science.sciencemag.org/content/367/6473/96/suppl/DC1

Supplementary Materials for

Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit

Candace S. Britton, Trevor R. Sorrells, Alexander D. Johnson†

†Corresponding author. Email: ajohnson@cgl.ucsf.edu

Published 3 January 2020, *Science* **367**, 96 (2020)
DOI: 10.1126/science.aax5217

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at science.sciencemag.org/content/367/6473/96/suppl/DC1)

Data S1 (.xlsx)

Data S1
Data S1 is an excel file that comprises information on the genotypes of yeast strains, plasmids, and oligonucleotide sequences that were used in this study.

Materials and Methods

Growth conditions and media

All strains were grown on yeast extract peptone dextrose media (YEPD) at 30°C unless otherwise noted.

Identification of the Phaffomycetaceae as candidate species

The Phaffomycetaceae form a monophyletic group that branches outside of the last common ancestor of the Saccharomycetaceae (the group that contains the *Saccharomyces* and *Kluyveromyces* species, also known as the "*S. cerevisiae* clade"), as well as the Saccharomycodaceae (an outgroup to the Saccharomycetaceae that includes *Hanseniaspora valbyensis* and other *Hanseniaspora* species). This clade consists of four species that had sequenced genomes at the time this study was conducted: *W. anomalus*, *W. ciferrii*, *C. jadinii*, and *C. fabianii* (*12*, *13*). Multiple phylogenomic analyses have placed the Phaffomycetaceae as branching outside of the Saccharomycetaceae and the Saccharomycodaceae with high confidence (*12*, *13*). Other Phaffomycetaceae species lacked sequenced genomes during the time this study was completed and were therefore excluded from this study. In this study we use the terms "*W. anomalus* clade" and Phaffomycetaceae interchangeably.

Identification of Phaffomycetaceae mating type transcriptional regulators

The genome sequenced *W. anomalus* strain (NRRL Y-366-8) is an **a** cell. The **a**-mating type locus (*MAT**a***), which encodes the **a**-mating type transcriptional regulators *MAT***a***1* and *MAT***a***2*, is located between the *DIC1* and *SLA2* genes, which are syntenic with the mating type locus in many Saccharomycotina species (*13*, *25*). Unlike *S. cerevisiae*, *W. anomalus* is not known to undergo mating-type switching, and there is no evidence for a silenced mating-type locus or HO endonuclease in its genome (*14*, *25*). To determine the sequence of the α-mating type locus (*MATα*), we designed oligonucleotides to PCR amplify the region between the *DIC1* and *SLA2* coding sequences of NRRL Y-2153-4, another *W. anomalus* isolate that has been reported to mate with the sequenced **a** cell. This sequencing revealed the genes for the α-mating type transcriptional regulators *MATα1* and *MATα2*. All of the *W. anomalus* mating transcriptional regulator genes except for *MATα1* contain introns, so we amplified and sequenced *MATα2*, *MAT***a***1* and *MAT***a***2* from cDNA to determine their exon coding sequences. Orthology of the *MAT* genes was further confirmed by best reciprocal BLAST to other Saccharomycotina sequences.

The genome sequenced strain of *W. ciferrii* (NRRL Y-1031) is also an **a** cell, so we used the same approach to determine the sequence of the *W. ciferrii MATα* locus. We amplified and sequenced the region between *W. ciferrii DIC1* and *SLA2* in the complementary mating types NRRL# Y-1031-11 and NRRL# Y-1031-27, and found them to be an **a** cell and an α cell, respectively. The *C. fabianii* genome sequenced strain, NRRL Y-1871, is an α cell, so we were able to directly determine the sequence of *MATα* from the existing sequenced genome. The sequenced *C. jadinii* isolate is homothallic and tetraploid, and the genome therefore encodes both

*MAT**a*** and *MATα*. All four Phaffomycetaceae *MATα2* genes contain introns, so for each gene we sequenced the cDNA transcript in order to determine its exonic coding sequence.

Identification of cell-type specific genes

Phaffomycetaceae haploid, **a**, and *α*-specific genes were identified by best reciprocal TBLASTN to other Saccharomycotina cell-type specific genes, and with the Yeast Genome Analysis Pipeline (YGAP) (*26*, *27*). The expression patterns of the cell-type specific genes in *W. anomalus* were confirmed by measuring transcript abundance using the NanoString nCounter system (method described below) (Fig. S2).

Strain construction in *W. anomalus*

The genome-sequenced strain of *W. anomalus*, NRRL Y-366-8, is also the *W. anomalus* type strain and an **a** cell. To generate an **a**/α strain of *W. anomalus*, NRRL Y-366-8 was plated on media containing 5-FOA to select for spontaneous uracil auxotrophs. An **a** cell auxotrophic for uracil was then crossed to NRRL Y-2153-4, an α cell with a naturally occurring arginine auxotrophy. The resulting **a**/α cell (yCSB 8) was selected for by its ability to grow on synthetic defined media with both arginine and uracil dropped out. The 2N diploid ploidy of yCSB 8 was confirmed by staining with SYTO™ 13 Green Fluorescent Nucleic Acid Stain followed by flow cytometry, and the presence of both *MAT**a*** and *MATα* mating type loci in yCSB 8 was confirmed by PCR.

*W. anomalus MAT**a*** gene deletions were made in NRRL Y-366-8, and *MAT***a**/α deletions were made in yCSB 8. We were unable to efficiently transform the NRRL Y-2153-4 α cell strain. In order to generate a more genetically tractable α cell, we sporulated yCSB 8 by growing it on a sterilized carrot slice at room temperature for at least five days (*14*). Carrots were peeled, cut into cylindrical wedges, then placed in glass tubes with 1 ml water and autoclaved at 115°C for 15 minutes. 500 µl of cells from liquid culture were collected, washed in water, then resuspended in 200 µl water, layered on top of a sterilized carrot wedge in a six well plate, and placed at room temperature. Spores were visible after five days, and sporulation can be started either from saturated overnight cultures, or mid-log cultures. Spores were then resuspended in 100 µl of 1 mg/ml zymolyase for 11 minutes at 37°C degrees, then dissected. A resulting α cell that proved to be genetically tractable (yCSB 342) was the parent strain for generating the α cell *MATα2* deletions. In addition, the dissected spores included **a** and α cells with arginine auxotrophies that were used as tester strains in quantitative mating assays (described below).

W. anomalus genes were deleted using a nourseothricin-resistance marker and a freeze-thaw transformation protocol optimized for use in *W. ciferrii* by Schorsch *et al*. (*28*). To construct deletion cassettes, we cloned homology arms into pCS.ΔLig4, a plasmid generated by Schorsch *et al*. (*28*). pCS.ΔLig4 contains *nat*, the nourseothricin-resistance gene, codon-optimized for expression in *W. ciferrii*, and under control of *W. ciferrii* P_{PDA1} and T_{TEF1}. It also contains a region of the *W. ciferrii LIG4* gene flanked by loxP sites, for inserting pCS.ΔLig4 into *LIG4* by ends-in recombination. To use the *W. ciferrii*-optimized *nat* marker to delete genes in *W. anomalus* we excised the *LIG4* fragment from pCS.ΔLig4 and ligated in 5' homology regions in between the SexAI and SalI restriction sites, and 3' homology regions between the SacI and SpeI restriction sites. Homology regions ranged from approximately 700 basepairs to 2000 basepairs in length. Transformed cells were selected on 50, 200, and 400 µg/ml nourseothricin for NRRL Y-366-8, yCSB 342, and yCSB 8, respectively. Deletions were screened and confirmed by PCR testing for the absence of the deleted ORF, as well as the presence of the junctions between the deletion

cassette and the surrounding genome. Approximately 1/500 transformed colonies contained gene deletions, because *W. anomalus*, like many other fungi, seems to favor taking up DNA by non-homologous end joining rather than homologous recombination (*28*).

pCS.ΔLig4 also served as the backbone for *MATα2* add-in experiments. A full-length wildtype *MATα2* gene was PCR amplified and inserted into pCS.ΔLig4 between the SexAI and SalI restriction sites, creating pCSB 147. This plasmid was subsequently modified to produce the different alleles of *MATα2* tested in Fig. S6B by digesting at either the PstI/XbaI restriction sites or the StyI/BsrGI restriction sites within the *MATα2* coding sequence and ligating in gBlock gene fragments (Integrated DNA Technologies) with specific mutations. These constructs were then linearized by PvuI digestion, transformed into yCSB 5, and allowed to randomly integrate into the genome. At least two independent integrants of each allele was assayed in any experiment involving a randomly integrated construct in *W. anomalus* in order to control for variability due to integration at different loci.

pCSB 147 was modified for chromatin immunoprecipitation of Matα2 by using a gBlock to introduce a 3x HA tag at the N-terminus of *MATα2* at the PstI/XbaI restriction sites. Similarly, to tag *MAT***a***1* we inserted a full length *MAT***a***1* construct with its promoter and terminator into pCS.ΔLig4 at the SexAI/SalI restriction sites. We then introduced a 3x HA tag at the N-terminus of *MAT***a***1* by inserting a gBlock between the XbaI/NcoI restriction sites. Tagged *HA-MATα2* and *HA-MAT***a***1* constructs were inserted into *W. anomalus* by random integration.

We also used pCS.ΔLig4 as the basis for a hygromycin-resistance marker optimized for expression in *W. anomalus*. *W. ciferrii* P_{PDA1} and T_{TEF1} were amplified from pCS.ΔLig4 and cloned into pUC19 at the BamHI/XbaI and PstI/SphI restriction sites, respectively, to create pCSB 137. The coding sequence of hph (a hygromycin-resistance gene) was then codon-optimized for expression in *W. anomalus* using a codon usage table generated by the Kazusa DNA Research Institute (http://www.kazusa.or.jp/codon/), synthesized as a gBlock (IDT), and cloned into pCSB 137 at the XbaI/PstI restriction sites to create pCSB 182. pCSB 182 was used as the backbone for all P_{RME1}-GFP experiments. First, a GFP coding sequence was also codon-optimized for expression in *W. anomalus*, synthesized as a gBlock with the *W. anomalus ACT1* terminator, and cloned into pCSB 182 at the BamHI/SalI restriction sites. 944 basepairs of the *RME1* regulatory region (immediately upstream of the beginning of the *RME1* ORF) were cloned upstream of GFP by Gibson assembly. Modifications of P_{RME1} were subsequently made by inserting gBlocks between the DraIII/SpeI restriction sites that naturally occurred within P_{RME1}. All P_{RME1}-GFP constructs were linearized by digestion with SphI and allowed to randomly integrate into the genome of a *W. anomalus* **a** cell (NRRL Y-366-8). Transformants were selected on YEPD plates with 300 μg/ml hygromycin.

Strain construction in *S. cerevisiae*

All *S. cerevisiae* experiments were performed in the W303 strain, and transformations were performed by a published lithium acetate/single-stranded carrier DNA/polyethylene glycol method (*29*). *MATα2* was deleted from the W303 α cell using a KanMX marker amplified with homology to *MATα2*.

MATα2 expression constructs for **a**-specific gene transcriptional repression assays were cloned into pNH 604, a plasmid that integrates into the *S. cerevisiae* genome in single copy at the *trp1* locus using a *TRP1* marker (*11*, *30*). pNH 604 was digested with SacII and KpnI to excise the Tef overexpression promoter. *MATα2* constructs were amplified from cDNA or synthesized as gBlock synthetic gene fragments (Integrated DNA Technologies), and fused to the *S. cerevisiae*

*MATA2*promoter by fusion PCR, then inserted into pNH 604 between the SacII/KpnI restriction sites. pNH 604 were linearized by PmeI digestion, transformed into the *MATα matα2-Δ* strain, and selected for on SD-trp plates.

The P_{CYC1}-GFP reporter was designed to integrate in single copy at the *ura3* locus using the hph hygromycin resistance marker. It was adapted from the GFP reporter pTS 61 and the beta-galactosidase reporter pLG699z (*31*, *32*). P_{CYC1} was inserted upstream of GFP, with the addition of a KpnI restriction site, allowing *cis*-regulatory elements to be inserted between the XhoI and KpnI sites in P_{CYC1}, between the upstream activating sequence and transcription start site of the promoter. The hph marker is up stream of P_{CYC1}-GFP, and the two together are flanked by homology to *URA3*. Overlapping oligos with the **a**-specific gene *cis*-regulatory elements and phosphorylated sticky ends were designed, annealed, and ligated into the XhoI/KpnI sites.

Design of *MATα2* mutant alleles

To construct chimeric alleles of *MATα2*, we defined the five regions of the Matα2 protein based on the *S. cerevisiae* sequence as previously described (region 1: amino acid 1-21, region 2: amino acid 22-108, region 3: amino acid 109-127, region 4: amino acid 128-188, region 5: amino acid 189-210) (*11*). To introduce a loss of function mutation into region 5 of *W. anomalus MATα2*, we introduced a L215A point mutation (*33*)

Gene expression profiling by mRNA-seq

For gene expression profiling by RNA-seq, single colonies were grown to saturation overnight in YEPD, diluted to OD_{600} = 0.1 in fresh YEPD, and then grown for 4-4.5 hours at 30°C. Cells were pelleted by centrifugation at 3000 rpm for 10 minutes, then flash frozen in liquid nitrogen and stored at -80°C. RNA was extracted, purified, and DNase-treated using an Ambion RiboPure RNA Purification kit for yeast. Quality of total RNA was evaluated on an Agilent Bioanalyzer using an Agilent RNA 6000 Pico Kit. Two rounds of polyA selection were performed using the Qiagen Oligotex mRNA Mini Kit, and mRNA quality was evaluated on an Agilent Bioanalyzer using an Agilent RNA 6000 Pico Kit. mRNA samples were then concentrated using a Zymo RNA Clean and Concentrator Kit. cDNA synthesis and library preparation were performed using a NEBNext Directional RNA-seq kit. Libraries were sequenced using single end 50 basepair reads on the Illumina HiSeq 4000 sequencing system in the UCSF Center for Advanced Technologies. Reads were pseudoaligned to all *W. anomalus* coding sequences using kallisto, and differential expression analyses were performed using sleuth (*34*, *35*). Coding sequences were downloaded from the Joint Genome Institute annotation of genes for *Wickerhamomyces anomalus* NRRL Y-366-8 v1.0 (*13*). For differential expression analyses, we focused on the expression patterns of candidate haploid, **a**, and α-specific genes, except in the case of *matα2-Δ* in α cells, where we used a false discovery rate cut off of $q < 0.05$ to search for differentially expressed genes. mRNA-seq data shas been deposited at the NCBI Gene Expression Omnibus under accession number GSE133191.

Gene expression profiling with NanoString nCounter

To measure transcript abundance with the NanoString nCounter system (www.nanostring.com), single colonies were grown to saturation overnight in YEPD, diluted to OD_{600} = 0.1 in fresh YEPD, and then grown for 4-4.5 hours at 30°C. Cells were pelleted by centrifugation for 10 minutes at 3000 rpm, then flash frozen in liquid nitrogen and stored at -80°. RNA was extracted, purified, and DNase-treated using an Ambion RiboPure RNA Purification kit

for yeast. NanoString quantification was performed by the UCSF Center for Advanced Technologies using a set of probes designed by NanoString to target the *W. anomalus* mating-type transcriptional regulators, haploid-specific genes, **a**-specific genes, α-specific genes, and a set of non-differentially expressed reference genes (*GAPDH, TBP1, SPC98 and MRPL49*). The NanoString probe for the **a**-mating pheromone gene (*MFA*) targets at least seven paralogs of this gene in *W. anomalus*, because their transcripts are too small and similar in sequence to distinguish from each other by this method. Transcript counts were normalized across samples to the quantities of the non-differentially expressed reference genes.

GFP reporter assays

Single colonies were grown to saturation overnight in YEPD, diluted to OD_{600} = 0.1 in SD supplemented with amino acids and uracil, and then grown for 4-4.5 hours at 30°C. GFP fluorescence was measured on a BD LSR II flow cytometer. 10,000 cells were measured per sample in each experiment. Cells were gated to exclude debris, and the mean GFP fluorescence was calculated for each reporter strain sample. Three independent isolates of each reporter strain were measured.

Quantitative mating assays

Quantitative mating assays were performed based on methods described for other yeast species (*36*). Strains of interest (for example, gene deletions and their parent strains) were plated on 5-FOA to select for uracil auxotrophs (as described above), and tester strains of the opposite mating type with arginine were isolated from sporulating the **a**/α strain yCSB 8. Single colonies were grown to saturation overnight in YEPD, diluted to OD_{600} = 0.15 in fresh YEPD, and then grown to mid-log phase and combined in a final volume of 5 ml. 1 ml (about $1x10^7$ cells) of the mating type present in excess was combined with 200 µl of the strain to be measured, for a final ratio of 5:1 strain in excess to limiting strain, by volume. The mating mixes were then deposited onto 0.8 µm nitrocellulose filters using a Millipore 1225 Vacuum Sampling Manifold. The filters were then placed on a YEPD agar plate and incubated for 5 hours at 30°C. The cells were resuspended by placing the filters in 5 ml water, and cells dispersed by vortexing. Dilutions were plated on SD-arg-ura plates to select for conjugants, and SD-ura to select for the limiting parent as well as to conjugants. After 2-3 days of growth at 30°C, colonies were counted, and the mating efficiency was calculated by the following formula: Mating efficiency = (mating products) / (mating products + limiting tested parents) x 100%. In other words, Mating efficiency = (number of colonies on SD-arg-ura)/(number of colonies on SD-ura) x 100%.

Cis-regulatory motif discovery and searching

A *cis*-regulatory motif for the Phaffomycetaceae **a**-specific genes was generated using MEME (*37*). We used sequences 500 basepairs upstream of the **a**-specific gene orthologs in *W. anomalus* and *W. ciferrii* and used MEME to search for motifs between 24 and 27 bases wide occurring zero or one time in those sequences. These sequences were from the *W. ciferrii STE2*, *STE6*, and *AXL1* genes, and the *W. anomalus STE2*, *STE6*, *ASG7*, *AXL1*, *BAR1*, and *STE14* genes. This returned a motif that resembled a Mat**a**2-Mcm1 binding site. We used the TOMTOM tool in the MEME suite to align and compare this *W. anomalus*/*W. ciferrii* **a**-specific gene motif to both the *S. cerevisiae* Matα2-Mcm1 motif and the *C. albicans* Mat**a**2-Mcm1 motif (*38*). We searched for Matα2, Mat**a**1, Mat**a**2, and Mcm1 *cis*-regulatory motifs using the MAST and FIMO tools in the MEME suite.

RT-qPCR

Single colonies were grown to saturation overnight in YEPD, diluted to OD_{600} = 0.1 in fresh YEPD, then grown for 4-4.5 hours at 30°C. Cells were pelleted by centrifugation for 10 minutes at 3000 rpm, then flash frozen in liquid nitrogen and stored at -80°. RNA was purified using the MasterPure™ Yeast RNA Purification Kit from Epicentre. cDNA was synthesized with random hexamer primers using SuperScript® IV reverse transcriptase. qPCR was performed using BioRad SYBR Green Master Mix on a BioRad CFX Connect™ Real-Time PCR Detection System. Each sample was quantified in three technical replicates, and each experiment was performed with two biological replicates per sample. Relative quantifications were determined using a standard curve, and quantifications for *RME1* and *GFP* were normalized to the non-differentially regulated housekeeping gene *TBP1*.

Chromatin-immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed as previously described (*32*, *39*). Briefly, single colonies were grown to saturation overnight in YEPD, diluted to OD_{600} = 0.1 in fresh YEPD, then grown at 30°C to OD_{600} = 0.4. Cells were then crosslinked with formaldehyde, quenched with glycine, washed with TBS (20mMTris/HCl pH 7.4, 150 mM NaCl), flash frozen in liquid nitrogen, and stored at -80°C. Cells were lysed at 4°C in lysis buffer (50 mM HEPES/KOH pH 7.5, 140 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate) plus protease inhibitors (Roche Complete Protease Inhibitor Cocktail EDTA-free) with 0.5 mm glass beads on a vortex for two hours. Recovered chromatin was then sheared via sonication on a Diagenode Biorupter (three ten minute periods of sonication on level 5 for 30 seconds, then off for 1 minute). 10 mg of anti-HA 12CA5 from Roche was used to bind the HA-tagged transcription factors overnight at 4°C in lysis buffer plus protease inhibitors, and protein-G sepharose beads were used to collect immunoprecipitated DNA. Immunoprecipitated DNA was then recovered by washing with lysis buffer, wash buffer (10mM Tris/HCl pH 8.0, 250mM LiCl, 0.5% NP-40, 0.5% Na-Deoxycholate, 1mM EDTA), and TE, then collected in elution buffer (50mM Tris/HCl pH8.0, 10mM EDTA, 1%SDS). Crosslinking was reversed by Proteinase K treatment for two hours at 37°C, then16 hours at 65°C. DNA was then cleaned up using the Qiagen MinElute kit.

*HA-MAT***a***1* and *HA-MATα2* constructs were inserted into the *W. anomalus* genome via random integration. Two to four independent isolates of each strain were processed, and two samples of untagged **a**/α cells, and one each of untagged **a** and untagged α cells were also processed as negative controls. Abundance of the *RME1* promoter region in tagged vs. untagged samples was measured by qPCR and compared to the promoter regions of the housekeeping genes *TBP1*, *ACT1*, and *RPS20*, and the **a**-specific gene *STE2*.

Supplementary Text

Yeast cell-type specific gene expression regulation and regulatory evolution

The expression of the cell-type specific genes in fungi provides a powerful system to mechanistically study the evolution of gene expression patterns across a period of several hundred million years (*8-11, 20-23*). Saccharomycotina yeast have two mating types termed **a** and α cells that are genetically identical except for a short DNA segment called the mating-type locus (*MAT*) (*22*). This segment codes for transcriptional regulators that ensure the proper expression of the **a**, α, and haploid-specific genes. Thus, **a** cells express the **a**-specific and haploid-specific genes (but not the α-specific genes), and α cells express the α-specific and haploid specific (but not the **a**-specific genes) (Fig. 1A). When an **a** cell successfully mates with an α cell, an **a**/α cell is formed. This cell type does not express the **a**, α, or haploid-specific genes but is competent to undergo meiosis. The proper regulation of these three sets of genes, which is directed by the regulatory proteins coded at the mating type locus, is thus essential not only for proper mating, but also for meiosis, which completes the fungal sexual life cycle.

The expression patterns of these three sets of genes are broadly conserved across extant yeasts (as are the genes themselves) and therefore must have been present in the last common ancestor of these species, some hundreds of millions of years ago. However, extant species accomplish the regulation of the cell-type specific genes using different mechanisms (Fig. 1B) (*10, 11, 23*). In particular, in one group of species (including *Candida albicans*), the **a**-specific genes are turned on by a transcriptional activator protein (Mat**a**2) that is made only in **a** cells and which binds a *cis*-regulatory sequence upstream of the **a**-specific genes (*10, 11, 23*). In another group of species (including *Saccharomyces cerevisiae*) the **a**-specific genes are turned on by a ubiquitous activator and turned off by binding of a repressor (Matα2) that is made only in α and **a**/α cells.

Both schemes ensure that the **a**-specific genes are expressed only in **a** cells, although the mechanism behind this pattern is completely different. Previous work established that the activation scheme (activating **a**-specific genes by Mat**a**2) is ancestral to the repression scheme (repression of the **a**-specific genes by Matα2) (Fig. 1B) (*10, 11, 23*). Species like *C. albicans* preserve the ancestral form of regulation in which Mat**a**2 activates the **a**-specific genes, Matα2 is not capable of repressing the **a**-specific genes, and the **a**-specific genes do not possess Matα2 binding sites. In species like *S. cerevisiae* that are under the more recently evolved form of regulation the **a**-specific gene promoters have Matα2-binding sites, Matα2 is capable of repressing the **a**-specific genes (due to the presence of protein-protein interactions with Tup1 and Mcm1), and Matα2 does repress the **a**-specific genes (*10, 11, 23*).

W. anomalus Matα2 has functional Mcm1 and Tup1-interaction regions

Given the similarity of the *W. anomalus* Matα2 protein sequence to that of *S. cerevisiae* (Fig. S1A-B), we hypothesized that the *W. anomalus* protein represented the derived form of Matα2. To test this hypothesis, we expressed the *W. anomalus* and, as a control, the *S. cerevisiae* *MATα2* coding sequences in a *S. cerevisiae* α cell that had the *MATα2* gene deleted and contained a P_{CYC1}-GFP reporter with an **a**-specific gene *cis*-regulatory sequence taken from the promoter of the *S. cerevisiae* **a**-specific gene *STE2*. As expected, the *S. cerevisiae* Matα2 represses transcription from this promoter by more than 99 percent. In contrast to our expectations, *W. anomalus* Matα2 failed to repress the reporter (Fig. S1C). By constructing a series of chimeric proteins, we mapped the "deficiency" of the *W. anomalus* Matα2 to its homeodomain, indicating that although the protein is in large part conserved, its homeodomain simply did not recognize the *S. cerevisiae* **a**-specific gene *cis*-regulatory sequence (Fig. 2B). To test this idea, we swapped the

S. cerevisiae homeodomain into *W. anomalus* Matα2 and we found that it could now carry out repression of the reporter construct, although not as well as the bona fide *S. cerevisiae* protein (Fig. S1D). *W. ciferrii* and *C. jadinii* Matα2 regions 1 and 2 also resemble the *S. cerevisiae* sequence, and their full-length sequences and the *C. jadinii* homeodomain also do not support **a**-specific gene repression (Fig. S1D).

W. anomalus **a**-specific genes are not directly regulated by Matα2

To determine whether the *W. anomalus* **a**-specific genes are repressed by Matα2, activated Mat**a**2, or both repressed by Matα2 and activated by Mat**a**2, we deleted *MATα2* from an α cell and assayed gene expression changes genome wide by mRNA-seq. None of the **a**-specific genes increased in expression (Fig. 2D, Fig. S3), indicating that the *W. anomalus* Matα2 does not directly repress the **a**-specific genes. We further tested this conclusion using a more classical approach. In *S. cerevisiae*, *MATα2* is necessary for proper mating of α cells; its deletion causes ectopic expression of the **a**-specific genes thereby preventing efficient mating. In *W. anomalus* a wildtype α cell and two independently derived deletions of *MATα2* from α cells all mated with similar efficiency, indicating, unlike in *S. cerevisiae*, that *MATα2* is not required for α cell mating (Fig. 2C). These experiments provide additional support for the conclusion that the *W. anomalus* Matα2 does not regulate the **a**-specific genes in *W. anomalus*.

W. anomalus **a**-specific genes are activated by Mat**a**2, the ancestral mode of gene expression

The fact that *W. anomalus* Matα2 does not directly repress the **a**-specific genes suggests that, by default, the transcriptional activator Mat**a**2 must positively regulate the **a**-specific genes in *W. anomalus*. To test this hypothesis, we deleted *MAT***a***2* in an **a** cell and measured the

expression levels of the **a**-specific genes with the NanoString nCounter system, a method for directly counting individual transcripts of choice in an RNA sample (*24*). Each of the **a**-specific genes was significantly under-expressed in the *MAT***a***2* deletion strain with respect to a wildtype **a** cell (Fig. 2E). We also tested the consequences of *MAT***a***2* deletion on **a**-specific mating in a quantitative mating assay and found that an **a** cell with *MAT***a***2* deleted fails to mate at any detectable level in our assay (Fig. 2C). Thus, *W. anomalus* **a**-specific genes are directly regulated by Mat**a**2-activation alone, and Matα2 repression is not involved. Consistent with this conclusion are bioinformatic and experimental analyses of the **a**-specific genes in *W. anomalus, C. jadinii and W. ciferrii,* which identifies *cis*-regulatory sequences that are similar to those controlled by Mat**a**2 in other species (Fig. S4A-B). Previous work has provided a detailed comparison of the functional and sequences differences between Mat**a**2-Mcm1 and Matα2-Mcm1 binding sites in **a**-specific gene promoters, and established the evolutionary order of events in which the regulatory logic of **a**-specific gene regulation evolved (*10*, *11*, *23*). As discussed above, the Mat**a**2 activation scheme of regulating the **a**-specific genes is also observed in *C. albicans* and other outgroup species and, by inference, is likely to represent the situation in the last common ancestor of the three clades (*10*, *11*, *23*).

<u>*W. anomalus* Matα2 retains the deep ancestral function of repressing the haploid-specific genes with Mat**a**1</u>

Matα2's more ancient function is repression of the haploid-specific genes with Mat**a**1. To test whether this ancestral role is retained in *W. anomalus*, we separately deleted either *MATα2* or *MAT***a***1* in an **a**/α cell background and measured gene expression changes by mRNA-seq. In both deletion strains the haploid-specific genes are significantly de-repressed compared to a wildtype

a/α cell (Fig. 3A, Fig. S5). We note that deletion of either *MAT***a***1* or *MATα2* also de-represses the *MAT***a***2* and *MATα1* genes, and the **a** and α-specific genes are also improperly expressed in the deletion strains. This result also explains how the **a** and α-specific genes are kept off in the **a**/α cells. From these results, we conclude that *W. anomalus* retains the deep ancestral function of Matα2, to repress the haploid-specific genes in combination with Mat**a**1.

Mcm1 is necessary for haploid-specific gene repression in *W. anomalus*

We introduced a series of *MATα2* constructs into an **a** cell and measured changes in haploid-specific gene expression using the NanoString nCounter system. As expected, addition of a wildtype allele of *MATα2* resulted in transcriptional repression of the haploid-specific genes relative to their expression levels in an **a** cell (Fig. S6B). *RME1*, a deeply conserved haploid-specific gene, was especially tightly repressed, to a level less than one percent of that in **a** cells. Addition of a *MATα2* allele with a point mutation in the Mat**a**1-binding region (region 5), on the other hand, failed to repress *RME1* and the other haploid-specific genes (Fig. S6B). Addition of *MATα2* chimeras where the Tup1 or Mcm1 interaction regions of Matα2 were replaced by the homologous region of the *C. albicans* (outgroup) Matα2 resulted in substantial loss of repression of many haploid-specific genes (Fig. S6B). We conclude that both the Tup1 and Mcm1-interacting regions of Matα2 are necessary for the complete repression of haploid-specific genes in *W. anomalus.*

The requirement for Matα2 to carry the Mcm1-interaction region in order to repress the haploid-specific genes in *W. anomalus* was surprising because this region is dispensable for haploid-specific gene repression in *S. cerevisiae* and *C. albicans* (whereas the Tup1-interaction region *is* necessary for haploid-specific gene repression in *S. cerevisiae*, and is not present in *C.*

albicans) (*11*). We therefore carried out an experiment to test whether Mcm1 has a direct role in repressing the haploid-specific genes in *W. anomalus*. Mcm1 controls the expression of hundreds of genes and is an essential gene in all fungal species tested to date (*40*). Therefore, experiments that alter levels of Mcm1 are difficult to rigorously interpret. Instead, we turned to the regulatory region of the *RME1* gene, the most tightly regulated haploid-specific gene, and found two high-scoring matches to the Mcm1 recognition sequence (a *cis*-regulatory sequence that is virtually invariant across these species) (Fig. 3B, Fig. S6) (*40*). We confirmed that both Mat**a**1 and Matα2 bind at the *RME1* locus by chromatin-immunoprecipitation followed by qPCR (Fig. S6C). To test whether Mcm1 is also required for *RME1* repression in an **a**/α cell we constructed a reporter driven by the DNA upstream of the *RME1* gene, showed that it was repressed in **a**/α cells, and tested the effect of deleting the Mcm1 recognition sequences. This experiment has the advantage of maintaining proper regulation of all the normal haploid specific genes, including the endogenous copies of *RME1*.

Deletion of the proximal, but not the distal Mcm1 recognition sequence did away with repression of the *RME1* reporter in **a**/α cells (Fig. 3C, Fig. S6D). Based on motif searching, we also identified several potential Mat**a**1-Matα2 binding sequences in the *RME1* control region (Fig. S6A). While the two best matches to the Mat**a**1-Matα2 binding site were not required for *RME1* repression, deletion of a possible Matα2 site that abuts the critical Mcm1 site also de-represses *RME1* in an **a**/α cell (Fig. S6E). From these results we conclude that Mat**a**1, Matα2, and Mcm1 are all required to repress *RME1* transcription in **a**/α cells. Also required are the regions of Matα2 that bind Mcm1 and Mat**a**1.

Supplementary Figures

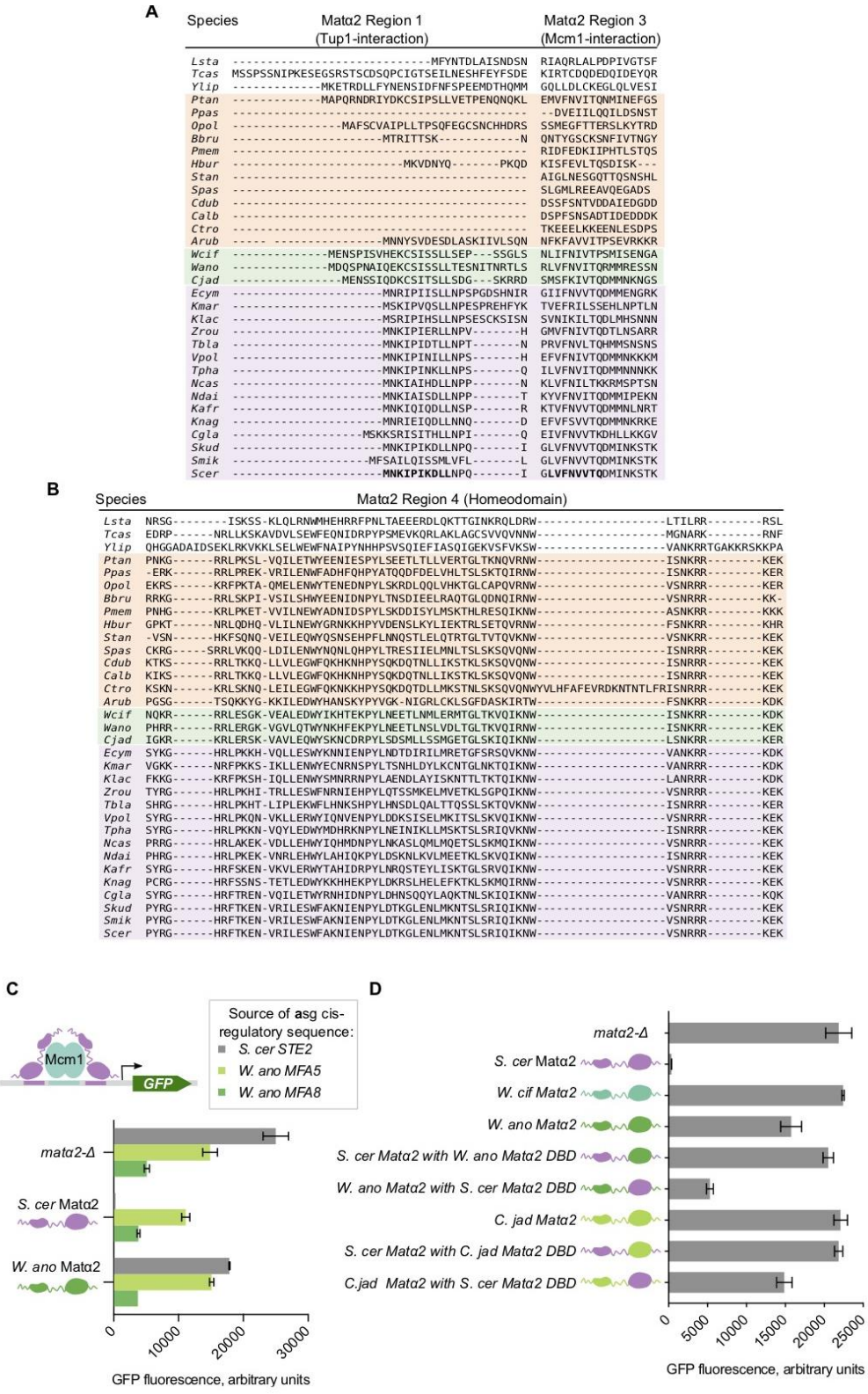

Fig. S1. *W. anomalus* clade Matα2 orthologs contain functional Tup1 and Mcm1-interaction regions, but cannot repress a**-specific gene transcription due to changes in homeodomain**

A. Multiple sequence alignment of Matα2 regions 1 and 3, generated by MUSCLE. *S. cerevisiae* clade (Saccharomycetaceae) sequences are shown in purple, *W. anomalus* clade (Phaffomycetaceae) sequences are shown in green, and *C. albicans* clade (Pichiacea and Debaryomycetaceae) sequences are shown in orange. Species are denoted with a four letter code where the first letter is the first letter of the genus name, and the next three letters are the first three letters of the species name (*33*, *41-43*). Residues known to be functionally important for binding Tup1 and Mcm1 in *S. cerevisiae* are bolded on the *Scer* line.

B. Multiple sequence alignment of Matα2 region 4, the homeodomain, generated by MUSCLE. As above, *S. cerevisiae* clade (Saccharomycetaceae) sequences are shown in purple, *W. anomalus* clade (Phaffomycetaceae) sequences are shown in green, and *C. albicans* clade (Pichiacea and Debaryomycetaceae) sequences are shown in orange.

C. The **a**-specific gene *cis*-regulatory elements from the *S. cerevisiae STE2* gene, and two *W. anomalus MFA* genes were placed into a reporter construct and used to test the ability of *S. cerevisiae* Matα2 (purple) and *W. anomalus* Matα2 (green) to repress **a**-specific gene transcription. *MATα2* constructs were driven by the *S. cerevisiae MATα2* promoter and inserted into *S. cerevisiae* α cells with *MATα2* deleted (*matα2-Δ*). Mean and SD of three independent isolates grown and tested in parallel are shown for all GFP reporter experiments.

D. *W. anomalus* clade Matα2 proteins, as well as chimeric Matα2 alleles with the *S. cerevisiae* Matα2 homeodomain DNA-binding domain ("DBD") (purple) swapped into *W. anomalus* clade Matα2 protein sequences (*W. anomalus*, *W. ciferrii*, and *C. jadinii*) (shades of green),

and vice versa, were tested for their ability to repress transcription of a reporter with the Matα2-Mcm1 binding site from *S. cerevisiae STE2* promoter. *W. anomalus* clade Matα2s consistently do not support **a**-specific gene transcriptional repression in this assay. For both *W. anomalus* and *C. jadinii*, this lack of repression is due to the DNA-binding domain, although the *S. cerevisiae* DNA-binding domain is not sufficient to confer **a**-specific gene repression to *C. jadinii* Matα2. Mean and SD of three independent genetic isolates grown and tested in parallel are shown.

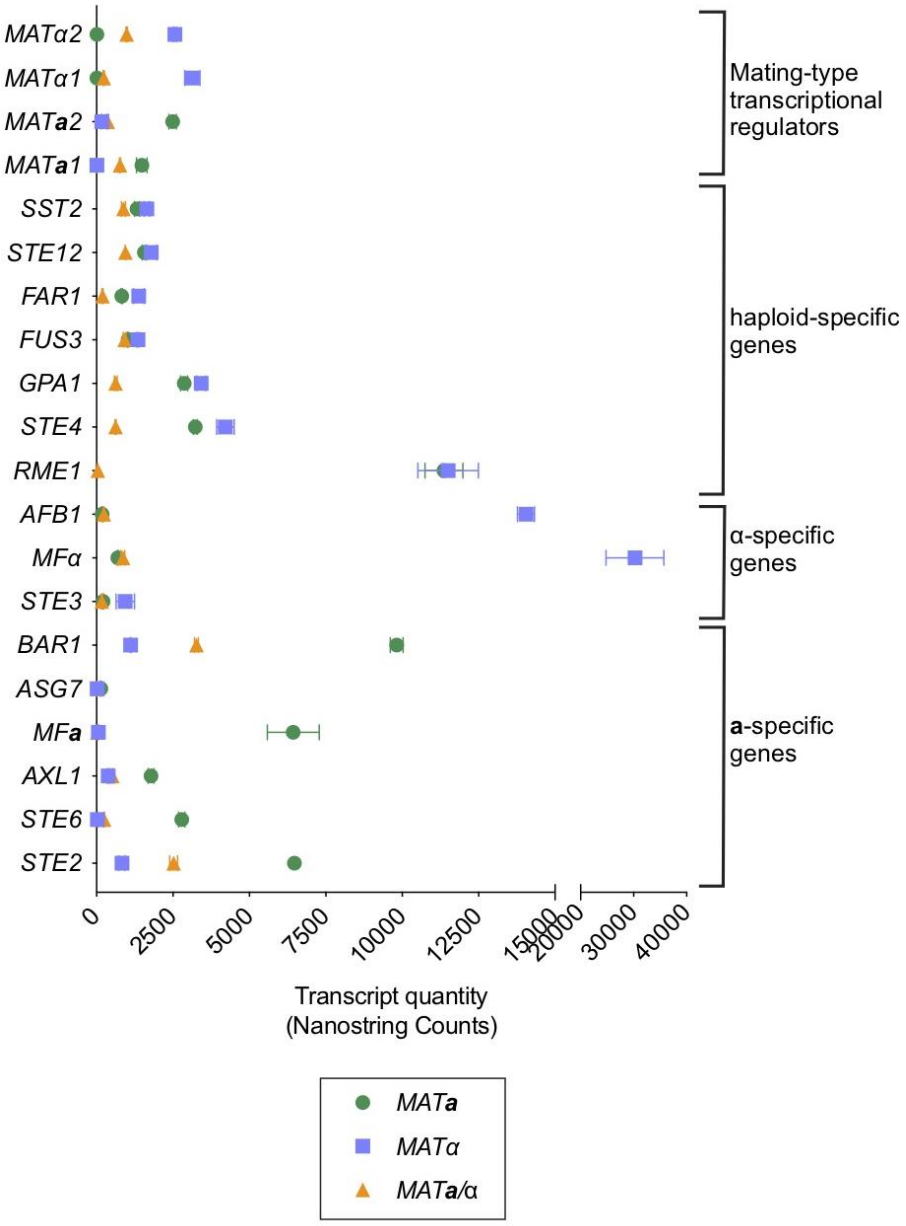

Fig. S2. Cell-type specific gene expression patterns in *W. anomalus*

Cell-type specific gene expression levels in wildtype *W. anomalus* **a** (green circles), α (blue squares), and **a**/α (orange triangles) cells measured with the NanoString nCounter system. Mean and SD of two samples are plotted, except for cases in which the error bars are smaller than the data point. This is not intended to be an exhaustive list, and there may be additional cell-type specific genes.

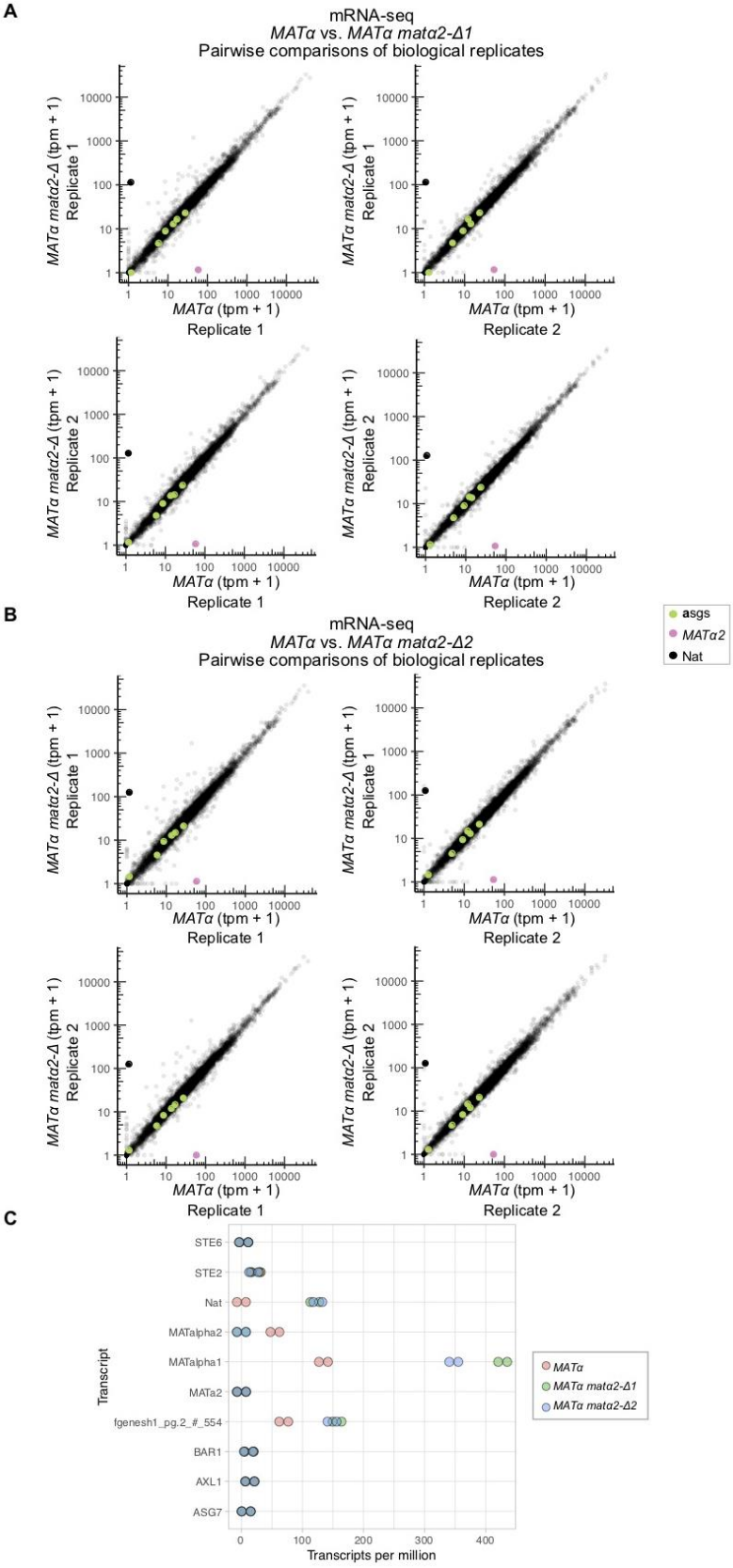
A
mRNA-seq
MATα vs. MATα matα2-Δ1
Pairwise comparisons of biological replicates
MATα matα2-Δ (tpm + 1)
Replicate 1
MATα (tpm + 1)
Replicate 1
MATα matα2-Δ (tpm + 1)
Replicate 1
MATα (tpm + 1)
Replicate 2
MATα matα2-Δ (tpm + 1)
Replicate 2
MATα (tpm + 1)
Replicate 1
MATα matα2-Δ (tpm + 1)
Replicate 2
MATα (tpm + 1)
Replicate 2
B
mRNA-seq
MATα vs. MATα matα2-Δ2
Pairwise comparisons of biological replicates
asgs
MATα2
Nat
C
STE6
STE2
Nat
MATalpha2
MATalpha1
MATa2
fgenesh1_pg.2_#_554
BAR1
AXL1
ASG7
Transcript
Transcripts per million
MATα
MATα matα2-Δ1
MATα matα2-Δ2

Fig. S3. *W. anomalus MATα2* does not directly repress the a**-specific genes**

A. mRNA-seq of wildtype *W. anomalus* α cells (*MATα*) vs. first isolate of α cells with *MATα2* deleted (*MATα2 matα2-Δ1). MATα2* expression in transcripts per million (tpm) is shown in pink, *Nat* expression is shown in opaque black, the **a**-specific genes *STE2*, *AXL1*, *ASG7*, *BAR1*, *STE6* and *MAT***a***2* are shown in green, and all other mRNAs are shown in translucent black. Replicate 1 vs. Replicate 1 plot is the same as shown in main text Fig. 2D, all other plots are pairwise comparisons of the other biological replicates performed in parallel in this experiment.

B. mRNA-seq of wildtype *W. anomalus* α cells (*MATα*) vs. second isolate of α cells with *MATα2* deleted (*MATα2 matα2-Δ2). MATα2* expression in transcripts per million (tpm) is shown in pink, *Nat* expression is shown in opaque black, the **a**-specific genes *STE2*, *AXL1*, *ASG7*, *BAR1*, *STE6*, and *MAT***a***2* are shown in green, and all other mRNAs are shown in translucent black. The wildtype α cell data is the same as in Fig. S3A.

C. Expression levels of genes of particular interest from mRNA-seq experiments plotted above, in transcripts per million (tpm). Shown are **a**-specific genes (*STE2*, *AXL1*, *ASG7*, *BAR1*, and *STE6),* mating-type transcriptional regulators (*MAT***a***2, MATα2, and MATα1)*, the drug-resistance marker used to delete *MATα2* (Nat), and the transcript "fgenesh1_pg.2_#_554," which, besides *MATα2, and MATα1* and Nat, is the only significantly differentially regulated transcript shared between both isolates of *MATα2 matα2-Δ vs. MATα2* (at a false discovery rate $q<0.05$, calculated using the RNA-seq analysis software sleuth).

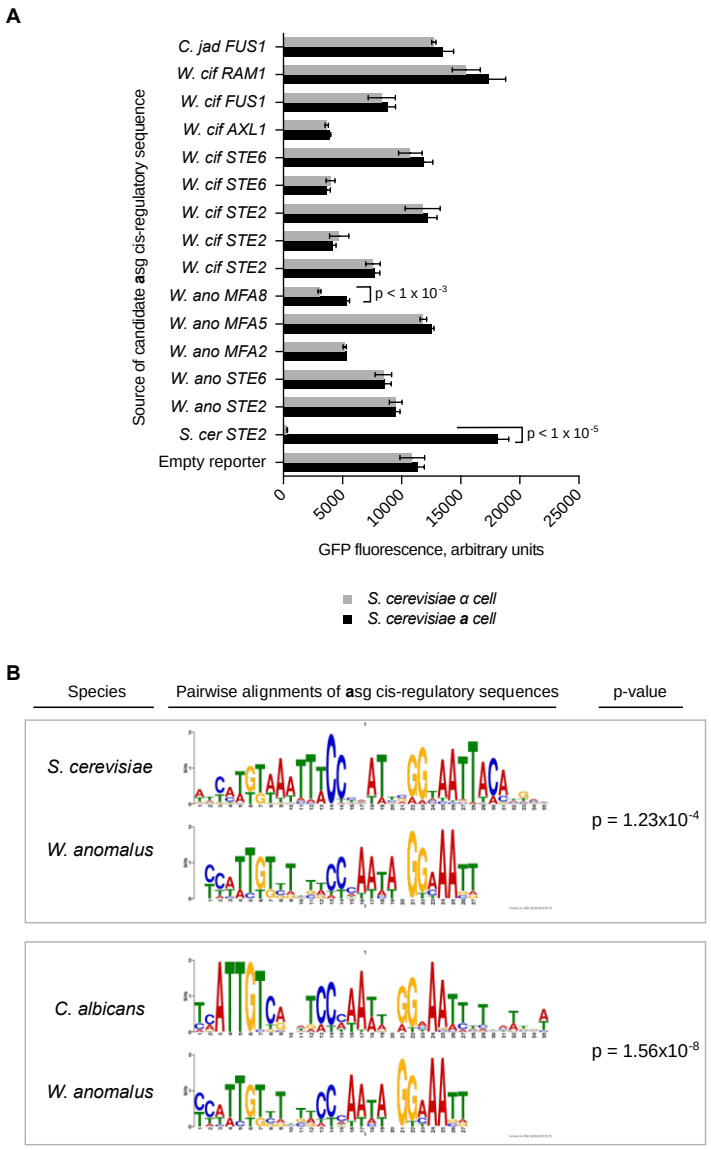

Fig. S4. *W. anomalus* a-specific genes are activated by Mata2, the ancestral mode of gene regulation

A. Previous work has detailed the differences between **a**-specific gene *cis*-regulatory sequences bound by Mat**a**2-Mcm1 (as in *C. albicans*) and Matα2-Mcm1 (as in *S. cerevisiae*) (*11*, *23*). We identified matches to both the *C. albicans* Mat**a**2-Mcm1 and the *S. cerevisiae* Matα2-Mcm1 binding sites in the promoters of the *W. anomalus*, *W. ciferrii*, and *C. jadinii* **a**-specific genes (*11*, *44*). These putative *cis*-regulatory sequences (typically

35 basepairs to match the *S. cerevisiae* Matα2-Mcm1 motif, plus five additional basepairs on either side) were cloned into the *S. cerevisiae* P_{CYC1}-GFP reporter and transformed into *S. cerevisiae* **a** and α cells to test for Matα2-dependent transcriptional repression. In the P_{CYC1}-GFP reporter, the promoter region of the *CYC1* gene, including an upstream activator sequence, drives expression of GFP. Only *cis*-regulatory sequences that support transcriptional repression by Matα2 result in the repression of GFP when cloned into the vector and inserted into a *S. cerevisiae* **a** cell. The *cis*-regulatory sequences were isolated from their endogenous promoter elements and inserted into the P_{CYC1}-GFP to test whether they are sufficient to support Matα2-dependent transcriptional repression. While the *S. cerevisiae STE2 cis*-regulatory sequence supports transcriptional repression by Matα2, none of the *W. anomalus* clade **a**-specific gene *cis*-regulatory sequences supported repression by Matα2 to the same extent. Only one of the 14 CREs tested, found upstream of a paralog of the *W. anomalus* **a**-mating pheromone gene *MFA (*"*MFA8*"*)*, showed a statistically significant difference between **a** and α cells with 1.75-fold repression ($p < 0.05$). This sequence is also the best match to the *S. cerevisiae* Matα2-Mcm1 binding site that we identified in the *W. anomalus* clade. Given the overall evidence for a lack of *W. anomalus* clade Matα2 repression of the **a**-specific genes, we do not believe that this single instance of partial repression by *S. cerevisiae* Matα2 reflects actual or vestigial Matα2-Mcm1 repression in *W. anomalus*.

B. *De novo* motifs for the *W. anomalus* clade **a**-specific genes were generated using the motif-elicitation software MEME, and compared to both the *C. albicans* **a**-specific gene *cis*-regulatory sequence (Mat**a**2-Mcm1 binding site) and the *S. cerevisiae* **a**-specific gene *cis*-regulatory sequence (Matα2-Mcm1 binding site) using Tomtom, a tool within the MEME

suite (*37*, *38*). While the *W. anomalus* clade **a**-specific gene *cis*-regulatory sequence shares homology with both the *S. cerevisiae* and *C. albicans* sequences due to the shared Mcm1-binding site, it is more similar to the *C. albicans* site than the *S. cerevisiae* one (*11*, *23*).

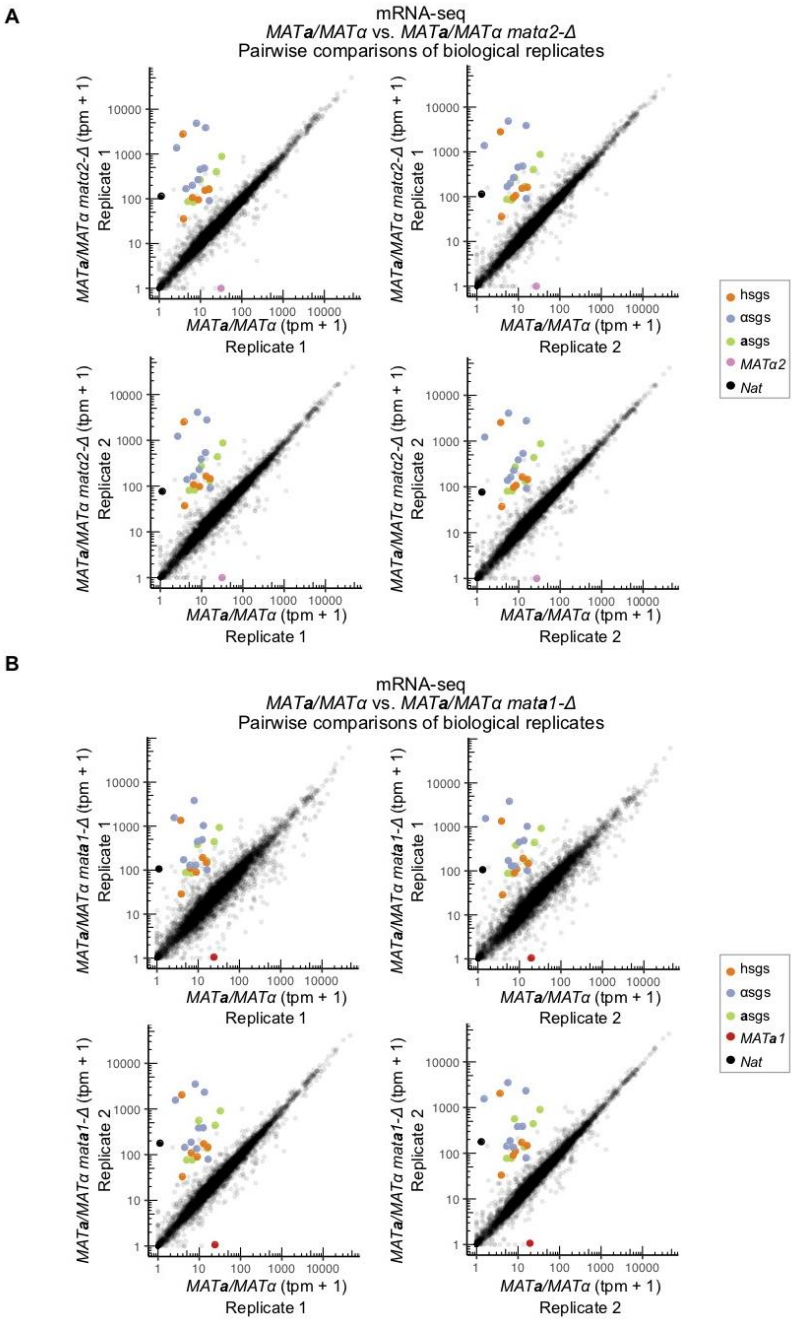

Fig. S5. The *W. anomalus* haploid-specific genes are repressed by Matα2 and Mata**1**

A. mRNA-seq of wildtype *W. anomalus* **a**/α cells (*MAT***a***/MATα*) vs. **a**/α cells with *MATα2* deleted (*MAT***a***/MATα matα2-Δ). MATα2* expression in transcripts per million (tpm) is shown in pink, the **a**-specific genes *STE2*, *AXL1*, *ASG7*, *BAR1*, *STE6* and *MAT***a***2* are shown in green, the haploid-specific genes *STE4*, *GPA1*, *FUS3*, *SST2*, *RME1*, and *FAR1*

are shown in orange, the α-specific genes *SAG1*, *STE3*, *STE13*, *MFα* (multiple copies), *AFB1* (multiple copies), and *MATα1* are shown in blue-gray, the selective marker used to delete *MATα2* is shown in opaque black, and all other mRNAs are shown in translucent black. The results show that the **a**-, α- and haploid-specific genes all increase in expression when *MATα2* is deleted from an **a**/α cell. The **a**- and α-specific genes are expressed because the proteins that activate them (Mat**a**2 and Matα1) are normally repressed in an **a**/α cell but are expressed when *MATα2* is deleted. Replicate 1 vs. Replicate 1 plot is the same as shown in main text Fig. 3A, all other plots are pairwise comparisons of the other biological replicates performed in parallel in this experiment.

B. mRNA-seq of wildtype *W. anomalus* **a**/α cells (*MAT***a***/MATα*) vs. **a**/α cells with *MAT***a***1* deleted (*MAT***a***/MATα mat***a***1-Δ)*. Genes are highlighted as above, except *MAT***a***1* is shown in red and *MATα2* is not highlighted. As when *MATα2* is deleted from an **a**/α cell, the **a**-, α- and haploid-specific genes are more highly expressed when *MAT***a***1* is deleted. Pairwise comparisons of the biological replicates performed in parallel in this experiment are shown.

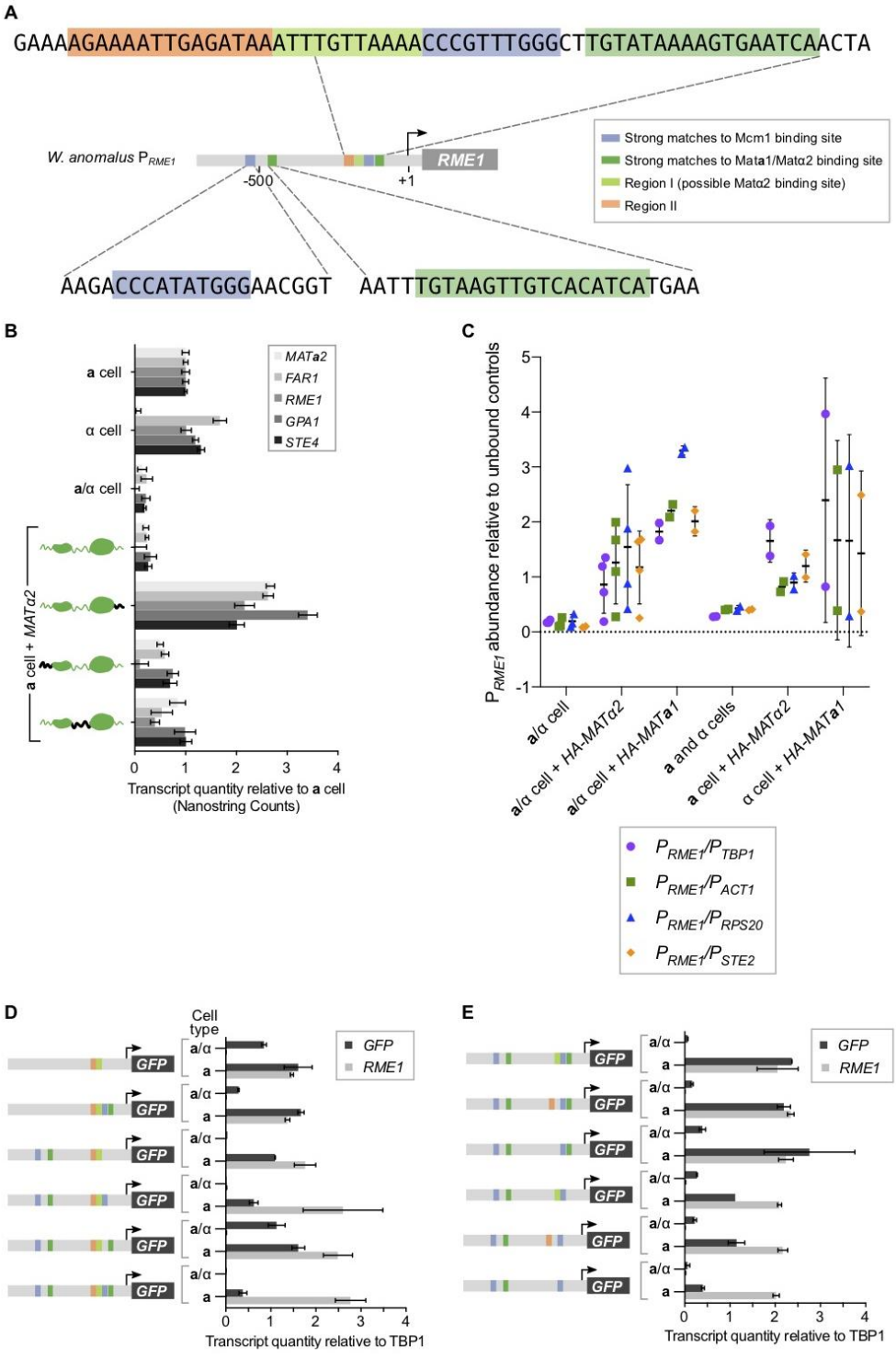

Fig. S6. Matα2, Mata1, and an Mcm1 *cis*-regulatory sequence are directly required for *RME1* repression in *W. anomalus* a/α cell

A. Diagram of the sequence upstream of the *W. anomalus RME1* coding sequence indicating the two strongest matches to Mat**a**1-Matα2 (green) and Mcm1 (blue) binding sites detected, as well as their location relative to the transcription start site. In addition, the smaller lime green box (Region I) indicates a possible Matα2 binding site, and Region II indicates another stretch of sequence to the left of the proximal Mcm1 site with which we experimented. The TGT residues within Region I may be a possible Matα2 binding site and are positioned relative to the adjacent Mcm1 binding site as they would be in a *S. cerevisiae* Matα2-Mcm1 **a**-specific gene *cis*-regulatory site.

B. Haploid-specific gene expression levels in *W. anomalus* **a** cells, α cells, **a**/α cells, and **a** cells with various alleles of Matα2 inserted, were measured by the NanoString nCounter system. Wildtype *W. anomalus* Matα2 sequences are shown in green, while "loss of function" sequences are shown with thicker black lines. For the Tup1- and Mcm1-interacting regions (regions 1 and 3), the loss of function sequences are homologous regions from *C. albicans* Matα2 swapped in (see Fig. S1A-B). The Mat**a**1-interaction region (region 5) loss of function sequence is a point mutation (*33*). Mean and SD of two cultures per genotype, grown and tested in parallel, were calculated and then normalized to expression levels in the *W. anomalus* **a** cell.

C. *W. anomalus* Matα2 and Mat**a**1 were N-terminally HA-tagged and transformed into *W. anomalus* cells. Chromatin immunoprecipitation (ChIP) using anti-HA antibodies was performed with wildtype *W. anomalus* cells as an untagged control, followed by qPCR to measure abundance of P_{RME1} relative to several unbound regions as controls (P_{TBP1} in purple circles, P_{ACT1} in green squares, P_{RPS20} in blue triangles, and P_{STE2} in orange diamonds).

Results from individual ChIP samples are plotted as dots, and lines show the mean and SD for each strain background.

D. Expression levels of endogenous *W. anomalus RME1* transcript and P_{RME1}-GFP reporter (*GFP*) in *W. anomalus* **a** and **a**/α cells measured by RT-qPCR. Mat**a**1-Matα2 (green) and/or Mcm1 (blue) binding sites were deleted from P_{RME1}-GFP individually and in combination, and the reporters were transformed into **a** cells. Orange and lime green boxes show the Region I and II *cis*-sequences to the left of the proximal Mcm1 start site that are unaltered in this experiment. The **a** cells were crossed to α cells to form **a**/α cells, and expression levels of the reporter and the endogenous transcript were measured in both cell types. The proximal Mcm1-binding site with respect to the transcription start site is necessary for *RME1* repression in the **a**/α cell. Quantities are mean and SD of two cultures grown and measured in parallel, normalized to expression of the housekeeping gene *TBP1*. This is a repeat of the experiment shown in Fig. 3C, with independent genetic isolates of the strains.

E. Expression levels of endogenous *W. anomalus RME1* transcript and P_{RME1}-GFP reporter (*GFP*) in *W. anomalus* **a** and **a**/α cells measured by RT-qPCR. Possible Mat**a**1-Matα2 binding sites (orange and lime green boxes) immediately to the left of the proximal Mcm1 site were deleted from P_{RME1}-GFP individually and in combination, and the reporters were transformed into **a** cells. The **a** cells were then crossed to α cells to form **a**/α cells, and expression levels of the reporter and the endogenous transcript were measured in both cell types. While the sequences to the right of the proximal Mcm1 site strongly resemble the Mat**a**1-Matα2 binding site from other yeast species, it is actually the sequences to the left of this Mcm1 *cis*-regulatory site that are necessary for complete *RME1* repression in *W.*

anomalus **a**/α cells (*45*). Quantities are mean and SD of two cultures grown and measured in parallel, normalized to expression of the housekeeping gene *TBP1*.

References and Notes

1. V. J. Lynch, G. P. Wagner, Resurrecting the role of transcription factor change in developmental evolution. *Evolution* **62**, 2131–2154 (2008). doi:10.1111/j.1558-5646.2008.00440.x Medline

2. A. M. Cheatle Jarvela, V. F. Hinman, Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. *EvoDevo* **6**, 3 (2015). doi:10.1186/2041-9139-6-3 Medline

3. D. L. Stern, V. Orgogozo, The loci of evolution: How predictable is genetic evolution? *Evolution* **62**, 2155–2177 (2008). doi:10.1111/j.1558-5646.2008.00450.x Medline

4. G. A. Wray, The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007). doi:10.1038/nrg2063 Medline

5. P. J. Wittkopp, G. Kalay, Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69 (2011). doi:10.1038/nrg3095 Medline

6. H. Li, A. D. Johnson, Evolution of transcription networks—Lessons from yeasts. *Curr. Biol.* **20**, R746–R753 (2010). doi:10.1016/j.cub.2010.06.056 Medline

7. C. Sayou, M. Monniaux, M. H. Nanao, E. Moyroud, S. F. Brockington, E. Thévenon, H. Chahtane, N. Warthmann, M. Melkonian, Y. Zhang, G. K.-S. Wong, D. Weigel, F. Parcy, R. Dumas, A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* **343**, 645–648 (2014). doi:10.1126/science.1248229 Medline

8. C. R. Baker, B. B. Tuch, A. D. Johnson, Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7493–7498 (2011). doi:10.1073/pnas.1019177108 Medline

9. X.-X. Shen, D. A. Opulente, J. Kominek, X. Zhou, J. L. Steenwyk, K. V. Buh, M. A. B. Haase, J. H. Wisecaver, M. Wang, D. T. Doering, J. T. Boudouris, R. M. Schneider, Q. K. Langdon, M. Ohkuma, R. Endoh, M. Takashima, R. I. Manabe, N. Čadež, D. Libkind, C. A. Rosa, J. DeVirgilio, A. B. Hulfachor, M. Groenewald, C. P. Kurtzman, C. T. Hittinger, A. Rokas, Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* **175**, 1533–1545.e20 (2018). doi:10.1016/j.cell.2018.10.023 Medline

10. A. E. Tsong, M. G. Miller, R. M. Raisner, A. D. Johnson, Evolution of a combinatorial transcriptional circuit: A case study in yeasts. *Cell* **115**, 389–399 (2003). doi:10.1016/S0092-8674(03)00885-7 Medline

11. C. R. Baker, L. N. Booth, T. R. Sorrells, A. D. Johnson, Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* **151**, 80–95 (2012). doi:10.1016/j.cell.2012.08.018 Medline

12. X.-X. Shen, X. Zhou, J. Kominek, C. P. Kurtzman, C. T. Hittinger, A. Rokas, Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. *G3* **6**, 3927–3939 (2016). doi:10.1534/g3.116.034744 Medline

13. R. Riley, S. Haridas, K. H. Wolfe, M. R. Lopes, C. T. Hittinger, M. Göker, A. A. Salamov, J. H. Wisecaver, T. M. Long, C. H. Calvey, A. L. Aerts, K. W. Barry, C. Choi, A. Clum, A.

Y. Coughlan, S. Deshpande, A. P. Douglass, S. J. Hanson, H.-P. Klenk, K. M. LaButti, A. Lapidus, E. A. Lindquist, A. M. Lipzen, J. P. Meier-Kolthoff, R. A. Ohm, R. P. Otillar, J. L. Pangilinan, Y. Peng, A. Rokas, C. A. Rosa, C. Scheuner, A. A. Sibirny, J. C. Slot, J. B. Stielow, H. Sun, C. P. Kurtzman, M. Blackwell, I. V. Grigoriev, T. W. Jeffries, Comparative genomics of biotechnologically important yeasts. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9882–9887 (2016). doi:10.1073/pnas.1603941113 Medline

14. C. Kurtzman, J. W. Fell, Eds., *The Yeasts - A Taxonomic Study* (Elsevier, 1998).

15. K. Komachi, A. D. Johnson, Residues in the WD repeats of Tup1 required for interaction with α2. *Mol. Cell. Biol.* **17**, 6023–6028 (1997). doi:10.1128/MCB.17.10.6023 Medline

16. A. Stoltzfus, On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181 (1999). doi:10.1007/PL00006540 Medline

17. J. Lukeš, J. M. Archibald, P. J. Keeling, W. F. Doolittle, M. W. Gray, How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* **63**, 528–537 (2011). doi:10.1002/iub.489 Medline

18. M. W. Gray, J. Lukes, J. M. Archibald, P. J. Keeling, W. F. Doolittle, Irremediable complexity? *Science* **330**, 920–921 (2010). doi:10.1126/science.1198594 Medline

19. A. Wagner, Robustness, evolvability, and neutrality. *FEBS Lett.* **579**, 1772–1778 (2005). doi:10.1016/j.febslet.2005.01.063 Medline

20. C. K. Dalal, A. D. Johnson, How transcription circuits explore alternative architectures while maintaining overall circuit output. *Genes Dev.* **31**, 1397–1405 (2017). doi:10.1101/gad.303362.117 Medline

21. T. R. Sorrells, A. D. Johnson, Making sense of transcription networks. *Cell* **161**, 714–723 (2015). doi:10.1016/j.cell.2015.04.014 Medline

22. I. Herskowitz, A regulatory hierarchy for cell specialization in yeast. *Nature* **342**, 749–757 (1989). doi:10.1038/342749a0 Medline

23. A. E. Tsong, B. B. Tuch, H. Li, A. D. Johnson, Evolution of alternative transcriptional circuits with identical logic. *Nature* **443**, 415–420 (2006). doi:10.1038/nature05099 Medline

24. G. K. Geiss, R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D. L. Dunaway, H. P. Fell, S. Ferree, R. D. George, T. Grogan, J. J. James, M. Maysuria, J. D. Mitton, P. Oliveri, J. L. Osborn, T. Peng, A. L. Ratcliffe, P. J. Webster, E. H. Davidson, L. Hood, K. Dimitrov, Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 (2008). doi:10.1038/nbt1385 Medline

25. K. H. Wolfe, G. Butler, Evolution of Mating in the Saccharomycotina. *Annu. Rev. Microbiol.* **71**, 197–214 (2017). doi:10.1146/annurev-micro-090816-093403 Medline

26. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990). doi:10.1016/S0022-2836(05)80360-2 Medline

27. E. Proux-Wéra, D. Armisén, K. P. Byrne, K. H. Wolfe, A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics* **13**, 237 (2012). doi:10.1186/1471-2105-13-237 Medline

28. C. Schorsch, T. Köhler, E. Boles, Knockout of the DNA ligase IV homolog gene in the sphingoid base producing yeast Pichia ciferrii significantly increases gene targeting efficiency. *Curr. Genet.* **55**, 381–389 (2009). doi:10.1007/s00294-009-0252-z Medline

29. R. D. Gietz, R. A. Woods, Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87–96 (2002). doi:10.1016/S0076-6879(02)50957-5 Medline

30. H. Youk, W. A. Lim, Secreting and sensing the same molecule allows cells to achieve versatile social behaviors. *Science* **343**, 1242782 (2014). doi:10.1126/science.1242782 Medline

31. L. Guarente, M. Ptashne, Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2199–2203 (1981). doi:10.1073/pnas.78.4.2199 Medline

32. T. R. Sorrells, L. N. Booth, B. B. Tuch, A. D. Johnson, Intersecting transcription networks constrain gene regulatory evolution. *Nature* **523**, 361–365 (2015). doi:10.1038/nature14613 Medline

33. T. Li, M. R. Stark, A. D. Johnson, C. Wolberger, Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science* **270**, 262–269 (1995). doi:10.1126/science.270.5234.262 Medline

34. H. Pimentel, N. L. Bray, S. Puente, P. Melsted, L. Pachter, Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **14**, 687–690 (2017). doi:10.1038/nmeth.4324 Medline

35. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016). doi:10.1038/nbt.3519 Medline

36. M. G. Miller, A. D. Johnson, White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* **110**, 293–302 (2002). doi:10.1016/S0092-8674(02)00837-1 Medline

37. T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S. Noble, MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009). doi:10.1093/nar/gkp335 Medline

38. S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, W. S. Noble, Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007). doi:10.1186/gb-2007-8-2-r24 Medline

39. I. Nocedal, E. Mancera, A. D. Johnson, Gene regulatory network plasticity predates a switch in function of a conserved transcription regulator. *eLife* **6**, e23250 (2017). doi:10.7554/eLife.23250 Medline

40. B. B. Tuch, D. J. Galgoczy, A. D. Hernday, H. Li, A. D. Johnson, The evolution of combinatorial gene regulation in fungi. *PLOS Biol.* **6**, e38 (2008). doi:10.1371/journal.pbio.0060038 Medline

41. J. Mead, H. Zhong, T. B. Acton, A. K. Vershon, The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. *Mol. Cell. Biol.* **16**, 2135–2143 (1996). doi:10.1128/MCB.16.5.2135 Medline

42. K. Komachi, M. J. Redd, A. D. Johnson, The WD repeats of Tup1 interact with the homeo domain protein alpha 2. *Genes Dev.* **8**, 2857–2867 (1994). doi:10.1101/gad.8.23.2857 Medline

43. S. Tan, T. J. Richmond, Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. *Nature* **391**, 660–666 (1998). doi:10.1038/35563 Medline

44. T. L. Bailey, M. Gribskov, Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics* **14**, 48–54 (1998). doi:10.1093/bioinformatics/14.1.48 Medline

45. L. N. Booth, B. B. Tuch, A. D. Johnson, Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature* **468**, 959–963 (2010). doi:10.1038/nature09560 Medline