Intersecting transcription networks constrain gene regulatory evolution

Trevor R. Sorrells1,2, Lauren N. Booth1,2, Brian B. Tuch1,3,† & Alexander D. Johnson1,2

Epistasis—the non-additive interactions between different genetic loci—constrains evolutionary pathways, blocking some and permitting others4–6. For biological networks such as transcription circuits, the nature of these constraints and their consequences are largely unknown. Here we describe the evolutionary pathways of a transcription network that controls the response to mating pheromone in yeast9. A component of this network, the transcription regulator Ste12, has evolved two different modes of binding to a set of its target genes. In one group of species, Ste12 binds to specific DNA binding sites, while in another lineage it occupies DNA indirectly, relying on a second transcription regulator to recognize DNA. We show, through the construction of various possible evolutionary intermediates, that evolution of the direct mode of DNA binding was not directly accessible to the ancestor. Instead, it was contingent on a lineage-specific change to an overlapping transcription network with a different function, the specification of cell type. These results show that analysing and predicting the evolution of cis-regulatory regions requires an understanding of their positions in overlapping networks, as this placement constrains the available evolutionary pathways.

Pathways of evolution are highly dependent on their starting points, a phenomenon often referred to as historical contingency7–9. For example, studies of protein evolution have shown that initial permissive mutations are often required before a second mutation causes an evolutionary shift in function8. Without the initial mutation, the second mutation could be deleterious. The structural requirements for protein folding and function underlie such behaviour and are a common source of epistasis5,6.

Considerably less is known about historical contingency in the evolution of transcription networks, which often comprise several transcription regulators and many target genes. It has been suggested

**Figure 1** Evolution of the pheromone response in budding yeast. a, During mating, cells sense a peptide pheromone (brown molecule) and transmit this signal to the transcription factor Ste12. b, *K. lactis* a cells were induced with 10μg ml–1 synthetic α-factor pheromone, and gene expression was measured by quantitative reverse transcriptase PCR (qRT–PCR) over time. Shown are mean values ± s.d. for three independent genetic isolates, each grown and analysed separately. c, Conserved pheromone responsive genes were scored for the presence of Ste12 cis-regulatory motifs in their upstream regulatory regions. The track indicates the −log10(P) for the enrichment of the Ste12 motif in each set of genes in each species, by hypergeometric distribution. The genes were divided into a-specific genes and general pheromone-activated genes. The general pheromone-activated genes were used to independently generate the Ste12 cis-regulatory motif in each of four clades using MEME.
that epistasis is prevalent in the evolution of regulatory networks\textsuperscript{8,10}, but detailed examples documenting its underlying causes are largely lacking. Here we studied the effects of contingency in the evolution of a transcription network in yeast that has diversified over several hundred million years. We found that network-level changes altered the evolutionary pathways available to individual cis-regulatory regions as these changes occurred at the intersection of two different but overlapping networks.

The transcription regulator Ste12 controls the response to mating pheromone in the budding yeast \textit{Saccharomyces cerevisiae}\textsuperscript{11} (Fig. 1a). When cells sense pheromone of the opposite mating type, Ste12 is activated by phosphorylation and transcriptionally upregulates many genes involved in mating\textsuperscript{12}. This function of Ste12 is conserved in the dairy yeast, \textit{Kluyveromyces lactis}\textsuperscript{13} (Fig. 1b, c and Extended Data Fig. 1), a species that diverged from \textit{S. cerevisiae} \textsim 100 million years ago\textsuperscript{14}, and in the more distantly related pathogen of humans, \textit{Candida albicans}\textsuperscript{15,16}. Thus, many aspects of pheromone-activated gene expression (including the DNA sequence recognized by Ste12) have been conserved across the \textsim 200 million years separating these three species from their common ancestor.

The genes induced by pheromone-activated Ste12 can be classified into two distinct sets, those genes specific to either the \textalpha or the \textv cell types (5–12 genes, depending on the species), and the general pheromone-activated genes (\textsim 100 genes), which are induced in both cell types. The three yeast species described above all have two mating cell types, \textalpha and \textv, that express complementary sets of genes allowing them to mate with the opposite mating type\textsuperscript{17}. This study focuses on the \textalpha-specific genes (\textalpha\textsuperscript{-sgs}), so named because they are expressed and respond to pheromone in \textalpha cells but not in \textv cells.

Using bioinformatics approaches, we quantified the number of Ste12 cis-regulatory sequences upstream of the \textalpha\textsuperscript{-sgs} and the general pheromone-activated genes in the genomes of 40 yeasts. For the general pheromone-activated genes, Ste12 binding sites were significantly enriched (relative to the rest of the genome) across the \textit{Candida}, \textit{Kluyveromyces}, and \textit{Saccharomyces} clades. Given the central, conserved role of Ste12 in the pheromone response, this enrichment was expected. However, the \textalpha\textsuperscript{-sgs} showed enrichment for Ste12 binding sites only in the \textit{Saccharomyces} clade (Fig. 1c and Extended Data Fig. 2). Although the \textalpha\textsuperscript{-sgs} respond to pheromone across all three clades, they seemed to lack Ste12 cis-regulatory sequences in the \textit{Kluyveromyces} and \textit{Candida} clades. We considered three scenarios for how Ste12 activates the \textalpha\textsuperscript{-sgs} in clades where these genes lack Ste12 cis-regulatory sequences: (1) Ste12 could bind to DNA sites in the \textalpha\textsuperscript{-sgs} that are sufficiently degenerate that they fall below our threshold of detection; (2) Ste12 could be recruited through protein–protein interactions with a second transcription regulator\textsuperscript{18} that binds directly to other cis-regulatory sites upstream of the \textalpha\textsuperscript{-sgs}; or (3) Ste12 could act ‘upstream’ by activating another regulator that binds to and activates the \textalpha\textsuperscript{-sgs} directly.

We distinguished among these models (or a combination of them) by identifying the regulatory regions bound by Ste12 in \textit{K. lactis} by chromatin immunoprecipitation followed by sequencing (ChIP-seq). As in \textit{S. cerevisiae}, \textit{K. lactis} Ste12 was bound to the upstream regions of the \textalpha\textsuperscript{-sgs} (Fig. 2a,b) as well as to the general pheromone-activated genes (Fig. 2c). Thus, \textit{K. lactis} Ste12 is recruited to the promoters of the \textalpha\textsuperscript{-sgs} and activates them in response to pheromone in the absence of recognizable Ste12 binding sites.

We considered the possibility that Ste12 is recruited to the \textalpha\textsuperscript{-sgs} in \textit{K. lactis} through a series of low-affinity cis-regulatory sequences or ‘mismatched’ sequences, a situation that has been documented for certain other transcription regulators\textsuperscript{19,20}. In our case, the mismatched Ste12 sites are ubiquitous across the \textit{K. lactis} genome (Extended Data Fig. 3); by mutating these mismatched sites, we showed they are not required for induction of the \textalpha\textsuperscript{-sgs} by Ste12 (Fig. 2d, e; Extended Data Fig. 4).

Figure 2 Ste12 is recruited to \textalpha\textsuperscript{-sgs} promoters in \textit{K. lactis} without its binding site. a–c Chromatin immunoprecipitation of Ste12–myc using anti-myc antibodies (orange tracks) and an untagged control (grey track). The \textalpha\textsuperscript{-sgs} and the general pheromone-inducible Ste12 sites are ubiquitous across the \textsim 200 million years separating three species from their common ancestor. Figure 3 Three experimental results show that, in the \textit{K. lactis} lineage, Ste12 is indirectly recruited to the \textalpha\textsuperscript{-sgs} by association with the transcription regulator \textalpha\textsuperscript{2}. \textalpha\textsuperscript{2} is encoded by the \textit{MAT\alpha2} gene (also known as \textit{MAT\alpha2}) in the \textalpha mating-type locus; it is thus expressed in \textalpha cells (but not in \textv cells), where it activates transcription of the \textalpha\textsuperscript{-sgs}\textsuperscript{21}. (1) The Ste12 ChIP peaks at the \textalpha\textsuperscript{-sgs} were positioned over the cis-regulatory motif for \textalpha\textsuperscript{2} (Fig. 2a, b). (2) Deletion of the \textit{MAT\alpha2} gene impairs pheromone-induction of the \textalpha\textsuperscript{-sgs} (Fig. 3b, c). (3) The cis-regulatory sequence for \textalpha\textsuperscript{2} is sufficient to mediate pheromone induction when moved into a naive promoter. \textalpha\textsuperscript{2} always works with a ubiquitously expressed regulator, \textit{Mcm1} (ref. 21; Extended Data Fig. 5), so we kept the cis-regulatory sites of these regulators together for the purpose of this experiment. The \textalpha\textsuperscript{-Mcm1} site, when inserted into a test reporter, activated green fluorescent protein (GFP) in response to pheromone (Fig. 3d and Extended Data Fig. 6). No \textalpha\textsuperscript{-regulated sites (even mismatched ones) were present in the DNA added to the reporter construct, and the reporter itself was not pheromone-inducible. ChIP-seq was performed in a \textit{K. lactis} strain that contained this reporter, and the \textalpha\textsuperscript{-Mcm1} site was sufficient for efficient recruitment of Ste12 in the presence of pheromone (Fig. 3e). We can rule out the possibility that \textit{Mcm1} alone recruits Ste12 from the fact that \textit{Mcm1} binds to hundreds of genes in \textit{K. lactis} and none of these are pheromone-inducible, except those also bound by \textalpha\textsuperscript{2} (refs 22, 23). Thus, the majority of the specificity for recruitment of Ste12 must come from \textalpha\textsuperscript{2}. We also note that Ste12 did not bind to the promoter region of the \textit{MAT\alpha2} gene (not shown), ruling out the possibility that during the pheromone response, Ste12 acts as an upstream activator of \textalpha\textsuperscript{2}, which then activates the \textalpha\textsuperscript{-sgs}.

We now turn to the question of how the two different recruitment modes of Ste12 (direct and indirect) arose. Prior experiments in
Figure 3 | a2 mediates the a-specific gene pheromone response in K. lactis.

a. Diagram of asg cell-type-specific regulation in K. lactis. b. c. Induction of the asgs in response to pheromone in WT and MATaΔA strains. Expression is shown as mean ± s.d. for three independent genetic isolates, each grown and analysed separately. for a-specific genes (b) and general pheromone-activated genes (c). STE2 is shown as a general pheromone-activated gene because in this strain, its promoter was replaced by the SST2 promoter. The differences between WT and MATaΔA pheromone induction are significant in b, P < 0.01, but not in c, P > 0.05, by ANOVA. d. The a2-Mcm1 sites were moved into the S. cerevisiae CYC1ΔUAS promoter and expression was measured in K. lactis in uninduced (solid bars) and pheromone-induced (striped bars) conditions. Shown is mean fluorescence ± s.d. of three independent genetic isolates, each grown and analysed separately ***P < 0.001, by ANOVA followed by Tukey’s honest significant difference. e. Ste12–myc ChIP-Seq signal for the genome-integrated reporter construct. Symbols are described in Fig. 2.

Several species combined with the phylogenetic distribution of Ste12 (Fig. 1c) and a2 (ref. 16) binding sites indicate that indirect recruitment by a2 at the asgs was ancestral, and that the Ste12 sites were gained in the Saccharomyces clade. At the general pheromone-activated genes, Ste12 is directly bound to DNA by cis-regulatory sequences in all three clades, providing an explanation for why Ste12 has retained the same DNA-binding specificity.

To understand why Ste12 cis-regulatory sites were gained at the asgs in the Saccharomyces clade and not other clades, we introduced these sites into an ancestral-like promoter (STE2 of K. lactis) and observed the effect on asg expression (Fig. 4a). Introducing strong Ste12 sites by point-mutating mismatched sites in the regulatory region increased the expression in a cells, but it also resulted in mis-expression of the gene in α cells. Expression of asgs in α cells interferes with sexual reproduction because the cells respond to their own pheromones and eventually become insensitive to both α and α pheromone4. Thus, we conclude—in the absence of other changes—that the evolution of high-affinity Ste12 sites disrupts the function of the ancestral form of regulation, causing misexpression of the asgs in α cells that would result in a loss of mating.

These results indicate that a prior permissiveness change in asg regulation was probably necessary for Ste12 sites to be gained in the Saccharomyces clade. In the Saccharomyces lineage, after the split from Kluyveromyces, the activator of the asgs (a2) was replaced by an α cell-type-specific repressor, α2 (ref. 23). To test whether this change permitted the Ste12 cis-regulatory sites to be gained upstream of the asgs, we performed an experiment in Lachancea kluyveri (Fig. 4b), a species in which all of the asgs are activated by a2, and (in contrast to K. lactis) two of them are also repressed by α2 (ref. 23). In other words, we chose a species in which both a2 and α2 are active. We used the L. kluyveri STE2 regulatory region, which is normally regulated by a2 but not α2, to test the effects of introducing the Ste12
and a2 cis-regulatory sites. Introducing Ste12 cis-regulatory sites by point mutation again resulted in ectopic expression of the a1s in a cells, as we observed for K. lactis (Fig. 4c). However, this mis-expression could be mitigated by adding cis-regulatory sequences for the repressor a2 (Fig. 4c), the situation found naturally in S. cerevisiae.4,25. These results indicate that the switch from a2-mediated activation to a2-mediated repression of the a1s was permissive for the gain of Ste12 cis-regulatory sequences in this lineage.

a2 repression was gained in the common ancestor of Kluyveromyces and Saccharomyces clades27, and the Ste12 cis-regulatory sites were gained millions of years later, in the Saccharomyces clade (Fig. 5a). These observations indicate that, while the gain of a2 repression was permissive for the subsequent gain of Ste12 sites, it did not directly cause it. The gain of Ste12 sites did occur, however, at roughly the same time as another evolutionary event, the loss of the activator MATa2 (ref. 21), suggesting that the gain of Ste12 sites may have permitted the loss of this gene (Fig. 5a). To test this possibility, we observed the expression of the K. lactis STE2 reporter in the absence of a2 (Fig. 4d and Extended Data Fig. 7). Adding the Ste12 cis-regulatory sites could indeed compensate for the deletion of MATa2 in both basal and pheromone-activated gene expression (Fig. 4c, d). If we assume that loss of MATa2—on its own—would have been deleterious, it is likely that at least some Ste12 sites were gained first, thus mitigating the potentially deleterious effects of losing MATa2.

Gene regulatory regions are shaped by evolutionary forces including selection and drift, each of which leave distinct patterns in the cis-regulatory regions of genes.28. To further understand the forces that shaped the evolution of a1s cis-regulatory regions, we employed two independent approaches25,28. Both methods suggested that Ste12 cis-regulatory sites in the a1s are maintained by purifying selection in the Saccharomyces clade, but, as predicted, the ‘mismatched’ Ste12 sites in both the Saccharomyces and Kluyveromyces clades are not (Extended Data Figs 8, 9). In particular, in the Saccharomyces clade, Ste12 sites were conserved with respect to the rest of the intergenic regions (P = 1.40 × 10−6 by likelihood ratio test (LRT)), while mismatched sites were not (P = 1 by LRT). In the Kluyveromyces clade, neither Ste12 sites (P = 0.26 by LRT) nor mismatched sites (P = 1 by LRT) show significant conservation in the a1s.

Our previous studies have shown how the transcription network controlling cell-type identity in yeast evolved25,27,28. Here, we have shown how the evolutionary history of this network affected the evolution of an overlapping but distinct transcription network controlling the response to mating pheromone. The two networks overlap by virtue of sharing a common set of target genes. The genes at the intersection of these two networks (the a1s) were constrained in the evolutionary pathways available to them. In the ancestral state, the gain of Ste12 cis-regulatory sites were not allowed, as this would lead to cell-type misexpression. The evolution of a2 repression, which occurred in the ancestor of the S. cerevisiae clade, made the gain of these sites possible, and the loss of the MATa2 gene at a much later time necessitated the maintenance of the Ste12 sites and prevented a return to the ancestral mode of regulation (Fig. 5b).

In protein evolution, neutral mutations can alter the effect of subsequent mutations, suggesting that evolution depends on chance events. Although we do not know whether the gain of a2 repression was strictly neutral, it did not change the overall a1s expression pattern. However, it did permit changes in the overlapping circuit responsible for pheromone induction. Stated another way, changes in the a1s circuitry altered the effect of introducing Ste12 sites to the a1s. Such epistatic interactions are the source of the historical contingency we observed in the a1s regulatory network, effects that are well documented for individual proteins.

Studies of enhancers across species have indicated that many molecular solutions are possible for a given output.25,27,28,30. Our results show that not all solutions are accessible, and those that are accessible are contingent on the functional interactions between overlapping transcription networks. These constraints are not apparent from observation of the individual components of networks (for example, a single enhancer), but only arise from complex interactions between networks. We propose that understanding the evolution of cis-regulatory regions (and predicting their behaviour) requires an understanding of the network structures in which they function.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 February; accepted 4 June 2015.
Published online 8 July 2015.

1. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proc. Natl Acad. Sci. USA 105, 7899–7906 (2008).
2. Harms, M. J. & Thornton, J. W. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. Nature 512, 203–207 (2014).
3. Mohrig, J. R. et al. Importance of historical contingency in the stereochemistry of hydrolase-dehydrogenase enzymes. Science 269, 527–529 (1995).
4. Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nature Rev. Genet. 14, 559–571 (2013).
5. Ortí, E. A., Bridgham, J. T., Redinbo, M. R. & Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. Science 317, 1544–1548 (2007).
6. Bershtein, S., Segal, M., Beikerman, R., Tokuriki, N. & Tawfik, D. S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature 444, 929–932 (2006).
7. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. Proc. Natl Acad. Sci. USA 103, 5869–5874 (2006).
8. Gerke, J., Lorenz, K. & Cohen, B. Genetic interactions between transcription factors cause natural variation in yeast. Science 323, 498–501 (2009).
9. Ren, B. et al. Genome-wide location and function of DNA binding proteins. Science 290, 2306–2309 (2000).
10. Payne, J. L. & Wagner, A. Constraint and contingency in multifunctional gene regulatory circuits. PLoS Comput. Biol. 9, e1003071 (2013).
11. Dolan, J. W., Kirkman, C. & Fields, S. The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. Proc. Natl Acad. Sci. USA 86, 5703–5707 (1989).
12. Herskowitz, I. MAP kinase pathways in yeast: for mating and more. Cell 80, 187–197 (1995).
13. Coria, R. et al. The pheromone response pathway of Kluyveromyces lactis. FEMS Yeast Res. 6, 336–344 (2006).
14. Kences, P. R., Ott, M., Duthil, B. E. & Huyten, M. A. Conservation of divergent transcription in fungi. Trends Genet. 24, 207–211 (2008).
15. Bennet, R. J., Uhl, M. A., Miller, M. G. & Johnson, A. D. Identification and characterization of a Candida albicans mating pheromone. Mol. Cell. Biol. 23, 8189–8201 (2003).
Tsong, A. E., Tuch, B. B., Li, H. & Johnson, A. D. Evolution of alternative transcriptional circuits with identical logic. *Nature* **443**, 415–420 (2006).

Herskowitz, I. A regulatory hierarchy for cell specialization in yeast. *Nature* **342**, 749–757 (1989).

Bornean, A. R. et al. Divergence of transcription factor binding sites across related yeast species. *Science* **317**, 815–819 (2007).

Ramos, A. I. & Barolo, S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Phil. Trans. R. Soc. Lond. B* **368**, 20130018 (2013).

Crocker, J. et al. Low affinity binding site clusters confer Hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).

Tsong, A. E., Miller, M. G., Raisner, R. M. & Johnson, A. D. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* **115**, 389–399 (2003).

Tuch, B. B., Galgoczy, D. J., Hernday, A. D., Li, H. & Johnson, A. D. The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* **6**, e38 (2008).

Baker, C. R., Booth, L. N., Sorrells, T. R. & Johnson, A. D. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* **151**, 80–95 (2012).

Strathern, J., Hicks, J. & Herskowitz, I. Control of cell type in yeast by the mating type locus. The +1→+2 hypothesis. *J. Mol. Biol.* **147**, 357–372 (1981).

Galgoczy, D. J. et al. Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **101**, 18069–18074 (2004).

Wray, G. A. et al. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419 (2003).

Otto, W. et al. Measuring transcription factor-binding site turnover: a maximum likelihood approach using phylogenies. *Genome Biol. Evol.* **1**, 85–98 (2009).

Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).

Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R. & Eisen, M. B. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet.* **4**, e1000106 (2008).

Swanson, C. I., Schwimmer, D. B. & Barolo, S. Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol.* **21**, 1186–1196 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Coyle, I. Nocedal, C. Dalal, C. Britton, C. Baker, and V. Hanson-Smith for valuable comments on the manuscript. We also thank K. Pollard, E. Mancera, C. Nobile, and Q. Mitrovich for advice and assistance with experiments and analysis. We thank K. Boundy-Mills of the Phaff Yeast Culture Collection, University of California, Davis for the *K. lactis* isolates. The work was supported by grant R01 GM037049 from the National Institutes of Health. T.R.S. was supported by a Graduate Research Fellowship from the National Science Foundation.

**Author Contributions** T.R.S. performed gene expression experiments, reporter assays, ChIP-Seq, and bioinformatics analyses. L.N.B. obtained and sequenced the *α*-sg promoters of the *K. lactis* isolates. B.B.T. conceived of, designed, and performed a preliminary analysis of the phylogenetic distribution of Ste12 sites across the asgs. T.R.S., L.N.B., and A.D.J. designed and interpreted experiments and edited the manuscript. T.R.S. and A.D.J. wrote the manuscript.

**Author Information** ChIP-Seq data has been deposited at the Gene Expression Omnibus (GEO) repository under accession number GSE65792. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.D.J. (ajohnson@cgl.ucsf.edu).
METHODS

The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Distribution of Ste12 cis-regulatory sites. Conserved, highly pheromone-induced genes from the genome of S. cerevisiae were divided into 100 500 bp segments. Each segment was used as a psi-blast query to identify orthologues in species whose genomes have been sequenced. These genes were also used as a psi-blast query to find orthologues in species whose genomes are currently being sequenced. The upstream 600 bp of intergenic regions were scored as described so that the motif was considered up to 600 bp upstream of the start codon, a distance that is likely to capture functional Ste12 regulatory sites.

Phenotype analysis by RT-qPCR. The promoter UAS was deleted from the corresponding MATa or MATα strains and the STE2 promoter was replaced by the STE2 promoter fused to GFP into a centromeric reporter, pRS412, using the KpnI and SacI sites.

RT–qPCR. K. lactis cells were grown overnight for 14 h, then starved in medium lacking phosphate as previously described. Cells were cultured in 6.25 μM α-factor pheromone (WSWTITLR and PGPQPIF >95% purity, 100 mg/mL in 100% DMSO; Genemed Synthesis) or an equivalent amount of DMSO. Cells were removed at several time points and pelleted, washed, and frozen in liquid nitrogen. Mutant strains were harvested at 4 h, the final time point. Each experiment was performed in triplicate. The RNA was extracted using the RiboPure RNA Purification kit (Ambion). The RNA was reverse transcribed into cDNA using SuperScript II as described previously. Transcript levels were measured using SYBR green on a StepOnePlus RT PCR machine (Applied Biosystems). Transcripts were normalized to that of TPS4, a gene that was found to be highly consistent across conditions in previous expression experiments. Statistics were performed using an ANOVA.

Chromatin immunoprecipitation. Strains YTS314 and YTS315 were grown with a phosphate starvation protocol modified from our previously published protocol36. Yeast were grown overnight for 14 h, then pelleted, resuspended in phosphate starvation medium, then diluted in 100–200 μl phosphate starvation media to OD600 = 0.25–0.3. These cultures were grown for 2 h, then 13-mer α-factor pheromone was added to 6.25 μM, and the cultures were grown for an additional 2 h. Cells were crosslinked with 1% formaldehyde for 15 min, quenched with glycine, pelleted, washed, and frozen in liquid nitrogen. Cells were lysed by 700–300 μl lysis buffer (50 mM HEPE/S/KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) with added EDTA-free protease inhibitor tablet (Roche). Cells were transferred to a fresh tube with 500 μl 0.4 mm glass beads, and were lysed for 45 min on a vortexer with glass beads (LabReps). Lysate was recovered by centrifugation, and sonicated 3–4× on a Diagenode BioRuptor (level 5, 30 s on, 1 min off). Lyse was cleared by centrifugation at max speed in a table-top centrifuge at 4 °C. Immunoprecipitation was carried out overnight with 300 μl cleared lysate, 200 μl fresh lysis buffer, and 10 μl 200 μg α-factor anti-myc antibody (Invitrogen). 50 μl of washed Protein G sepharose beads 50% slurry was added and incubated for 2 h. The beads were washed and eluted as previously described37. Crosslinking was reversed by incubating 16 h at 65 °C, and immunoprecipitated DNA was recovered with the MinElute kit (Qiagen). ChIP-Seq libraries were prepared using the NEBNext ChiP-Seq Library Prep kit for Illumina (New England Biolabs) and AMPure XP magnetic beads (Beckman Coulter, Inc.). Size distribution of the libraries was determined with a BioAnalyzer 2100 (Agilent). Libraries were pooled and sequenced on a HiSeq 2500 (Illumina) through the UCSF Center for Advanced Technology (cat.ucsf.edu). Quality of reads was checked using FastQC (http://www.bioinformatics.babraham.ac.uk) http://www.bioinformatics.babraham.ac.uk). Reads were mapped to the genome using Bowtie 2*. Type filters were manipulated using SAMtools38, and peaks were called using MACS39. Coverage and peak call visualization was visualized in MochiView40. The ChIP experiment was performed with a total of three replicates on two separate days, and a single replicate was performed of the untagged control.

Reporter assays. K. lactis reporter strains were grown as 1 ml cultures in SD medium overnight for 14 h in a 96-well plate. Cells were pelleted, resuspended in phosphate starvation medium, and diluted to OD600 = 0.25. Cells were grown 6 h, then induced with 6.25 μM α-factor (or DMSO for control cells) for 4 h. Cells were diluted tenfold into the same media before measuring GFP fluorescence on a BD LSR II Flow Cytometer. A total of 10,000 cell measurements were taken for each strain. Cells were gated to exclude debris and the mean of the cell population was used for further analysis. The autofluorescence level of a cell containing no reporter was subtracted from each reporter strain, and the standard deviation of each reporter strain was added to the standard deviation of the autofluorescence products.
strain. Values were divided by the mean autofluorescence to give an indication of the signal relative to noise. Multiple independent isolates were tested for each reporter strain, and each experiment was performed separately two or more times. In rare cases where one isolate displayed a drastic difference in expression from the others, it was tested to see whether it was a statistical outlier, and if so, was removed for subsequent experimental repetitions. Statistics for the reporter assays were conducted using an ANOVA followed by Tukey’s HSD test.

Selection on Ste12 cis-regulatory sites. The intergenic sequences for the a-specific genes were obtained from YGOB for the species S. cerevisiae, S. paradoxus, S. mikatae, S. uvarum, and S. kudriavzevii. 19 K. lactis strains were obtained from the Phaff Yeast Culture Collection (Davis, California). From these and 3 additional K. lactis strains, a-specific gene intergenic sequences were amplified from both directions using ExTaq (Takara) and Sanger sequenced. The intergenic regions from the Saccharomyces sensu stricto and K. lactis species complex were then analysed separately as follows. For each species or isolate, the a-specific gene intergenic regions were scored for the presence of strong and weak Ste12 cis-regulatory sites. The intergenic regions were aligned with MUSCLE\(^\text{48}\), and concatenated for tree building. The appropriate nucleotide evolution model was selected as previously described\(^\text{44}\); the HKY + G model was selected for the sensu stricto species, and the GTR + G model was selected for K. lactis. These models output the estimated evolutionary rate for each site in the alignment and each site was categorized as being part of a consensus or mismatched Ste12 cis-regulatory site in at least one species, or no site at all. Then, the trees generated by this method were used as input to test for different evolutionary rates among the different categories using the RPHAST package. The null model was generated using phyloFit on the intergenic regions lacking mismatched and consensus Ste12 sites using the ‘HKY85’ model for the sensu stricto species and the ‘REV’ model for K. lactis. Then, the mismatched and consensus Ste12 sites were tested for increased conservation in comparison the null model using phyloP.

31. Tirosi, I., Weinberger, A., Bezaei, D., Kaganovich, M. & Barkai, N. On the relation between promoter divergence and gene expression evolution. Mol. Syst. Biol. 4, 159 (2008).

32. Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Natural history and evolutionary principles of gene duplication in fungi. Nature 449, 54–61 (2007).

33. Gordon, J. L. et al. Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. Proc. Natl Acad. Sci. USA 108, 20024–20029 (2011).

34. Lusk, R. W. & Eisen, M. B. Spatial promoter recognition signatures may enhance transcription factor specificity in yeast. PLoS ONE 8, e53778 (2013).

35. O’haeseeleer, P. What are DNA sequence motifs? Nature Biotechnol. 24, 423–425 (2006).

36. Wach, A. PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in S. cerevisiae. Yeast 12, 259–265 (1996).

37. Cormack, B. P. et al. Yeast-enhanced green fluorescent protein (YEGFP): a reporter of gene expression in Candida albicans. Microbiology 143, 303–311 (1997).

38. Tagwerker, C. et al. Hb tag modules for PCR-based gene tagging and tandem affinity purification in Saccharomyces cerevisiae. Yeast 23, 623–632 (2006).

39. Kooistra, R., Hooykaas, P. J. J. & Steensma, H. Y. Efficient gene targeting in Klyveromyces lactis. Yeast 21, 781–792 (2004).

40. Gojkovic, Z., Jahnke, K., Schnackerz, K. D. & Piskur, J. PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. J. Mol. Biol. 295, 1073–1087 (2000).

41. Mitrović, Q. M., Tuch, B. B., Guthrie, C. & Johnson, A. D. Computational and experimental approaches double the number of known introns in the pathogenic yeast Candida albicans. Genome Res. 17, 492–502 (2007).

42. Booth, L. N., Tuch, B. B. & Johnson, A. D. Intercalation of a new tier of transcription regulation into an ancient circuit. Nature 468, 959–963 (2010).

43. Borneman, A. R. et al. Target hub proteins serve as master regulators of development in yeast. Genes Dev. 20, 435–448 (2006).

44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359 (2012).

45. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

46. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).

47. Homann, O. R. & Johnson, A. D. MochiView: versatile software for genome browsing and DNA motif analysis. BMC Biol. 8, 49 (2010).

48. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004).

49. Posada, D. & Crandall, K. A. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). Mol. Biol. Evol. 18, 897–906 (2001).
Extended Data Figure 1 | Conservation of Ste12 function. Expression of the asg STE2 and the general pheromone-activated gene FUS3 in the presence of K. lactis Ste12, in the absence of Ste12, and in the presence of S. cerevisiae Ste12. Expression under uninduced and pheromone-induced conditions are shown as mean fluorescence ± s.d. of three independent genetic isolates, grown and analysed separately.
Extended Data Figure 2 | Distribution of Ste12 cis-regulatory sites. Individual a-specific genes and general pheromone-activated genes were scored for the presence of Ste12 cis-regulatory motifs in their upstream regulatory regions. Shown is the number of consensus Ste12 sites within 600 bp of the translation start site. Orthologues that could not be identified in a given species are shown in grey. The enrichment for the Ste12 motif in each set of genes in each species is shown in Fig. 1c.
Extended Data Figure 3 | Distribution of mismatched Ste12 cis-regulatory sites. Asgs were scored for the presence of mismatched Ste12 binding sites in their upstream regulatory regions. The enrichment of the Ste12 site in the asgs in each species is shown in the bottom row.
Extended Data Figure 4 | Ste12 sites are required in *S. cerevisiae* asgs. The *S. cerevisiae* STE2 promoter was fused to GFP and the role of Ste12 cis-regulatory sites in expression. Expression under uninduced and pheromone-induced conditions are shown as mean fluorescence ±s.d. of three independent genetic isolates, grown and analysed separately. Symbols are described in Fig. 2d.
Extended Data Figure 5 | Mcm1 sites are required in *K. lactis* asgs. The *K. lactis* STE2 GFP reporter was tested with a mutated Mcm1 *cis*-regulatory site. Shown is mean fluorescence ± s.d. of three independent genetic isolates, grown and analysed separately. Symbols are described in Fig. 3d.
Extended Data Figure 6 | a2 sites are sufficient for pheromone activation. The *K. lactis* STE2 GFP reporter and heterologous reporter containing the a2-Mcm1 binding site were transformed into wild type and ste12Δ cells. Shown is mean fluorescence ±s.d. of three independent genetic isolates, grown and analysed separately.
Extended Data Figure 7 | Mcm1 sites do not compensate for the loss of a2.

a, The presence of the Mcm1 binding site is unchanged across the Candida, Kluyveromyces, and Saccharomyces clades.\textsuperscript{22,24} a\textsubscript{gs} were scored for the strength of the Mcm1 binding site in their upstream regulatory regions. The enrichment of the strength of the Mcm1 site in the a\textsubscript{gs} in each species is shown in the bottom row. b, A construct with increased Mcm1 binding site strength, marked with an ‘S’, was tested for its ability to compensate for the deletion of a2. Shown is the mean fluorescence ± s.d. of three independent genetic isolates, grown and analysed separately.
Extended Data Figure 8 | Loss of Ste12 sites decreased in the *Saccharomyces* clade. The yeast phylogeny along with the number of Ste12 binding sites in extant species was used to estimate the gain and loss rate of the sites over evolutionary time. The *Saccharomyces* clade was allowed to have different rates (orange) \((\text{LRT}, \chi^2 = 49.33, \text{df} = 2, P = 1.94 \times 10^{-11})\). The gain and loss rate units are cis-regulatory sites per amino acid substitution in the species phylogeny.
Extended Data Figure 9 | Saccharomyces Ste12 binding sites evolve under purifying selection. The evolutionary rate of nucleotides in Ste12 binding sites was compared to rates in the rest of the upstream regulatory regions of the α-specific genes, shown as violin plots. Promoters between closely related species were aligned and the evolutionary rate of each base pair in the alignment was determined after model selection.