ARTICLES

Evolution of alternative transcriptional circuits with identical logic

Annie E. Tsong[1,2]*†, Brian B. Tuch[1,2]*, Hao Li[1] & Alexander D. Johnson[1,2]

Evolution of gene regulation is an important contributor to the variety of life. Here, we analyse the evolution of a combinatorial transcriptional circuit composed of sequence-specific DNA-binding proteins that are conserved among all eukaryotes. This circuit regulates mating in the ascomycete yeast lineage. We first identify a group of mating genes that was transcriptionally regulated by an activator in a fungal ancestor, but is now transcriptionally regulated by a repressor in modern bakers' yeast. Despite this change in regulatory mechanism, the logical output of the overall circuit remains the same. By examining the regulation of mating in modern yeasts that are related to different extents, we deduce specific, sequential changes in both *cis*- and *trans*-regulatory elements that constitute the transition from positive to negative regulation. These changes indicate specific mechanisms by which fitness barriers were traversed during the transition.

Darwinian evolution posits that natural selection, acting on heritable, random, 'successive, slight variations' in organisms over billions of years, can result in new biological features[1]. Although recent work has revealed that biological novelty is often attributable to changes in transcriptional regulation[2–6], detailed analyses of such changes are often limited to a subset of the *cis*- or *trans*-elements involved[7–13]. Here, we present a step-by-step analysis of evolution in a combinatorial transcriptional circuit that regulates mating in many yeast species of the ascomycete lineage.

Mating type in the yeasts *Saccharomyces cerevisiae* and *Candida albicans* is controlled by a segment of DNA called the MAT locus[14–16]. The MAT locus exists in two versions, MAT**a** and MATα, each of which encodes unique sequence-specific DNA-binding proteins that direct an extensive program of gene transcription. Cells that express only the MAT**a**- or MATα-encoded DNA-binding proteins are **a**-cells and α-cells, respectively, and are specialized for mating. The **a**-cells express **a**-specific genes (**a**sgs), which are required for **a**-cells to mate with α-cells. Likewise, α-cells express the α-specific genes (αsgs). The third cell type, **a**/α, is formed when an **a**-cell mates with an α-cell. These cells do not mate, because the **a**sgs and αsgs are turned off (Fig. 1a).

Although this strategy is the same for both *S. cerevisiae* and *C. albicans*, the molecular details differ in a remarkable way[16]. In *S. cerevisiae*, the **a**sgs are on by default, and are repressed in α- and **a**/α-cells by a homeodomain protein (α2) that is encoded by MATα. In *C. albicans*, however, the **a**sgs are off by default, and are activated in **a**-cells by an HMG-domain protein (**a**2) that is encoded by MAT**a** (Fig. 1a). Both molecular mechanisms give the same logical output: **a**sgs are expressed only in **a**-cells. As the **a**2-activation mode is found over a broad phylogenetic range of yeasts, this strategy most likely represents the ancestral state[16–21] (Fig. 1b). By contrast, the **a**2 gene was recently lost in the *S. cerevisiae* lineage, which now uses the α2-repressing mode of **a**sg regulation[22], indicating that α2-mediated repression of **a**sgs is a recent innovation.

The evolutionary transition from positive to negative regulation of the **a**sgs has necessarily included at least two steps: (1) **a**sg expression becoming independent of the activator **a**2, and (2) **a**sgs coming under negative control by α2. We have used experimental and informatic approaches to identify multiple changes in *cis*- and *trans*-elements that underlie these steps; we also infer the order in which these steps probably occurred.

Identification of ancestral a-specific genes

To understand how α2 came to repress the **a**sgs in *S. cerevisiae*, we first sought the ancestral *cis*-element that was responsible for positive regulation of **a**sgs by **a**2. We reasoned that extant yeasts that retain the ancestral regulatory logic, such as *C. albicans*, might also have retained *cis*-elements close to the ancestral form. *C. albicans*, a fungal pathogen of humans, last shared a common ancestor with *S. cerevisiae* 200–800 million years ago[16,23,24].

We first experimentally identified the **a**sgs in *C. albicans* by comparing the transcriptional profiles of pheromone-induced **a**-cells to that of pheromone-induced α-cells (Fig. 2; for experimental details, see Methods and Supplementary Fig. 1)[16,25,26]. This comparison revealed a group of six genes that were induced only in **a**-strains (Fig. 2c). Below, we show that expression of the gene *STE2* is also specific to **a**-cells. Of these seven genes, four have orthologues previously classified as **a**-specific in *S. cerevisiae* (*ASG7*, *BAR1*, *STE2* and *STE6*), indicating that they were **a**-specific in the common ancestor of *S. cerevisiae* and *C. albicans*.

Identification of *C. albicans* asg regulatory sequences

To identify *cis*-elements that are involved in activation by **a**2, we submitted *C. albicans* **a**sg promoters (1,000 base pairs (bp)) to the motif-finding program MEME[27]. In the promoters of six **a**sgs, we found a regulatory element with several distinctive features (Fig. 3a). First, at 26-bp long, the element is more specified than the typical eukaryotic *cis*-acting sequence. Second, the sequence contains a region that closely resembles the binding site of Mcm1, a MADS box sequence-specific DNA-binding protein that is expressed equally in all three mating types, and is required for the regulation of both **a**sgs and αsgs in *S. cerevisiae*. The Mcm1 residues that contact DNA[28,29] are fully conserved between *C. albicans* and *S. cerevisiae*, strongly implicating this region of the element as a binding site for Mcm1 in *C. albicans*. Third, the putative Mcm1 site in *C. albicans* **a**sg

[1]Department of Biochemistry & Biophysics, [2]Department of Microbiology & Immunology, University of California San Francisco, San Francisco, California 94143-2200, USA.
†Present address: Department of Molecular and Cell Biology, University of California Berkeley, Lawrence Berkeley National Labs, 1 Cyclotron Road, Mailstop 84-355, Berkeley, California 94703, USA.
*These authors contributed equally to this work.

promoters lies next to a motif of the consensus sequence CATTGTC (Fig. 3a). The spacing between this motif and the Mcm1 site is always 4 bp. This motif is similar to demonstrated binding sites for **a**2 orthologues in *Schizosaccharomyces pombe* and *Neurospora crassa*, and to the α2 monomer site of *S. cerevisiae*[20,30–32] (Fig. 3b).

Experimental validation of *C. albicans* asg regulatory sequence

To test whether the motif upstream of *C. albicans* **a**sgs is functional, we fused a wild-type or mutant fragment of the *STE2* promoter to a green fluorescent protein (GFP) reporter[33] (Fig. 3c). In the mutant promoter, the conserved motif was mutated from CATTGTC to CATAATC, a change that is predicted to destroy the **a**2-binding site. The wild-type promoter activated GFP on exposure to α-factor (Fig. 3c), whereas the mutant promoter showed no induction of GFP (Fig. 3d), demonstrating that this *cis*-element is required for **a**2-dependent activation of **a**sgs.

Analysis of *cis*- asg regulation across species

For ancestral **a**sgs to undergo the transition from positive to negative regulation, **a**2-bound *cis*-elements were probably lost, whereas α2-bound elements must have been gained. To investigate when this transition occurred, we first inferred a phylogeny of 16 yeast species whose genomes have been sequenced, then identified orthologues of the **a**sgs of *C. albicans* and *S. cerevisiae* in all 16 yeasts[34,35] (Fig. 4b, see Methods). Position-specific scoring matrices (PSSMs) constructed from the *S. cerevisiae* or *C. albicans* **a**sg operators (Fig. 4a) were used to scan the promoters of each **a**sg orthologue. Maximum $\log_{10}$-odds scores are shown in Fig. 4c, d.

S. cerevisiae-like **a**sg operators (an Mcm1 site flanked by two α2-binding sites) were clearly found in orthologous promoters of organisms as far diverged as *Saccharomyces castellii*. In further diverged organisms, the presence of an *S. cerevisiae*-like **a**sg operator was diminished, although it was found in some *Candida glabrata*, *Kluyveromyces lactis*, *Eremothecium gossypii* and *Kluyveromyces waltii* promoters (Fig. 4c). The *C. albicans* PSSM yielded a nearly converse pattern (Fig. 4d). Organisms that branch with *C. albicans* have *C. albicans*-like **a**sg operators (an Mcm1 site flanked by a single **a**2 site); however, this matrix recovered no significant matches in species close to *S. cerevisiae*, correlating with the loss of **a**2 (ref. 22). These results are unchanged by recently proposed alternative phylogenetic topologies[36].

Identification of the asg operator in the *K. lactis* branch

Neither the *C. albicans* matrix nor that of *S. cerevisiae* elicited strong matches in the *K. lactis*-branch yeasts, which share a more recent common ancestor with *S. cerevisiae* than does *C. albicans* (Fig. 4b). To determine independently whether this lineage has a unique **a**sg operator, we submitted promoters of the ancestral **a**sg orthologues (*ASG7*, *BAR1*, *STE2* and *STE6*) from the *K. lactis* branch yeasts to MEME. The highest scoring hit was a DNA motif with features in

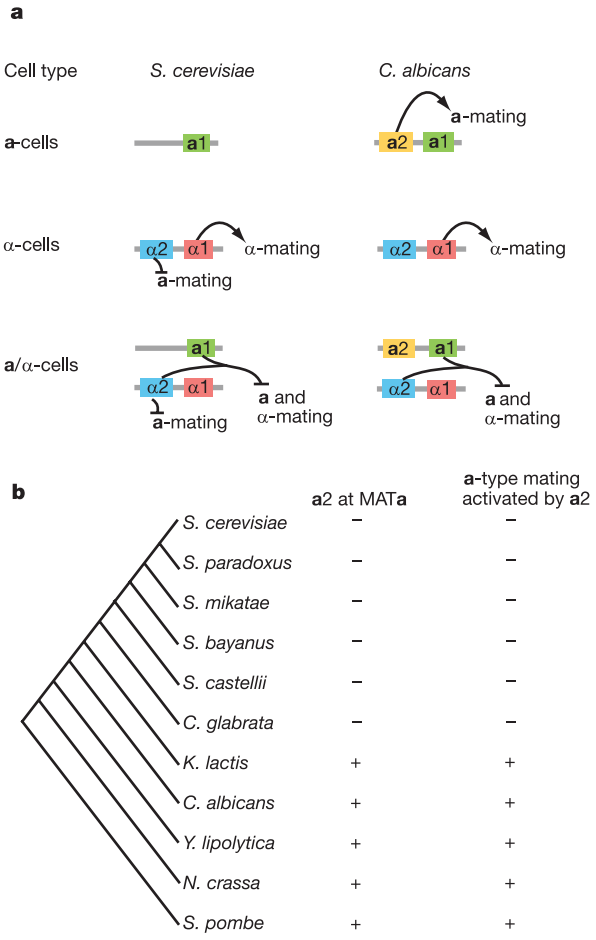

Figure 1 | a-type mating is negatively regulated in modern *S. cerevisiae*, but was positively regulated in its ancestor. **a**, *S. cerevisiae* and *C. albicans* transcribe their genes according to one of three programs, which produce the **a**-, α- and **a**/α-cells. The particular cell type produced is determined by the MAT locus, which encodes sequence-specific DNA-binding proteins (coloured blocks). Regulation of **a**-type mating differs substantially between *S. cerevisiae* and *C. albicans*. In *S. cerevisiae*, **a**-type mating is repressed in α-cells by α2. In *C. albicans*, **a**-type mating is activated in **a**-cells by **a**2. In both organisms, **a**-cells mate with α-cells to form **a**/α-cells, which cannot mate. **b**, **a**2 is an activator of **a**-type mating over a broad phylogenetic range of yeasts[16,18–21,47]. In *S. cerevisiae* and close relatives, **a**2 is missing and α2 has taken over regulation of the **a**sgs[22].

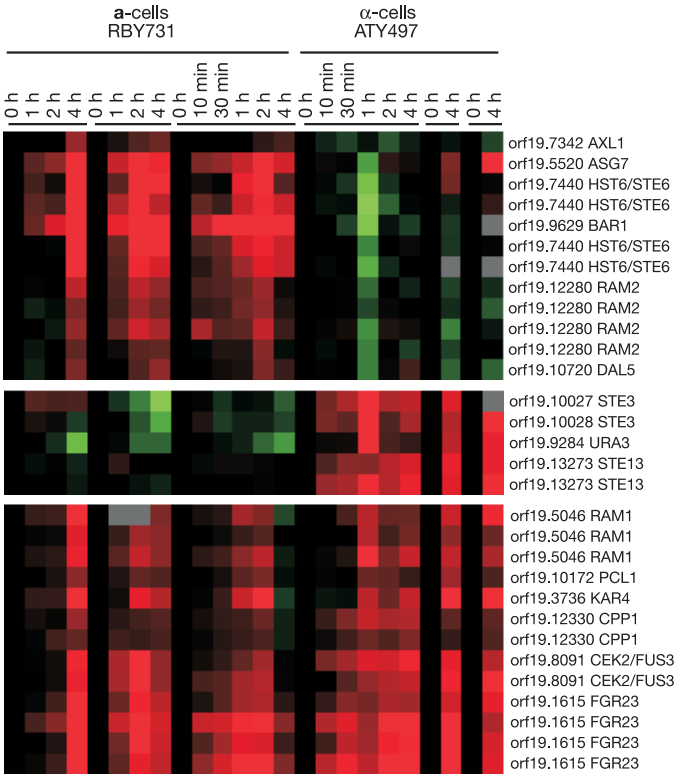

Figure 2 | Identification of a-specific genes in *C. albicans*. Pheromone induction profiles of **a**-cells (RBY731) and α-cells (ATY497) in six pheromone induction time-courses are compared. Top: genes upregulated only in **a**-cells. Middle: genes upregulated only in α-cells. *URA3* is induced because it is under the control of the *STE3* promoter. Bottom: subset of genes upregulated in both **a**- and α-cells. The first two time-courses were previously published by Bennett *et al.*[26].

common with both the *S. cerevisiae* and *C. albicans* **a**sg operators, indicating that it might be a transitional form. As in *C. albicans*, this motif contains an Mcm1 site flanked by an **a**2 site on one side. However, it is also defined on the opposite side, resembling the tripartite structure of the *S. cerevisiae* operator. This additional sequence information is similar to both the *S. cerevisiae* α2- and *C. albicans* **a**2-site consensus binding sequences; moreover, the spacing from the Mcm1 binding sequence is also similar to that found in *S. cerevisiae* and *C. albicans* **a**sg operators (Fig. 4e). An independent clustering analysis of putative **a**sg operators supports the idea that there is a transitional form in the *K. lactis* branch (see Supplementary Fig. 2).

Because of low genome sequence coverage[36] we did not systematically incorporate the yeast *Saccharomyces kluyveri*, which branches near *K. lactis* and retains **a**2 (refs 11, 22, 37, 38), into our studies. However, the available sequences of **a**sg promoters from *S. kluyveri* also contain operators similar to those of the *K. lactis* branch (not shown), indicating that transitional forms of the operator might also exist in this species.

Emergence of the α2–Mcm1 interaction

Repression of the **a**sgs in *S. cerevisiae* requires a cooperative interaction between the *trans*-factors α2 and Mcm1. To determine when this interaction arose, we aligned orthologues of α2 and Mcm1 across several yeast species, then searched for conservation of the interaction interface[28,29,39] (Fig. 5a, b). The region of Mcm1 that contacts α2 is highly conserved across all species analysed (Fig. 5a). Many proteins besides α2 contact this region, so the high degree of conservation is not surprising. By contrast, the portion of α2 that contacts Mcm1 varies considerably across yeasts (Fig. 5b). A critical nine-residue 'linker' region that is required for the interaction between α2 and Mcm1 in *S. cerevisiae*[39] is highly conserved from *S. cerevisiae* to *C. glabrata*, and is also somewhat conserved in *K. lactis* and *S. kluyveri*; however, this region shows no conservation in yeasts that branch with *C. albicans*, consistent with observations that α2 is not involved in **a**sg expression in *C. albicans*[16] (Fig. 1a).

Structural homology modelling of *K. lactis* α2 and Mcm1 using the *S. cerevisiae* crystal structure[29] as a template shows that, despite several substitutions, the α2–Mcm1 interaction interfaces in *K. lactis* are fully compatible[40] (Fig. 5c); thus, the appearance of the α2–Mcm1 interaction coincides with the emergence of the tripartite, *S. cerevisiae*-like **a**sg operator in the *K. lactis* branch (Fig. 4). This suggests that the *K. lactis* **a**sg operators are bound by α2–Mcm1. We also know that *K. lactis* **a**2 is required for wild-type levels of **a**-type mating (A.E.T., unpublished work). Together, our data indicate that *K. lactis* **a**sgs are controlled by both α2 and **a**2 through one of three possible scenarios: (1) some operators are bound exclusively by **a**2–Mcm1 and others are bound exclusively by α2–Mcm1, (2) hybrid operators are bound by both **a**2–Mcm1 and α2–Mcm1, or (3) a combination of these.

Discussion

In this work, we identify a group of genes (the **a**sgs) that was positively regulated in an ancestral yeast, but is negatively regulated in modern *S. cerevisiae*. Orthologues of these genes are required for sexual differentiation in fungal lineages that are proposed to span up to 1.3 billion years of evolution[24,41]. We identify specific changes in *cis*- and *trans*-elements that underlie the two critical steps in this transition: (1) **a**sg expression becoming independent of the activator **a**2, and (2) **a**sg expression coming under negative control of α2. The nature of these changes provides a plausible explanation for how fitness barriers were overcome during the regulatory transition, both in terms of the smaller-scale challenges of evolving individual protein–protein and protein–DNA interactions, and as regards the larger-scale challenge of maintaining appropriate **a**sg regulation throughout the transition.

Independence of asg expression from the activator a2. During the transition from positive to negative regulation of the **a**sgs, **a**sg expression became independent of the activator **a**2. We have

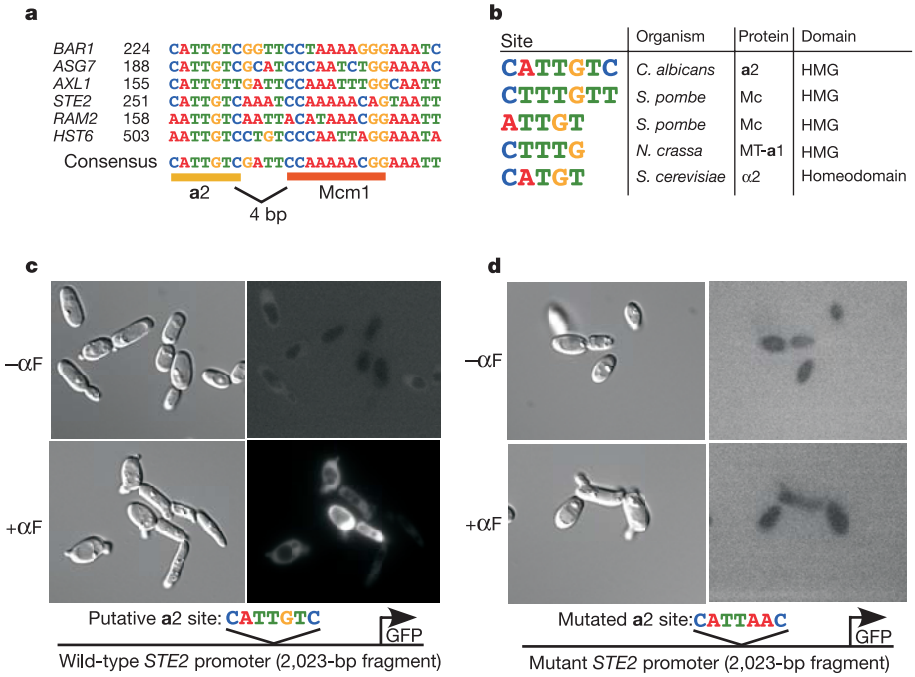

Figure 3 | Identification and validation of the *C. albicans* asg operator. **a**, 1,000 bp of each *C. albicans* asg promoter were submitted to MEME[27]. The motif shown was present in six asg promoters. Distance from the translation start site is indicated. The element contains a conserved 7-bp site (yellow) and a putative Mcm1 binding site (orange), separated by 4 bp. **b**, The 7-bp motif is similar to binding sites of a2 orthologues from *N. crassa* and *S. pombe*, as well as α2 from *S. cerevisiae*[20,30–32]. *S. pombe* MatMc binds the two indicated sites equally[30]. **c**, **d**, A wild-type (**c**) or mutant (**d**) 2,023-bp fragment of the *STE2* promoter was fused to a GFP reporter and integrated at the RP10 locus of *C. albicans*[33]. Top panels: uninduced cells. Bottom panels: α-factor induction. Only the wild-type *STE2* promoter activates GFP expression (bottom right panels).

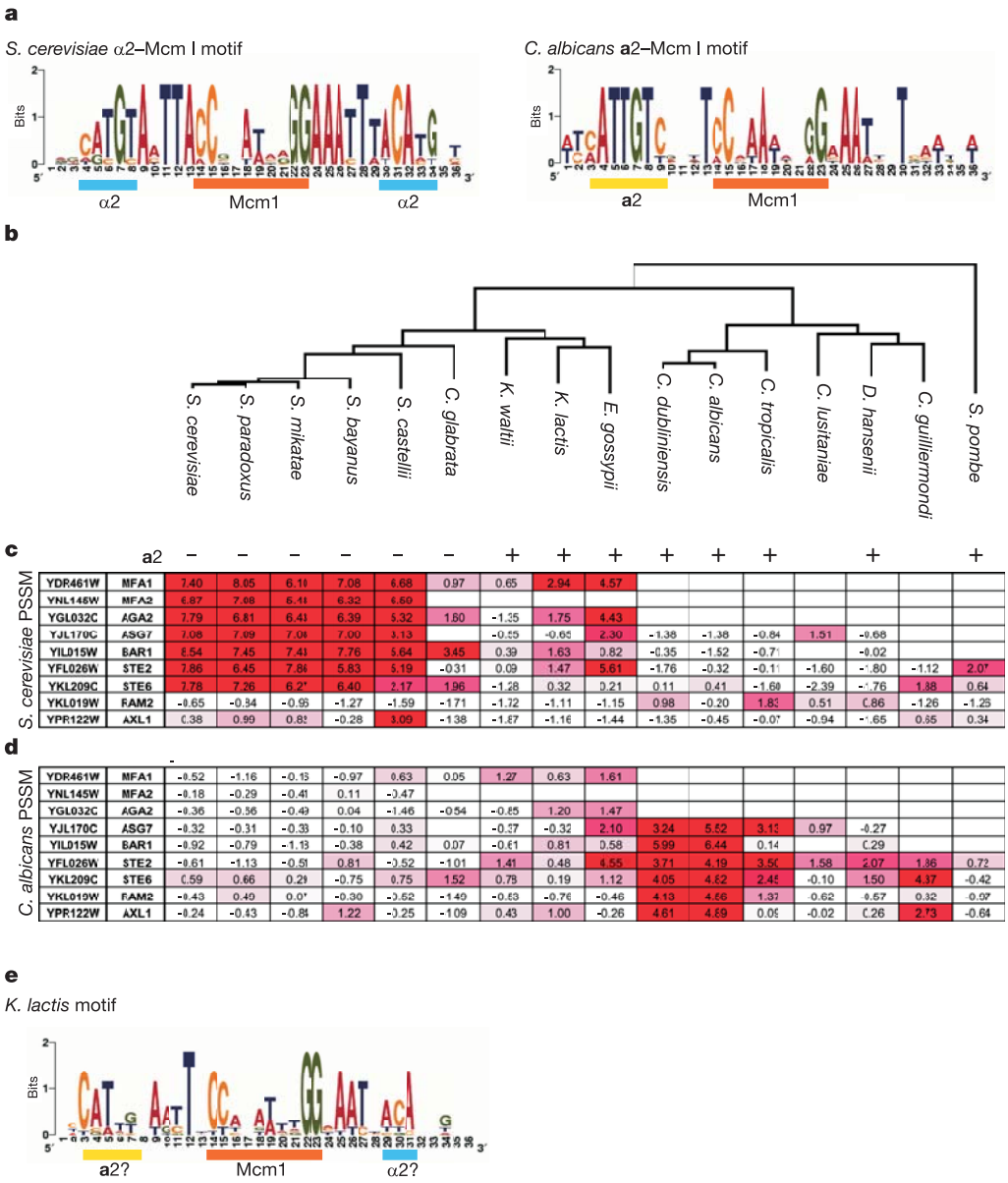

Figure 4 | **Analysis of *cis*- asg regulation across species. a**, *S. cerevisiae* α2–Mcm1 and *C. albicans* a2–Mcm1 position-specific scoring matrices (PSSMs) were derived from the seven *S. cerevisiae* asg operators, or six *C. albicans* asg operators. **b**, A phylogeny of 16 sequenced yeasts was inferred using methods similar to those of Rokas *et al.*[34]. asg orthologue promoters were scanned with the *S. cerevisiae* PSSM (**c**) or *C. albicans* PSSM (**d**). Maximum $\log_{10}$-odds scores are shown. Darker shades of red indicate stronger matches. **e**, Promoters from the *K. waltii*, *K. lactis* and *E. gossypii* orthologues of *ASG7*, *BAR1*, *STE2* and *STE6* were pooled and submitted to MEME[27]. The recovered motif has elements of both the *S. cerevisiae* and *C. albicans* asg operators: an a2-like site resembles that of *C. albicans*, and the tripartite structure resembles the *S. cerevisiae* operator.

shown that the transcriptional regulator Mcm1 was present at ancestral **a**sg promoters as a co-activator with **a**2; in *S. cerevisiae*, Mcm1 is also present at **a**sg promoters, serving as both an activator (on its own) and a co-repressor (with α2). In *S. cerevisiae*, high A/T content surrounding the Mcm1 binding site allows Mcm1 to function without a cofactor[42]. Therefore, a simple increase in the A/T content surrounding the ancestral Mcm1 binding site could 'tune up' existing Mcm1 activity so that it no longer requires the cofactor **a**2 to activate transcription. Consistent with this idea, the A/T content flanking Mcm1 sites in *S. cerevisiae* **a**sg operators is far higher than that flanking Mcm1 sites in *C. albicans* **a**sg operators (Fig. 4a).

Establishment of asg **repression by α2.** On its own, an increase in A/T content flanking the Mcm1 site would lead to inappropriate constitutive activation of the **a**sgs, as Mcm1 is expressed equally in all cell types. However, **a**sg regulation could be maintained if this increase were accompanied by evolution of α2-mediated repression. Indeed, this is precisely what we observe: *cis*- and *trans*-changes, signifying the emergence of α2-mediated repression in the *K. lactis* branch, accompany the increase in A/T content surrounding the Mcm1 site (Figs 4e, 5). Previous involvement of Mcm1 in **a**sg regulation probably assisted in the evolution of α2-mediated repression by increasing the number of surfaces available for α2-promoter interaction to include both protein and DNA.

The similarity of the **a**2-binding site (CATTGTC) to the α2-binding site (CATGT), in both sequence and spacing from the Mcm1 site, no doubt contributed to the evolution of the *S. cerevisiae* **a**sg operator (Figs 3, 4); a small change to the *cis*-element could convert it from an **a**2- to an α2-recognition sequence. The similarity of the sites is particularly striking, given that **a**2 and α2 belong to different protein families (the HMG and homeodomain families, respectively).

Ordering the pathway. An important clue as to the order in which individual *cis*- and *trans*- changes occurred comes from *K. lactis* and

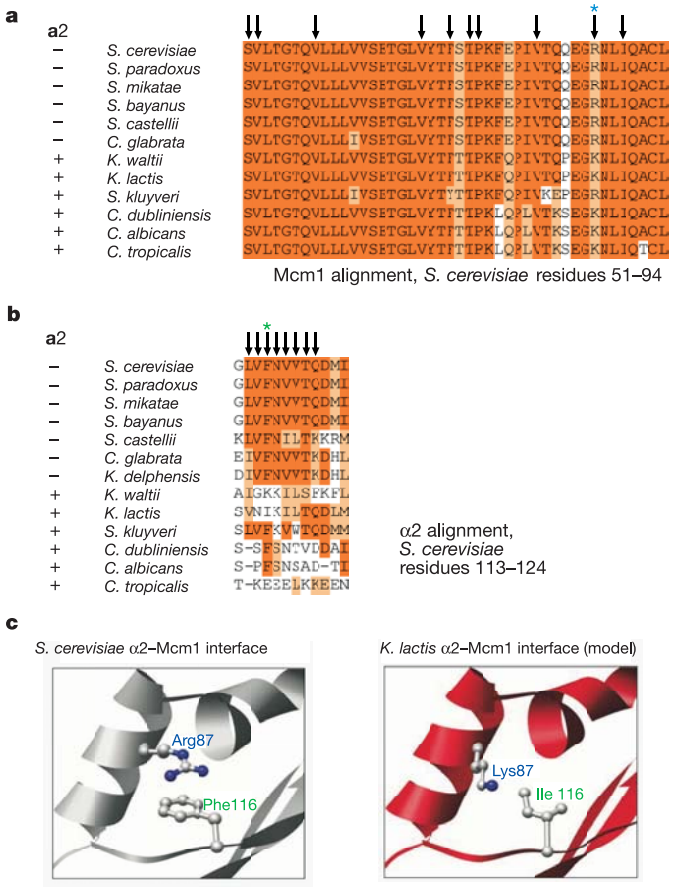

Figure 5 | Evolution of the α2–Mcm1 interaction. **a**, Mcm1 sequences from 12 species are aligned. Arrows denote residues of Mcm1 that contact α2 in *S. cerevisiae*[28,29]. **b**, α2 sequences from 13 species are aligned. Arrows indicate residues of α2 that contact Mcm1, and are required for α2–Mcm1 repression[29,39]. This region is well conserved out to *C. glabrata*, with *K. lactis* and *S. kluyveri* α2 also showing significant conservation. **c**, The *K. lactis* α2–Mcm1 complex was modelled using the crystal structure of the *S. cerevisiae* α2–Mcm1 complex (PDB ID: 1MNM; Tan S 1998) as a template. Left: *S. cerevisiae* α2 linker region and Mcm1 interface. Mcm1–Arg87 (blue asterisk of **a**) and α2–Phe116 (green asterisk of **b**) form a favourable pi-stacking interaction. Right: *K. lactis* model. The Arg87–Phe116 interaction is not present, indicating that the *K. lactis* interaction is probably weaker than that of *S. cerevisiae*.

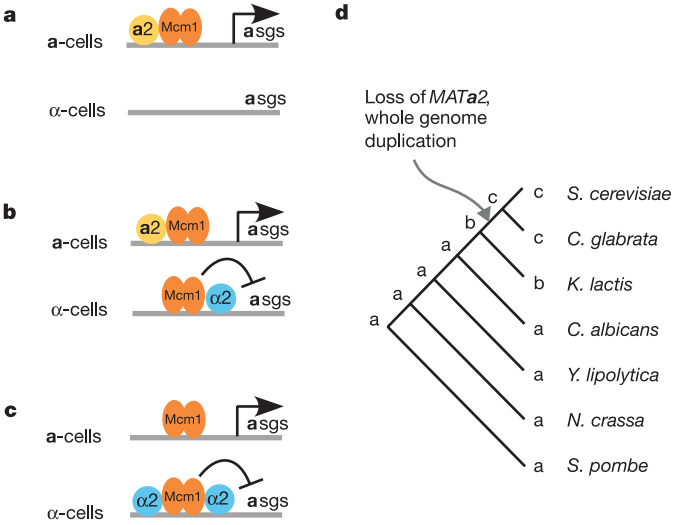

Figure 6 | Ordering the changes in *cis*- and *trans*-regulatory elements. **a**, In an ancestral yeast, a2–Mcm1 activated asgs in a-cells. This scheme persists in modern *C. albicans*. **b**, *cis*- and *trans*-elements in the *K. lactis* branch suggest that asgs are positively regulated by a2–Mcm1 in a-cells and negatively regulated by α2–Mcm1 in α-cells. **c**, In modern *S. cerevisiae*, asgs are activated by Mcm1 in a-cells and repressed by α2–Mcm1 in α-cells. **d**, The regulatory schemes shown in **a**–**c** are mapped onto extant species and ancestral nodes. Species from *C. albicans* to *S. pombe* most closely resemble **a**[16–19], whereas *K. lactis* fits **b** and *S. cerevisiae* and *C. glabrata* fit **c**. The most parsimonious evolutionary scenario maps scheme **a** as the ancestral state. Scheme **b** is transitional, first appearing in the ancestor of *K. lactis* and *S. cerevisiae*. Scheme **c** is the most derived, appearing in the ancestor of *C. glabrata* and *S. cerevisiae*.

S. kluyveri. Both yeasts have retained **a**2 at their MAT**a** loci, but in both yeasts an α2–Mcm1 interaction interface and a tripartite **a**sg operator similar to the *S. cerevisiae* α2–Mcm1 binding site have emerged. By examining the data in a phylogenetic context (Fig. 6d), we can tentatively define the succession of events leading to repression of **a**sgs in modern *S. cerevisiae* as follows. First, **a**2–Mcm1 activated **a**sgs in an ancestor (Fig. 6a, d). Subsequently, the α2–Mcm1 protein interaction evolved, coincident with evolution of an α2 site and a strengthening of the Mcm1 binding site in the **a**sg operator (Fig. 6b, d). After the divergence of *K. lactis*, the α2–Mcm1 *cis*-operator specificity and A/T content were increased, and **a**2 was lost, completing the hand-off from positive to negative control (Fig. 6c, d). A crucial feature of this model is that **a**sgs are appropriately regulated throughout each stage of circuit evolution, a condition made possible by the continued presence of Mcm1. Intriguingly, both the loss of **a**2 and the conversion of **a**sg regulation to an exclusively negative regulatory scheme coincide with a whole-genome duplication[43,44]. The evolution of the **a**sg regulatory circuit might have been facilitated in part by greater flexibility in **a**sg regulation conferred by duplication of its component *cis*- and *trans*-elements.

Conclusion. Our analysis shows how a concerted series of subtle changes in *cis*- and *trans*-elements can lead to a profound evolutionary change in the wiring of a combinatorial circuit. These changes include: (1) 'tuning up' of a binding site for a ubiquitous activator, making gene expression independent of a cell-type-specific activator; (2) a small change in an existing DNA-binding site, converting its recognition from one protein to that of an unrelated protein; (3) a small change in the amino-acid sequence of a sequence-specific DNA-binding protein, allowing it to bind DNA cooperatively with a second protein. Significantly, the coordinated optimization of protein–DNA and protein–protein interactions that we have described allows regulation of the target genes to be maintained throughout a major evolutionary transition. Because the proteins that have participated in this transition represent several highly conserved and prominent protein families, including the MADS box family (Mcm1), the HMG-domain family (**a**2) and the homeodomain family (α2), the types of change we have described are likely to apply to other examples of transcriptional circuit evolution.

METHODS

Detailed information on Methods is described in the Supplementary Information.

Strain construction. The pheromone **a**-factor has not yet been identified in *C. albicans*. To compare the pheromone response of **a**-cells to that of α-cells, we 'fooled' α-cells into responding to α-factor by ectopically expressing the α-factor receptor (strain ATY497), a strategy previously employed in *S. cerevisiae*[25]. Constructs and primers used are listed in Supplementary Information.

Induction of α-factor. Strains were grown to an optical density (OD_{600}) of 1.0 in YEPD plus 55 μg ml^{-1} adenine, then induced with 10 μg ml^{-1} α-factor from a stock dissolved in either dimethylsulphoxide (DMSO) or water. Sample preparation and microarrays were as previously described[26]. All microarray data are available online (http://genome.ucsf.edu/asg_evolution/).

Yeast phylogeny. Briefly, groups of orthologous genes (see Supplementary Information) with one and only one representative from each of the 16 yeasts were multiply aligned with ClustalW[45], then concatenated to yield a single alignment. A maximum-likelihood species tree was inferred from this alignment using the TREE-PUZZLE algorithm[46]. Trees with identical topologies were also

generated using additional algorithms (see Supplementary Information).

Structural modelling. The *K. lactis* α2–Mcm1 interaction was modelled using the Protein Local Optimization Program, by M. P. Jacobson, Department of Pharmaceutical Chemistry, University of California San Francisco, USA (http://francisco.compbio.ucsf.edu/~jacobson/), using the crystal structure of the *S. cerevisiae* α2–Mcm1 complex (PDB ID: 1MNM; Tan S 1998) as a template.

Received 24 February; accepted 21 July 2006.

1. Darwin, C. R. *The Origin of Species* (Gramercy, New York, 1859).
2. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity* (Blackwell Science, Malden, Massachusetts, 2001).
3. Davidson, E. H. *Genomic Regulatory Systems* (Academic, San Diego, California, 2001).
4. Gerhart, J. & Kirschner, M. *Cells, Embryos, and Evolution* (Blackwell Science, Malden, Massachusetts, 1997).
5. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424,** 147–151 (2003).
6. Doebley, J. & Lukens, L. Transcriptional regulators and the evolution of plant form. *Plant Cell* **10,** 1075–1082 (1998).
7. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433,** 481–487 (2005).
8. Ihmels, J. *et al.* Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science* **309,** 938–940 (2005).
9. Ludwig, M. Z., Patel, N. H. & Kreitman, M. Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* **125,** 949–958 (1998).
10. Ronshaugen, M., McGinnis, N. & McGinnis, W. Hox protein mutation and macroevolution of the insect body plan. *Nature* **415,** 914–917 (2002).
11. Tanay, A., Regev, A. & Shamir, R. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc. Natl Acad. Sci. USA* **102,** 7203–7208 (2005).
12. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430,** 85–88 (2004).
13. Galant, R. & Carroll, S. B. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415,** 910–913 (2002).
14. Hull, C. M., Raisner, R. M. & Johnson, A. D. Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. *Science* **289,** 307–310 (2000).
15. Herskowitz, I., Rine, J. & Strathern, J. *Mating-type determination and Mating-type interconversion in* Saccharomyces cerevisiae (eds Jones, E. W., Pringle, J. R. & Broach, J. R.) 583–656 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1992).
16. 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).
17. Staben, C. & Yanofsky, C. *Neurospora crassa a* mating-type region. *Proc. Natl Acad. Sci. USA* **87,** 4917–4921 (1990).
18. Kelly, M., Burke, J., Smith, M., Klar, A. & Beach, D. Four mating-type genes control sexual differentiation in the fission yeast. *EMBO J.* **7,** 1537–1547 (1988).
19. Kurischko, C., Schilhabel, M. B., Kunze, I. & Franzl, E. The MATA locus of the dimorphic yeast *Yarrowia lipolytica* consists of two divergently oriented genes. *Mol. Gen. Genet.* **262,** 180–188 (1999).
20. Philley, M. L. & Staben, C. Functional analyses of the *Neurospora crassa* MT a-1 mating type polypeptide. *Genetics* **137,** 715–722 (1994).
21. Turgeon, B. G. *et al.* Cloning and analysis of the mating type genes from *Cochliobolus heterostrophus*. *Mol. Gen. Genet.* **238,** 270–284 (1993).
22. Butler, G. *et al.* Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proc. Natl Acad. Sci. USA* **101,** 1632–1637 (2004).
23. Calderone, R. A. *Candida and Candidiasis* (ed. Calderone, R. A.) (ASM, Washington DC, 2002).
24. Hedges, S. B. The origin and evolution of model organisms. *Nature Rev. Genet.* **3,** 838–849 (2002).
25. Bender, A. & Sprague, G. F. Jr. Yeast peptide pheromones, **a**-factor and α-factor, activate a common response mechanism in their target cells. *Cell* **47,** 929–937 (1986).
26. Bennett, 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).
27. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 28–36 (1994).
28. Acton, T. B., Mead, J., Steiner, A. M. & Vershon, A. K. Scanning mutagenesis of Mcm1: residues required for DNA binding, DNA bending, and transcriptional activation by a MADS-box protein. *Mol. Cell. Biol.* **20,** 1–11 (2000).
29. Tan, S. & Richmond, T. J. Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. *Nature* **391,** 660–666 (1998).
30. Kjaerulff, S., Dooijes, D., Clevers, H. & Nielsen, O. Cell differentiation by interaction of two HMG-box proteins: Mat1-Mc activates M cell-specific genes in *S. pombe* by recruiting the ubiquitous transcription factor Ste11 to weak binding sites. *EMBO J.* **16,** 4021–4033 (1997).
31. Smith, D. L. & Johnson, A. D. A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an α2 dimer. *Cell* **68,** 133–142 (1992).
32. van Beest, M. *et al.* Sequence-specific high mobility group box factors recognize 10–12-base pair minor groove motifs. *J. Biol. Chem.* **275,** 27266–27273 (2000).
33. Care, R. S., Trevethick, J., Binley, K. M. & Sudbery, P. E. The MET3 promoter: a new tool for *Candida albicans* molecular genetics. *Mol. Microbiol.* **34,** 792–798 (1999).
34. Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425,** 798–804 (2003).
35. Byrne, K. P. & Wolfe, K. H. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* **15,** 1456–1461 (2005).
36. Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301,** 71–76 (2003).
37. Belloch, C., Querol, A., Garcia, M. D. & Barrio, E. Phylogeny of the genus *Kluyveromyces* inferred from the mitochondrial cytochrome-c oxidase II gene. *Int. J. Syst. Evol. Microbiol.* **50,** 405–416 (2000).
38. Wong, S., Butler, G. & Wolfe, K. H. Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proc. Natl Acad. Sci. USA* **99,** 9272–9277 (2002).
39. Mead, J., Zhong, H., Acton, T. B. & Vershon, A. K. The yeast α2 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).
40. Jacobson, M. P. & Friesner, R. A. Protein local optimization program. ⟨http://francisco.compbio.ucsf.edu/~jacobson/plop_manual/plop_overview.htm⟩ (2006).
41. Lengeler, K. B. *et al.* Signal transduction cascades regulating fungal development and virulence. *Microbiol. Mol. Biol. Rev.* **64,** 746–785 (2000).
42. Acton, T. B., Zhong, H. & Vershon, A. K. DNA-binding specificity of Mcm1: operator mutations that alter DNA-bending and transcriptional activities by a MADS box protein. *Mol. Cell. Biol.* **17,** 1881–1889 (1997).
43. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428,** 617–624 (2004).
44. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387,** 708–713 (1997).
45. Higgins, D. G. & Sharp, P. M. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73,** 237–244 (1988).
46. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18,** 502–504 (2002).
47. Debuchy, R., Arnaise, S. & Lecellier, G. The *mat*$^-$ allele of *Podospora anserina* contains three regulatory genes required for the development of fertilized female organs. *Mol. Gen. Genet.* **241,** 667–673 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Jacobson for advice provided in modelling the *K. lactis* α2 and Mcm1 structures. We also thank S. Åstrøm for providing unpublished *K. lactis* strains and advice on their handling, P. Sudbury for providing the GFP reporter construct, M. Lorentz and G. Fink for the collaboration that produced the DNA microarrays used in this paper, and the Broad Institute (http://www.broad.mit.edu/annotation/fungi/fgi/), the Sanger Center (http://www.sanger.ac.uk/Projects/Fungi/), and the Pathogen Sequencing Unit at the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/) for making sequence data available. M. Ptashne provided valuable comments on the manuscript. We thank B. Hromatka for overseeing microarray printing, and R. Bennett for microarray data and discussions. R. Zordan, A. Uhl, M. Lohse, M. Miller, R. Wu, C. Chaivorapol and other members of the Johnson and Li labs provided useful discussions. This work was supported by grants from the NIH to A.D.J. A.E.T was supported by a Howard Hughes Medical Institute Predoctoral Fellowship. B.B.T. is an NSF Predoctoral Fellow. B.B.T. and H.L. acknowledge partial support from a Packard Fellowship in Science and Engineering (to H.L.) and an NIH grant.

Author Contributions A.E.T. determined the **a**sgs of *C. albicans*, and validated the **a**sg operator site. B.T. constructed the phylogenetic tree, analysed **a**sg operators across multiple species, and modelled the *K. lactis* α2–Mcm1 interaction. H.L. and A.D.J. oversaw the work.

Author Information Microarray data are available in Supplementary Information and at http://genome.ucsf.edu/asg_evolution/. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.D.J. (ajohnson@cgl.ucsf.edu).