# **Supporting Information**

## Baker et al. 10.1073/pnas.1019177108

 $Ca-ATTCCTTTCGCAGCCAAATTGGGTAACA\\Ca-\Delta 2-ATTCCTTTCGCAGCCAAATTGTTCAACA$ 



Fig. S1. Mcm1-binding site within C. albicans  $\alpha$ sg cis-regulatory sequence is necessary for C. ablicans (Ca)-Mat $\alpha$ 1-dependent transcriptional activation. In S. cerevisiae, Mat $\alpha$ 1 synergistically binds the  $\alpha$ sg cis-regulatory sequence with Mcm1 (1). Disruption of the Mcm1-binding site causes a complete loss of Mat $\alpha$ 1-dependent transcriptional activation (1). The co-occurrence of Mat $\alpha$ 1- and Mcm1-binding sites at a fixed distance in C. albicans strongly suggests that Mat $\alpha$ 1 and Mcm1 are also acting synergistically at C. albicans  $\alpha$ sg cis-regulatory sequences. To test whether, like S. cerevisiae, the Mcm1-binding site within C. albicans  $\alpha$ sg cis-regulatory sequence promoter (Ca- $\Delta$ 2). Activity was monitored by  $\beta$ -galactosidase assays. Consistent with a synergistic interaction between Mcm1 and C. albicans Mat $\alpha$ 1, disruption of the Mcm1-binding site eliminated all C. albicans Mat $\alpha$ 1.

1. Jarvis EE, Clark KL, Sprague GF Jr. (1989) The yeast transcription activator PRTF, a homolog of the mammalian serum response factor, is encoded by the MCM1 gene. Genes Dev 3: 936–945.



Fig. 52. Species-specific binding to a second set of  $\alpha$ sg *cis*-regulatory sequences. In the reporter construct and gel shift experiments, the  $\alpha$ sg *cis*-regulatory sequences used were derived from the *C. albicans* (Ca) and *S. cerevisiae* (Sc)  $\alpha$ -mating pheromone *cis*-regulatory sequences. Thus, it was a concern that the results were dependent on this specific set of  $\alpha$ sg *cis*-regulatory sequences. To address this directly, electrophoretic mobility gel shift assays were performed on a second set of  $\alpha$ sg *cis*-regulatory sequences taken from the promoter sequences for the *C. albicans* and *S. cerevisiae* 4-factor receptor genes (*STE3*). The labeled oligonucleotide used in this experiment was either the *C. albicans STE3*  $\alpha$ sg *cis*-regulatory sequence (lanes 1 and 3) in both images. Extracts were prepared from either the *S. cerevisiae* MAT $\alpha$  1 (p415GAL). Galactose inducible copy of *C. albicans* erformed on swernight on samples in lanes 3 and 4 (lanes 1 and 2 are grown in glucose) in both images. Each lane contains 5 mg of protein from cell extracts.



Fig. S3. Functional conservation of Mat $\alpha$ 1 DNA-binding specificity between *S. cerevisiae* (SC) and *K. lactis* (KL). The  $\alpha$ sg *cis*-regulatory sequence of the promoter for the  $\alpha$ -mating pheromone gene from *K. lactis* was inserted into a basal promoter construct (pLG669z), and the *K. lactis* Mat $\alpha$ 1 was expressed from a 415-translation elongation factor promoter (TEF) plasmid. Plasmids were transformed into an *S. cerevisiae* MAT $\alpha \Delta mat\alpha$ 1 strain. Reporter activity was monitored using  $\beta$ -galactosidase assays. CA, *C. albicans*.



**Fig. 54.** Evolution of the Mat $\alpha$ 1 HMG DNA-binding domain. A multiple protein sequence alignment for the Mat $\alpha$ 1 HMG DNA-binding domain was computed using ClustalW2. (A) Quantification of Mat $\alpha$ 1 HMG DNA-binding domain divergence relative to the *S. cerevisiae* (*S. cer*) Mat $\alpha$ 1 sequence. Percent protein sequence similarity, identity, and gaps were calculated using BLAST2. (*B*) Quantification of Mat $\alpha$ 1 HMG DNA-binding domain divergence relative to the *S. cerevisiae* (*S. cer*) Mat $\alpha$ 1 sequence. Percent protein sequence similarity, identity, and gaps were calculated using BLAST2. (*B*) Quantification of Mat $\alpha$ 1 HMG DNA-binding domain divergence relative to the *C. albicans* (*C. alb*) Mat $\alpha$ 1 sequence. Note that the *C. lusitaniae* (*C. lus*) and *S. cerevisaie* Mat $\alpha$ 1 sequences have comparable similarity scores to *C. albicans* Mat $\alpha$ 1 despite their extensive difference in phylogenetic proximity to *C. albicans*. (*C) C. lusitaniae* (Clus) Mat $\alpha$ 1 sequence does not branch with the other CTG-clade species in a tree constructed using the Mat $\alpha$ 1 HMG DNA-binding domain. The tree was generated using the ClustalW2 alignment of the Mat $\alpha$ 1 HMG DNA-binding domain. The bootstrap values in support of this particular branching configuration are shown in red. Notably, in contrast to the ascomycete phylogenetic tree, the *C. lusitaniae* Mat $\alpha$ 1 branches outside the CTG-clade. This result is consistent with the divergent DNA-binding specificity of *C. lusitaniae* relative to *C. albicans*.



**Fig. 55.** *Y. lipolytica*  $\alpha$ sg *cis*-regulatory PSSM. In Fig. 4, neither the *S. cerevisiae* nor the *C. albicans* PSSMs identified close matches in the promoters of *Y. lipolytica*  $\alpha$ sgs. Using MEME, six sequences were located in the promoters of *Y. lipolytica*  $\alpha$ sgs. The PSSM built from these six sequences contains a Mcm1binding site; corroborating the results of Fig. 4, the sequence at the position of the Mat $\alpha$ 1-binding site does not resemble either the *C. albicans* or *S. cerevisiae* Mat $\alpha$ 1 site. Unfortunately, no genomes have been sequenced for species closely related to *Y. lipolytica*, which could provide further support for the possibility of yet another change in Mat $\alpha$ 1 DNA-binding specificity. Table S1. Sequence sets for *S. cerevisiae*, *K. lactis*, *C. albicans*, *Y. lipolytica*, and the filamentous fungi (sg cis-regulatory sequence PSSMs

S. cerevisiae clade sequence set 6323070\_Ylr040cp\_[S\_cere\_+\_234 6325069\_Mf(alpha)1p\_[S\_c\_-\_368 22168\_YPL187W\_[S\_paradox\_-\_313 6325069\_Mf(alpha)1p\_[S\_c\_-\_315 14612\_YLR040C\_[S\_mikatae\_+\_230 15773 YLR040C [S bayanus + 246 21565\_YPL187W\_[S\_mikatae\_-\_315 6321349\_Mf(alpha)2p\_[S\_c\_+\_498 21565\_YPL187W\_[S\_mikatae\_-\_366 13309\_YJR004C\_[S\_bayanus\_+\_415 6321349\_Mf(alpha)2p\_[S\_c\_-\_319 22168 YPL187W [S paradox - 365 12315\_YJR004C\_[S\_paradox\_+\_371 14144\_YKL178C\_[S\_paradox\_+\_345 13094\_YKL178C\_[S\_mikatae\_+\_330 12196\_YJR004C\_[S\_mikatae\_+\_381 6322463 Sag1p [S cerevis + 357 14076\_YKL178C\_[S\_bayanus\_+\_356 8057\_YGL089C\_[S\_bayanus]\_-\_330 14679\_YLR040C\_[S\_paradox\_+\_217 24677\_YPL187W\_[S\_bayanus\_-\_384 7908\_YGL089C\_[S\_mikatae]\_+\_496 6322671 Ste3p [S cerevis + 339 8663\_YGL089C\_[S\_paradoxu\_+\_502 24677\_YPL187W\_[S\_bayanus\_-\_333 C. albicans clade sequence set CTRG\_04755.3\_[C\_tropical\_-\_353 7.85 [C dubliniensis] - 534 CTRG\_04808.3\_[C\_tropical\_-\_257 CTRG\_00576.3\_[C\_tropical\_-\_335 orf19.4481\_[C\_albicans]\_-\_258 9.rc.176\_[C\_dubliniensis\_-\_234 9.rc.176 [C dubliniensis + 308 orf19.2492\_[C\_albicans]\_+\_340 7.85\_[C\_dubliniensis]\_+\_426 orf19.2492\_[C\_albicans]\_-\_248 7.85\_[C\_dubliniensis]\_-\_264 orf19.4481\_[C\_albicans]\_-\_504 K. lactis clade sequence set 00082.8\_+\_32434-33745\_[K\_+\_424 00034.rc.7\_-\_67410-67020\_-\_276 50306787\_[K\_lactis]\_-\_241 50309605\_[K\_lactis]\_-\_465 00082.8\_+\_32434-33745\_[K\_+\_336 00215.4\_+\_23082-25671\_[K\_-\_183 50302709\_[K\_lactis]\_+\_416 50311651\_[K\_lactis]\_+\_134 00201.23\_+\_24401-24854\_[\_+\_465 00648.rc.5\_-\_9278-8468\_[\_+\_133 00834.2\_+\_1798-2458\_[K\_w\_+\_123 00201.23\_+\_24401-24854\_[\_-\_302 00008.rc.4\_-\_23286-21879\_+\_252 00008.rc.4\_-\_23286-21879\_+\_368 Filamentous fungi sequence set CIMG\_06159.2\_[C\_immitis]\_+\_580 SS1G\_04155.1\_[S\_scleroti\_+\_272 CIMG\_06159.2\_[C\_immitis]\_+\_446 FG07270.1\_[F\_graminearum\_+\_273 UREG\_04230.1\_[U\_reesii]\_+\_463 UREG\_04230.1\_[U\_reesii]\_+\_596 AN7743.3 [A nidulans] + 517 FG07270.1\_[F\_graminearum\_-529

TTCATTGAACGCCCAAATAGGGAAAA TCTATTGAAGGACTAATTAGGAAAGT GTCGTTGATGGCCTAATTAGGAAAGT GTCGTTGATGGCCTAATTAGGAAGGT TTCATTGAAGGCCCAAATAGGGAAAA TCTATTGAACGCCCAAATAGGAAAAA GTCGTTGATGGCCTAATTAGGAAAGT CTCATTGATGTCCCTAAAAGGTAATG TCAATTGGAGGACTAAATAGGAAAGT ATCATTGATAACCTAATTAGGAAGGT GGCATTGATGTACCAATGAAGAAATA TTTATTGAAGGACTAAATAGGAAAGT GTCATTGATGACCCAATTAGGAAGGT GTCATTGTGACACTAATTAGGAAACT GTCATTGTGACACTAATTAGGAAACT TTCATTGATGACCCAATTAGGAAGTT GTCATTGATGACCTAATTAGGAAGGT GTCATTGTGACACTAATTAGGAAACT GGCATTGATGTACCGATTAGGAAATT TTCATTGAACGCCCAAATAGGGAAAA TTTATTGAAGGACTGAATAGGAAAGT ACCATTGATGTCCAAAATAGGTAATA GTCATTGTGACACTAATTAGGAAACT CCCATTGATGTCCAGAATAGGTAATG GTCGTTGATGGCCTAATTAGGAAAGT AATCCTTTCGCAGCCAAATCAGGAAA

TCCATTTAAA CCCTAATCAG GAAGTT TGCTTTGTTT GACAGATTAG GAAAGT ATCATTTGAA CCCAGATTAG GAAAGA AGCTTTACAT GACAGTTTAG GAAAGA ACCATTGAAG GACCAAATAG GAAAGT TGCATTTAAA CACTAAATAG GAAAAC TGCATTAAAG GCTTAATCCG GAAATC ATCATATATT GCCGAAAAAG GAAACC TCCTTTAATA CCCAGAAGAG GAAAGT TGCATTGCGT GCCTAAAACG GAAAGG ATTATTAGTC GCTCGAAAAG GAAACC TACCTTGATC GACAGATTAG GAAACT AACATTGATG TTCGAAAGGG GAAACA CTCATTAAAT CCTCGAACAG GAAATT

CCTTATTGAT ACCCAAATCG GGTTAG ATCTATTGAG ACCCTCAATA GGAAAA CGTTATTGAG GCCCTGTATA GGAAGC CCTTATTGAC AGGCTGAACT GGCCTA CGTTATTGAG GCCCTGTATA GGAATC CCTTATTGAT ACCCAAATCG GGTTAG ATCTATTGAC ACCCAAATCA GGCAAT GTCTATTGAC AACCCGATCG GGTAAG

### Table S1. Cont.

VAC DNAS

MGG_06452.5_[M_grisea]_–_600	CTCTATTCAG ACCCGAATAG GGAATT
Y. lipolytica sequence set	
49649895_[Y_lipolytica]_–_464	TCGATACTGT CCCAAATAGG TAATTT
49651178_[Y_lipolytica]_+_179	TGTATGGTGT CCAAAGGGGG CAATTC
49650459_[Y_lipolytica]_–_105	TGGAAGGGCT CCCATGGTGT TTATAT
49649895_[Y_lipolytica]_–_191	CGTATGGATT CTGCAAAGGG TAATTG
49651178_[Y_lipolytica]_+_230	TGTATAGTAT CCTAAAACAG TAATAT
49650459_[Y_lipolytica]_+_600	TGTTTGAATG CCCAAATAGG AAATAT

## d C albi

Table S2. Oligonucleotides and codon-changed C. al	bicans Matal sequence
<i>C. albicans</i> α-pheromone cis-regulatory sequence	tcgacAAAATTCCTTTCGCAGCCAAATTGGGTAACAAATAATg
C. albicans $\alpha$ -pheromone cis-regulatory sequence (RC)	tcgacATTATTTGTTACCCAATTTGGCTGCGAAAGGAATTTTg
C. albicans Matα1 codon-changed	ATGGGAAATGAGAAAAAGACAAGAAAAACTGTTCCAAAAGAGTTT
	ATCTCATTGTTTAGAGTACATTCAGGGAGAGATGCTCCTAGACG
	TGATACTAGGGAAGTTCAAAAGAGTAAGAAGCACGGCTTCCGTTTT
	ACTAGTTTGCCTGATTTGCCAGTCGCCAGTAATGCTTTACAAGAGTTATTGTT
	AGAGTATGGATTACTGAACGATATCAAATGGGATTCCAAGGG
	CTTAAAGGGTTCCAAAAACAAAAAGACAAAGTTGAAGCCAATCAA
	CTCTTTCATAGCATTCAGATCCTTCTATTCAAGAACTATTTCTAA
	CCCAGAACATCAGAGAGAATTGTCATCTAAGCTAGCCGATGTAT
	GGACACAAGAATCAAATCAGGAAGTGTGGAAACAATACACAA
	TCTTACAATAACTACCTACTTTTACCTGATGCTAAACTAAACTTCG
	ΤΤGΑCTGGCTGTGTGAAGCATTGGACTACACTATTGACAACAC
	ΤΑΓΑΓΓΑΔΑΤΑΓΑΔΟΓΑΤΟΤΟΛΟΤΑΛΟΤΟΛΟΛΟΛΟΛΟΛΟΛΟΛΟΛΟΛΟΛΟΛΟΛΟΛΟΛ
	CTTTCTGGTACCATCGAAGATGTCTACATTATGAAGTAA
C albicans Mater BamH1 3'	GGGTTTGGATCCTTACTTCATAATGTAGACATCTTCGATGGTACCAG
C albicans Mater Damin' 5	
C. albicans where more A Materia site	
C. albicans a pheromone cis regulatory sequence (PC)	
C. albicaris a-pheromone AMata1 site	
S. corovisiae $\alpha$ -pheromone $\Delta Mata1$ site	
SUCA	
SC/Ca (RC)	
Ca/Sc	
Ca/Sc	
S. cerevisiae Mata1 Xba1 5'	AGAGAGTCTAGAATGTTTACTTCGAAGCCTGCTTTC
<i>S. cerevisiae</i> Matα1 BamH1 3′	AGAGAGTCTAGAGGGTTTGGATCCCTATAAGGCCAAATGTACAAACACATC
U. reesii Ste3 cis-regulatory sequence	ACGCTATATACACGCCTGGCGGATCTGgtctaacccgatttgggtatcaataaggagCCTAAAAA TTTTTTCCTTTCCTGCTTT
U. reesii Ste3 cis-regulatory sequence (RC)	AAAGCAGGAAAGGAAAAAATTTTTAGGctccttattgatacccaaatcgggttagacCAGATCCG
C albicans « phoromono AMcm1 sito	
C. albicans $\alpha$ -pheromone AMcm1 site (PC)	
C. albicans $\alpha$ -pheromone $\Delta$ mentri site (RC)	
C. albicans ste3 cis-regulatory sequence	
C. albicaris step cis-regulatory sequence (RC)	
S. cerevisiae ste3 cis-regulatory sequence (RC)	
S. cerevisiae sted cis-regulatory sequence	GCIGGIACAATTTCICIGICATIGIGACACTAATAGGAAACTICICGACTATCAATGI
C. Iusitaniae $\alpha$ -pheromone cis-regulatory site	
C. Iusitaniae $\alpha$ -pheromone cis-regulatory site (RC)	
K. lactis $\alpha$ -pheromone cis-regulatory site	tcgacAATTAACTTICCTAATTAGTCCTTAATGAGAACATg
K. lactis $\alpha$ -pheromone cis-regulatory site (RC)	tcgacAIGIICICAIIAAGGACIAAIIAGGAAAGTTAATTg
C. Iusitaniae Mata1 Xba1 5'	
C. Iusitaniae Mata1 Bamh1 3'	GGGTTTGGATCCTTAGTCAACAAAGACATCCTCCACATTCTT
<i>K. lactis</i> Matα1 Sal1 5′	CCC TTT GTC GAC TCA GAC TGA GTT CAT CAA GGA ATA GC
<i>K. lactis</i> Matα1 Bamh1 3′	AGA GAG GGA TCC ATG AAA TCG AAT GCT CCA ACC TTT AAA GTG

Ca, C. albicans; RC, reverse complement; Sc, S. cerevisiae.

#### Table S3. S. cerevisiae strains constructed

PNAS PNAS

Fig. 2	
a Lane 1	W303 MATa p415TEF pLG669z-Cyc1-∆UAS
a Lane 2	W303 MATa p415TEF pLG669z-Cyc1-∆UAS-Ca
a Lane 3	W303 MATa p415TEF pLG669z-Cyc1-∆UAS-Ca∆
a Lane 4	W303 MATa p415TEF-codon-changed-caMatα1 pLG669z-Cyc1-ΔUAS-Ca
a Lane 5	W303 MATa p415TEF-codon-changed-CaMatα1 pLG669z-Cyc1-ΔUAS-CaΔ
<b>b</b> Lane 1	W303 MATa p415GAL-codon-changed-caMatα1
<b>b</b> Lane 2	W303 MATa p415GAL-codon-changed-caMatα1
<b>b</b> Lane 3	W303 MATa p415GAL-codon-changed-caMatα1
<b>b</b> Lane 4	W303 MATa p415GAL-codon-changed-caMatα1
Fig. 3	
a Lane 1	EG123 MATα ΔMatα1 p415TEF pLG669z-Cyc1-ΔUAS-Sc
a Lane 2	EG123 MATα ΔMatα1 p415TEF-scMatα1 pLG669z-Cyc1-ΔUAS-Sc
a Lane 3	EG123 MATα ΔMatα1 p415TEF-codon-changed caMatα1 pLG669z-Cyc1-ΔUAS-Sc
a Lane 4	EG123 MATα ΔMatα1 p415TEF pLG669z-Cyc1-ΔUAS-Ca
a Lane 5	EG123 MATα ΔMatα1 p415TEF-scMatα1 pLG669z-Cyc1-ΔUAS-Ca
a Lane 6	EG123 MATα ΔMatα1 p415TEF-codon-changed caMatα1 pLG669z-Cyc1-ΔUAS-Ca
<b>b</b> Lane 1	W303 MATa p415GAL-scMata1
<b>b</b> Lane 2	W303 MATa p415GAL-scMata1
<b>b</b> Lane 3	W303 MATa p415GAL-codon-changed-caMatα1
<b>b</b> Lane 4	W303 MATa p415GAL-codon-changed-caMatα1
<b>b</b> Lane 5	W303 MATa p415GAL-scMata1
<b>b</b> Lane 6	W303 MATa p415GAL-codon-changed-caMatα1
c Lane 1	W303 MATα p415TEF pLG669z-Cyc1-ΔUAS-Sc/Ca
c Lane 2	W303 MATa p415TEF-codon-changed caMatα1 pLG669z-Cyc1-ΔUAS-Sc/Ca
c Lane 3	W303 MATα p415TEF pLG669z-Cyc1-ΔUAS-Ca/Sc
<b>c</b> Lane 4	W303 MATa p415TEF-codon-changed caMatα1 pLG669z-Cyc1-ΔUAS-Ca/Sc
Fig. 4	
c Lane 1	EG123 MATα ΔMatα1 p415TEF pLG669z-Cyc1-ΔUAS-Kl
<b>c</b> Lane 2	EG123 MATα ΔMatα1 p415TEF-klMatα1 pLG669z-Cyc1-ΔUAS-Kl
<b>c</b> Lane 3	EG123 MATα ΔMatα1 p415TEF-klMatα1 pLG669z-Cyc1-ΔUAS-Sc
<b>c</b> Lane 4	EG123 MATα ΔMatα1 p415TEF-klMatα1 pLG669z-Cyc1-ΔUAS-Ca
c Lane 5	EG123 MATα ΔMatα1 p415TEF-scMatα1 pLG669z-Cyc1-ΔUAS-KI
<b>c</b> Lane 6	EG123 MATα ΔMatα1 p415TEF-codon-changed-caMatα1 pLG669z-Cyc1-ΔUAS-Kl
e Lane 1	W303 MATα p415TEF pLG669z-Cyc1-ΔUAS-FF
e Lane 2	W303 MAT $\alpha$ p415TEF pLG669z-Cyc1- $\Delta$ UAS-FF $\Delta$
e Lane 3	W303 MATa p415TEF-codon-changed caMatα1 pLG669z-Cyc1-ΔUAS-FF
e Lane 4	W303 MATa p415TEF-codon-changed caMata1 pLG669z-Cyc1- $\Delta$ UAS-FF $\Delta$
Fig. 5	
b Lane 1	EG123 MATα ΔMatα1 p415TEF pLG669z-Cyc1-ΔUAS-Cl
<b>b</b> Lane 2	EG123 MATα ΔMatα1 p415TEF-ClMatα1 pLG669z-Cyc1-ΔUAS-Cl
<b>b</b> Lane 3	EG123 MATα ΔMatα1 p415TEF-ClMatα1 pLG669z-Cvc1-ΔUAS-Ca
<b>b</b> Lane 4	EG123 MATα ΔMatα1 p415TEF-klMatα1 pLG669z-Cyc1-ΔUAS-Sc
<b>b</b> Lane 5	EG123 MAT $\alpha \Delta$ Mat $\alpha$ 1 p415TEF- codon-changed-caMat $\alpha$ 1 pLG669z-Cyc1- $\Delta$ UAS-Cl
<b>b</b> Lane 6	EG123 MAT $\alpha \Delta$ Mat $\alpha$ 1 p415TEF-scMat $\alpha$ 1 pLG669z-Cyc1- $\Delta$ UAS-Cl
Fia. S1	
Lane 1	W303 MATa p415TEF-codon-changed-caMata1 pLG669z-Cyc1-ΔUAS
Lane 2	W303 MATa p415TEF-codon-changed-caMatα1 pLG669z-Cyc1-ΔUAS-Ca
Lane 3	W303 MATa p415TEF-codon-changed-caMat $\alpha$ 1 pLG669z-Cyc1- $\Delta$ UAS-Ca $\Delta$ 2
	(Mcm1 site mutant)
Fia. S2	
Lane 1	W303 MATα p415GAL-scMatα1
Lane 2	W303 MATa p415GAL-codon-changed-caMatα1
Lane 3	W303 MATa p415GAL-scMata1
Lane 4	W303 MATa p415GAL-codon-changed-caMatα1
Lane 5	W303 MAT $\alpha$ p415GAL-scMat $\alpha$ 1
Lane 6	W303 MATa p415GAL-codon-changed-caMatα1
Lane 7	W303 MATa p415GAL-scMata1
Lane 8	W303 MATa p415GAL-codon-changed-caMatα1
	,

Ca, C. albicans; Ca∆2; Sc, S. cerevisiae.