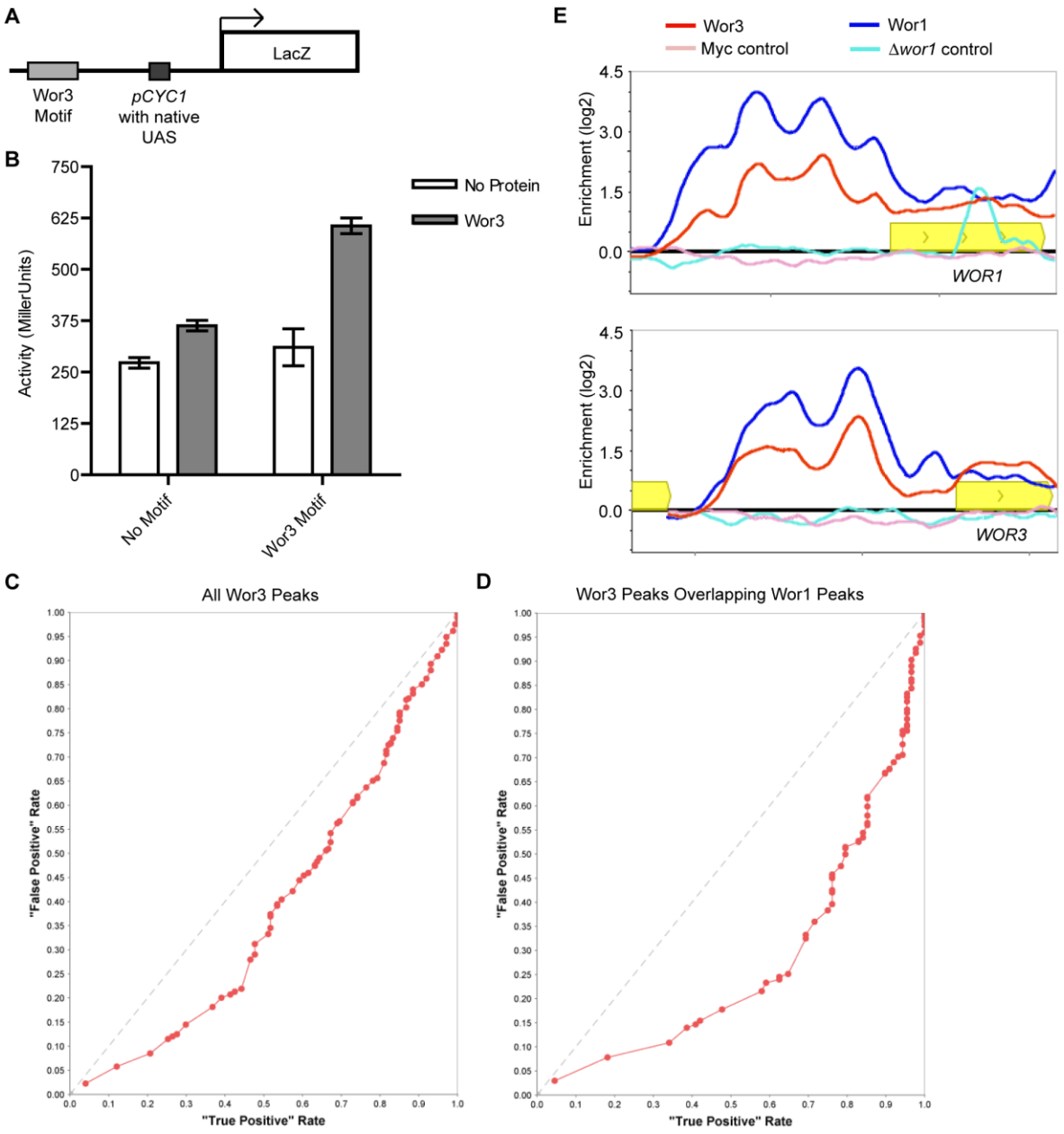


Figure S6: Functional relevance of the Wor3 DNA motif *in vivo*, as determined by transcriptional activation assays and ChIP-chip. (A) Activation assay setup; the Wor3 motif is introduced to a version of the *CYC1* promoter that still contains the native UAS. This setup tests whether the presence of Wor3 and the Wor3 motif have an impact on existing transcription levels. (B) Wor3 enhances transcription of a UAS-containing *CYC1* promoter in a sequence specific manner. Error bars represent standard error of the

mean; experiments were performed with 12 samples for each strain on the same day. (C) ROC Enrichment plot for the Wor3 motif at all Wor3 binding sites; the fraction of the experimental set (174 Wor3 binding sites) with a given motif score is plotted on the x-axis ("True Positive"). The fraction of a control set (1,506 500bp regions randomly selected from intergenic regions not bound by Wor3 or Wor1) with the same motif score is plotted on the y-axis ("False Positive"). (D) ROC Enrichment plot for the Wor3 motif; the experimental set in this panel consists of locations where both Wor3 and Wor1 are bound (88 binding sites). The control for this panel is a set of 1,318 602bp regions randomly selected from intergenic regions not bound by Wor3 or Wor1, the 602bp size reflects the mean size of the experimental set. Plots were made in MochiView using an approach similar to that previously reported (1, 2). (E) Wor3 binding enrichment shows a similar profile to that seen for Wor1; upstream regions of *WOR1* (top) and *WOR3* (bottom) are shown. ChIP-chip binding data shown for Wor3-myc (red), untagged control for Wor3 (pink), Wor1 (blue) and a *wor1* deletion mutant control (light blue). Open reading frames are represented as yellow boxes. Binding enrichment (log_2) is plotted on the y-axis. Data were mapped and plotted using MochiView. Wor1 ChIP-chip data from Zordan et al., 2007 (3).

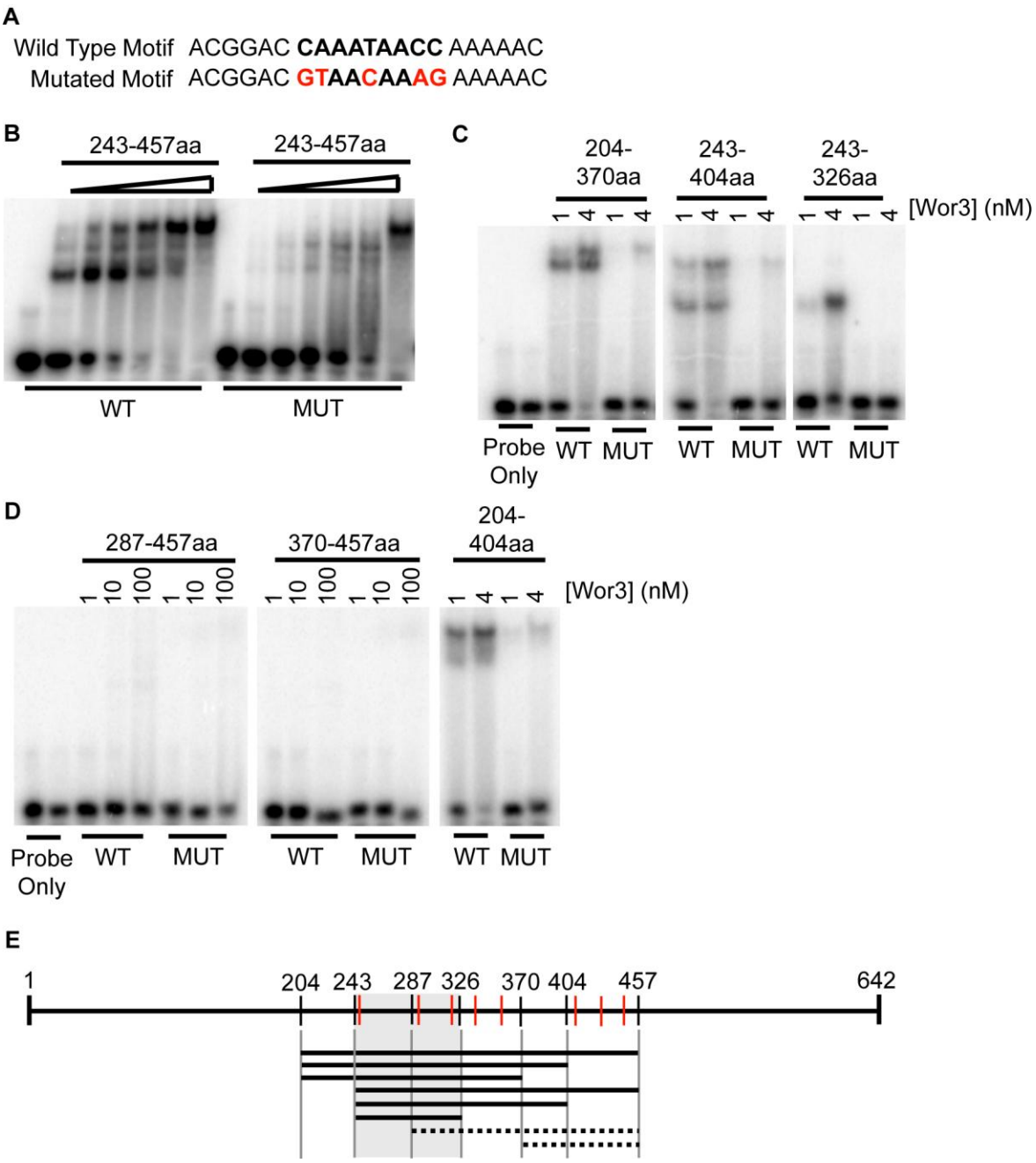


Figure S7: Identification of an 84 amino acid Wor3 sequence sufficient for sequence-specific DNA binding. (A) Wild Type and Mutated oligonucleotide sequences used in electrophoretic mobility shift assays (EMSAs). Wor3 motif portion of sequences bolded, mutations are colored in red. (B) EMSAs using DNA fragments containing the Wor3 motif (WT) or a mutated version of the motif (MUT) were performed with the Wor3 243-

457aa truncation. From left to right, protein concentrations are 0, 0.5, 1, 2, 4, 8, and 32nM. (C) EMSAs performed with the Wor3 204-370aa, 243-404aa, and 243-326aa truncations. DNA fragments containing the Wor3 motif (WT) or a mutated version of the motif (MUT) were used. Protein concentrations indicated above images. All portions of the panel are taken from the same gel. (D) EMSAs performed with the Wor3 287-457aa, 370-457aa, and 204-404aa truncations. DNA fragments containing the Wor3 motif (WT) or a mutated version of the motif (MUT) were used. Protein concentrations indicated above images. All portions of the panel are taken from the same gel. (E) Summary of Wor3 truncation EMSA results. The top line represents full length Wor3, drawn to scale, with truncation locations (black) and instances of the “CxxC” motif (red) indicated with vertical lines. Lower lines represent different Wor3 truncations, Wor3 truncations that bound DNA are indicated with solid lines, ones that did not bind DNA are indicated with dashed lines. The 84aa region sufficient for binding to DNA (243-326aa) is indicated with a grey box.

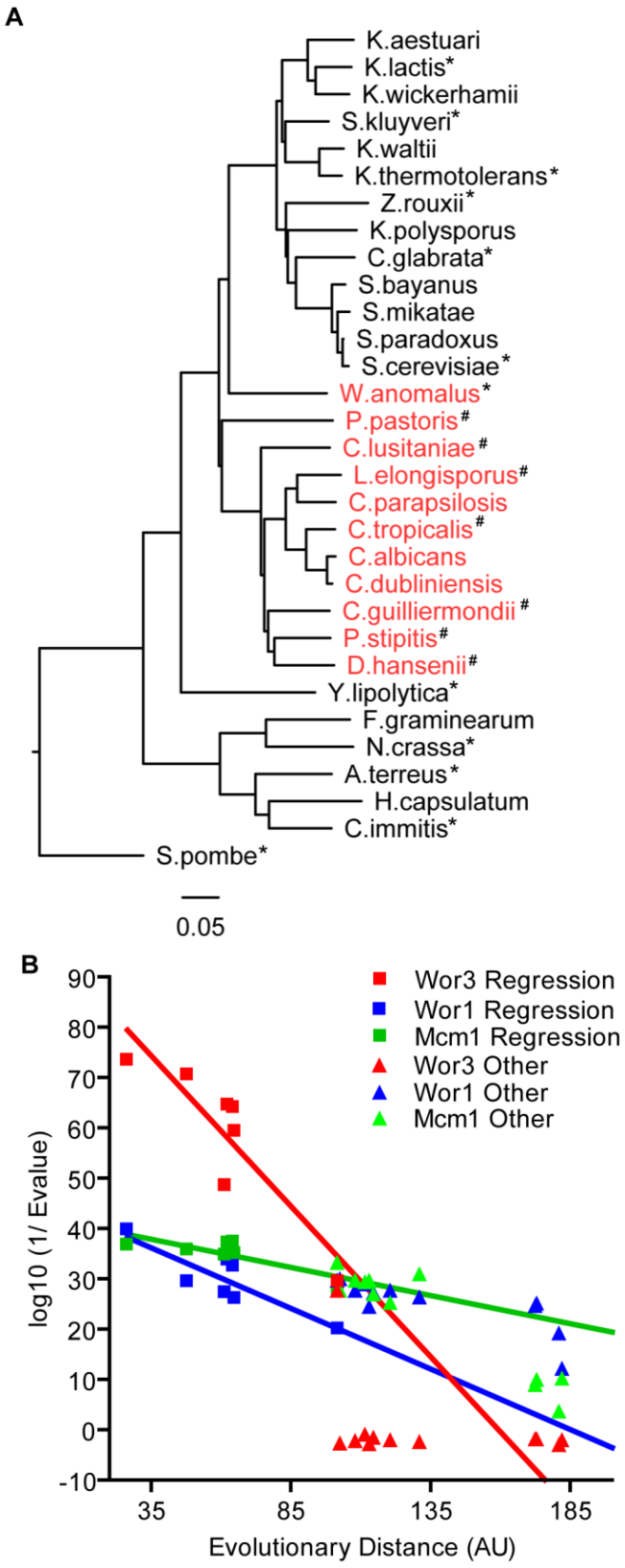

A
K.aestuari
K.lactis*
K.wickerhamii
S.kluyveri*
K.waltii
K.thermotolerans*
Z.rouxii*
K.polysporus
C.glabrata*
S.bayanus
S.mikatae
S.paradoxus
S.cerevisiae*
W.anomalus*
P.pastoris#
C.lusitaniae#
L.elongisporus#
C.parapsilosis
C.tropicalis#
C.albicans
C.dubliniensis
C.guilliermondii#
P.stipitis#
D.hansenii#
Y.lipolytica*
F.graminearum
N.crassa*
A.terreus*
H.capsulatum
C.immitis*
S.pombe*
0.05
B
Wor3 Regression
Wor1 Regression
Mcm1 Regression
Wor3 Other
Wor1 Other
Mcm1 Other
log10 (1/ Evalue)
90
80
70
60
50
40
30
20
10
0
-10
35
85
135
185
Evolutionary Distance (AU)

Figure S8: Comparison of observed and expected divergence in the Wor3 sequence as a function of increasing distance from *C. albicans*. (A) Phylogenetic tree of 31 fungal species, adapted from Figure 5. The scale bar represents the number of amino acid substitutions per site for this set of 79 conserved genes. Species containing a Wor3 homolog are indicated in red, 7 species used to develop the linear regressions in panel B are indicated with a "#" symbol, and 12 other species used in panel B are indicated with a "*" symbol. (B) Comparison of observed and expected divergence in the Wor3, Wor1, and Mcm1 sequences as a function of increasing distance from *C. albicans*. Log10 transformed inverse E-values of the highest scoring BLAST hit to the conserved regions of *C. albicans* Wor3 (red), Wor1 (blue), and Mcm1 (green) are plotted on the y-axis; evolutionary distance (arbitrary units) from *C. albicans* based on the phylogeny in panel A is plotted on the x-axis. Values for best hits from *C. tropicalis* to *P. pastoris* are represented with squares; values for more distant species are represented with triangles. The linear regressions shown (Wor3 R^2=0.80 (red), Wor1 R^2=0.75 (blue), Mcm1 R^2=0.62 (green)) are based on the Wor3, Wor1, and Mcm1 homologs from *C. tropicalis* to *P. pastoris*, respectively.

Supplemental Tables

Table S1: White-opaque switch frequencies for ectopic overexpression of Wor3 or Czf1 in Wild-type and various deletion backgrounds on both repressing ("OFF>OFF") and inducing ("OFF>ON") plates. White-to-opaque and opaque-to-white switching frequencies for wild-type and *wor3* deletion strains under normal laboratory conditions are also shown.

		OFF→OFF		OFF→ON	
Strain	Ectopic Expression Construct	Switching Frequency	n	Switching Frequency	n
Wild Type	Blank	<0.51%	197	<0.35%	283
Wild Type	*WOR3*	<0.43%	231	99.84%	621
Δ/Δ*wor1*	Blank	<0.45%	222	<0.32%	308
Δ/Δ*wor1*	*WOR3*	<0.69%	145	<0.16%	608
Δ/Δ*wor2*	Blank	<0.47%	215	<0.29%	339
Δ/Δ*wor2*	*WOR3*	<0.40%	252	<0.13%	762
Δ/Δ*czf1*	Blank	<0.77%	130	<0.50%	201
Δ/Δ*czf1*	*WOR3*	<1.19%	84	<0.25%	397
Δ/Δ*efg1*	Blank	16.86%	334	40.26%	668
Δ/Δ*efg1*	*WOR3*	13.21%	280	100%	854
Wild Type	*CZF1*	0.51%	198	98.31%	593
Δ/Δ*wor3*	*CZF1*	<0.59%	169	1.36%	514
		White→Opaque		Opaque→White	
Strain		Switching Frequency	n	Switching Frequency	n
Wild Type		2.79%	1289	2.06%	678
Δ/Δ*wor3*		1.30%	1000	1.71%	933

Table S2: HHPred search results for Wor3, using five sets of search criteria. The first search used the full *C. albicans* Wor3 sequence as the search seed and default settings for HHPred, The second search was a refinement of the first search to the restricted region of amino acids 296 to 434, as suggested by HHPred, otherwise using default settings for HHPred. The third search was based on a hidden Markov Model (hMM) generated for the family of Wor3 sequences. The fourth search was based on an hMM using a restricted sequence region as suggested by HHPred. The final search used the 84 amino acid sequence (243-326aa) sufficient for binding to DNA *in vitro.*

Search 1: Full length C. albicans Wor3 sequence, default settings for HHPred

PDB Code	Match Description	Probability	E-Value	P-Value
1zbd_B	Rabphilin-3A; G protein, effector	73.6	1.5	5.10E-05
3gox_A	Restriction endonuclease HPY99I; endonuclease-DNA complex	73.5	1.1	3.80E-05
1tjl_A	DNAK suppressor protein; DKSA, transcription factor	69.8	1.5	5.00E-05
2zet_C	Melanophilin; complex, GTP-binding protein, GTPase, G-protein	69.1	2	6.90E-05
2kgo_A	Uncharacterized protein YBII; Zn finger	62.2	1.1	3.80E-05
3na7_A	HP0958; flagellar biogenesis, flagellum export, C4 Zn-ribbon	61.9	1.7	5.80E-05
2qgp_A	HNH endonuclease	58.6	4.4	1.50E-04

Search 2: C. albicans Wor3 296-434aa sequence, default settings for HHPred

PDB Code	Match Description	Probability	E-Value	P-Value
2jrp_A	Putative cytoplasmic protein; two-zinc binding protein, structural genomics, PSI-2	95.1	0.0057	2.00E-07
2jne_A	Hypothetical protein YFGJ; zinc fingers, two zinc	92.6	0.021	7.20E-07
1twf_L	ABC10-alpha, DNA-directed RNA polymerases I, II, and III	88.4	0.058	2.00E-06
1tjl_A	DNAK suppressor protein; DKSA, transcription factor	88.3	0.14	4.70E-06

3h0g_L	DNA-directed RNA polymerases I, II, and III subunit rpabc4	87.8	0.19	6.70E-06
2kq9_A	DNAK suppressor protein; zinc binding protein	87.3	0.1	3.60E-06
2jrp_A	Putative cytoplasmic protein; two-zinc binding protein	85.5	0.15	5.40E-06
3a43_A	HYPD, hydrogenase nickel incorporation protein HYPA	85.4	0.11	3.80E-06
2pk7_A	Uncharacterized protein; NESG, PLR1, putative tetraacyldisaccharide-1-P 4-kinase	85	0.39	1.40E-05
1pft_A	TFIIB, PFTFIIBN; N-terminal domain, transcription initiation factor	84.7	0.29	1.00E-05

Search3: Using hidden Markov Model (hMM) generated for the family of Wor3 sequences

PDB Code	Match Description	Probability	E-Value	P-Value
3gox_A	Restriction endonuclease HPY99I; endonuclease-DNA complex	78.2	1	3.50E-05
3na7_A	HP0958; flagellar biogenesis, flagellum export, C4 Zn-ribbon	67	1.4	4.80E-05
1tjl_A	DNAK suppressor protein; DKSA, transcription factor	65.5	2.1	7.20E-05
2zet_C	Melanophilin; complex, GTP-binding protein, GTPase, G-protein	59.2	5.3	1.80E-04
3ttc_A	HYPF, transcriptional regulatory protein; Zn finger	54.7	6.8	2.40E-04

Search 4: Using hidden Markov Model (hMM) generated for the family of Wor3 sequences, region suggest by HHPred

PDB Code	Match Description	Probability	E-Value	P-Value
3na7_A	HP0958; flagellar biogenesis, flagellum export, C4 Zn-ribbon	57.7	3	0.0001
2m0e_A	Zinc finger and BTB domain-containing protein 17; C2H2 zinc fingers	47.5	6.9	0.00024
1znf_A	31ST zinc finger from XFIN; zinc finger DNA binding domain	46.3	8.4	0.00029
2lvu_A	Zinc finger and BTB domain-containing protein 17; C2H2 zinc finger	42.4	7.4	0.00026
2m0f_A	Zinc finger and BTB domain-containing protein 17; C2H2 zinc fingers	35.5	11	0.00037

Search 5: Wor3 243-326aa sequence, default settings for HHPred

PDB Code	Match Description	Probability	E-Value	P-Value
2kq9_A	DNAK suppressor protein; zinc binding protein	72.4	0.46	1.60E-05
1tjl_A	DNAK suppressor protein; DKSA, transcription factor	69.8	0.65	2.20E-05
1znf_A	31ST zinc finger from XFIN; zinc finger DNA binding domain	62.1	1.8	6.00E-05
2kgo_A	Uncharacterized protein YBII; Zn finger	61.9	1.7	6.00E-05
2lvu_A	Zinc finger and BTB domain-containing protein 17; C2H2 zinc finger	63.2	2	6.90E-05

Table S3: List of strains used in this study.

Description	Number	Genotype	Reference
HIS- LEU- auxotrophic a/a, White	RZY47	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm434 IRO1/iro1D::imm434*	1
Wor1 Deletion	RZY219	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor1Δ::C.m.LEU2/wor1Δ::C.d.HIS1*	1
Wild Type, White	AHY135	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}*	This Study
Wild Type, Opaque	AHY136	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434}*	This Study
Wor3 Deletion, White	AHY207	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor3Δ::C.m.LEU2/wor3Δ::C.d.HIS1*	This Study
Wor3 Deletion, Opaque	AHY212	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor3Δ::C.m.LEU2/wor3Δ::C.d.HIS1*	This Study
Wor3-myc, White	AHY224	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR3/WOR3-13myc*	This Study
Wor3-myc, Opaque	AHY231	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR3/WOR3-13myc*	This Study
Wor3-GFP, white	MLY695	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR3/WOR3-GFP*	This Study
Wor3-GFP, opaque	MLY698	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} WOR3/WOR3-GFP*	This Study
Wild Type, pMET3-Wor3	AHY219	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} pMET3-Wor3; SAT1; RP10*	This Study
Wild Type, pMET3-blank	AHY214	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} pMET3-blank1; SAT1; RP10*	This Study
Wor1 deletion, pMET3-Wor3	AHY223	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor1Δ::C.m.LEU2/wor1Δ::C.d.HIS1 pMET3-Wor3; SAT1; RP10*	This Study

Wor2 deletion, pMET3-Wor3	AHY220	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor2Δ::C.m.LEU2/wor2Δ::C.d.HIS1 pMET3-Wor3; SAT1; RP10*	This Study
Efg1 deletion, pMET3-Wor3	AHY222	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a efg1Δ::C.m.LEU2/efg1Δ::C.d.HIS1 pMET3-Wor3; SAT1; RP10*	This Study
Czf1 deletion, pMET3-Wor3	AHY221	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a czf1Δ::C.m.LEU2/czf1Δ::C.d.HIS1 pMET3-Wor3; SAT1; RP10*	This Study
Wor1 deletion, pMET3-blank	AHY218	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor2Δ::C.m.LEU2/wor2Δ::C.d.HIS1 pMET3-blank; SAT1; RP10*	This Study
Wor2 deletion, pMET3-blank	AHY215	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor2Δ::C.m.LEU2/wor2Δ::C.d.HIS1 pMET3-blank; SAT1; RP10*	This Study
Efg1 deletion, pMET3-blank	AHY217	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor2Δ::C.m.LEU2/wor2Δ::C.d.HIS1 pMET3-blank; SAT1; RP10*	This Study
Czf1 deletion, pMET3-blank	AHY216	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor2Δ::C.m.LEU2/wor2Δ::C.d.HIS1 pMET3-blank; SAT1; RP10*	This Study
Wild Type, pMET3-Czf1	AHY555	*a/a C.m.LEU2/leu2Δ C.d.HIS1/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} pMET3-Czf1; SAT1; RP10*	This Study
Wor3 deletion, pMET3-Czf1	AHY556	*a/a leu2Δ/leu2Δ his1Δ/his1Δ URA3/ura3Δ::imm^{434} IRO1/iro1Δ::imm^{434} a/a wor3Δ::C.m.LEU2/wor3Δ::C.d.HIS1 pMET3-Czf1; SAT1; RP10*	This Study
pTEF-blank, CYC1 blank	MLY939	*MATa ura3 his3 leu2 met. PCYC1-UAS::PFLO11 (empty vector)-LacZ; URA3; 2μ pTEF-(empty vector) Leu2; integrated*	This Study
pTEF-blank, CYC1 Wor3	MLY944	*MATa ura3 his3 leu2 met. PCYC1-UAS::WOR3 site-LacZ; URA3; 2μ pTEF-(empty vector) Leu2; integrated*	This Study
pTEF-Wor3, CYC1 blank	MLY937	*MATa ura3 his3 leu2 met. PCYC1-UAS::PFLO11 (empty vector)-LacZ; URA3; 2μ pTEF-Wor3 Leu2; integrated*	This Study

pTEF-Wor3, CYC1 Wor3	MLY942	*MATa ura3 his3 leu2 met. PCYC1-UAS::WOR3 site-LacZ; URA3; 2μ pTEF-Wor3 Leu2; integrated*	This Study
References			
	1	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.	

Table S4: List of plasmids used in this study.

Description	Name	Reference
Ectopic Expression, C. albicans		
pMET3-blank-SAT1	pADH33	This Study
PMET3 Wor3-SAT1	pADH41	This Study
pMET3-Czf1-SAT1	pADH37	This Study
Wor3 Codon Optimization and Protein Expression		
Full Length Wor3	pMBL348	This Study
Wor3 204aa-457aa	pMBL405	This Study
Wor3 243aa-457aa	pMBL406	This Study
Wor3 287aa-457aa	pMBL407	This Study
Wor3 243aa-404aa	pMBL540	This Study
Wor3 243aa-326aa	pMBL542	This Study
Wor3 204aa-404aa	pMBL537	This Study
Wor3 204aa-370aa	pMBL538	This Study
Wor3 370aa-457aa	pMBL536	This Study
Ectopic Expression, S. cerevisiae		
Integrated pTEF Control	pJW404	1
Integrated pTEF-Wor3	pMBL429	This Study
S. cerevisiae Reporter Plasmids		
UAS-CYC1-LacZ Control	CB195	2
UAS-CYC1-LacZ Wild Type Motif	pMBL436	This Study
C. albicans Tools		
LEU2 Knock Out	pSN40	3
HIS1 Knock Out	pSN52	3
13-myc Source	pADH34	3
GFP Surce	pMBL162	4
References		
1		Zalatan JG, Coyle SM, Rajan S, Sidhu SS, Lim WA (2012) Conformational control of the Ste5 scaffold protein insulates against MAP kinase misactivation. *Science* 337(6099):1218-1222.

2		Baker CR, Booth LN, Sorrells TR, Johnson AD (2012) Protein modularity, cooperative binding, and hybrid regulatory States underlie transcriptional network diversification. *Cell* 151(1):80-95.
3		Hernday AD, Noble SM, Mitrovich QM, Johnson AD (2010) Genetics and molecular biology in Candida albicans. *Methods Enzymol* 470:737-758.
4		Lohse MB, Johnson AD (2010) Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans. *Mol Microbiol* 78(2):331-343.

Table S5: List of oligonucleotide sequences used in this study, including ones for the Wor3-specific library of oligonucleotides containing systematic substitutions of all possible nucleotides at each position within the Wor3 *cis*-regulatory sequence.

Name	Description	Sequence
pMET3-SAT1 plasmid construction		
AHO251	NEW MCS 5'	gatcccccgggctgcaggaattcgatatca
AHO252	NEW MCS 3'	AGCTTGATATCGAATTCCTGCAGCCCGGGG
AHO249	SAT1 5' Amplification	aaagaacatgtgagtgaaattctggaaatctgg
AHO250	SAT1 3' Amplification	TGCTCACATGTGCAGGACCACCTTTGATTG
Ectopic Expression Plasmid Construciton		
AHO268	Czf1 5' BglII	ccaAGATCTatgagttcaatacccaatatcaattg
AHO269	Czf1 3' XmaI	AAACCCGGGTTATTTACTTCTGTATTCAACAATACCTCTC
AHO324	Wor3 5' BamHI	ccccggatccATGGATCAAACATATTTGGACCAAC
AHO325	Wor3 3' XmaI	GCAGCCCGGGTTAATTGTTTGGATACTCTTGGTGG
GFP tagging of Wor3		
MBL 453	GFP-SAT1 Cassette 5'	caacagcagcagcagcaccaacagcaacaaccataccaccaagagtatccaaacaat GGT GGT GGT TCT AAA GGT GAA GAA TTA
MBL 454	GFP-SAT1 Cassette 3'	CTG ACT ATG CAA GCA AAA CGT GTT ATT TAA AAG TAG TAA TGT AGG TTT TAA AAA CTA TTA GCG GCC GCT CTA GAA CTA GTG GAT CT
MBL 455	Wor3 5' Check Primer	CGC CAA TTC AAA ACC AAT ATG GAA TGA ACA TGT C
MBL 456	Wor3 3' Check Primer	GCT CAA GAG GGG CAT ACC TCA TCA GAG
WOR3 Knock Out		
AHO314	5' flank, external	AAGCCGTACATTCTTTCAAGTTAC
AHO315	5' flank, internal	CACGGCGCGCCTAGCAGCGGGGTGGATACTATGTTCTATTATATGTGG
AHO316	3' flank, internal	GTCAGCGGCCGCATCCCTGCaaataacacgttttgcttgc

AHO317	3' flank, external	CTCAGACGAGATTAAAAACACTTTCATG
AHO320	ORF check 1	ATGGATCAAACATATTTGGACC
AHO321	ORF Check 1	ATTGAAGATTGCTCGTAACG
AHO318	5' Flank Check	TTTATGGTTCCTTTTCAGTTCAAGC
AHO319	3' Flank Check	CCAACCAAAACGGAAATAATAC
AHO322	ORF check 1	AACAGCAACAGCCTCAGCAAC
AHO323	ORF check 1	TGCAGGCTTTTCTTCTTGGC
WOR2 Knock Out		
AHO140	5' flank, external	TTTAACCTGTAAGACTCATCCTTC
AHO141	5' flank, internal	CACGGCGCGCCTAGCAGCGGTAGCTTCACACTTGATTTTG
AHO142	3' flank, internal	GTCAGCGGCCGCATCCCTGCTAATAAATCCAATATATTCATACTTTTG
AHO143	3' flank, external	TTAACAATAGTCAATATATGTGTTCTC
AHO144	5' Flank Check	TTATACTATGATCTCTCGATTTCCG
AHO145	3' Flank Check	AAGAATTTTGAGTTTGTGGG
AHO166	ORF check 1	ATGACACAATTACCTTCTGTTTCAG
AHO167	ORF check 1	TGTACTGGCAATTGTGACTC
AHO168	ORF Check 2	CACAACAACAACAACTCCAG
AHO169	ORF Check 2	CCTTGGTGGTAATTTCAGTAAATC
CZF1 Knock Out		
AHO146	5' flank, external	TATAGCAAAATTCAAAGGGC
AHO147	5' flank, internal	CACGGCGCGCCTAGCAGCGGCCAGATAGTTTTCGTTTGAATG
AHO148	3' flank, internal	GTCAGCGGCCGCATCCCTGCTAAGCTTCTCTGTGTTGGAGG
AHO149	3' flank, external	CAAGTAATATGGCCAACAAAC
AHO150	5' Flank Check	CCTCAACATATTCTATATACCCAAC
AHO151	3' Flank Check	CTTTACACACGACACCAATTAC
AHO158	ORF check 1	ACCCAATATCAATTGGAATGACCCTAAC
AHO159	ORF check 1	ACATCATGGCATTTTGCTCG
AHO160	ORF Check 2	CTGCCTCGACTCAACAATATC
AHO161	ORF Check 2	GCACTCAGTACACCTTGGTC
EFG1 Knock Out		
AHO152	5' flank,	TTGATTTAGTGTATTACATCCAGCC

	external	
AHO153	5' flank, internal	CACGGCGCGCCTAGCAGCGGAATGGGTTAAGGGTTGGTTG
AHO154	3' flank, internal	GTCAGCGGCCGCATCCCTGCAGGTTCAGTTCACCCTTCAC
AHO155	3' flank, external	CCATCGAGTAAAATATACTTGTTCG
AHO156	5' Flank Check	CTGACACAGTCAAAAAGTTAGCAGAG
AHO157	3' Flank Check	ACGCCACAAACTATCATCTC
AHO162	ORF check 1	CTATACCCTATTACAATCAAATGAACGG
AHO163	ORF check 1	TGCATTGTCGATACATGTGG
AHO164	ORF Check 2	TCAGGATACGTTGAACGCCTC
AHO165	ORF Check 2	TTACTGCCACCACTGGTAGC
13 x myc Tagging of Wor3		
AHO354	5' Myc Tagging Primer	tgctcaacaacaacagcagcagcagcaccaacagcaacaaccataccaccaagagtatccaaacaa tcggatccccgggttaattaacgg
AHO355	3' Myc Tagging Primer	GTAAAACCTGACTATGCAAGCAAAACGTGTTATTTAAAAGTAGTAATGTAGG TTTTAAAAACTAGGCGGCCGCTCTAGAACTAGTGGATC
AHO346	5' check primer	AATACATTCAACCACCTCCACC
AHO349	3' check primer	GCATACCTCATCAGAGAAAACTTACATG
MITOMI Template Preparation		
MLP118	Full Length Wor3, 5'	CTCGAGAATTCGCCACC ATG GAT CAA ACA TAT TTG GAC CAA C
MLP119	Full Length Wor3, 3'	GTA GCA GCC TGA GTC GTT ATT AAT GAT GAT GAT GAT GAT GAT TGT TTG GAT ACT CTT GGT GGT ATG
MLP224	204aa Wor3 Truncation, 5'	CTCGAGAATTCGCCACC atg agtagtgcatatcatccagttgat
MLP 225	243aa Wor3 Truncation, 5'	CTCGAGAATTCGCCACC atg ACT AGA TCC ATT TGT ACC AGG
MLP 226	457aa Wor3 Truncation, 3'	GTA GCA GCC TGA GTC GTT ATT AAT GAT GAT GAT GAT GAT GTG ATC TTT TCT TTT TAG TTC GAC AAT TAA AAC
Wor3 CTG Codon Optimization		
MLP112	XmaI 5' End	Gttaca cccggg ATG GAT CAA ACA TAT TTG GAC CAA C

MLP113	Stop-XhoI 3' End	CTC GGA CTC GAG TTA ATT GTT TGG ATA CTC TTG GTG GTA TG
MLP114	Codon A Change, Forward	CAA GAT ACT AAT TCC ATT CCA CAA CAA C
MLP115	Codon A Change, Reverse	GTT GTT GTG GAA TGG AAT TAG TAT CTT G
MLP116	Codon B Change, Forward	CAA AAT AAC GAG TCC AGA AGA GGA AG
MLP117	Codon B Change, Reverse	CTT CCT CTT CTG GAC TCG TTA TTT TG
Wor3 Expression Cloning		
MLP197	204aa XmaI 5'	Gttacacccggg ATG agtagtgcatatcatccagttgat
MLP198	243aa XmaI 5'	Gttacacccggg ATG ACT AGA TCC ATT TGT ACC AGG
MLP199	287aa XmaI 5'	Gttacacccggg ATG AGA ACT TTC AAA CTA TGT GAT CAT TGT C
MLP468	326aa XmaI 5'	Gttacacccggg TTG GCA GAA AGA AGA TTT GTT TTG T
MLP469	370aa XmaI 5'	Gttacacccggg AAC GAA GGT AAC AAT AAT GGT GAT GA
MLP470	404aa Stop-XhoI 3'	CTC GGA CTC GAG TTA GGC ACC TCG TGC AG
MLP471	370aa Stop-XhoI 3'	CTC GGA CTC GAG TTA GTT TTC TTT GGA TAA AAT AGA TGC ATC C
MLP472	326aa Stop-XhoI 3'	CTC GGA CTC GAG TTA CAA TGG TAT TTC TGA CCC ACA TC
MLP200	457aa Stop-XhoI 3'	CTC GGA CTC GAG TTA TGA TCT TTT CTT TTT AGT TCG ACA ATT AAA AC
S. cerevisiae Ectopic Expression Cloning		
MLP112	Integrated, pJW404 5' XmaI	Gttaca cccggg ATG GAT CAA ACA TAT TTG GAC CAA C
MLP251	Integrated, pJW404 3' Stop BamHI	CTC GGA GGA TCC TTA ATT GTT TGG ATA CTC TTG GTG GTA TG
S. cerevisiae LacZ reporter cloning		

MLP240	Wild Type Motif, Forward, XhoI	TCGAG Agtc ACGGAC CAAATAACC AAAAACagtc C
MLP241	Wild Type Motif, Reverse, XhoI	TCGAG GAC TGT TTT TGG TTA TTT GGT CCG TGA CT C
Gel Shifts		
MLP212	Wild Type Motif, Forward	ACGGAC CAAATAACC AAAAAC
MLP213	Wild Type Motif, Reverse	GTT TTT GGT TAT TTG GTC CGT
MLP220	Mutated Motif, Forward	ACGGAC GTAACAAAG AAAAAC
MLP221	Mutated Motif, Reverse	GTT TTT CTT TGT TAC GTC CGT
MITOMI binding Curves		
33	Wor3_canonical	CGCAATTGCGAGTACACTCATAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
34	Wor3_0mutA	CGCAATTGCGAGTACAC**A**CATAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
35	Wor3_0mutC	CGCAATTGCGAGTACAC**C**CATAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
36	Wor3_0mutG	CGCAATTGCGAGTACAC**G**CATAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
37	Wor3_1mutA	CGCAATTGCGAGTACACT**A**ATAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
38	Wor3_1mutG	CGCAATTGCGAGTACACT**G**ATAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
39	Wor3_1mutT	CGCAATTGCGAGTACACT**T**ATAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
40	Wor3_2mutC	CGCAATTGCGAGTACACTC**C**TAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
41	Wor3_2mutG	CGCAATTGCGAGTACACTC**G**TAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
42	Wor3_2mutT	CGCAATTGCGAGTACACTC**T**TAACCAGATCGACCTTCCTCTCCGGCGGTATGAC
43	Wor3_3mutA	CGCAATTGCGAGTACACTCA**A**AACCAGATCGACCTTCCTCTCCGGCGGTATGAC
44	Wor3_3mutC	CGCAATTGCGAGTACACTCA**C**AACCAGATCGACCTTCCTCTCCGGCGGTATGAC

45	Wor3_3mutG	CGCAATTGCGAGTACACTCA**G**AACCAGATCGACCTTCCTCTCCGGCGGTATGAC
46	Wor3_4mutC	CGCAATTGCGAGTACACTCAT**C**ACCAGATCGACCTTCCTCTCCGGCGGTATGAC
47	Wor3_4mutG	CGCAATTGCGAGTACACTCAT**G**ACCAGATCGACCTTCCTCTCCGGCGGTATGAC
48	Wor3_4mutT	CGCAATTGCGAGTACACTCAT**T**ACCAGATCGACCTTCCTCTCCGGCGGTATGAC
49	Wor3_5mutC	CGCAATTGCGAGTACACTCATA**C**CCAGATCGACCTTCCTCTCCGGCGGTATGAC
50	Wor3_5mutG	CGCAATTGCGAGTACACTCATA**G**CCAGATCGACCTTCCTCTCCGGCGGTATGAC
51	Wor3_5mutT	CGCAATTGCGAGTACACTCATA**T**CCAGATCGACCTTCCTCTCCGGCGGTATGAC
52	Wor3_6mutA	CGCAATTGCGAGTACACTCATAA**A**CAGATCGACCTTCCTCTCCGGCGGTATGAC
53	Wor3_6mutG	CGCAATTGCGAGTACACTCATAA**G**CAGATCGACCTTCCTCTCCGGCGGTATGAC
54	Wor3_6mutT	CGCAATTGCGAGTACACTCATAA**T**CAGATCGACCTTCCTCTCCGGCGGTATGAC
55	Wor3_7mutA	CGCAATTGCGAGTACACTCATAAC**A**AGATCGACCTTCCTCTCCGGCGGTATGAC
56	Wor3_7mutG	CGCAATTGCGAGTACACTCATAAC**G**AGATCGACCTTCCTCTCCGGCGGTATGAC
57	Wor3_7mutT	CGCAATTGCGAGTACACTCATAAC**T**AGATCGACCTTCCTCTCCGGCGGTATGAC
58	Wor3_8mutC	CGCAATTGCGAGTACACTCATAACC**C**GATCGACCTTCCTCTCCGGCGGTATGAC
59	Wor3_8mutG	CGCAATTGCGAGTACACTCATAACC**G**GATCGACCTTCCTCTCCGGCGGTATGAC
60	Wor3_8mutT	CGCAATTGCGAGTACACTCATAACC**T**GATCGACCTTCCTCTCCGGCGGTATGAC
61	Wor3_9mutA	CGCAATTGCGAGTACACTCATAACCA**A**ATCGACCTTCCTCTCCGGCGGTATGAC
62	Wor3_9mutC	CGCAATTGCGAGTACACTCATAACCA**C**ATCGACCTTCCTCTCCGGCGGTATGAC
63	Wor3_9mutT	CGCAATTGCGAGTACACTCATAACCA**T**ATCGACCTTCCTCTCCGGCGGTATGAC

Table S6: List of red flagged locations in ChIP-chip data. Called peaks were filtered by subtraction of this set of likely artifactual peaks, based on the fact that these loci showed variable but substantial enrichment in many deletion (control) ChIP-chip experiments.

SEQ_NAME	START	END	STRAND
Ca21chr1	1725	2224	+
Ca21chr1	3082	3581	+
Ca21chr1	4875	5374	+
Ca21chr1	16323	17328	+
Ca21chr1	31842	32546	+
Ca21chr1	126390	128619	+
Ca21chr1	263281	264720	+
Ca21chr1	269104	270022	+
Ca21chr1	289712	290461	+
Ca21chr1	332739	334096	+
Ca21chr1	334662	336535	+
Ca21chr1	486942	488232	+
Ca21chr1	488543	489042	+
Ca21chr1	521342	522202	+
Ca21chr1	630176	631681	+
Ca21chr1	632142	633313	+
Ca21chr1	643289	644215	+
Ca21chr1	650129	651390	+
Ca21chr1	662242	664402	+
Ca21chr1	748185	748684	+
Ca21chr1	751347	752645	+
Ca21chr1	765540	767039	+
Ca21chr1	867157	868489	+
Ca21chr1	870737	871236	+
Ca21chr1	871888	872400	+
Ca21chr1	880839	881402	+
Ca21chr1	882239	882769	+
Ca21chr1	886084	888099	+
Ca21chr1	890449	892837	+
Ca21chr1	949400	951722	+
Ca21chr1	1168042	1168541	+
Ca21chr1	1195221	1197023	+
Ca21chr1	1213902	1214902	+

Ca21chr1	1278633	1280738	+
Ca21chr1	1281025	1282994	+
Ca21chr1	1337672	1340068	+
Ca21chr1	1375772	1376741	+
Ca21chr1	1377882	1379293	+
Ca21chr1	1409046	1409812	+
Ca21chr1	1468534	1469344	+
Ca21chr1	1489443	1489942	+
Ca21chr1	1563038	1565967	+
Ca21chr1	1728613	1729112	+
Ca21chr1	1834354	1835814	+
Ca21chr1	1864168	1865548	+
Ca21chr1	1952161	1953340	+
Ca21chr1	2022581	2024443	+
Ca21chr1	2104743	2105242	+
Ca21chr1	2299778	2300710	+
Ca21chr1	2301286	2302525	+
Ca21chr1	2364348	2369345	+
Ca21chr1	2406592	2407384	+
Ca21chr1	2432455	2433574	+
Ca21chr1	2434130	2435635	+
Ca21chr1	2435875	2437043	+
Ca21chr1	2492138	2495200	+
Ca21chr1	2513569	2514777	+
Ca21chr1	2588693	2589192	+
Ca21chr1	2698172	2699181	+
Ca21chr1	2782583	2785508	+
Ca21chr1	2830955	2831454	+
Ca21chr1	2839943	2840956	+
Ca21chr1	2883066	2886263	+
Ca21chr1	2891695	2893019	+
Ca21chr1	2911732	2913486	+
Ca21chr1	2955233	2955732	+
Ca21chr1	2960191	2960690	+
Ca21chr1	3045995	3048182	+
Ca21chr1	3061043	3063251	+
Ca21chr1	3074225	3075665	+
Ca21chr1	3105892	3107870	+
Ca21chr1	3113806	3114951	+
Ca21chr1	3120465	3120964	+
Ca21chr1	3146380	3147648	+

Ca21chr1	3179728	3180818	+
Ca21chr2	5665	8043	+
Ca21chr2	19691	21205	+
Ca21chr2	28566	28802	+
Ca21chr2	28805	30051	+
Ca21chr2	174599	176239	+
Ca21chr2	180800	182863	+
Ca21chr2	201472	203088	+
Ca21chr2	287869	288374	+
Ca21chr2	327699	331928	+
Ca21chr2	496141	496640	+
Ca21chr2	603543	604784	+
Ca21chr2	623592	626320	+
Ca21chr2	749879	751801	+
Ca21chr2	790034	790533	+
Ca21chr2	804502	805314	+
Ca21chr2	805800	806857	+
Ca21chr2	810413	810913	+
Ca21chr2	830516	831313	+
Ca21chr2	831684	832648	+
Ca21chr2	833913	835946	+
Ca21chr2	857630	858129	+
Ca21chr2	961154	961997	+
Ca21chr2	1110528	1111502	+
Ca21chr2	1112105	1112604	+
Ca21chr2	1141457	1142393	+
Ca21chr2	1143556	1144055	+
Ca21chr2	1146357	1146856	+
Ca21chr2	1167060	1168075	+
Ca21chr2	1302340	1304152	+
Ca21chr2	1363522	1364021	+
Ca21chr2	1395602	1396519	+
Ca21chr2	1497955	1498454	+
Ca21chr2	1506984	1508930	+
Ca21chr2	1620567	1621885	+
Ca21chr2	1633110	1634121	+
Ca21chr2	1637927	1638426	+
Ca21chr2	1690195	1691935	+
Ca21chr2	1887962	1888461	+
Ca21chr2	1921934	1923103	+
Ca21chr2	1925905	1930214	+

Ca21chr2	2030706	2032361	+
Ca21chr2	2166626	2167987	+
Ca21chr2	2169743	2171046	+
Ca21chr2	2186591	2187484	+
Ca21chr2	2189961	2192530	+
Ca21chr2	2229463	2229962	+
Ca21chr3	1752	8097	+
Ca21chr3	15852	16357	+
Ca21chr3	16407	19533	+
Ca21chr3	133121	135808	+
Ca21chr3	195793	196292	+
Ca21chr3	317436	317996	+
Ca21chr3	318066	320137	+
Ca21chr3	361668	362167	+
Ca21chr3	458414	458913	+
Ca21chr3	538653	539506	+
Ca21chr3	590523	592247	+
Ca21chr3	817856	818355	+
Ca21chr3	818882	819381	+
Ca21chr3	823335	826483	+
Ca21chr3	845635	846134	+
Ca21chr3	866133	866632	+
Ca21chr3	941388	942396	+
Ca21chr3	973884	975470	+
Ca21chr3	991647	992664	+
Ca21chr3	993121	994596	+
Ca21chr3	1023174	1024521	+
Ca21chr3	1030736	1036390	+
Ca21chr3	1038787	1039286	+
Ca21chr3	1056566	1057274	+
Ca21chr3	1067948	1068926	+
Ca21chr3	1096418	1097684	+
Ca21chr3	1130362	1131166	+
Ca21chr3	1153194	1154044	+
Ca21chr3	1191826	1192587	+
Ca21chr3	1193137	1194313	+
Ca21chr3	1232514	1233013	+
Ca21chr3	1279191	1282919	+
Ca21chr3	1366734	1368201	+
Ca21chr3	1469441	1469940	+
Ca21chr3	1574975	1576045	+

Ca21chr3	1608042	1611284	+
Ca21chr3	1636486	1637102	+
Ca21chr3	1714840	1715339	+
Ca21chr3	1715616	1716965	+
Ca21chr3	1730059	1732305	+
Ca21chr3	1793428	1793927	+
Ca21chr3	1796668	1797167	+
Ca21chr4	39465	41389	+
Ca21chr4	207805	208304	+
Ca21chr4	227585	229004	+
Ca21chr4	245550	246049	+
Ca21chr4	255674	257030	+
Ca21chr4	310023	312647	+
Ca21chr4	315421	316861	+
Ca21chr4	317039	318349	+
Ca21chr4	326309	326808	+
Ca21chr4	336838	337601	+
Ca21chr4	470771	471464	+
Ca21chr4	480635	481443	+
Ca21chr4	548009	549178	+
Ca21chr4	674364	678707	+
Ca21chr4	678838	680277	+
Ca21chr4	779961	780460	+
Ca21chr4	780911	781410	+
Ca21chr4	782356	782855	+
Ca21chr4	837806	841339	+
Ca21chr4	845981	846480	+
Ca21chr4	936910	937879	+
Ca21chr4	962583	963769	+
Ca21chr4	992579	997715	+
Ca21chr4	1053028	1054625	+
Ca21chr4	1055449	1056829	+
Ca21chr4	1362564	1363934	+
Ca21chr4	1426618	1429261	+
Ca21chr4	1574337	1575542	+
Ca21chr4	1585505	1586137	+
Ca21chr5	23548	29661	+
Ca21chr5	30709	32873	+
Ca21chr5	134693	135645	+
Ca21chr5	217346	217845	+
Ca21chr5	219884	220383	+

Ca21chr5	342916	344807	+
Ca21chr5	347919	351071	+
Ca21chr5	468716	471745	+
Ca21chr5	594542	596396	+
Ca21chr5	637707	638822	+
Ca21chr5	640411	641235	+
Ca21chr5	642637	644991	+
Ca21chr5	683005	683943	+
Ca21chr5	769369	769868	+
Ca21chr5	776891	778356	+
Ca21chr5	785599	787547	+
Ca21chr5	793466	794083	+
Ca21chr5	856605	857263	+
Ca21chr5	890964	891463	+
Ca21chr5	891567	892770	+
Ca21chr5	916885	918405	+
Ca21chr5	919805	921556	+
Ca21chr5	993756	995592	+
Ca21chr5	997037	997796	+
Ca21chr5	1153238	1154134	+
Ca21chr5	1154156	1155352	+
Ca21chr6	23430	24815	+
Ca21chr6	54387	56126	+
Ca21chr6	64400	66035	+
Ca21chr6	111161	111660	+
Ca21chr6	114731	115590	+
Ca21chr6	309668	310167	+
Ca21chr6	331257	332767	+
Ca21chr6	357012	358099	+
Ca21chr6	363878	365166	+
Ca21chr6	402823	403696	+
Ca21chr6	425041	425969	+
Ca21chr6	463176	466025	+
Ca21chr6	489921	491447	+
Ca21chr6	496091	499011	+
Ca21chr6	520867	521713	+
Ca21chr6	772990	775210	+
Ca21chr6	971454	978542	+
Ca21chr6	980040	983792	+
Ca21chr6	992851	995441	+
Ca21chr6	1027172	1028091	+

Ca21chr7	20	1155	+
Ca21chr7	111461	111960	+
Ca21chr7	131585	132105	+
Ca21chr7	132567	134273	+
Ca21chr7	202217	203315	+
Ca21chr7	211801	212739	+
Ca21chr7	257914	259126	+
Ca21chr7	299186	299685	+
Ca21chr7	301198	301697	+
Ca21chr7	303017	305451	+
Ca21chr7	384489	385385	+
Ca21chr7	425812	428712	+
Ca21chr7	439531	441654	+
Ca21chr7	480373	482487	+
Ca21chr7	519137	519938	+
Ca21chr7	710719	712392	+
Ca21chr7	918848	919780	+
Ca21chrR	155204	162065	+
Ca21chrR	172397	172896	+
Ca21chrR	339478	339977	+
Ca21chrR	368753	371056	+
Ca21chrR	403787	404286	+
Ca21chrR	456397	458440	+
Ca21chrR	515659	516158	+
Ca21chrR	516764	517873	+
Ca21chrR	561933	564818	+
Ca21chrR	566055	568721	+
Ca21chrR	646775	647274	+
Ca21chrR	692396	693534	+
Ca21chrR	717483	719056	+
Ca21chrR	773211	775418	+
Ca21chrR	777788	779491	+
Ca21chrR	839137	840810	+
Ca21chrR	920253	921025	+
Ca21chrR	955134	957305	+
Ca21chrR	1019888	1020387	+
Ca21chrR	1041023	1041886	+
Ca21chrR	1095525	1096646	+
Ca21chrR	1158728	1159227	+
Ca21chrR	1164086	1164655	+
Ca21chrR	1190811	1191385	+

Ca21chrR	1229910	1230409	+
Ca21chrR	1233115	1233614	+
Ca21chrR	1274197	1276674	+
Ca21chrR	1297252	1298587	+
Ca21chrR	1316661	1318437	+
Ca21chrR	1322922	1325071	+
Ca21chrR	1386474	1388619	+
Ca21chrR	1388917	1389695	+
Ca21chrR	1422007	1424679	+
Ca21chrR	1460904	1461403	+
Ca21chrR	1577043	1577798	+
Ca21chrR	1606104	1607579	+
Ca21chrR	1607784	1608283	+
Ca21chrR	1611848	1614199	+
Ca21chrR	1656905	1657927	+
Ca21chrR	1662337	1662836	+
Ca21chrR	1712520	1713910	+
Ca21chrR	1723451	1725166	+
Ca21chrR	1742933	1747664	+
Ca21chrR	1762547	1764388	+
Ca21chrR	1772309	1773150	+
Ca21chrR	1783560	1786123	+
Ca21chrR	1787033	1787541	+
Ca21chrR	1788653	1791743	+
Ca21chrR	1791747	1792251	+
Ca21chrR	1808535	1809537	+
Ca21chrR	1813236	1813735	+
Ca21chrR	1830846	1831363	+
Ca21chrR	1881876	1882958	+
Ca21chrR	1884874	1885373	+
Ca21chrR	1885575	1886074	+
Ca21chrR	1888541	1889561	+
Ca21chrR	1890215	1890714	+
Ca21chrR	1890785	1896981	+
Ca21chrR	2049137	2049636	+
Ca21chrR	2050368	2052509	+
Ca21chrR	2227016	2229053	+
Ca21chrR	2260510	2265696	+

Supplemental Data Files

File S1: Compilation of microarray, RNA-seq, and ChIP-chip data presented in this study. From left to right in the Excel spreadsheet, columns are as follows. (A) Orf19 number. (B) Gene name, where applicable. (C) CGD description of the gene. (D) Whether the gene is a transcriptional regulator, based on Homann *et al.*, 2009 (4), “1” represents yes. (E) Whether the upstream region for the gene is bound by either Wor1 or Wor3 in opaque cells, “1” represents yes. (F) Whether the gene was excluded from our analysis based on a lack of observed transcription in previously published RNA-seq experiments (5), “1” represents exclusion. (G) Maximum Wor1 enrichment in the upstream region for the gene, based on reanalysis of previously published data (3), values are on a log2 scale. (H) Maximum Wor3 enrichment in the upstream region for the gene, values are on a log2 scale. (I) Whether RNAseq data for a gene is considered trustworthy. (J) Previously published RNAseq of opaque versus white cells (5), values are on a log2 scale. (K) Microarray analysis of opaque versus white cells, values are on a log2 scale. (L) Microarray analysis of a white *wor3* deletion strain versus wild-type white cells, values are on a log2 scale. (M) Microarray analysis of an opaque *wor3* deletion strain versus wildtype opaque cells, values are on a log2 scale. (N-P) Same as columns K-M, with only changes greater than 2-fold (log2>1) shown. (Q) Top 20 opaque enriched genes as determined by RNAseq, “1” represents yes. (R) Top 20 opaque enriched genes as determined by gene expression array, “1” represents yes. (S) Genes with top twenty enriched upstream regions for Wor1 binding, “1” represents yes.

File S2: Plots of 15kb regions centered on the set of 174 Wor3 binding sites. Smoothed enrichment data for the Wor3-myc strain is shown in red, for the untagged control strain in pink. The 500bp called peaks of Wor3 enrichment are indicated by the green boxes in the lower track in each image. Peaks are arranged in order of decreasing Wor3 enrichment. Enrichment (log2) is indicated on the left y-axis. Chromosomal locations and specific enrichment levels for the peak are indicated in the strip above each panel, when multiple peaks are present the enrichment value corresponds to the peak at the center of the plot. Yellow boxes correspond to genes; both the coding sequence and any 5' or 3' untranslated regions are included in the box. Plots produced using the SnapShot Function in MochiView v1.46 (6).

File S3: List of all oligonucleotide sequences for the revised MITOMI 2.0 Random 8mer Library used in this study.

File S4: Series of concentration-dependent binding curves for a set of all possible single systematic mutations within the Wor3 *cis*-regulatory sequence. For each nucleotide, the measured fluorescence intensity ratios (y-axis, expressed as DNA/Protein, red circles) are plotted as a function of soluble DNA concentration (x-axis). The solid red lines reflect global fits to a single-site binding model.

File S5: HHPred search results for a library of 50 artificially-generated 100 amino acid sequences, each containing four instances of the "CxxC" motif and random inter-motif spacers of ten to twenty amino acids. For each sequence, an HHPred search was

conducted using the default settings. Of the fifty artificially-generated sequences, thirty-one (62%), had a significant hit to at least one protein, as defined by an P value less than 1e-4. Significant hits are indicated in bold.

File S6: Position Specific Affinity Matrices for selected Wor3 Motifs. This file contains PSAMs for the Wor3 motifs developed using the Wor3 204-457aa truncation presented in Fig. 4A (7bp GGTTAKN, 8bp DGGTTHNN) and Fig. S3C (9bp NAKAACCAG) as well as a 6bp version of this motif (GGTTAK). The PSAM in Fig. 4C, calculated from the relative binding affinities shown in Fig. 4B, is also included (10bp, TCATAACCAG). Also included is the PSAM for the 8bp GGTTATKW motif developed using the full length Wor3 (Fig. S3B, 8bp motif). This file also includes the summaries of the Matrix Reduce results for all motifs developed using the Wor3 204-457aa and full length Wor3 data sets.

Supplemental Materials and Methods

Strain Construction

A list of strains used in this study can be found in Table S3. *wor3* deletion strains were generated using the His and Leu selectable marker deletion system as previously described (7). Briefly, the DNA sequences flanking the *WOR3* ORF were amplified and fused by stitching PCR to the *HIS3* and *LEU2* selectable marker cassettes. Following transformation of the fusion PCR constructs into AHY2, correct chromosomal integration and deletion of the *WOR3* ORF was confirmed by colony PCR. Similar methods were used to generate the *czf1*, *wor2*, and *efg1* deletion strains. Construction of the *wor1* deletion strain has been previously described (8).

Ectopic expression and “empty-vector” control strains were generated by integrating the p*MET3*-driven expression constructs contained within pADH41 (*WOR3*), pADH37 (*CZF1*) or pADH33 (empty vector) into the recipient strains listed in Table S3. Correct integration of these expression constructs at the *RP10* locus was confirmed by colony PCR using the oligonucleotides listed in Table S5. Epitope tagging of Wor3 was performed using the C-terminal myc-tagging system previously described (7). The oligonucleotides used to amplify the tagging construct are listed in Table S5.

To GFP tag Wor3, PCR was used to amplify *GFP* with the *SAT1* flipper cassette (9) from the previously reported pMBL162 (10), adding approximately 60bp homology to the end of the *WOR3* ORF and the region immediately 3’ of the ORF. The standard background addback strain (AHY135) was transformed with the PCR cassette and selected for growth on YPD supplemented with 200 µg/mL nourseothricin (clonNAT,

WERNER BioAgents). Insertion was verified by PCR against both flanks, after which the *SAT1* marker was removed by growth in YPD media supplemented with 2% Maltose. Cells were then plated on YPD supplemented with 25 µg/mL nourseothricin to identify colonies that had lost the nourseothricin resistance marker. Both proper flank integration and loss of the *SAT1* marker were then verified with a further round of PCR checks.

All *Saccharomyces cerevisiae* strains were generated in a MATa *his leu ura met* derivative of the BY4741/BY4742 S288c deletion library strain background (Open Biosystems).

Plasmid construction

Lists of plasmids and oligonucleotides used in this study can be found in Tables S4 and S5, respectively.

Ectopic expression constructs were built using the pADH33 plasmid backbone. The sequence of pADH33 is available at GenBank (http://www.ncbi.nlm.nih.gov/genbank), accession # KC202163. This plasmid contains the p*MET3* promoter, an *RP10* integration targeting sequence, and a *SAT1* selectable marker. *CZF1* and *WOR3* ORFs were amplified with primers that introduce either a BamHI or BglII restriction site at the 5' end of the ORF, and an XhoI cut site at the 3' end of the ORF. These fragments were digested and cloned into pADH33 to generate pADH41 (*WOR3*) and pADH37 (*CZF1*).

The two CTG codons in Wor3 were changed using two rounds of PCR amplification with primers corresponding to the codons to be mutated. The second round of PCR added 5' XmaI and 3' XhoI restriction sites to allow for insertion into the

pET28b derivative pLIC-H3 (11). pLIC-H3 was a gift of Oren Rosenberg and Jeff Cox. The resulting plasmid was sequenced and three mutations relative to the Wor3 sequence at the Candida Genome Database (CGD, http://www.candidagenome.org/) were observed. Specifically, we observed T81N, D131Y, and loss of Q547. All three of these mutations were also observed when non-codon-optimized Wor3 was cloned into plasmids and sequenced. We believe that these mutations reflect natural variation in the Wor3 sequence in the SC5314 background.

Working from the codon-optimized full length pLIC-H3-Wor3 plasmid (pMBL348), portions of the *WOR3* ORF (204-457aa and 243-457aa), corresponding to the conserved regions, were PCR amplified. These truncations were amplified with 5' XmaI and 3' XhoI restriction sites to allow for insertion into pLIC-H3. Further truncations were produced using similar methods.

Constitutively active *WOR3* plasmids were constructed in the genome integration capable pJW404 background using BamHI and XmaI sites. A version of this plasmid lacking *WOR3* was used as a negative control. The pJW404 plasmid was a gift of Jessica Walter and Wendell Lim and is a variant of the previously reported pNH605 plasmid (12) with the *TEF1* promoter added. pJW404 and its derivatives were linearized by digestion with PmeI prior to transformation. Activation assays used derivatives of the LacZ reporter plasmid with the intact p*CYC1* UAS, CB195 (13). 29bp oligonucleotide pairs containing the motif were added at the XhoI site using a previously described method (1). Plasmids were sequenced to verify the correct orientation of the insertion as well as proper sequence.

Switching Assays

Plate based quantitative white-opaque switching assays were performed as previously described (3, 14). In brief, strains were grown at room temperature for seven days. Three entirely white or opaque colonies were separately resuspended in H_2O, diluted, plated on SD+aa+Uri, and allowed to grow for 1 week. We then examined colonies and counted switch events (for white to opaque switching, opaque colonies or white cells with one or more opaque sections, for opaque to white switching white colonies or opaque colonies with one or more white sectors).

Plate based ectopic expression assays using the p*MET3* ectopic expression system (15) were performed as previously described (3, 8). In brief, strains were grown on repressing media (+Met+Cys) for seven days. Colonies were resuspended in H_2O, diluted, and then plated on either inducing (-Met-Cys) or repressing media. After seven days, plates were scored for colony phenotypes. Strains transformed with the unmodified p*MET3* plasmid were used as negative controls.

MITOMI 2.0 molding master fabrication

Both flow and control layer molding masters were fabricated on 4” test-grade silicon wafers (University Wafer). To improve adhesion of subsequent photoresist layers, all wafers were initially coated with a 5 µm thick layer of SU-8 2005 photoresist (Microchem Corp.) according to the manufacturer’s instructions. Control layer molding master features were then fabricated from SU-8 2025 photoresist (Microchem Corp.) according to the manufacturer’s instructions to produce ~20-25 µm features. For flow molding masters, valves were created from AZ50 XT photoresist (Capitol Scientific)

using the following protocol: [1] spin (500 rpm for 5s with a 5s ramp to spread photoresist, followed by 4,750 rpm for 60s with a 15s ramp to achieve the desired thickness); [2] soft bake (50-112°C at full speed for 18 mins followed by slow cooling to room temperature on an aluminum contact hot plate); [3] rehydrate (overnight); [4] expose (19s exposure at 18.9 mW/cm^2 using a standard i-line photolithography mask aligner); [5] develop (2-5 mins in a 1:3 dilution of AZ400k developer and water); and [6] hard bake (65-190°C at a ramp of 15°C per hour for 14 hrs followed by slow cooling to room temperature on a contact hot plate). Flow channels (~15 µm tall) were then created from SU-8 2015 photoresist (Microchem Corp.) according to the manufacturer's instructions.

MITOMI 2.0 device fabrication

Devices were fabricated from poly(dimethylsiloxane) (PDMS) using standard multilayer soft lithography techniques. Briefly, molding masters were first exposed to trichloro(1H,1H,2H,2H-perfluorooctyl)silane vapors under vacuum for 15-30 mins to prevent adhesion of PDMS to molding masters. Control molds were coated with a ~5 mm thick layer of RTV 615 (R.S. Hughes) at a ratio of 1:5 (cross-linker:base), mixed with a planetary centrifugal mixer (Thinky USA) (mixed for 5 min at 2,000 rpm, de-bubbled for 4 min at 2,200 rpm), and degassed in a vacuum chamber for 45 mins. Flow molds were coated with a thin layer of a similarly mixed 1:20 mixture of RTV 615 using a spin coater (Specialty Coating Systems) (500 rpm for 5s with a 5s ramp, followed by 1,950 rpm for 60s with a 15s ramp). Both molds were then baked at 80°C for 1 hr in a convection oven. After baking, PDMS control layers were peeled from the molds, cut

into individual devices, and aligned to flow layers remaining on their molds. Aligned devices were baked for an additional hour, cut from flow molds, punched to create control access ports, and aligned to DNA arrays printed on SuperChip epoxysilane-coated 2” x 3” slides (ThermoFisher Scientific). This final assembly was baked on a ceramic hotplate for at least 8 hrs before devices were used for experiments.

β-galactosidase assays

β-galactosidase assays were performed using a standard protocol (16). Strains were grown overnight, diluted back, and allowed to reach log phase before harvesting for assays. Strains were grown in selectable media. Data shown are from the same day.

Intergenic Region Overlap Comparison

In order to compare the degree of overlap in intergenic regions bound by Wor1 and Wor3, we started by generating a list of all *Candida albicans* intergenic regions in MochiView by subtracting out a list of transcribed ORFs from the *C. albicans* genome using the “Merge Location Set (Subtraction)” function. The transcribed ORF list used includes the transcribed regions (including UTRs) of 6,018 genes as well as 8 centromeres but excludes 178 ORFs that have not been observed to be transcribed as well as the recently described novel transcriptionally active regions (5). We then used the “Merge Location Set (Union)” function (with “Only keep locations intersected by all contributing location sets” selected) to create location sets of intergenic regions bound by either Wor1 or Wor3. We then used the same function to merge the Wor1 and Wor3 bound intergenic regions to create the list of intergenic regions bound by both Wor1 and

Wor3, which was then subtracted out of the full Wor1 and Wor3 binding set lists using the “Merge Location Set (Subtraction)” function to give the intergenic regions bound by only Wor1 or Wor3.

Binding site overlap was determined in MochiView using the 500bp Wor1 and Wor3 peak enrichment location sets. We used the “Merge Location Set (Union)” function (with “Only keep locations intersected by all contributing location sets” selected) to create a location set of all binding sites where both Wor1 and Wor3 were bound. The Wor1 and Wor3 binding sites were each of a fixed length, specifically 500bp. These locations do not perfectly overlap, and rather than arbitrarily choosing a 500bp portion of the combined region, we considered the entire region defined by the overlapping sites. As such, this results in a distribution of binding region sizes with a mean size of 602bp, larger than the fixed 500bp of the two inputs. In one location, we observed a chain effect where two distinct Wor3 binding peaks overlap different ends of a single Wor1 binding peak.

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 (1, 2, 17). In short, we used MochiView to generate a random set of locations of equivalent size to the experimental peak calls (500bp) or to the mean size for the Wor1+Wor3 binding set (602bp). These random sets were developed from the list of all intergenic regions not bound by either Wor1 or Wor3. These location sets were then analyzed in MochiView using the “Motif Enrichment Plot” utility to generate ROC Plots for the motif at various location sets

relative to the equivalently sized random control set. The 8bp “GGTTATKW” motif developed using full length Wor3 (Fig. S3B, 8bp motif) was used for this analysis. A PSAM for this motif can be found in File S6.

Transcriptional Regulator List

We based our transcriptional regulator list on the set of 283 in Supplemental Dataset 1 from Homann et al. (4). We manually added Wor1 (Orf19.4884) and Wor3 (Orf19.467) to this list based on previous reports and this study (1, 3). We removed MTLα2 (Orf19.10708) from this list because we were working with MTLa/a cells. The two additions and one subtraction bring the number of genes in our Transcriptional Regulator list to 284.

Identification of Wor3 homologs

We performed the first of these analyses using five Wor3 sequences chosen to represent the breadth of the Wor3 family and four reconstructed “ancestral” Wor3 sequences. We reconstructed maximum likelihood ancestral sequences for Wor3, using the software PAML version 4.2 (18) and the Lazarus user interface (19). We consider *Cyberlindnera jadinii* (*Candida utilis*) and *Wickerhamomyces anomalus* to fall outside of the CTG/*Candida* clade and *Dekkera bruxellensis* to fall as an outgroup near *Pichia pastoris* at the base of the CTG/*Candida* clade based on previously published phylogenetic trees (20-25) as well as the newly developed tree from Figure 5 in the case of *W. anomalus*.

In order to determine whether Wor3 had similarities to any known protein family, we conducted several searches using full length and conserved region Wor3 sequences from five species (*C. albicans*, *Debaryomyces hansenii*, *P. pastoris*, *D*. *bruxellensis*, and *W. anomalus*) as well as four different ancestral reconstructions of Wor3. Searches using the PFAM database (26) failed to return any significant matches, although we did observe several non-significant hits to various zinc-ribbon and phorbol esters/diacylglycerol binding domains. The scores of these non-significant hits were similar to those seen for false positive hits to several other transcriptional regulators tested as controls, and well below the scores seen for the correct domains for this control set.

PHYRE2 (27) searches were performed for each protein to see if it could be modeled to a known protein structure. The highest scoring of these sequences returned a maximum modeling confidence of 77.2% limited to 74 residues of one of the full length ancestral reconstructions; however, modeling of the conserved region from this reconstruction covered only 21 residues at 69% confidence. PHYRE2 searches with the other conserved regions produced similar results, roughly 20-25 residues, or 10-15% of the region, mapped at less than 75% confidence.

We next performed a series of PSI-BLAST (28) searches with the nine conserved regions in an attempt to find more distant homologs of Wor3. Only one of the nine first round searches (*W. anomalus*) produced an unexpected result, a predicted protein similar to a WD repeat domain 19 from *Taeniopygia guttata* (E-value of 5e-4 versus a 6e-30 for the “worst” Wor3 homolog); however, a reciprocal blast search with this protein against the *W. anomalus* genome did not return Wor3 as a match. Given this

lack of a reciprocal blast hit, we left this protein out when conducting the second round of PSI-BLAST searches. The second round searches returned four additional unexpected hits; again, searches with these proteins failed to produce a reciprocal blast hit in the relevant fungal species. These four hits were to proteins from *Geobacter daltonii*, *Methanococcus voltae*, *Capnocytophaga* sp. CM59, and delta proteobacterium NaphS2. We conducted a third round of PSI-BLAST searches including the five unexpected hits in the model development, resulting in a large number of new hits. These hits, however, varied from search to search and a manual examination of the alignments suggests they are spurious marches to other proteins with multiple cysteines.

Next, we conducted a series of blastP and TblastN searches (29) versus the microbial, plant, fungal, and eukaryotic genomes available at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). As above, searches were conducted using the conserved regions from the nine proteins. The best hits for the nine members in a given search (e.g. blastP versus plant genomes) were not consistent among the different searches (no protein was the best hit for a majority of searches for a genome set) and a majority of the best hits had E-scores greater than 1. The best hit in all of these searches was the previously mentioned WD repeat domain 19 from *Taeniopygia guttata*, only one other hit scored below 1E-2 and this protein was the top hit in only a single protein's search (*Trichinella spiralis* phosphatidylinositol 4-kinase beta, 0.005).

We searched for Wor3 homolouges in the protein data bank (30) using the program HHPred (31) in five configurations. First, we searched using the *C. albicans* Wor3 sequence as the search seed with default settings for HHPred. Second, we

refined the search to the restricted region of sites 296 to 434, as suggested by HHPred. Third, we searched PDB using a hidden Markov Model (hMM) generated for the family of Wor3 sequences. Fourth, we searched with the hMM using a restricted sequence region as suggested by HHPred. Finally, we searched with the 84 amino acid sequence (243-326aa) sufficient for binding to DNA *in vitro.* Results of these searches are listed in Table S2.

An additional round of searches was performed with the 84 amino acid sequence (243-326aa) sufficient for binding to DNA *in vitro*. No significant Pfam results were detected; the top insignificant hit was to the Microtubule-associated protein CRIPT family. PYHRE2 searches produced a best hit of 43.2% confidence to the PRC-barrel domain involved in photosynthesis. Multiple rounds of PSI-BLAST searches with this region failed to produce any unexpected results and a series of blastP and TblastN searches produced results with significance scores similar to those discussed above for larger portions of Wor3.

Finally, we conducted a comparison of observed and expected divergence in Wor3 sequences as a function of distance from *C. albicans*. Evolutionary distances between species were determined based on the phylogeny presented in Figure 5. The conserved regions of *C. albicans* Wor3 (aa 246-456), Wor1 (aa 9-89), or Mcm1 (aa 60-145) were used as the basis for tblastn searches against the nineteen target genomes indicated in Figure S8A, the highest scoring hit was taken for each genome. Inverse E-scores were log10 transformed and plotted against the estimated evolutionary distance. To estimate the rate of divergence due to evolution, we determined the linear regression for data points corresponding to seven species indicated in Figure S8A, giving a range

of species close to *C. albicans* as well as ones nearly as distantly related as *S. cerevisiae* and *Kluyveromyces lactis* (e.g. *P. pastoris*). The abrupt decrease in the score of the best BLAST hit to Wor3 beyond *W. anomalus* is much greater than would be predicted to arise from drift from an ancient protein (for example, the decreases seen for ancient transcription factors such as Wor1 or Mcm1) (Fig. S8B).

References

1. 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.
2. 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.
3. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD (2007) Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol* 5(10):e256.
4. Homann OR, Dea J, Noble SM, Johnson AD (2009) A phenotypic profile of the Candida albicans regulatory network. *PLoS Genet* 5(12):e1000783.
5. Tuch BB, et al. (2010) The Transcriptomes of Two Heritable Cell Types Illuminate the Circuit Governing Their Differentiation. *PLoS Genet* 6(8):e1001070.
6. Homann OR, Johnson AD (2010) MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol* 8:49.
7. Hernday AD, Noble SM, Mitrovich QM, Johnson AD (2010) Genetics and molecular biology in Candida albicans. *Methods Enzymol* 470:737-758.
8. 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.
9. Reuss O, Vik A, Kolter R, Morschhäuser J (2004) The SAT1 flipper, an optimized tool for gene disruption in *Candida albicans*. *Gene* 341:119-127.
10. Lohse MB, Johnson AD (2010) Temporal anatomy of an epigenetic switch in cell programming: the white-opaque transition of C. albicans. *Mol Microbiol* 78(2):331-343.
11. Hammon J, Palanivelu DV, Chen J, Patel C, Minor DL, Jr. (2009) A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms. *Protein Sci* 18(1):121-133.
12. Zalatan JG, Coyle SM, Rajan S, Sidhu SS, Lim WA (2012) Conformational control of the Ste5 scaffold protein insulates against MAP kinase misactivation. *Science* 337(6099):1218-1222.
13. Baker CR, Booth LN, Sorrells TR, Johnson AD (2012) Protein modularity, cooperative binding, and hybrid regulatory States underlie transcriptional network diversification. *Cell* 151(1):80-95.
14. Miller MG, Johnson AD (2002) White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* 110(3):293-302.
15. Care RS, Trevethick J, Binley KM, Sudbery PE (1999) The MET3 promoter: a new tool for Candida albicans molecular genetics. *Mol Microbiol* 34(4):792-798.
16. Rupp S (2002) *LacZ* assays in yeast. *Methods Enzymol* 350:112-131.
17. Nobile CJ, et al. (2012) A recently evolved transcriptional network controls biofilm development in Candida albicans. *Cell* 148(1-2):126-138.

18. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586-1591.
19. Hanson-Smith V, Kolaczkowski B, Thornton JW (2010) Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol Biol Evol* 27(9):1988-1999.
20. 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.
21. Wang H, Xu Z, Gao L, Hao B (2009) A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evol Biol* 9:195.
22. Kurtzman CP (2012) Phylogeny of the ascomycetous yeasts and the renaming of Pichia anomala to Wickerhamomyces anomalus. *Antonie Van Leeuwenhoek* 99(1):13-23.
23. Curtin CD, Borneman AR, Chambers PJ, Pretorius IS (2012) De-novo assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast Dekkera bruxellensis AWRI1499. *PLoS One* 7(3):e33840.
24. Woolfit M, Rozpedowska E, Piskur J, Wolfe KH (2007) Genome survey sequencing of the wine spoilage yeast Dekkera (Brettanomyces) bruxellensis. *Eukaryot Cell* 6(4):721-733.
25. Tomita Y, Ikeo K, Tamakawa H, Gojobori T, Ikushima S (2012) Genome and transcriptome analysis of the food-yeast Candida utilis. *PLoS One* 7(5):e37226.
26. Punta M, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40(Database issue):D290-301.
27. Kelley LA, Sternberge MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc* 4(3):363-371.
28. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410.
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.