UCSF
UC San Francisco Previously Published Works

Title

Protein Modularity, Cooperative Binding, and Hybrid Regulatory States Underlie Transcriptional Network Diversification

Permalink
https://escholarship.org/uc/item/1pw6m213

Journal
Cell, 151(1)

ISSN
0092-8674

Authors
Baker, Christopher R
Booth, Lauren N
Sorrells, Trevor R
et al.

Publication Date
2012-09-01

DOI
10.1016/j.cell.2012.08.018

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

NIH Public Access
Author Manuscript
Cell. Author manuscript; available in PMC 2013 September 28.

Published in final edited form as:
Cell. 2012 September 28; 151(1): 80–95. doi:10.1016/j.cell.2012.08.018.

Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification

Christopher R. Baker[*], **Lauren N. Booth**[*], **Trevor R. Sorrells**, and **Alexander D. Johnson**
Department of Microbiology and Immunology and Department of Biochemistry, The University of California, San Francisco, San Francisco, CA 94102, USA

Summary

We examine how different transcriptional network structures can evolve from an ancestral network. By characterizing how the ancestral mode of gene regulation for genes specific to **a-**type cells in yeast species evolved from an activating paradigm to a repressing one, we show that regulatory protein modularity, conversion of one *cis*-regulatory sequence to another, distribution of binding energy among protein-protein and protein-DNA interactions, and exploitation of ancestral network features all contribute to the evolution of a novel regulatory mode. The formation of this derived mode of regulation did not disrupt the ancestral mode and thereby created a hybrid regulatory state where both means of transcription regulation (ancestral and derived) contribute to the conserved expression pattern of the network. Finally, we show how this hybrid regulatory state has resolved in different ways in different lineages to generate the diversity of regulatory network structures observed in modern species.

Introduction

In many organisms, gene regulatory networks have been shown to undergo significant divergence over evolutionary time (reviewed by Carroll, 2005; Davidson and Erwin, 2006; Doebley and Lukens, 1998; Tuch et al., 2008; Wohlbach et al., 2009; Wray, 2007). In the simplest cases, the gain or loss of a *cis*-regulatory sequence upstream of a single gene can produce changes in coloration, losses of ancestral anatomical features, or altered ability to digest sugars (Chan et al., 2010; Gompel et al., 2005; Tishkoff et al., 2007). Yet, it seems likely that the evolution of complex biological innovations requires concerted evolution across entire networks of genes (Lavoie et al., 2010; Lynch et al., 2011; Tuch et al., 2008). Two considerations suggest that network evolution requires mechanisms in addition to the loss and gain of single *cis*-regulatory sequences. First, the adaptive value of acquiring coordinated expression of a large set of genes may not be realized until all or at least a large fraction of the gene set acquires the new regulatory input. Second, expression of only a portion of the gene network could be detrimental to the fitness of the organism, for example, through the non-stoichiometric expression of components of a protein complex.

© 2012 Elsevier Inc. All rights reserved.

Corresponding Author: ajohnson@cgl.ucsf.edu.
[*]These authors contributed equally and are listed alphabetically

C.R.B. performed reporter assays and the Mcm1 ancestral gene reconstruction. L.N.B. performed experiments in *L. kluyveri* and *K. wickerhamii*. L.N.B. and T.R.S. performed *K. lactis* experiments. Data was analyzed and computation experiments performed by C.R.B., L.N.B., and T.R.S. All authors contributed to the design of the study and wrote the paper.

The authors declare no conflict of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

$watermark-text

$watermark-text

$watermark-text

To understand the molecular events that underlie changes in the regulation of groups of genes, we investigated a transcriptional network that determines cell-type in a wide variety of fungal species. This network—comprised of the **a**-specific genes (**a**sgs) and their regulators—underwent a major circuit rewiring in the hemiascomycete yeasts (Tsong et al., 2003; Tsong et al., 2006). This group of yeast includes *Saccharomyces cerevisiae* (the baker's yeast), *Kluyveromyces lactis* (a dairy yeast), *Candida albicans* (the most common human fungal pathogen), and over 30 additional genome-sequenced species (Figure 1A). This lineage has been estimated to represent at least 300 million years of evolutionary time (Taylor and Berbee, 2006). Virtually all of yeast species in the hemiascomycete lineage exist in three cell types—the mating competent **a** and α cells and the product of their mating, the **a**/α cell (Figure 1B). Mating cell-type is controlled by transcriptional regulators that are encoded at the mating-type (*MAT*) locus (Herskowitz, 1989). These regulators control the expression of genes that are responsible for the specialized properties of each of the three cell types. The **a**sgs are a group of seven to ten genes (depending on the species) whose key regulatory characteristic is that they are expressed in the **a** cell-type but not in the α and **a**/α cell-types (Galgoczy et al., 2004; Herskowitz, 1989; Tsong et al., 2003) (Figure 1B). The **a**sgs encode proteins (e.g. α mating pheromone receptor, **a** mating pheromone, agglutinins and exporters) that are necessary for the specific properties of **a** cells (Herskowitz, 1989) (Madhani, 2007).

In principle, there are two ways that the **a**sgs could be expressed in **a** cells but not in the other two cell types: (1) the **a**sgs could be activated by a regulatory protein present only in **a** cells or (2) the **a**sgs could be repressed by a regulator made only in α and **a**/α cells. In fact, both schemes are observed, the latter in *S. cerevisiae* and the former in *C. albicans* and (Strathern et al., 1981; Tsong et al., 2003). In *C. albicans*, the HMG domain protein **a**2 binds to and activates the **a**sgs. In *S. cerevisiae*, the homeodomain protein α2 binds to and represses the **a**sgs (Johnson and Herskowitz, 1985). We previously showed that the activation mode of regulation (by **a**2) was present in the ancestor of *C. albicans* and *S. cerevisiae* and that the switch to the repression mode (mediated by α2) occurred along the branch to *S. cerevisiae* (Tsong et al., 2006). Indeed, the gene encoding the **a**2 protein was lost from the genome in an ancestor of *S. cerevisiae* (Butler et al., 2004). (Figure 1C)

Here we define the evolutionary path for the switch in regulation of the **a**sg network using a combination of bioinformatic analysis, direct experiments in the yeasts *Kluyveromyces wickerhamii, Kluyveromyces lactis*, and *Lachancea kluyveri,* ancestral protein reconstruction, and *trans*-species reporter gene analysis in *S. cerevisiae*. Our principle conclusions are as follows: First, regulatory protein modularity was crucial for the change in network regulation. In particular, protein modularity accounts for the cooption of an existing repressor for a new function (repression of the **a**sgs) while maintaining its ancestral function. Second, the cooperative binding of transcriptional regulators facilitated the gain of the repression mode of regulation across this gene set by stabilizing early evolutionary intermediates. Third, the conversion of one *cis*-regulatory sequence into another occurred through an "intermediate" *cis*-regulatory sequence that was recognized by regulators of both the ancestral and derived regulatory modes. Fourth, the evolution of **a**sg repression in the common ancestor of *K. lactis* and *S. cerevisiae* did not disrupt the ancestral (positive) mode of regulation, and thereby formed a "hybrid" regulatory state (Tsong et al., 2006). Finally, we show that once the hybrid regulatory network formed, it resolved in different ways along the branches to the modern yeast species: in *S. cerevisiae* the ancestral form was discarded, leaving only the derived form; in *K. lactis* the derived form was inactivated, reverting to the ancestral mode of regulation; in *L. kluyveri* and *K. wickerhamii,* aspects of the hybrid regulatory state have been maintained. Because the regulatory proteins studied here are conserved in all eukaryotes, the evolution of **a**sg regulation can serve as a model for

$watermark-text $watermark-text $watermark-text

understanding the molecular mechanisms underlying the extraordinary flexibility of transcriptional circuits over evolutionary time.

Results

α2 repression of the asgs evolved prior to the divergence of Saccharomyces and Kluyveromyces

We determined the time at which repression of the **a**sgs arose during evolutionary time. To do this, we moved the **a**sg regulatory sequences (from the conserved **a**sg *STE2*) and the α2 proteins from a variety of species into *S. cerevisiae* and determined their abilities to support repression (Fig. 2A). In *S. cerevisiae*, α2 binds **a**sg *cis*-regulatory sequences cooperatively with a MADS-box transcription regulator, Mcm1 (Figure 1C). Both proteins bind with high affinity to DNA sequences and their cooperative binding results from a relative weak protein-protein interaction (Tan and Richmond, 1998; Vershon and Johnson, 1993). The *cis*-regulatory sequence consists of an Mcm1 homodimer site flanked by two α2 binding sites (Keleher et al., 1988). Removal of any these four binding sites from an **a**-specific *cis*-regulatory sequence, or disruption of the protein-protein interaction, severely compromises repression (Smith and Johnson, 1994; Vershon and Johnson, 1993).

The *STE2 cis*-regulatory sequences from species that branch from the *S. cerevisiae* lineage prior to the loss of the **a**2 gene—such as *Zygosaccharomyces rouxii, K. lactis,* and *Ashbya gossypii*— supported levels of α2 repression comparable to the *S. cerevisiae* site (Figure 2A). *STE2 cis*-regulatory sequences taken from the *Candida* clade (*C. albicans* and *Pichia membranifaciens*) and the out-group species *Yarrowia lipolytica* failed to support repression in this assay (Figure 2A), consistent with the inference that in *C. albicans* and the *C. albicans-S. cerevisiae* ancestor, α2 does not repress the **a**sgs (Tsong et al., 2006).

Full-length α2 ORFs from 8 species were fused to the *S. cerevisiae* α2 promoter and integrated into the genome in single copy (Figure 2B). α2 orthologs from species within the *Kluyveromyces* group repressed the **a**sg reporter comparable to levels observed for the *S. cerevisiae* protein (Figure 2B). In addition, the α2 ortholog of a species (*Z. rouxii*) that branches within the *Saccharomyces* group, but prior to the loss of **a**2, (Figure 1A) efficiently repressed the **a**sg reporter (Figure 2B). In contrast, α2 orthologs from *Candida* clade species failed to repress the reporter. The *C. albicans* α2 protein also failed to repress the *C. albicans* **a**sg *cis*-regulatory sequence (Figure 2C). These results show that changes in both the **a**sg *cis*-regulatory sequences and the α2 protein were both necessary for the switch in regulation and that the gain of α2 repression of the **a**sgs clearly preceded the loss of the **a**2 gene.

The clear trend from these experiments is that **a**sg *cis*-regulatory sequences and α2 proteins from the *Saccharomyces* and *Kluyveromyces* clades (Figure 1A) are competent to bring about repression, whereas those outside these clades are not. However, there is an important exception to this observed pattern. The *K. lactis* α2 protein failed to repress in this assay even though its *STE2 cis*-regulatory sequence is competent to bring about repression in this same assay (Figure 2B). To rule out the trivial possibility that α2 was misfolded or poorly expressed, we carried out a series of control experiments (Figure S1A). We will return to this unique feature of *K. lactis* later in this paper.

The evolution of a new function for α2

To investigate the molecular events that gave rise to α2 repression of the **a**sgs, we considered first the contribution of *trans* changes (coding sequence mutations in α2 or Mcm1).To identify regions of the α2 protein that may have been critical for the gain of α2-mediated repression, we quantified the levels of conservation across the α2 protein (Figure

$watermark-text
$watermark-text
$watermark-text

3B). The α2 protein sequences from the hemiascomycete yeasts were divided into two groups: those that diverged prior to and those that diverged after the gain of α2 repression of the **a**sgs. In Figure 3B, high scores indicate conservation of those residues in the species group, whereas low scores indicate unconserved regions. Regions where the scores for the two groups are dissimilar reflect positions within α2 that experienced different levels of purifying selection in these two groups.

Much of the α2 protein has similar levels of conservation between the clades. This includes the 60 amino acid homeodomain (which mediates the sequence specific DNA-binding) (Hall and Johnson, 1987) and the 15 amino acid region of α2 that interacts with **a**1 (Mak and Johnson, 1993). DNA-binding and the interaction with **a**1 are functions of α2 that are required in all the clades considered, and their high sequences conservation reflects their high functional conservation. The α2 conservation traces diverged at two regions within the α2 protein, regions 1 and 3 (Figure 3A–C). Both regions displayed high levels of conservation in the *Saccharomyces-Kluyveromyces* lineages and low levels in the *Candida* lineage, implicating these regions in the evolution of α2 repression of **a**sgs. In fact, both regions are critical for α2 repression of the **a**sgs in *S. cerevisiae*; region 1 is responsible for recruiting the general repressor Tup1 (Komachi et al., 1994), and region 3 forms the interaction with Mcm1 (Tan and Richmond, 1998; Vershon and Johnson, 1993). The importance of the evolution of the Mcm1 interaction region in α2 (region 3) to the evolution of **a**sg repression is consistent with previous work using structural homology modeling (Tsong et al., 2006).

To test these predictions directly, we designed a series of genetic swaps between the *C. albicans* and *S. cerevisiae* α2 proteins. The *S. cerevisae* α2 protein can be divided into five functional and structural regions (Figure 3A). We individually replaced each of these five regions of *S. cerevisiae* α2 with the homologous region of the *C. albicans* α2 protein and integrated (in single copy) the fusion proteins driven by the *S. cerevisiae* α2 promoter (Figure 3D). The ability of the modified α2 protein to repress expression was monitored using a reporter with a *S. cerevisiae* **a**sg or haploid specific gene *cis*-regulatory site in the promoter.

As predicted by the bioinformatic analysis, replacement of *S. cerevisiae* region 1 (Tup1 interaction) or region 3 (Mcm1 interaction) by the equivalent *C. albicans* sequences eliminated **a**sg repression. Also, as predicted, the swap of region 3 eliminated **a**sg repression, but left intact the protein's capacity for repression of the haploid specific genes. In contrast, the α2 functional region 1 swap protein (Tup1 interaction) failed to repress either the **a**sg reporter or the haploid specific gene reporter (Figure 3D). Replacing either functional region 1 or 3 with aligning sequence from another species *(Pichia pastoris)* that diverged prior to the gain of α2 repression at the **a**sgs gave similar results (Figure S1B). These observations show that the gain of **a**sg repression required the creation of two new functional regions within α2—a region that interacts with Mcm1 and a region that interacts with Tup1. In contrast to these two regions, the rest of the *S. cerevisiae* α2 protein sequence could be swapped for the homologous sequence from *C. albicans* α2 without a substantial effect on **a**sg repression. (Figure 3D).

Are the acquisition of the Tup1 and Mcm1 interaction regions was sufficient for α2 to acquire the capability to repress the **a**sgs? We swapped these functional regions from *S. cerevisiae* α2 into the *C. albicans* α2 protein and measured the ability of these hybrids to repress an **a**sg reporter. Neither region alone "rescued" the *C. albicans* protein; however, swapping both regions into *C. albicans* α2 together conferred the ability to repress the **a**sg reporter onto the hybrid protein (Figure 3E). These results demonstrate that the failure of the *C. albicans* α2 protein to repress the **a**sg reporter in *S. cerevisiae* reflect the inability of the

$watermark-text

$watermark-text

$watermark-text

protein to productively interact with both Tup1 and Mcm1. Consistent with this conclusion, swapping both of these regions into another *Candida*-group α2 protein (this one from *P. pastoris*) also conferred the ability to repress the **a**sgs onto that hybrid protein (Figure S1C). In summary, while two regions of α2 (regions 4 & 5) have been functionally conserved over large evolutionary distances (Figure 3B & D), two other regions (regions 1 & 3) evolved more recently in the ancestor of the *Saccharomyces/Kluyveromyces* groups (Figure 3B–C). These two recent additions are sufficient for α2 to gain its new function. This analysis illustrates how the evolutionary history of the α2 protein gave rise to its modular structural organization.

We also determined whether changes in Mcm1—the binding partner of α2—contributed to the evolution of **a**sg repression. To do this, we relied on ancestral gene reconstruction, an approach proven useful for testing evolutionary predictions (Thornton, 2004). The strategy depends on the accurate protein alignments of the ortholog group of interest, followed by the calculation of amino acid probabilites at each position within the ancestral protein using a species or gene tree as a guide (Figure S2). Given the strong conservation of the Mcm1 MADS-box domain, all amino acid positions could be reconstructed within this domain with high accuracy in each ancestral protein. We synthesized a series of ancestral Mcm1 proteins and replaced the endogenous *S. cerevisiae* Mcm1 with them. Ancestral Mcm1 proteins dating back to the divergence of *S. cerevisiae*-*C. albicans* supported repression at levels equivalent to the modern *S. cerevisiae* Mcm1 (Figure S2). Thus, the gain of a new interaction between α2 and Mcm1 did not require changes in Mcm1. Instead, it appears that the evolution of the new protein-protein interaction was one-sided, with all the changes occurring in a short module of α2.

Integration of a new regulator into an existing regulatory network

Although the evolution of new protein-protein interaction modules in α2 was critical for the rewiring of the **a**sg network, the *cis*-regulatory sequences of the **a**sgs also evolved to become efficiently recognized by the α2 protein (Figure 2A). The similarities and differences between the **a**2-regulated (ancestral) and α2-regulated (derived) **a**sg *cis*-regulatory sequences have been described (Tsong et al., 2006). The most striking similarities are the presence of a binding site for Mcm1 and the close relationship between the *cis*-regulatory sequences recognized by **a**2 and α2. Despite belonging to different transcription regulator superfamilies (HMG domain for **a**2 versus homeodomain for α2), both proteins recognize a core TGT sequence, with the outer nucleotides differing in their respective binding sites (Figure 3G). A major difference between the two regulatory sequences is in their symmetries. The *C. albicans* **a**2-regulated **a**sg binding sequence contains information specifying **a**2 binding on only one side of Mcm1. The *S. cerevisiae* α2 binding sequence, however, contains information on both sides of the Mcm1 binding site, specifying the binding of an α2 monomer on either side (Johnson and Herskowitz, 1985).

In our next set of experiments, we examined in more detail the differences between the **a**2 and α2 recognition sequence and how the ancestral **a**2 site evolved to be recognized by α2. We found that *S. cerevisiae* α2 could repress *Kluyveromyces* group species **a**sg *cis*-regulatory sequences even though they varied significantly from the *S. cerevisiae* sites (Figure 3F). In fact, α2 efficiently repressed **a**sg *cis*-regulatory sequences (such as *Z. rouxii STE6* and *K. lactis STE2*) that contained precise **a**2 binding sites, as assessed by the Position Specific Scoring Matrix for **a**2 in the *Candida* clade (Figure 3G). In contrast, each **a**sg *cis*-regulatory sequence from a *Candida* group species failed to be repressed by *S. cerevisiae* α2 (Figure 3F), even when α2 was overexpressed (Figure S3). Thus, the ancestral **a**sg *cis*-regulatory sequences (recognized by **a**2) must have been converted to sites recognized by α2 along the *Saccharomyces-Kluyveromyces* lineage. To determine the minimum number

of mutations necessary to convert an **a**2 site to a functional α2 site, we mutated three positions (positions 6, 26 and 27), from the *C. albicans RAM2 cis*-regulatory site, to their counterpart in the *S. cerevisiae* consensus sequence. Mutation of two of these nucleotides generated a construct that could be repressed by *S. cerevisiae* α2 (Figure 3H). Neither of these positions is highly constrained within the *Candida* group (Figure 3F–G). This conversion could occur without compromising the ancestral, positive regulatory mode because both proteins recognize the same core sequence (TGT). Specific bases to the "left" of the core are required for efficient **a**2 binding while specific bases to the "right" are required by α2 (Figure 3F). From these experiments we conclude that (1) *Candida* clade **a**-specific cis-regulatory sequences are recognized efficiently by **a**2, but not α2, (2) a small number of mutations (2) can convert an **a**2 site to an α2 site, and (3) these mutations occurred at positions that were likely under weak constraint in the ancestor.

The contribution of non-specific protein interactions to early intermediates

It is simple to envision how a couple of mutations could "convert" a single ancestral **a**sg *cis*-regulatory sequence into a sequence that can be recognized by α2. However, there are at least 7 **a**sgs in each species. And, as we discussed above, targeting of α2 to **a**sg *cis*-regulatory sequences also required the evolution of a new protein-protein interaction with Mcm1. How, then, did all of the gains required for this novel regulatory scheme arise? Did the Mcm1-α2 interaction evolve before or after the *cis*-regulatory changes? Or, did these events occur in concert?

To explore these questions, we mimicked two possible and extreme intermediate states in this evolutionary transition: the presence of the α2-Mcm1 protein-protein interaction without the *cis*-regulatory changes and the *cis*-regulatory changes without the α2-Mcm1 interaction. To create the first state, we replaced the *S. cerevisiae* **a**sg reporter with an **a**sg *cis*-regulatory sequence from the *Candida* clade (*C. albicans RAM2*). For the second state, we compromised the region of the *S. cerevisiae* α2 protein that binds Mcm1 by substituting it with the aligning sequence in the *C. albicans* protein. When the *C. albicans RAM2 cis*-regulatory sequence was tested with wild-type *S. cerevisiae* α2, we did not observe repression, even when α2 was over-expressed. However, when the Mcm1 interaction region was disrupted but the *S. cerevisiae cis*-regulatory sequence was used, we did observe repression when α2 was overexpressed. (Figure 4A)

We next determined how the α2 protein lacking the Mcm1 interaction region could still repress an **a**sg reporter, albeit weakly. In principle, either the "ancestral" α2 could bind the **a**sg reporter independently of Mcm1 or Mcm1 could stabilize ancestral α2 binding through non-specific protein-protein interactions. To distinguish between the models, we tested for repression of an **a**-specific *cis*-regulatory sequence in which the Mcm1 *cis*-regulatory site was destroyed by mutation (Figure 4B). (Mcm1, an essential protein, cannot be deleted from the cell.) Using this reporter, overexpression of a modified α2 protein that lacks the Mcm1 interaction region failed to show any detectable repression (Figure 4B). Thus, it appears that the second model best accounts for our results: even before the evolution of a specific Mcm1-interaction region, binding of the "ancestral" α2 was stabilized by its proximity to Mcm1. These results suggest a model where the effects of fortuitous *cis*-mutations, which stabilized α2 binding to DNA, would have been amplified by the contribution of non-specific interactions with Mcm1 during the earliest steps in the evolution of α2 repression at the **a**sgs.

We hypothesize that once a more optimized Mcm1-α2 protein interaction formed, α2 could have occupied *cis*-regulatory sequences that deviate from its preferred sequences. These types of sites may have occurred in intermediates and we modeled such an intermediate by

$watermark-text
$watermark-text
$watermark-text

$watermark-text

$watermark-text

$watermark-text

mutating a single, key base pair in the *S. cerevisiae STE2 cis*-regulatory sequence. Even with a mutated α2 binding site, we find that when α2 is overexpressed, it can mediate repression, but only if the Mcm1 interaction region of α2 is present (Figure 4C). Thus, a protein-protein interaction with Mcm1 can stabilize the binding of α2 to imperfect *cis*-regulatory sequences; such sequences may have been present in early, evolutionary intermediates.

If these ideas are correct, then the changes in *cis*-regulatory sequences and the evolution of this new protein-protein interaction are linked and must have evolved together. An attractive feature of this co-evolution model is that the interaction energy needed for the α2 and Mcm1 proteins to occupy an **a**sg *cis*-regulatory sequence can be distributed between the protein-protein and protein-DNA interactions, enabling all the **a**sgs to come under weak influence by α2 and then tuned individually through changes in each gene's *cis*-regulatory sequence.

Hybrid regulation of asgs by both a2 and α2 occurs in modern species

The experiments described here and by Tsong et al., 2006 indicate that the control of **a**sg expression passed through a hybrid regulatory state in which positive control by **a**2 and negative control α2 operated together. One can envision two, non-mutually exclusive types of such hybrid regulation. In the first, a given **a**sg would be both repressed by α2 in α cells and activated by **a**2 in **a** cells. In the second, regulation would be at the network level; some **a**sgs would be activated by **a**2 in **a** cells and other **a**sgs would be repressed by α2 in α cells. Both types of hybrid regulation would ensure that each **a**sg is expressed only in **a** cells. We next investigated the possibility that some form of hybrid regulation still exists in modern species. We chose to examine *L. kluyveri* and *K. wickerhamii* because both have an intact **a**2 gene (Butler et al., 2004), and the α2 protein of both species is able to repress a *S. cerevisiae* **a**sg *cis*-regulatory site (Figure 1A and 2B).

In *L. kluyveri,* a genome-wide ChIP of **a**2 was performed in **a** cells (Figure 5A, C, E and S4). Ten peaks of **a**2 binding met our enrichment cut-offs, and six of these peaks were upstream of genes whose orthologs are **a**sgs in either *C. albicans* or *S. cerevisiae* (*AGA2*, *ASG7*, *AXL1*, *BAR1*, *STE2*, and *STE6*) (Galgoczy et al., 2004; Tsong et al., 2003). To determine if these genes and the genes associated with the remaining four peaks are expressed in an **a**-specific pattern, RT-qPCR was performed using wild-type **a** cells and wild-type α cells (Figure S5A). We also tested the gene *RAM1* because *RAM1* is an **a**sg in *C. albicans* (Tsong et al., 2003), and its peak of **a**2 binding fell just below our significance threshold. Using this data, we defined the following nine genes as *L. kluyveri* **a**sgs: *AGA1, AGA2, ASG7, AXL1, BAR1, RAM1, STE2, STE6,* and *STE14*. Two of these genes, *STE14* and *AGA1* are **a**sgs in *L. kluyveri* but not in either *S. cerevisiae* or *C. albicans*; the others are **a**sgs in at least two of the three species. (Three genes associated with **a**2 binding in *L. kluyveri* (*ELA1*, *TID3*, and *SAKL0E14784g*) did not show **a**sg expression under any condition we tested and were excluded from further tests.) Transcript levels of all nine *L. kluyveri* **a**sgs were decreased when **a**2 was deleted (Δ*MATa2*), indicating that **a**2 activates these genes by binding to their *cis*-regulatory sequences (Figure 5G).

Next, full genome ChIP of myc-tagged α2 in α cells was used to ascertain its role, if any, in the regulation of **a**sgs, in *L. kluyveri* (Figure 5B, D, F and Figure S4). In α cells, binding peaks were observed upstream of two genes—the **a**sgs *AGA1* and *AGA2* (Figure 5B and D). These peaks are centered over the same region of DNA as the **a**2 binding peaks observed in **a** cells, showing that the two regulators associate with the same region of DNA but in different cell types. This result is consistent with the analysis described above showing that the two regulators have overlapping DNA binding specificities and each forms a protein interaction with Mcm1 (Figure 3G). To test whether *AGA1* and *AGA2* are repressed by α2,

$watermark-text

$watermark-text

$watermark-text

we performed RT-qPCR in wild-type α cells and in α2-deletion α cells (Δ*M α2*) (Figure 5H). The transcript abundance of *AT* both of these genes increased indicating that α2 represses these genes in α cells. The remaining seven **a**sgs were also tested by RT-qPCR and determined not to be targets of α2 repression in these conditions (Figure 5H). Taken together, these results indicate that all nine of the *L. kluyveri* **a**sgs are targets of direct **a**2 activation in **a** cells and that two of them are also targets of direct α2 repression in α cells. Thus, in *L. kluyveri,* two of the **a**sgs are regulated in a hybrid fashion. The results also show that, for these two genes, **a**2 and α2 act through association with the same DNA sequence in the two cell types.

The other species chosen for this analysis, *K. wickerhamii*, is described in Figure S6. The results indicate that at least two **a**sgs are regulated in a hybrid fashion in *K. wickerhamii*. We note that the genes that are hybrid-regulated in *K. wickerhamii* are not the same genes that are hybrid-regulated in *L. kluyveri* (summarized in Figure 7C).

Gains and losses in the asg network

In addition to changes in the overall form of regulation, we find that the **a**sg network has gained and lost individual target genes over the hemiascomycete lineage. We believe this can be accounted for by the formation and destruction of *cis*-regulatory sequences. For instance, we found that *STE14* is an **a**sg in *L. kluyveri* but not in the other species examined and that *AXL1* is an **a**sg in many species but not *S. cerevisiae* (Figure 7C, Table S2 and S3 and (Booth et al., 2010; Galgoczy et al., 2004; Tsong et al., 2003)).

K. lactis α2 lost the ability to repress asgs

The dairy yeast *K. lactis* diverged from *S. cerevisiae* after the gain of **a**sg repression, and it retains many of the *cis* and *trans* characteristics indicative of a hybrid form of regulation where both **a**2 with α2 are active (Tsong et al., 2006). Yet, as noted above, the *K. lactis* α2 protein is unable to repress the **a**sgs when moved into *S. cerevisiae* (Figure 2B–C).

To determine whether α2 represses the **a**sgs in *K. lactis* itself, we utilized gene expression profiling to compare transcript levels of wild-type **a** and wild-type α cells to Δ**a**2 **a** cells and Δα2 α cells, respectively. Deletion of α2 in α cells did not have an effect on transcript levels of any of the *K. lactis* **a**sgs (Figure 6E and Figure S5B) nor did it affect the expression of other genes in *K. lactis* (data not shown). We confirmed this result by measuring transcript levels of **a**sgs by RT-qPCR (data not shown). In contrast, deleting **a**2 in **a**-cells resulted in decreased expression of nearly all of the *K. lactis* **a**sgs (Figure 6E). Consistent with these results, **a**2 was found to be bound upstream of the *K. lactis* **a**sgs (Figure 6A, C and data not shown) but α2 binding was not detected at the **a**sgs or any other gene in α cells (Figure 6B, D and data not shown). (As a control, *K. lactis* α2 binding is observed at the haploid specific genes when α2 and **a**1 are expressed together (Booth et al., 2010).) Thus, although *K. lactis* has many of the hallmarks of hybrid regulation (in particular, its **a**sg *cis*-regulatory sequences support repression by *S. cerevisiae* (Figure 2A), α2 does not repress the **a**sgs in this species.

Comparison of the α2 sequences from multiple species pointed to a likely cause of the inability of the *K. lactis* α2 to repress the **a**sgs: amino acid residue 136 in *K. lactis* is an asparagine, but in all repressing-competent α2 proteins it is a small, hydrophobic residue, either a valine or leucine (Figure 3C). This position has been shown to be important for the interaction between α2 and Mcm1 (Mead et al., 1996; Tan and Richmond, 1998). Using the *S. cerevisiae* reporter assay, we tested this idea explicitly and found that mutating this single residue in the *K. lactis* α2 protein to a valine (N136V) restored its function as a repressor (Figure 6G). The simplest interpretation of these observations is that the *K. lactis* α2 protein

$watermark-text

$watermark-text

$watermark-text

recently acquired a mutation that compromised its ability to interact with Mcm1 thereby destroying the derived (repression) mode of **a**sg regulation and reverting to the ancestral (positive) mode. The evolutionary path by which this amino acid substitution likely occurred is explored in detail in Figure S7.

Discussion

The regulation of a set of cell-type specific genes, the **a**sgs, has changed over evolutionary time in the hemiascomycete branch of the fungal lineage. Based on data from numerous approaches, we describe the likely evolutionary path for the change in the mechanism by which the **a**sgs are regulated. We provide strong experimental evidence for an intermediate hybrid regulatory state in which **a**2 and α2 both participated in the cell-type regulation of the **a**sgs, and we show that this hybrid state resolved in several distinct ways along the lineages to modern species, generating a diversity of network structures (summarized in Figure 7A).

The gain of α2 repression at the **a**sgs required that α2 navigate a constrained regulatory landscape. As a result, this evolutionary path exploited multiple features of the existing network that both stabilized early intermediates and limited the number of mutations required to evolve this new function. We also show that protein modularity minimized the pleiotropy of the evolved features of the new regulatory mode. This work provides both a mechanistic account of how a particular transcription regulator evolved a new function and insights into the molecular origins of the extraordinary flexibility of transcriptional regulatory network architectures that appear across modern species.

In this discussion we first outline the key features of the ancestral network that were exploited (that is, exaptations) in the evolution of α2-repression of the **a**sgs. We next discuss the concerted changes in the *cis*-regulatory sequences and the *trans* regulators that enabled formation of the new mode of regulation. Third, we consider the consequences of the intermediate hybrid regulatory state and its role in the network diversity observed in modern species. Finally, we discuss the relative importance of adaptation and neutral drift to the diversification of gene regulatory networks.

Exploitation of ancestral network components

Several key features of the derived form of regulation (repression of the **a**sgs) were in place prior to its evolution. For instance, the new mode of regulation requires that the repressor be expressed in α and **a**/α cells, but not in **a** cells. For α2, this is true for virtually every species in the hemiascomycetes and reflects its deeply conserved function: it forms a heterodimer with **a**1 to regulate the haploid specific genes in **a**/α cells (Booth et al., 2010; Strathern et al., 1981; Tsong et al., 2003). Thus, the expression pattern necessary for α2 to act as a repressor of the **a**sgs was already present in the ancestor.

In contrast to the popular model wherby new *cis*-regulatory sequences arise *de novo* in unused regions of promoters, α2 exploited features of the existing **a**sg *cis*-regulatory sequences (Tsong et al., 2006). The monomers of **a**2 and α2 have related DNA-binding specificities (Figure 3G) despite belonging to different transcription regulator families (HMG box vs. homeodomain, respectively). This intrinsic overlap in DNA-binding specificities minimized the number of *cis*-regulatory mutations required for the transition: only two point mutations are required to convert an optimal **a**2 recognition sequence to an optimal α2 recognition sequence (Figure 3H). Moreover, we have shown that sequences exist in modern species that are efficiently recognized by both proteins (Figures 5, S4 and S6), thus further reducing the potential fitness barriers to this transition.

$watermark-text

$watermark-text

$watermark-text

In addition to the exploitation of **a**2 *cis*-sequences, the binding of α2 to the ancestral sequences was stabilized by the presence of a neighboring DNA-bound protein, Mcm1. We provide evidence for a model where the ancestral presence of Mcm1 at the *cis*-regulatory sites of the **a**sgs stabilized α2 DNA binding in early evolutionary intermediates through weak, relatively non-specific protein-protein contacts (Figure 4A and B). Subsequently, the protein-protein interaction became stronger and more specific through changes in the α2 protein, which stabilized the binding of Mcm1 and α2 to each other and to DNA. We have shown that the evolution of this specific interaction between Mcm1 and α2 was asymmetric: the α2 protein underwent numerous changes in a previously unconstrained region allowing it to recognize an existing surface of the ancestral Mcm1; therefore, no changes were necessary in Mcm1 (Figure 3B–E). Thus, from the earliest steps in this evolutionary transition, the interaction energy necessary to stabilize α2 binding was shared out between protein-protein and protein-DNA contacts. The exploitation of ancestral *cis* and *trans* features strongly guided the evolutionary trajectory of α2 (through stabilizing early intermediates) by minimizing the number of changes necessary.

Constraint and the evolution of novelty by *cis* and *trans* changes

Although several key network features needed for the evolution of α2-repression of the **a**sgs were already present in the ancestor, changes in both the *cis*-regulatory sequences and the α2 protein needed to occur for efficient **a**sg repression. The gain and loss of *cis*-regulatory sequences are readily acknowledged as major contributors to evolutionary novelty, but changes in the transcription regulators themselves are often described as less prevalent, particularly in the absence of gene duplication (Carroll, 2005; Wray, 2007). For example, it is frequently said that changes in transcription regulators will tend to be rare because they are pleiotropic—affecting the regulation of many genes simultaneously and likely disrupting existing networks.

The gain of function of α2 described here occurred within the context of a pre-existing, deeply conserved regulatory landscape: the regulation of the haploid specific genes by the **a**1-α2 heterodimer (Booth et al., 2010; Herskowitz, 1989; Hull and Johnson, 1999). The modularity of the α2 protein made it possible to gain a new function (repression of the **a**sgs) without compromising its ancestral function (repression of the haploid specific genes). Indeed, it seems likely that the only permissible evolutionary trajectories for the α2 protein to gain a new function would require that its ancestral function be preserved. How did this occur?

Two regions of the α2 protein—the DNA-binding homeodomain and the **a**1 interaction region—are needed for its ancestral function and are preserved, in sequence and function, through stabilizing selection across the entire hemiascomycete lineage (Figure 3B & D). The protein modules that more recently evolved to make **a**sg repression possible (regions 1 and 3, Figure 3B, C, and E) are short (~10) stretches of amino acids that developed within unconstrained regions of the ancestral protein (Figure 3B and C). The evolution of short, linear protein interaction regions spatially isolated from the ancestral functions bypassed the potential pleiotropic constraints on regulator evolution. We note that the gain of new functional modules in unused portions of the ancestral protein is akin to the acquisition of new *cis*-regulatory sequences at unconstrained positions in non-coding sequence. More generally, the modular structure of modern transcription regulators is likely the result of the sequential addition of new functions in previously unconstrained regions of the proteins, as described here.

$watermark-text
$watermark-text
$watermark-text

Hybrid intermediates and the diversification of regulatory networks

As we have described, the path to the gain of α2-repression of the **a**sgs occurred while the ancestral form of **a**-specific regulation (activation by **a**2) was still extant (Tsong et al., 2006). Thus, both forms of regulation existed together in the ancestor of the *Kluyveromyes* and *Saccharomyces* clades. We propose that this hybrid regulatory intermediate made possible the subsequent diversification of the **a**sg regulatory network architectures without a loss in regulation. Based on evidence from several modern species, we found that the hybrid regulatory state has diversified (resolved) in three directions:

- *Retention of both modes of regulation:* We showed that two modern species, *K. wickerhamii* and *L. kluyveri*, have retained both the ancestral (**a**2 activation) and derived (α2 repression) modes of regulation of the **a**sgs (Figures 5 & S6). Two additional species, *Z. rouxii* and *A. gossypii*, also possess α2 proteins that repress **a**sg expression (Figure 2B) and both appear to have functional **a**2 genes. Thus, we favor the hypothesis that these two species also retain some form of the hybrid regulatory state.
- *Loss of the ancestral mode of regulation: S. cerevisiae* and other post-whole genome duplication species regulate their **asg**s using the repressor α2 exclusively. Indeed, the gene coding for the activator **a**2 (the ancestral regulator) has been lost from these species (Butler et al., 2004); thus, the ancestral mode has been discarded.
- *Loss of the derived mode of regulation: K. lactis* appears to have lost α2 repression of the **a**sgs through a recent, single amino acid change in the α2 protein. The α2 protein of the nearby branching species *Kluyveromyces marxianus* also has a mutation at this same position (Figure 3C), although the substituted amino acid is different in the two species. In *K. lactis* (and presumably *K. marxianus*), the **a**sgs appear to be regulated by **a**2 alone, with the derived mode no longer in use.

We suggest hybrid regulatory states, such as the state described here, represent ‘high potential states’ for evolutionary change as they have the ability to resolve in several directions without destroying the overall logic of regulation (Figure 7B). Akin to gene duplication, the formation of a hybrid regulatory state generates a partially redundant intermediate that allows for diversification without a loss of the original function or regulatory logic (Tanay et al., 2005). Within the hybrid regulatory state, network reversion remains a permissible evolutionary trajectory. The reversion to an ancestral regulatory mode that we have described in *K. lactis* is not a strict molecular reversal. Instead, the *K. lactis* α2 protein acquired a mutation that inactivates the derived function while maintaining its ancestral function, haploid specific gene repression as a heterodimer with **a**1.

Our results also show that, over the evolutionary time period considered in this paper, a subset of **a**sgs moved in and out of the network through the gains and losses of *cis*-regulatory sequences (summarized in Figure 7C). Although some genes are expressed **a**-specifically in all species (e.g. those encoding pheromones and pheromone receptors), others are not. This implies that for the **a**sgs to undergo a transition from one regulatory mode to another, not all genes within the network would need to experience this switch in regulation. The looser requirements for the regulation of some genes in a network may facilitate changes in the mode of regulation of a network, as not all genes would have to be carried along during the initial phases of the switch.

Adaptive and neutral forces in regulatory evolution

Selection can only act on the output of a transcription regulatory network; if an evolutionary path exists between different regulatory architectures with near-identical spatial pattern,

$watermark-text $watermark-text $watermark-text

dynamic range, and kinetics of expression, then the network can be predicted to drift between these different solutions over evolutionary time (Lynch, 2007). The hybrid state we have described spawned a range of evolutionary outcomes (activation, repression or hybrid), each with different regulatory circuit architectures. In all cases, however, the overall logic of regulation (**a**sgs ON in **a** cells and OFF in the other two cell types) has been preserved. It is possible that each of the different forms of regulation we observed produce different dynamic ranges or kinetics of expression and that these qualities have been selected for on a gene-by-gene basis as different yeast species diversified. However, we favor the simpler model where the regulatory diversification following the formation of the hybrid regulatory state occurred largely through neutral, non-adaptive, drift. In other words, the network could drift between states where the dynamic range of regulation generally remained the same but the relative contributions of the ancestral and derived modes differed through the strengthening and weakening of protein-protein and protein-DNA interactions. The range of network structures observed in modern species would simply reflect the "breathing" of the hybrid regulatory network.

In contrast to the neutral model we favor for network diversification from the hybrid state, we currently favor the idea that the formation of the hybrid state was itself adaptive. For one thing, the gain of **a**sg repression to form the hybrid state required a reasonably large number of mutational events, both in *cis* and *trans*. For instance, the gain of two new protein interaction modules within α2 (one for Tup1 and one for Mcm1) involved greater than two-dozen amino acid changes and it seems unlikely that such a large number of amino acid changes that produce a new biochemical function could have reached fixation without directional selection. We cannot know for certain what adaptive value the invention of **a**sg repression had, if any, for the ancestor of the *Kluyveromyces* and *Saccharomyces* clades. However, in the supplemental text, we discuss a possible scenario in which the gain of repression at this gene set may have been a necessary regulatory response to another newly evolved trait in this ancestor, the gain of silent mating cassettes (Butler et al., 2004). These arguments are not conclusive, but they are consistent with the idea that positive selection played a role in the gain of α2 repression of the **a**sgs and the formation of the hybrid intermediate, and that the successive circuit diversification was nonadaptive.

Irrespective of the potential role of selection, a hybrid regulatory state can be short-lived (as in the ancestor of *S. cerevisiae*) or exceedingly long-lived (as in *L. kluyveri* and *K. wickerhamii*). We propose that the creation of hybrid regulatory states serves as a general model to rationalize the many examples of network-wide transcriptional regulatory divergence that have been observed among species.

Experimental Procedures

Identification of Gene Orthologs and Upstream Regulatory Sequences

Orthologs of experimentally identified **a**sgs (Galgoczy et al., 2004) (Tsong et al., 2003) were identified and confirmed using BLAST. To identify a Position Specific Scoring Matrix (PSSM) for α2-repression (derived), we submitted to MEME the 600 base pairs upstream of the **a**sgs from *S. cerevisiae*, *Saccharomyces mikatae, Saccharomyces paradoxus,* and *Saccharomyces bayanus.* Similarly, sequences from *C. albicans*, *Candida dubliniensis,* and *Candida tropicalis* were used to calculate a PSSM for **a**2-activation (ancestral). The 600 base pairs upstream of each **a**sg were scanned to identify the **a**sg *cis*-regulatory sequences of all genome sequenced hemiascomycetes using MAST (Bailey et al., 2009). See Extended Experimental Procedures for details.

$watermark-text

Strain Construction

A complete list of all strains used in this study can be found in Table S5. The primers used to generate and confirm these strains are listed in Table S6. For details regarding strain and plasmid construction see Extended Experimental Procedures.

β-galactosidase Assays

β-galactosidase assays were performed using a standard protocol (Guarente and Ptashne, 1981). Strains were grown in selective media to maintain transformed plasmids. For each strain, colonies were grown overnight, diluted, and allowed to reach late log phase. Cells were harvested and permeabilized, and activation assays were performed.

Quantification of Conservation Scores within α2

α2 orthologs were aligned using MUSCLE (Edgar, 2004). The genetic diversity spanned by the *Saccharomyces-Kluyveromyces* and *Candida* clade is similar (Taylor and Berbee, 2006), however, we removed from our analysis a subset of closely related sequences from the *Saccharomyces-Kluyveromyces* species to normalize the levels of conservation between the two groups. The displayed amino-acid conservation was calculated using the PAM250 amino-acid substitution matrix (Henikoff and Henikoff, 1992). The displayed curve (Figure 3B) has been smoothed by averaging each conservation score with the scores of adjacent residues. See Extended Experimental Procedures for details.

RNA Isolation and cDNA Preparation

RNA was isolated from yeast cultures using hot phenol/chloroform extraction. cDNA was prepared using SuperScript II (Invitrogen). Additional details can be found in the Extended Experimental Procedures.

Gene Expression Arrays

K. lactis cDNA was hybridized to a custom Agilent array. All data has been deposited in NCBI GEO at accession number (GSE39027). cDNA labeling, hybridization and data analysis are described in the Extended Experimental Procedures.

$watermark-text

Chromatin Immunoprecipitation

C-terminally myc tagged **a**2 and α2 proteins were created for ChIP. Tagged (experimental) and untagged (control) strains were grown, harvested and lysed. Chromatin was precipitated with commercially available anti-myc or anti-HA antibodies. The DNA was amplified, labeled and competitively hybridized to custom Agilent tiling oligonucleotide arrays. Display, analysis and identification of binding events were performed with MochiView (Homann and Johnson, 2010). Details are found in the Extended Experimental Procedures. Data has been deposited in NCBI GEO at accession numbers GSE38919 for *K. lactis* and (GSE39007) for *L. kluyveri*.

$watermark-text

Quantitative PCR

A complete list of all primers used for qPCR is found in Table S6.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Chiraj Dalal, Clarissa Nobile, and Brian Tuch for critically reading the manuscript and Jennifer Garcia, Victor Hanson-Smith, Oliver Homann, Quinn Mitrovich, and Brian Tuch for their generous and helpful advice concerning experimental and computational procedures. We thank Peter Philippsen for kindly sharing genomic DNA and sequences from *Ashbya gossypii*. This work was supported by RO1 GM057049 from the National Institute of Health.

References

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research. 2009; 37:W202–208. [PubMed: 19458158]

Booth LN, Tuch BB, Johnson AD. Intercalation of a new tier of transcription regulation into an ancient circuit. Nature. 2010; 468:959–963. [PubMed: 21164485]

Butler G, Kenny C, Fagan A, Kurischko C, Gaillardin C, Wolfe KH. Evolution of the MAT locus and its Ho endonuclease in yeast species. Proc Natl Acad Sci USA. 2004; 101:1632–1637. [PubMed: 14745027]

Carroll SB. Evolution at two levels: on genes and form. PLoS Biol. 2005; 3:e245. [PubMed: 16000021]

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010; 327:302–305. [PubMed: 20007865]

Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. Science. 2006; 311:796–800. [PubMed: 16469913]

Doebley J, Lukens L. Transcriptional regulators and the evolution of plant form. Plant Cell. 1998; 10:1075–1082. [PubMed: 9668128]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004; 32:1792–1797. [PubMed: 15034147]

Galgoczy DJ, Cassidy-Stone A, Llinás M, O'Rourke SM, Herskowitz I, DeRisi JL, Johnson AD. Genomic dissection of the cell-type-specification circuit in Saccharomyces cerevisiae. Proc Natl Acad Sci USA. 2004; 101:18069–18074. [PubMed: 15604142]

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature. 2005; 433:481–487. [PubMed: 15690032]

Guarente L, Ptashne M. Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. Proc Natl Acad Sci USA. 1981; 78:2199–2203. [PubMed: 6264467]

Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA. 1992; 89:10915–10919. [PubMed: 1438297]

Herskowitz I. A regulatory hierarchy for cell specialization in yeast. Nature. 1989; 342:749–757. [PubMed: 2513489]

Homann OR, Johnson AD. MochiView: versatile software for genome browsing and DNA motif analysis. BMC biology. 2010; 8:49. [PubMed: 20409324]

Hull CM, Johnson AD. Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans. Science. 1999; 285:1271–1275. [PubMed: 10455055]

Johnson AD, Herskowitz I. A repressor (MAT alpha 2 Product) and its operator control expression of a set of cell type specific genes in yeast. Cell. 1985; 42:237–247. [PubMed: 3893743]

Keleher CA, Goutte C, Johnson AD. The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. Cell. 1988; 53:927–936. [PubMed: 3289753]

Komachi K, Redd MJ, Johnson AD. The WD repeats of Tup1 interact with the homeo domain protein alpha 2. Genes & Development. 1994; 8:2857–2867. [PubMed: 7995523]

Lavoie H, Hogues H, Mallick J, Sellam A, Nantel A, Whiteway M. Evolutionary tinkering with conserved components of a transcriptional regulatory network. PLoS Biol. 2010; 8:e1000329. [PubMed: 20231876]

$watermark-text $watermark-text $watermark-text

Lynch M. The evolution of genetic networks by non-adaptive processes. Nat Rev Genet. 2007; 8:803–813. [PubMed: 17878896]

Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet. 2011; 43:1154–1159. [PubMed: 21946353]

Madhani, H. Yeast as a Model for Cellular Differentiation. 1. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2007. From a to alpha.

Mead J, Zhong H, Acton TB, Vershon AK. The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. Mol Cell Biol. 1996; 16:2135–2143. [PubMed: 8628280]

Smith DL, Johnson AD. Operator-constitutive mutations in a DNA sequence recognized by a yeast homeodomain. EMBO J. 1994; 13:2378–2387. [PubMed: 7910796]

Strathern J, Hicks J, Herskowitz I. Control of cell type in yeast by the mating type locus. The alpha 1-alpha 2 hypothesis. J Mol Biol. 1981; 147:357–372. [PubMed: 7031257]

Tan S, Richmond TJ. Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature. 1998; 391:660–666. [PubMed: 9490409]

Tanay A, Regev A, Shamir R. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. Proc Natl Acad Sci USA. 2005; 102:7203–7208. [PubMed: 15883364]

Taylor JW, Berbee ML. Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia. 2006; 98:838–849. [PubMed: 17486961]

Thornton JW. Resurrecting ancient genes: experimental analysis of extinct molecules. Nat Rev Genet. 2004; 5:366–375. [PubMed: 15143319]

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet. 2007; 39:31–40. [PubMed: 17159977]

Tsong AE, Miller MG, Raisner RM, Johnson AD. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. Cell. 2003; 115:389–399. [PubMed: 14622594]

Tsong AE, Tuch BB, Li H, Johnson AD. Evolution of alternative transcriptional circuits with identical logic. Nature. 2006; 443:415–420. [PubMed: 17006507]

Tuch BB, Li H, Johnson AD. Evolution of eukaryotic transcription circuits. Science. 2008; 319:1797–1799. [PubMed: 18369141]

Vershon AK, Johnson AD. A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. Cell. 1993; 72:105–112. [PubMed: 8422672]

Wohlbach DJ, Thompson DA, Gasch AP, Regev A. From elements to modules: regulatory evolution in Ascomycota fungi. Curr Opin Genet Dev. 2009:1–8. [PubMed: 19208465]

Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007; 8:206–216. [PubMed: 17304246]

$watermark-text $watermark-text $watermark-text

$watermark-text

$watermark-text

$watermark-text

Research Highlights

- Protein modularity & ancestral feature exploitation bypass evolutionary constraint
- Gain of new regulator-regulator interaction transformed a transcription network
- This gain resulted in a hybrid state with ancestral & derived regulatory features
- Partial redundancy of the hybrid state enables regulatory network diversification

$watermark-text
$watermark-text
$watermark-text

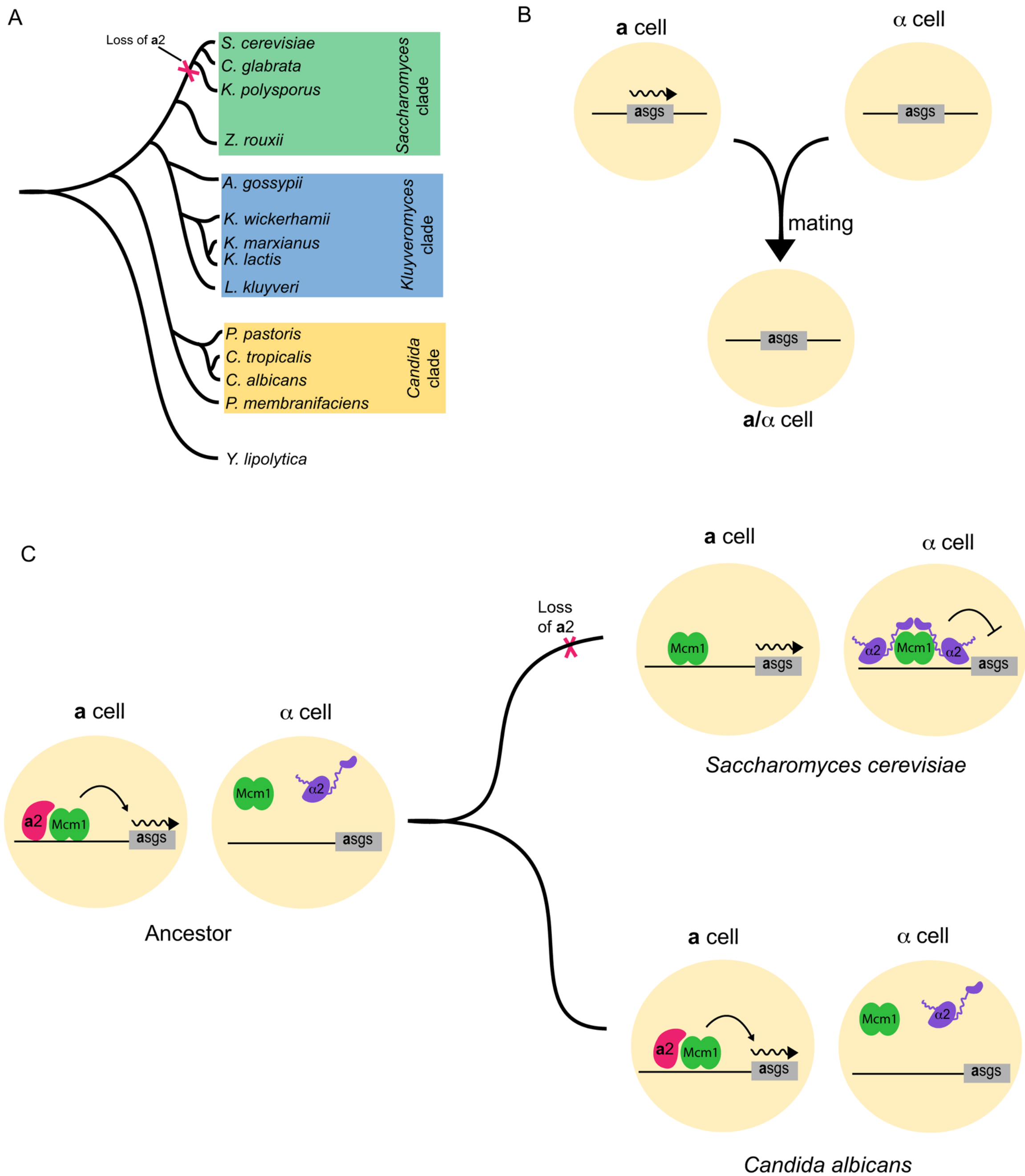

Figure 1. Cell-type specification in the hemiascomycetes

(A) Three hemiascomycete clades are considered—*Candida*, *Kluyveromyces* and *Saccharomyces*. The *Saccharmoyces* clade includes the pre-whole genome duplication species *Zygosaccharomyces rouxii* and the post-whole genome duplication species that lack an **a**2 gene (loss event indicated by a pink X). (B) The hemiascomycete yeasts have three cell types; the mating competent **a** and α cells and the product of their mating, an **a**/α cell. **a** cells express a set of genes called the **a**sgs (**a**sgs) (Herskowitz, 1989). (C) In *C. albicans* and the ancestor, the **a**sgs are activated by Mcm1 (present in all cell types) and **a**2 (present only in **a**-cells) (Tsong et al., 2003). In *S. cerevisiae*, the **a**sgs are specified using Mcm1 a cell-type specific repressor, α2 (Johnson and Herskowitz, 1985; Keleher et al., 1988).

$watermark-text

$watermark-text

$watermark-text

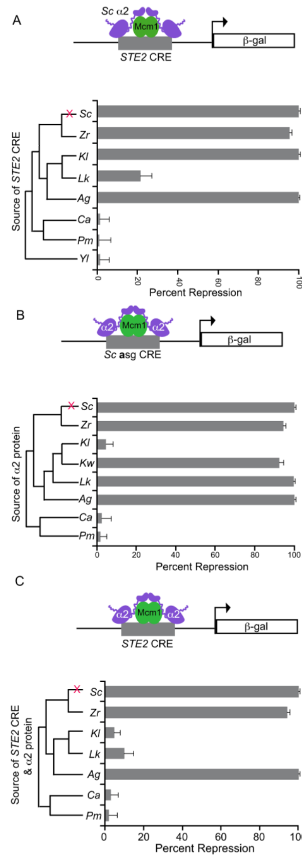

Figure 2. α2 repression of the asgs evolved prior to the divergence of *Saccharomyces* and *Kluyveromyces*

(A) The **a**sg *cis*-regulatory sequence of the α-pheromone receptor gene *STE2* from *S. cerevisiae* (*Sc)* and species that branch prior to the loss of the **a**2 gene, *Z. rouxii (Zr), K. lactis (Kl), L. kluyveri (Lk), A. gossypii (Ag), C. albicans (Ca), P. membranificians (Pm),* and *Y. lipolytica (Yl)* were inserted into a reporter construct to assay repression. Percent repression was determined by transforming constructs into *S. cerevisiae* **a**-cells (no α2) and α-cells (α2 present). (B) α2 protein coding sequence from a variety of hemiascomycete species including *K. wickerhamii (Kw)* were fused to the endogenous *S. cerevisiae* α2 promoter and integrated into the genome of a *S. cerevisiae MATΔ* strain. "Trans-species" α2 proteins were then assayed for their ability to repress the *S. cerevisiae STE2* **a**sg reporter. (C) Trans-species α2 proteins were combined with the *STE2 cis*-regulatory sequence reporter constructs from the same species and assayed for repression in a *MAT*Δ background. All values reported are a mean (*n*=3) and standard error of the mean.

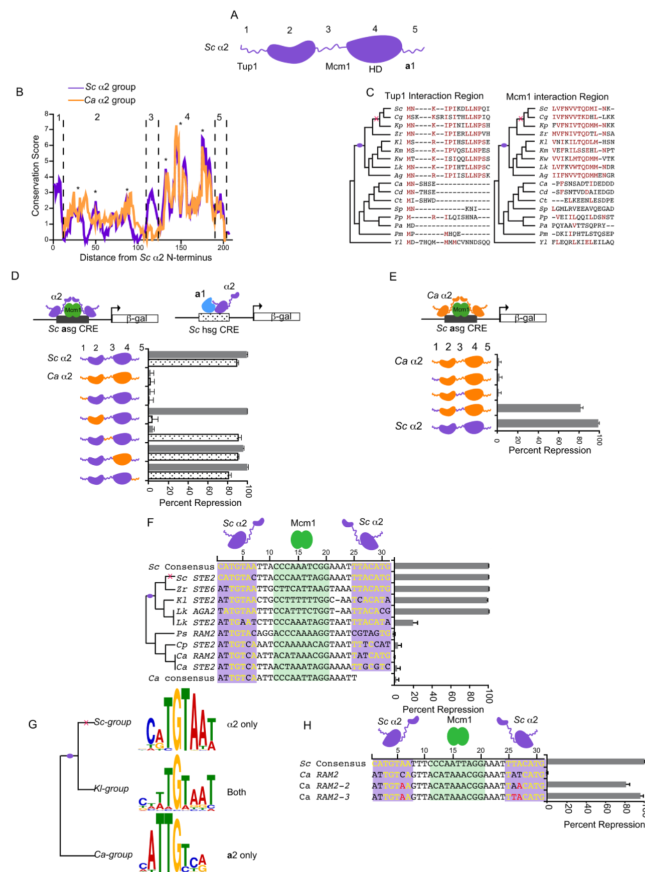

Figure 3. The *cis* and *trans*-evolution underlying the gain of a new function for α2

(A) Structured regions of *S. cerevisiae* α2 are displayed as globular, whereas, unstructured regions are displayed as curved lines. (B) Conservation scores for the α2 protein across the *Saccharomyces-Kluyveromyces* group (*Sc)* or the *Candida*-group *(Ca).* The vertical dashed lines correspond to the edges of the modular regions within the α2 protein. The positions of the three structurally predicted helices within regions 2 and 4 are marked (*) (C) The MUSCLE alignment for regions 1 and 3 are displayed. (D) *S. cerevisiae* α2 modules were swapped for the homologous regions from the *C. albicans* α2 protein. Each construct was genome-integrated in a *MATΔ* background and assayed for the ability to repress the *S. cerevisiae STE2* **a**sg (*Sc* **a**sg) and *STE4* haploid specific gene (*Sc* hsg) reporter constructs. (E) *S. cerevisiae* α2 regions 1 and 3 were swapped for the aligning sequence in the *C. albicans* **a**2 protein, genome-integrated in a *MATΔ* background, and assayed for repression of the *Sc* **a**sg reporter construct. (F) An array of **a**sg *cis*-regulatory sequences were selected from the *Kluyveromyces* and *Candida* clades based on their distribution across a range of similarity values to the *S. cerevisiae* **a**sg PSSM (Table S3). Purple shading indicates where α2 binds in *S. cerevisiae* and green shading indicates where Mcm1 binds. Yellow text highlights nucleotides that appear in the consensus binding-sites for *S. cerevisiae* α2. (G) PSSM for α2 alone site, **a**2/α2 site, and **a**2 site alone. (H) The *C. albicans RAM2* was mutated at key residues for α2 binding and tested for their ability to support repression. All values reported in bar graphs are a mean (*n*=3) and standard error of the mean. In each phylogenetic tree, the purple circle marks the gain of α2-mediated repression of **a**sgs and the pink X marks the loss of **a**2.

$watermark-text $watermark-text $watermark-text

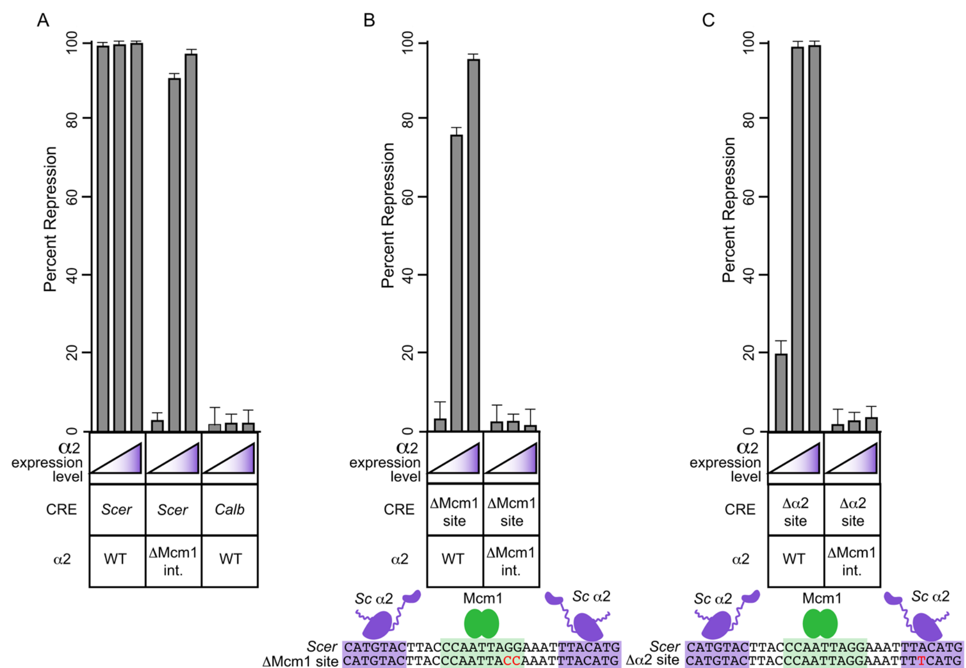

Figure 4. The contribution of non-specific protein interactions to early intermediates

(A) Wild-type *S. cerevisiae* α2 (WT) or mutant *S. cerevisiae* α2 with its Mcm1 interaction region replaced by the aligning sequence from *C. albicans* (ΔMcm1 int.) were tested for the ability to repress the *S. cerevisiae STE2* (*Scer)* or *C. albicans RAM2 (Calb)* **a**sg reporter. The α2 proteins were tested either at the endogenous level, using a strong promoter (*TEF1*), or using a very strong promoter (*TDH3*). (B) Both α2 constructs from (A) were tested for the ability to repress a modified *S. cerevisiae STE2* **a**sg *cis*-regulatory reporter construct where the Mcm1 binding site was compromised (ΔMcm1 site). (C) Both α2 constructs from (A) were tested for the ability to repress a modified *S. cerevisiae STE2* **a**sg *cis*-regulatory reporter construct where the α2 binding site was compromised (Δα2 site). In all panels, the purple and green shading represents the binding site of α2 and Mcm1, respectively. All values reported in bar graphs are a mean (n=3) and standard error of the mean.

$watermark-text
$watermark-text
$watermark-text

$watermark-text
$watermark-text
$watermark-text

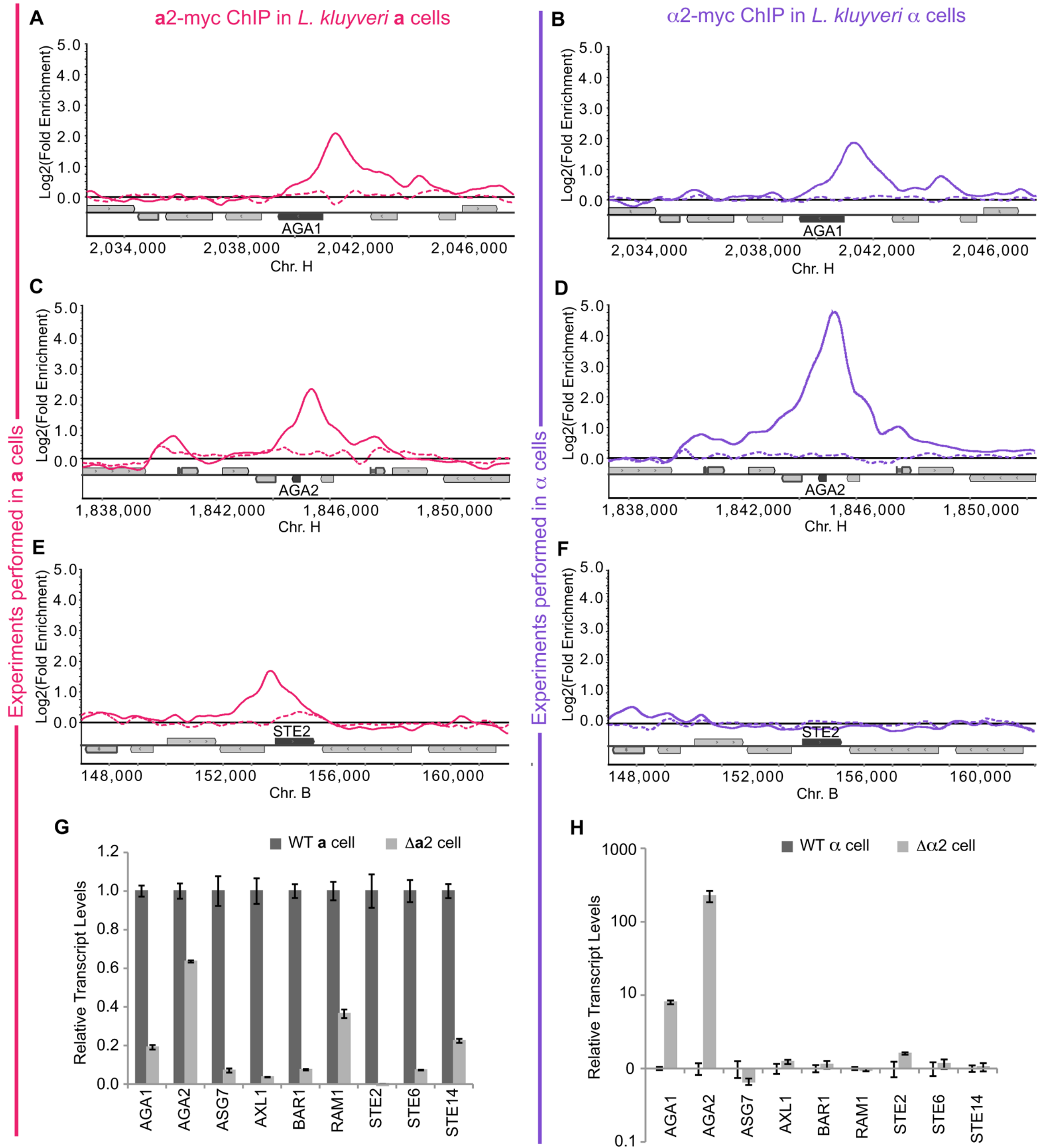

Figure 5. Regulation of the asgs in *Lachancea kluyveri*

(A–F) ChIP-chip was performed using anti-cMyc antibodies in a C-terminal myc-tagged *MATa2* **a** cells (A, C, and E solid, pink lines), wild-type **a** cells (A, C, and E dotted, pink lines), C-terminal myc-tagged MATα2 α cells (B, D, and F solid, purple lines) or wild-type α cells (B, D, and F dotted, purple lines). Wild-type cells serve as untagged controls. ChIP-chip enrichment profiles are shown for *AGA1* (A and B), *AGA2* (C and D) and *STE2* (E and F). Genes (grey rectangles) are displayed below the line if transcribed to the left and above the line if transcribed to the right. (G, H) The transcript levels of the **a**sgs in a wild-type or Δ*MATa2* **a** cell (G) and in a wild-type or Δ*MATα2* α cell (H) were measured relative to *ACT1* by RT-qPCR. The relative transcript abundance for each gene was

normalized to the abundance in wild-type **a** cells (G) or in wild-type α cells (H). Displayed is the mean (*n*=3) and standard error of the mean.

$watermark-text
$watermark-text
$watermark-text

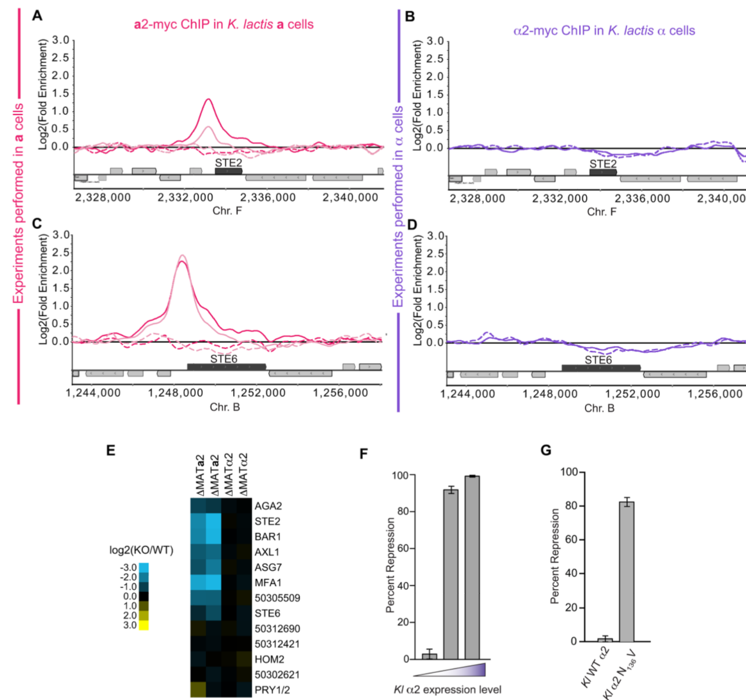

Figure 6. Regulation of the asgs in *Kluyveromyces lactis*

(A–D) ChIP-chip was performed using anti-cMyc antibodies in a C-terminal myc-tagged *MATa2* **a** cells (A and C solid, pink lines), wild-type **a** cells (A and C dotted, pink lines), C-terminal myc-tagged *MATα2* α cells (B and D solid, purple lines) or wild-type α cells (B and D dotted, purple lines). Wild-type cells serve as untagged controls. For ChIP performed in **a** cells (A and C), two conditions were used: one with pheromone induction (dark pink) and one without (light pink). ChIP-chip enrichment profiles are shown for *STE2* (A and B), and *STE6* (C and D). Genes (grey rectangles) are displayed below the line if transcribed to the left and above the line if transcribed to the right. (E) Results for orthologs of the **a**sgs from an expression array comparing mRNA levels from Δ*MATa2* **a** cells to wild-type **a** cells (two left columns) or mRNA levels from Δ*MATα2* α cells to wild-type α cells (two right columns). (F, G) The *K. lactis* α2 protein was assayed for its ability to repress a *S. cerevisiae STE2* operator sequence using a β-gal reporter. (F) Wild-type *K. lactis* α2 was expressed in a *S. cerevisiae MATΔ* cell using promoters of increasing strength. (G) Wild-type *K. lactis* α2 or *K. lactis* α2 with a single point mutation ($N_{136}V$) was expressed in a *S. cerevisiae MATΔ* cell using the endogenous *S. cerevisiae* α2 promoter. Displayed are the mean (*n*=3) and standard error of the mean.

$watermark-text
$watermark-text
$watermark-text

$watermark-text
$watermark-text
$watermark-text

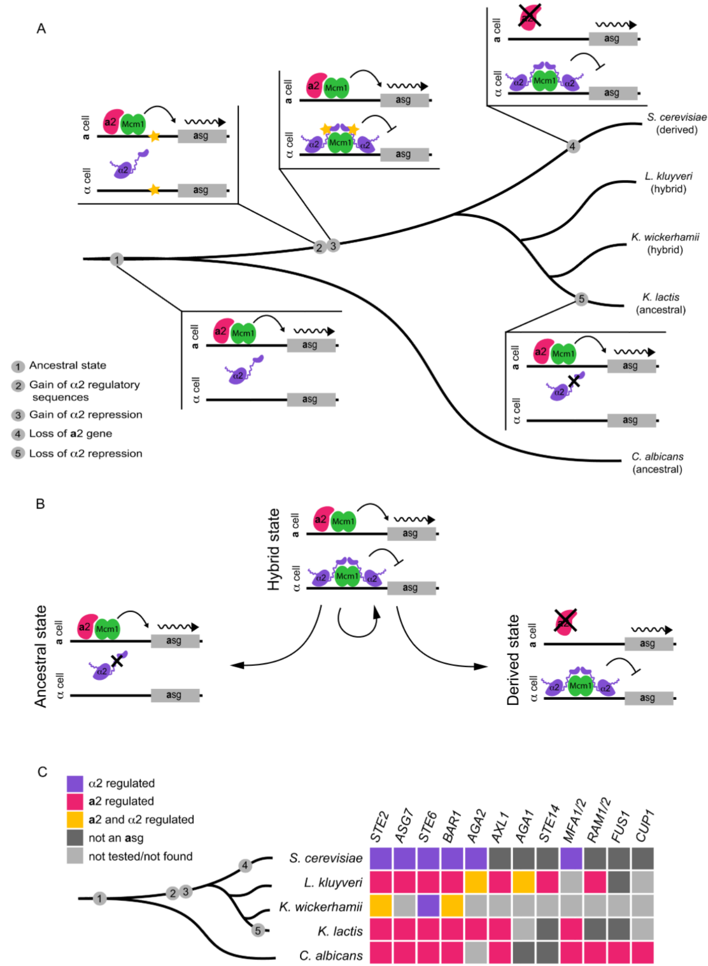

Figure 7. The gain of the hybrid regulatory state facilitated diversification of asg regulation

(A) The evolutionary trajectory of the gain of repression by α2 is shown for a representative **a**sg. Major evolutionary events are indicated by numbered, grey circles. Gains, either in *cis* or *trans* are indicated by yellow stars and losses by a black "x". The regulatory state of the extant yeast are shown (ancestral indicates **a**2 activation only, derived indicates α2 repression only and hybrid indicates both modes of regulation). (B) The hybrid intermediate can "resolve" in different ways. It can revert to the ancestral mode of regulation through loss of the derived mode (left arrow; *K. lactis*), maintain the hybrid in some fashion (circular, center arrow; *K. wickerhamii* and *L. kluyveri*), or lose the ancestral mode of regulation (right arrow; *S. cerevisiae*). (C) Individual genes are regulated differently between and within species. On the left is a recapitulation of part A of this figure. **a**sgs are listed by the *S. cerevisiae* orthologs on the top of the figure and their mode of regulation (if available) are indicated for each species by a colored square (see key in figure).