

Intrinsic cooperativity potentiates parallel *cis*-regulatory evolution

Trevor R Sorrells¹, Amanda N Johnson², Conor J Howard¹, Candace S Britton¹, Kyle R

Fowler¹, Jordan T Feigerle², P Anthony Weil², Alexander D Johnson^{1,3}

1. Department of Biochemistry and Biophysics, Department of Microbiology and Immunology, Tetrad Graduate Program, University of California, San Francisco 94158
2. Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0615
3. Lead Contact

Abstract

Convergent evolutionary events in independent lineages provide an opportunity to understand why evolution favors certain outcomes over others. We studied such a case, where a large set of genes—those coding for the ribosomal proteins—gained *cis*-regulatory sequences for a particular transcription regulator (Mcm1) in independent fungal lineages. We present evidence that these gains occurred because Mcm1 shares a mechanism of transcriptional activation with an ancestral regulator of the ribosomal protein genes, Rap1. Specifically, we show that Mcm1 and Rap1 have the inherent ability to cooperatively activate transcription through contacts with the general transcription factor TFIID. Because the two regulatory proteins share a common interaction partner, the presence of one ancestral *cis*-regulatory sequence can “channel” random mutations into functional sites for the second regulator. At a genomic scale, this type of intrinsic cooperativity can account for a pattern of parallel evolution involving the fixation of hundreds of substitutions.

Introduction

Contingency is widespread in evolution, with chance historical changes dictating the repertoire of future possibilities (Bershtein et al., 2006; Bloom et al., 2006; Blount et al., 2012; Ortlund et al., 2007; Sorrells et al., 2015; Starr et al., 2017). Nevertheless, repeated evolutionary events in different lineages demonstrate that evolution is, at least in some instances, predictable. Repeatability is broadly referred to as convergent evolution, and when a trait arises repeatedly by a similar molecular mechanism, it is referred to as parallel evolution (Stern, 2013). Beginning with Darwin (Darwin, 1883), convergence has been taken as evidence for adaptation, but it can also be caused by drift within the constraints that arise from the properties of biological systems (Losos, 2011). Instances of repeated events can be thought of as natural experimental replicates for finding general principles that result in convergent evolution, and this information could be used to predict which evolutionary paths are most probable.

Genomic studies have found sets of genes that underlie convergent evolution of a wide variety of traits from across the tree of life (Bellott et al., 2010; Denoeud et al., 2014; Gallant et al., 2014; Kohler et al., 2015; Marvig et al., 2015; McCutcheon et al., 2009; Nagy et al., 2014; Pfenning et al., 2014; Sherwood et al., 2014; Soria-Carrasco et al., 2014). Such sets of genes tend to evolve in functionally related groups and are controlled by *cis*-regulatory sequences for particular transcription regulators, forming a transcription network. One way that an entire network can be up- or down-regulated is through alterations in the expression of a “master” regulator, propagating a new expression level to all of its downstream target genes (Chan et al., 2010; McGregor et al., 2007; Rebeiz et al., 2009; Reed et al., 2011). However, in many cases—instead of

changing the expression pattern of a master regulator—each gene in the network independently acquires the same *cis*-regulatory sequence, requiring hundreds of mutations across the genome (Booth et al., 2010; Borneman et al., 2007; Gasch et al., 2004; Kasowski et al., 2010; Lowe et al., 2011; Paris et al., 2013; Piasecki et al., 2010; Schmidt et al., 2010; Tanay et al., 2005; Tuch et al., 2008b). An outstanding question is how this process occurs.

Several molecular mechanisms for gene-by-gene rewiring have been proposed (Britten and Davidson, 1971; Tuch et al., 2008b). Hitchhiking of a *cis*-regulatory site on a transposable element is a mechanism commonly found in plant and animal evolution, and can rapidly bring genes under new regulatory control (Bourque et al., 2008; Chuong et al., 2016; Kunarso et al., 2010; Lynch et al., 2011; Schmidt et al., 2012). Alternatively, one transcription regulator can gain a protein-protein interaction with another, followed by the evolution of *cis*-regulatory sequences (Baker et al., 2012; Lynch et al., 2008; Perez et al., 2014; Tsong et al., 2006). This latter mechanism is able, at least in principle, to rewire an entire set of genes at once upon evolution of the new protein-protein interaction; the individual gains of binding sites could occur secondarily. Nevertheless, many—if not most—examples of network rewiring seem to have occurred in the absence of evidence for either of these two mechanisms (Borneman et al., 2007; Gasch et al., 2004; Kasowski et al., 2010; Lowe et al., 2011; Paris et al., 2013; Piasecki et al., 2010; Schmidt et al., 2010; Tanay et al., 2005; Tuch et al., 2008a).

Here, we studied an example of transcription network rewiring in which ~100 ribosomal protein genes gained binding sites for the Mcm1 transcription regulator, and we describe evidence supporting a new mechanism for the concerted gains. The gain of

Mcm1 binding sites in the ribosomal protein genes occurred in parallel in two respects: 1) the *cis*-regulatory sites were gained upstream of a large proportion of the ribosomal proteins in each species; and 2) they were gained independently in seven different Ascomycete lineages. At each gene, several point mutations were probably necessary to produce a close match to the 16-basepair Mcm1 binding site, thus requiring several hundred mutations across the entire gene set.

Based on results from a variety of experimental approaches, we argue that the gain of Mcm1 sites was potentiated in several different clades by the presence of an ancestral transcription regulator of the ribosomal protein genes, Rap1. We demonstrate that Rap1 and Mcm1 have the intrinsic ability to cooperate in the activation of transcription, even when artificially introduced in species in which their sites are not found together. Biochemical and genetic experiments show that both regulators interact with the general transcription factor TFIID. We propose that the intrinsic, ancient ability of both proteins to interact with a common component of the general transcription machinery facilitated the repeated evolution of the Mcm1 sites in independent lineages.

Results

Gains of functional Mcm1 *cis*-regulatory sites

Previously, genome-wide chromatin immunoprecipitation and bioinformatics experiments revealed that Mcm1 *cis*-regulatory sites evolved independently at the ribosomal protein genes (RPGs) in several yeast lineages (Tuch et al., 2008a). To expand this analysis, we used 161 sequenced fungal genomes, identified RPG regulatory regions, and searched for *cis*-regulatory sequences for Mcm1 and 11 other transcription regulators

that are known to regulate ribosomal components in at least one species (Figure 1A, Supplemental Figure 1). Mcm1 sites were found highly enriched ($-\log_{10}(P) > 6$) at the RPGs in six different monophyletic groups including the clades represented by *Kluyveromyces lactis*, *Candida glabrata*, and *Yarrowia lipolytica*, as well as the individual species *Kazachstania naganishii*, *Pachysolen tannophilus*, and *Arthrotrrys oligospora*. Mcm1 sites were found moderately enriched ($6 > -\log_{10}(P) > 3$) in the RPGs in seven additional lineages.

The pattern of yeast lineages containing Mcm1 binding sites upstream of the RPGs could occur through two general scenarios: the Mcm1 binding sites were gained multiple times, or they were present in an ancient ancestor and lost multiple times. Three lines of evidence indicate that the pattern of Mcm1 *cis*-regulatory sequences in the RPGs represent at least nine independent gains, rather than many more losses. First, the gain model is more parsimonious. Assuming that gains are equally or somewhat less likely than losses (as they both require tens to hundreds of point mutations), the scenarios involving the fewest changes include nine to thirteen independent gains of the Mcm1 sites, along with one to eight losses (Figure 1B). Any scenario in which there are fewer than nine gains substantially increases the number of losses because of the sparse distribution of clades with Mcm1 sites (e.g. a single gain with 29 losses). The numbers of gains and losses are largely unchanged when taking into account (1) the ancient interspecies hybridization that occurred in an ancestor of the *Saccharomyces* clade (Marcet-Houben and Gabaldón, 2015) and (2) ambiguities in the placement of certain branches in the species tree (not shown). Second, other *cis*-regulatory sequences show similar patterns of evolution to that of Mcm1. The Dot6/Tod6 and Rim101 motifs are

also found upstream of the RPGs in several distantly related clades, although fewer clades than for Mcm1 (Supplemental Figure 1). Other regulators show clear single gains in the common ancestor of all Saccharomycotina yeasts (e.g. Rap1), or Pezizomycotina and Saccharomycotina (e.g. Cbf1) as well as losses in sparsely distributed individual clades and species. These results show that gains and losses of *cis*-regulatory sequences are common upstream of the RPGs, and that they can be distinguished from each other based on their distinct distributions among species (Lavoie et al., 2010; Tanay et al., 2005). Third, it was previously shown that the entire set of genes that Mcm1 regulates changes extensively over the timescale of Ascomycete evolution (Tuch et al., 2008a), suggesting that conservation from a distant ancestor would be a marked exception to this trend.

To test whether the Mcm1 *cis*-regulatory sites we identified upstream of the RPGs are functional, we linked several full-length RPG upstream intergenic regions to the fluorescent reporter GFP. We chose *RPL37* and *RPS18* from *Kl. lactis*, each of which has Mcm1 binding sites and measured the expression of these reporters under nutrient-replete conditions. To test whether they contributed to expression, we scrambled the Mcm1 sites. In both cases, the scrambled site reduced expression of the reporter (*RPS18* $P = 0.017$; *RPL37* $P = 0.027$; Welch's *t*-test) but left the cell-to-cell variability unchanged (Figure 1C&D; Supplemental Figure 2). Given that there are many known regulators of ribosomal protein transcription, this demonstrates that Mcm1 plays a non-redundant role in activating these genes in *Kl. lactis* under conditions that require high expression of the translational machinery.

Selection on RPG expression levels

Yeast ribosomes are composed of four ribosomal RNAs and 78 ribosomal proteins, assembled in equal stoichiometry (Woolford and Baserga, 2013). In rapidly growing cells, ribosomal protein transcripts are among the most highly expressed, and have short half-lives, leading to the estimation that approximately 50% of all RNA polymerase II initiation events occur at ribosomal proteins (Warner, 1999). These observations suggest that the expression of these genes is under strong selection as it plays a major role in energy expenditure in the cell.

Given that the Mcm1 *cis*-regulatory sites result in transcriptional activation, there are at least two plausible hypotheses for their appearance upstream of the RPGs. 1) In those species that have acquired Mcm1 *cis*-regulatory sequences, the expression of the RPGs is higher than in species without these sequences. 2) The gain of Mcm1 *cis*-regulatory sequences compensate for other *cis*-regulatory changes that lower expression of the genes, with no net gain in expression levels. This second hypothesis is plausible, in principle, because RPGs are known to lose regulator binding sites as well as gain them (Ihmels et al., 2005; Lavoie et al., 2010; Tanay et al., 2005; Tuch et al., 2008a).

To distinguish between these hypotheses, we examined directly whether the RPGs have a higher expression level in species that have acquired Mcm1 sites compared to those that have not. To accurately measure differences in RPG expression, we mated different species pairs to form interspecies hybrids; we then measured mRNA levels by RNA-seq and assigned each sequencing read to one genome or the other (Figure 2A). Comparing the expression of orthologous genes in this way controls for differences in

“*trans*-acting” factors like transcription regulators and therefore reflects only differences caused by *cis*-regulatory changes (Wittkopp et al., 2004).

For this allele-specific expression experiment, we constructed hybrids between *Kl. lactis* and two other species, *Kl. marxianus* and *Kl. wickerhamii*. These two additional species are relatively closely related to *Kl. lactis*, but have fewer Mcm1 sites at their RPGs (Figure 2A). mRNA reads for each gene were normalized to genomic DNA reads to control for mappability and gene length (see methods). Expression levels and allele-specific expression between genes in the two species’ genomes were reproducible between replicates (Supplemental Figure 3). At a false discovery rate of 0.05, we identified 2925 genes showing allele-specific expression out of a total of 4343 orthologs in the *lactis-marxianus* hybrid, and 3432 out of 4319 in the *lactis-wickerhamii* hybrid.

To ask broadly whether RPGs have experienced concerted *cis*-regulatory evolution, in each hybrid we asked whether the RPGs were more likely to show allele-specific expression in one direction relative to all genes in the genome (Figure 2B). In both hybrids, the RPGs showed evidence for *cis*-regulatory evolution (hypergeometric test, $P = 2.36e-3$ for *lactis-marxianus* hybrid; $P = 1.52e-6$ for *lactis-wickerhamii* hybrid). We conclude that in these species, the expression of the ribosomal proteins has evolved directionally as a group through *cis*-regulatory changes.

However, in both hybrids, the ribosomal protein genes were expressed, on average, lower in the *Kl. lactis* genome than in either of the other genomes. This observation rules out the hypothesis that the gain of Mcm1 sites was simply due to directional selection to increase expression of the ribosomal proteins as a whole (see also Supplemental Figure 4). Instead, it favors the second hypothesis: as Mcm1 sites evolved

in the *Kluyveromyces* clade, they compensated (perhaps incompletely) for other *cis*-regulatory changes that reduced RPG transcript levels.

We also revisited previously published yeast interspecies hybrid experiments performed with species from the *Saccharomyces* clade. Surprisingly, in all cases, the ribosomal proteins were among the sets of genes with reported directional *cis*-regulatory evolution (Bullard et al., 2010; Clowers et al., 2015; Lee et al., 2013; Martin et al., 2012). These observations demonstrate that RPG expression may frequently experience different evolutionary forces between closely related yeast species, and they are consistent with the high rate of gains and losses of *cis*-regulatory sequences that control RPG transcription.

Why were Mcm1 sites gained repeatedly?

Although RPGs have experienced different selection pressures across different species, this observation does not explain why Mcm1 sites, as opposed to *cis*-regulatory sequences for many other transcription regulators, are repeatedly gained in the lineages we examined. One possibility is that Mcm1 expression changes under certain environmental conditions, conferring optimal RPG expression levels for the *Kluyveromyces* clade (as well as the other clades that gained Mcm1 sites). The fact that Mcm1 is expressed across all yeast cell types and regulates many different genes that are part of different expression programs (Tuch et al., 2008a) makes this possibility unlikely, but we cannot explicitly rule it out.

We therefore investigated whether Mcm1 has other characteristics that allow it to gain *cis*-regulatory sites especially easily at the RPGs. We previously noted that, in *Kl. lactis*, the Mcm1 sites were gained a fixed distance away from the binding site for

another transcription regulator, Rap1, whose sites at the RPGs were ancestral to *Kl. lactis* and *S. cerevisiae* (Tuch et al., 2008a). By pooling data from additional species with Rap1-Mcm1 sites, we discovered that the spacing between Mcm1 and Rap1 sites was particularly precise with peaks at 54, 65, and 74 basepairs apart, favoring a configuration with the two proteins on the same side of the DNA helix (Figure 3A-C).

In order to understand whether the spacing between Mcm1 and Rap1 sites affects transcriptional activation, we moved a segment of the *Kl. lactis RPS23* upstream region that contains the Rap1 and Mcm1 sites into a heterologous reporter containing a basal promoter from *S. cerevisiae CYC1* (Guarente and Ptashne, 1981) (Figure 3D). This construct allowed us to study the Rap1 and Mcm1 *cis*-regulatory sequences independent of the additional regulators of the RPGs. We systematically varied the spacing between these two *cis*-regulatory sequences in 2bp increments and found that the transcriptional output remained similar in most constructs (Figure 3E). In these experiments the Mcm1 and Rap1 sites were both close to optimal, and we hypothesized that weaker sites would be more likely to reveal a spacing preference. To test this idea, we performed a second series of experiments with the Rap1 and Mcm1 *cis*-regulatory sequences from the *Kl. lactis RPS17* gene. The binding sites from this gene are weaker matches to the Rap1 and Mcm1 position weight matrices (9.34 versus 12.48 for Rap1 and 7.24 versus 14.33 for Mcm1). The transcriptional output of this second series of constructs showed sensitivity to spacing between the sites (Figure 3E).

We hypothesized that the observed sensitivity of transcriptional output to the strength and spacing of Rap1 and Mcm1 *cis*-regulatory sequences reflected cooperative activation of transcription. Deletion of the Mcm1 and Rap1 binding sites individually and

in combination showed that the proteins *KlMcm1* and *KlRap1* activated transcription cooperatively in *Kl. lactis*: their combined expression was approximately four times the sum of their individual contributions (Figure 3F; $P = 5.5e-4$, one sample *t*-test).

One likely explanation for these observations is the evolution of a favorable protein-protein interaction between Rap1 and Mcm1 that occurred around the time that Mcm1 sites appeared at the RPGs in the *Kluyveromyces* and *glabrata* clades (Tuch et al., 2008b), i.e. a “derived cooperativity” model (Figure 3G). To test this model, we placed the *Kl. lactis* Rap1-Mcm1 reporter into the genome of *S. cerevisiae*, a species that lacks Mcm1 sites at the RPGs. We found that *ScRap1* and *ScMcm1* activated expression of the reporter cooperatively ($P = 3.9e-3$, one sample *t*-test), thereby demonstrating that these proteins have the capacity to work together even in a species where their binding sites did not evolve close proximity (Figure 3H). This result rejects the derived cooperativity model and strongly suggests that the ability of Rap1 and Mcm1 to activate transcription cooperatively was ancestral, existing even in species that did not take advantage of it in regulating the RPGs.

Mechanism of intrinsic Mcm1-Rap1 cooperativity

Having established that cooperativity of Rap1 and Mcm1 was likely ancestral, we considered three possible mechanisms that could explain ancestral cooperativity: (1) Rap1 and Mcm1 could bind DNA cooperatively through an ancient, favorable protein-protein interaction, (2) they could bind nucleosomal DNA cooperatively through displacement of histones, or (3) they could bind a third transcription regulator resulting in cooperative transcriptional activation. To test the first possibility, we used a gel-mobility

shift assay with a radiolabeled DNA sequence from the *Kl. lactis* *RPS23* gene (Figure 3D). We asked whether purified full-length *S. cerevisiae* and *Kl. lactis* proteins as well as cell lysates from three additional species bound to Rap1 and Mcm1 binding sites cooperatively. In all cases, each protein bound to the RPG promoter DNA independently and did not appear to increase the other's affinity, indicating that Rap1 and Mcm1 do not, on their own, bind DNA cooperatively (Figure 4A, Supplemental Figure 5).

A second mechanism explaining cooperative activation is through displacement of nucleosomes. Many transcription regulators have the inherent property of competing for binding to DNA with nucleosomes, with some more effective than others (Zaret and Carroll, 2011). However, it is unlikely that cooperative nucleosome displacement is the primary mechanism for cooperative activation by Rap1 and Mcm1 at the RPGs. Although Rap1 can indeed bind nucleosomal Rap1 binding sites (Koerber et al., 2009; Rossetti et al., 2001), and displace nucleosomes (Kubik et al., 2015; Lickwar et al., 2012; Platt et al., 2013), it remains bound at the RPGs even during stress conditions when the genes are repressed and show higher levels of nucleosome occupancy (Bernstein et al., 2004; Lee et al., 2004). Furthermore, a general mechanism for the cooperative assembly of all transcription regulators has little explanatory power for why *cis*-regulatory sites for Mcm1 were repeatedly gained across RPGs rather than sites for many of the ~250 other regulators coded in the yeast genome.

We next considered the third possibility, namely that the cooperative transcriptional activation we observe for Rap1 and Mcm1 occurs through the interaction of both with a third factor that catalyzes a rate-limiting step in transcription activation (Lin et al., 1990). In principle, a regulator could activate transcription through contacts

with general transcription factors, Mediator, SAGA, various chromatin remodeling complexes, or RNA polymerase itself. To identify possible factors that might directly bind both Rap1 and Mcm1, we searched the *S. cerevisiae* BioGRID database for common interaction targets of both proteins (Oughtred et al., 2016). Two complexes, SWI/SNF and TFIID, fit this criterion. SWI/SNF is not required for ribosomal protein transcription (Sudarsanam et al., 2000), so it is unlikely that binding to SWI/SNF plays a role in Rap1 and Mcm1 cooperativity at the RPGs.

TFIID is a general transcription factor whose direct interaction with Rap1 is required for RPG transcription in *S. cerevisiae* (Garbett et al., 2007; Layer et al., 2010; Mencia et al., 2002; Papai et al., 2010; Reja et al., 2015). Furthermore, TFIID activates transcription through contacts with RNA Polymerase II, and its binding to the promoter is a rate-limiting step in the activation of TFIID-dependent genes (Wu and Chiang, 2001), such as the RPGs. The mechanism of how activators increase transcription rate through contacts with TFIID is not entirely understood but occurs either through a structural rearrangement of the complex or simply by an increase in its occupancy on DNA (Coleman et al., 2016; Fuda et al., 2009; Nogales et al., 2016; Papai et al., 2011).

The interaction of Rap1 with TFIID has been extensively documented (Garbett et al., 2007). Of particular importance is the interaction between the “activation domain” of Rap1 and the Taf5 subunit of TFIID, as mutations that compromise this interaction strongly reduce ribosomal protein gene transcription (Johnson and Weil, 2017). To test directly whether Mcm1 also binds to TFIID (as suggested by mass spectrometry-based analyses of TFIID and associated proteins (Sanders et al., 2002)), we performed a Far-Western protein-protein binding assay using purified *S. cerevisiae* TFIID, ScMcm1, and

individual TFIID subunits. We found that Mcm1 binds directly to the Taf4 subunit (Figure 4B,C); using deletions of Taf4, we further mapped the interaction to the N-terminal region of Taf4 (Figure 4D,E). Although Rap1 also binds to Taf4 (in addition to Taf5 and Taf12), its target is in the C-terminus of this subunit, distinct from the Mcm1 interaction site.

The finding that Rap1 and Mcm1 both interact with distinct domains of a common component of the transcription machinery, one whose assembly at the promoter is rate limiting, provides a simple explanation for their ability to activate transcription cooperatively. To test this idea explicitly, we performed a series of experiments *in vivo*. Because Rap1 is an essential gene, we took advantage of an altered DNA-binding specificity Rap1 (Rap1^{AS}) that binds to a non-natural *cis*-regulatory sequence and confers expression of a reporter (Johnson and Weil, 2017). Rap1^{AS} could then be manipulated without compromising the function of the endogenous Rap1. As shown in Figure 5A-C, Rap1^{AS} shows cooperative transcriptional activation with Mcm1. When the activation domain of Rap1^{AS} was mutated by introducing 7 point mutations (Rap1^{AS7Ab}) its ability to activate transcription strongly decreased, but was not entirely eliminated (Johnson and Weil, 2017). However, these mutations destroyed the ability of Rap1 to activate transcription of the reporter cooperatively with Mcm1 (Figure 5D-F).

Taken together, these experiments indicate that the cooperative transcriptional activation by Rap1 and Mcm1 at the RPGs in these clades is due to both proteins interacting with TFIID, a component of the general transcriptional machinery. This idea explains how Mcm1—and not a random mixture of other regulatory proteins—came to be repeatedly gained at the ribosomal protein genes.

Implications and predictions of the intrinsic cooperativity model

We tested several evolutionary and molecular predictions of this model. One prediction is that Rap1 and Mcm1 *cis*-regulatory sequences would occur at the prescribed distance apart but located at other genes besides the RPGs. We searched a subset of hemiascomycete yeast genomes for Rap1-Mcm1 sites with similar spacing and orientation to that observed in the RPGs (Supplemental Figure 6). Most species analyzed had only a few such genes of questionable significance, but *Ka. naganishii* had 33, showing that Rap1 and Mcm1 sites can evolve at genes other than the RPGs.

A second prediction of our model is that even a sub-optimal Mcm1 site would activate transcription of a regulatory region with an ancestral, strong Rap1 site. Because sub-optimal sites can evolve *de novo* with much higher probability than optimal sites (Dermitzakis and Clark, 2002; Stone and Wray, 2001), this property would increase the ease by which a large number of functional Mcm1 sites could arise during evolution. To test this prediction, we created a series of GFP reporter constructs with Mcm1 binding sites that drive different levels of expression (Acton et al., 1997). We tested the expression level of each of these Mcm1 sites in the presence and absence of a neighboring Rap1 site in *Kl. lactis* (Figure 6A). Consistent with the prediction, most of the sub-optimal Mcm1 sites displayed near wild-type levels of expression in the presence—but not in the absence—of neighboring Rap1 sites.

Finally, we asked whether our conclusions are generalizable to other pairs of transcription factors besides Mcm1 and Rap1. In particular, some of the clades identified in Figure 1A gained Mcm1 sites at RPGs that lack ancestral Rap1 sites. To address this

question, we centered each RPG intergenic region on the best Mcm1 site and plotted the location of other RPG regulators relative to Mcm1 (Figure 6B). This analysis revealed that, in some of these clades, Mcm1 sites were gained varying distances away from the pre-existing sites of other regulators (Tbf1, Rrn7, and Fhl1), suggesting that the evolutionary mechanism we identified with Rap1 and Mcm1 might be generalizable to other instances of *cis*-regulatory evolution (Figure 6C). Consistent with this idea, all three of these regulators are reported to interact (directly or indirectly) with TFIID (Gavin et al., 2002; Knutson et al., 2014; Mallick and Whiteway, 2013; Zhong and Melcher, 2010).

In summary, the experiments we have presented describe a special relationship between Rap1 and Mcm1 by virtue of their interaction with different surfaces of the same rate-limiting component of transcription. This relationship between Rap1 and Mcm1 is ancient, and in the next section, we discuss how this property can predispose transcription networks to evolve repeatedly along the same trajectory.

Discussion

Here we have investigated an example of parallel evolution where binding sites for a particular transcription regulator (Mcm1) were gained in a large group of genes (the ribosomal protein genes, RPGs). These gains occurred repeatedly in several independent fungal lineages. In three of these lineages, Mcm1 binding sites were gained a fixed distance from the sites for another transcription regulator, Rap1, and we show that these newly acquired Mcm1 sites are required for full activation of the RPGs. We also show that Mcm1 and Rap1 cooperatively activate these genes, at least in part through direct interactions of both regulators with a common target, the general transcription factor

TFIID. We demonstrate that the ability of Rap1 and Mcm1 to work together was ancestral to the more recent gains of Mcm1 sites adjacent to Rap1 sites at the RPGs.

How do these observations account for the fact that Mcm1 sites (as opposed to sites for other regulators) were repeatedly gained in parallel next to the Rap1 site at the RPGs? And how do they account for the distance constraints? One common explanation for parallelism is a specific environmental adaptation that occurs through a similar molecular mechanism. However, the yeast species with Mcm1 sites at the RPGs are from diverse ecological niches (e.g. plant leaves, mangrove sediment, the human body, soil) and utilize different nutrient sources (e.g. lactate, xylose, feline skin, nematode predation), defying a specific environmental adaptation explanation. Consistent with this view is our observation, based on analyzing interspecies hybrids, that a species in which the Mcm1 sites were gained at the RPGs does not express these genes at higher levels than related species that evolved fewer Mcm1 sites.

The model that best fits all of our data holds that the parallel gains arose from the ease with which the functional Mcm1 sites (and not the sites of other regulators) appeared in evolution, rather than selective pressure for particular adaptation. Specifically, we propose that, in the *Kluyveromyces-Saccharomyces* ancestor (before the parallel gains of Mcm1 sites) Rap1 bound to the ribosomal protein genes and activated transcription through interactions with TFIID, as it does in extant species. We further propose that, in the ancestor, Mcm1 activated non-ribosomal genes by a conserved interaction with a second site on TFIID, as it does in the extant species *S. cerevisiae*. Any suboptimal Mcm1 site that arose by chance point mutation at a specified distance from a Rap1 site would immediately be functional because even a weak Mcm1 DNA interaction would be

stabilized by Mcm1's intrinsic ability to directly bind TFIID (see experiments in Figures 4C & 6A). In this way, even suboptimal *cis*-regulatory sequences (which are much more likely to appear by chance than optimal sites (Dermitzakis and Clark, 2002; Stone and Wray, 2001)) could form under selection. The appearance of Mcm1 sites likely occurred concomitantly with the gradual losses of other *cis*-regulatory sites in the RPGs; in other words, the Mcm1 *cis*-regulatory sites would fall under selection as other *cis*-regulatory sites deteriorated by mutation (Figure 7). In essence, we propose that the free energy gained from the intrinsic interaction between Mcm1 and TFIID would favor formation of new Mcm1 sites at the expense of other *cis*-regulatory sequences, particularly since the latter provide a larger mutational target. This model can account for why Mcm1 sites (and not those of other transcription regulators) were repeatedly gained at the RPGs and why the distance between Rap1 and Mcm1 sites is constrained in those species in which the gains occurred.

Numerous experimental observations support this model and rule out alternative explanations: 1) In extant species, Rap1 and Mcm1 both interact with TFIID; 2) They interact with different parts of TFIID; 3) cooperative transcriptional interaction between Mcm1 and Rap1 requires the activation domain of Rap1, which is known to interact with TFIID; 4) the spacing between Rap1 and Mcm1 sites in the ribosomal protein genes places the proteins on the same side of the helix but at least 50bp apart, consistent with a physical interaction with a large complex; 5) Engineered suboptimal Mcm1 sites are functional as long as they are adjacent to Rap1 sites; 6) Mcm1 and Rap1 have the intrinsic ability to cooperate (through interactions with TFIID) even in a species where Mcm1 sites were not gained at the RPGs.

We note that this model does not require any change in Rap1 or Mcm1 during the gain of Mcm1 sites at the RPGs. Presumably, Rap1 and Mcm1 activated many genes separately in the ancestor, thus preserving by stabilizing selection their ability to interact with TFIID. This idea is consistent with the observation that the two proteins were able to cooperate on artificial constructs introduced into *S. cerevisiae* even though their binding sites are not found together at the RPGs. Mutations that alter the function of transcription regulators (for example, creating a new protein-protein interaction) can be pleiotropic, decreasing the likelihood that they can arise without disrupting the proteins ancestral function (Carroll, 2005; Stern and Orgogozo, 2008). (We note that the transcriptional output was slightly more cooperative in *Kl. lactis* than in *S. cerevisiae*, leaving open the possibility that additional fine-scale evolutionary changes may have occurred in how these proteins interact.) According to our model, the ability of the two regulators to work together was ancestral, part of each protein's intrinsic mechanism of transcriptional activation; therefore, their coupling at the RPGs avoided such pleiotropic changes.

How does this model account for the gains of Mcm1 sites observed in clades where Rap1 does not regulate the RPGs? In these cases, Mcm1 *cis*-regulatory sequences also show preferred spacing relative to known regulators in the RPGs, Tbf1 and Rrn7, and we propose that the same type of cooperativity with TFIID can also account for these cases. These two regulators are reported to interact with TFIID (Knutson et al., 2014; Mallick and Whiteway, 2013). Indeed, TFIID occupies the promoters of RPGs in human cell lines as well (ENCODE Project Consortium et al., 2012), raising the possibility that TFIID is a conserved general activator of the RPGs across fungi and animals, while the

specific transcription regulators that interact with TFIID simply interchange over this timescale.

While our cooperative activation model provides an explanation for the parallel acquisition of Mcm1 *cis*-regulatory site evolution, selection must have operated to preserve the Mcm1 sites as they arose in the population. Because they code for abundant, stoichiometric proteins, RPG expression levels must be under strong stabilizing selection. Based on the allele-specific expression data from multiple yeast species reported here and elsewhere, selective forces appear to be different among closely related yeast species and have resulted in subtly different relative expression levels from one species to the next. Possible causes include a relaxation of purifying selection in one lineage or directional selection to increase or decrease expression, with these forces varying over time. Thus over time, the gain of Mcm1 sites would compensate for other changes in the regulation of ribosomal gene transcript levels (Zeevi et al., 2014; 2011). This model helps to account for the surprising observation that transcriptional regulators that control the RPGs vary substantially across species (Gasch et al., 2004; Lavoie et al., 2010; Mallick and Whiteway, 2013; Tanay, 2004; Tuch et al., 2008a).

In conclusion, we have proposed a mechanism, supported by multiple lines of experimental evidence, to account for convergent regulatory evolution of a large set of genes. Although, on the surface, the parallel gains of Mcm1 sites at the ribosomal genes would seem to require a special evolutionary explanation, our model does not require an extraordinary mechanism beyond individual point mutations in the *cis*-regulatory region of each gene. However, the ancestral ability of the two key regulators to activate transcription simplifies the path to gaining these sites by producing a phenotypic output

from even non-optimal sites. Thus Mcm1, because of its intrinsic ability to cooperate with Rap1, can significantly activate transcription at the ribosomal proteins more easily than it would elsewhere in the genome; likewise Mcm1 (or another regulator that interacted with TFIID) would be preferred at the RPGs over regulators that did not share this common direct protein interaction. Thus, the intrinsic cooperativity of Rap1 and Mcm1 “channels” random mutations into functional Mcm1 *cis*-regulatory sequences, accounting for the observed parallel evolution. We speculate that gene activation through intrinsic cooperativity may be a general mechanism to explain the rapid and ubiquitous rewiring of transcription networks.

Acknowledgements

We thank K. Verba, L. Pack, S. Liang, C. Zhou, K. Pollard, and E. Chow for experimental and analysis guidance. We thank M. Lohse, I. Nocedal, S. Singh-Babak, V. Hanson-Smith and other members of the Johnson lab for technical guidance and comments on the manuscript. We thank S. Åstrom for use of the *Kl. dobzhanskii* genome prior to publication. We thank I. Grigoriev, J. K. Magnuson, P. Inderbitzin, M. Nowrousian, A. Grum-Grzhimaylo, K. O’Donnell, G. Bonito and the Metatranscriptomics of Forest Soil Ecosystems project, D. Greenshields, J. Crouch, F. Martin and the Mycorrhizal Genomics Initiative, and R. Gazis for providing access to unpublished genome data produced by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This research was supported by the grants R01 GM115892, R01 GM037049, and T32NS00749 from

the National Institutes of Health. T.R.S. was supported by a Graduate Research Fellowship from the National Science Foundation.

Author Contributions

TRS and ADJ conceived of the project. TRS designed and performed computational genomics analyses and allele-specific expression in interspecies hybrids. TRS and ANJ designed reporter experiments and TRS, ANJ, CJH, CSB, KRF performed them. ANJ, TRS, and JTF purified proteins. ANJ designed and performed Far Western experiments. TRS, ANJ, and JTF designed and performed gel-mobility shift assays. TRS, ANJ, and ADJ wrote the manuscript. PAW, CSB, and KRF edited the manuscript, and all authors approved of the final publication.

Data and code availability

Allele-specific expression data is available at the Gene Expression Omnibus (GEO) repository under accession number GSE108389. Flow cytometry data is available at Flow Repository under accession numbers FR-FCM-ZYWS, FR-FCM-ZYWT, FR-FCM-ZYWU, FR-FCM-ZYWV, FR-FCM-ZYJZ, FR-FCM-ZYJY, and FR-FCM-ZYJ2. Code used in computational analyses is available at the URL: https://github.com/trevorsorrells/RPG_evolution and will be available using a DOI for final publication.

Methods

Computational genomics of *cis*-regulatory sequences

Genomes were compiled from the Yeast Gene Order Browser (YGOB) (Gordon et al., 2011), the Department of Energy Joint Genome Institute MycoCosm portal (Grigoriev et al., 2014), and numerous individual genome releases. For *Kl. wickerhamii*, *Kl. marxianus*, and *H. viniae*, their genomes were annotated using the Yeast Gene Annotation Pipeline associated with YGOB. The annotations of genes and proteins in other genomes were obtained from the source of the genome sequence. RPGs were defined as any gene starting with “Rps” or “Rpl” and were identified using the ortholog annotation in YGOB. These genes were identified in genomes from other sources through psi-blast in BLAST+ (Camacho et al., 2009), as they are highly conserved.

Intergenic regions were extracted upstream of each gene using the python script `intergenic.py`. For scoring of potential transcription factor binding sites, motifs were obtained from the ScerTF database (Spivak and Stormo, 2012) as position weight matrices. Each intergenic region (in the ribosomal proteins or genome-wide, depending on the question) was scored using the script `TFBS_score.py` by adding up the log likelihood values of each base at each position in the motif, forward and backward, and then repeating for each position in the intergenic region. It is important to note that the motifs from ScerTF are corrected for the GC-content of the *S. cerevisiae* genome, but not individually for the genomes of each species in other parts of the tree. The motif was left as-is instead of correcting for the GC-content in each genome, because the purpose of this scoring was to identify DNA sequences that are most similar to the Mcm1 binding site, not those that are most statistically enriched given the GC-content.

The location of a binding site was defined as the midpoint of the motif. Previous studies have indicated that many features of the RPGs are defined by their relative location to the Rap1 binding site (Reja et al., 2015). To identify the relative locations, the best hit for the Rap1 site in each species was identified, then a cutoff was set (usually about half of the maximum potential score for a given position weight matrix) for the motif of the second transcription factor. This analysis was carried out using the scripts `TFBS_score.py` and `rel_locs_RPs.py`.

Genome-wide scoring of relative motif locations using `rel_loc_RPs.py` was used to identify additional genes beyond the RPGs that show a similar pattern of Rap1 and Mcm1 sites near to each other. Rap1 sites that faced toward the gene and had a score greater than 6.0 were identified, as were Mcm1 sites with a score above 6.0. Then, genes that had both a forward-facing Rap1 site and an Mcm1 site between 52 and 78bp downstream (the spacing of most sites at the RPGs) were identified. The order of Rap1 and Mcm1 *cis*-regulatory site appearance in evolution was inferred by the distribution of the sites in closely related species with available genome sequences.

Strain and reporter construction

The GFP and *HIS3* reporters used in *K. lactis* and *S. cerevisiae* have been previously described (Garbett et al., 2007; Mencía et al., 2002; Sorrells et al., 2015). These reporters allow full intergenic regions to be cloned upstream of GFP or *HIS3*, with the break between the original gene sequence and the reporter gene occurring at the start codon. A second version of the GFP reporter uses the *CYC1* promoter from *S. cerevisiae* (Guarente and Ptashne, 1981) with its upstream activation sequence replaced by two

restriction enzyme sites. Different versions of these vectors were made to integrate into the *K. lactis* genome and into the *S. cerevisiae* genome.

To make the full-length RPG GFP reporters, the wild-type intergenic regions were obtained by PCR with ExTaq (Takara) from genomic DNA, flanked by directional restriction enzyme sites for SacI and AgeI. These were cloned using a 2:1 ratio into pTS16 digested with the same restriction enzyme sites, and ligated using Fastlink ligase (Epicentre) to make pTS170 and pTS174.

To scramble Mcm1 and Rap1 binding sites, these sites were put into a text scrambler, then the resulting sequences were queried in ScerTF (Spivak and Stormo, 2012) to see if they contained matches to any other known transcription factor binding site motifs. If not, they were used for further experiments.

To make pTS171, pTS175, and pTS243-246 DNA sequences were synthesized containing scrambled Mcm1, Rap1 sites, or both. The insert for pTS171 and pTS243-244 were cloned into pTS170 using the restriction enzymes SacI and Bsu36I. The insert for pTS175 and pTS245-246 was cloned into pTS174 using SacI and EcoRV.

The vectors pTS176-179 contain the *K. lactis RPS23* Rap1-Mcm1 operator upstream of the *CYCI* promoter. These vectors were made by annealing oligos and ligating them in a 50:1 ratio into pTS26 digested with NotI and XhoI. The equivalent vectors for *S. cerevisiae* are pTS181-184 and were made by cloning into pTS180 with NotI and XhoI. The constructs testing the spacing between Rap1 and Mcm1 sites (pTS189-203 and pTS209-224) were cloned using the same approach. For pTS189-203 the intervening sequence between the sites was partially duplicated for some constructs to

increase the spacing. For pTS209-224 the endogenous spacing was 80bp so the entire series was made with deletions starting immediately downstream from the Rap1 site.

The reporters testing how weak Mcm1 sites cooperate with Rap1 were cloned by first adding a BamHI site along with a palindromic Mcm1 site (Acton et al., 1997) into the reporter containing the *K. lactis* *RPS23* Rap1-Mcm1 operator to make pTS247 and pTS248 (with a scrambled Rap1 site). Then variants of the palindromic site containing point mutations were cloned into these two vectors using BamHI and XhoI. For each variant, two point mutations were made to preserve the palindromic nature of the Mcm1 binding site (Acton et al., 1997).

These reporters were digested with KasI and HindIII and integrated into the *K. lactis* genome by transformation as previously described (Kooistra et al., 2004) and into the *S. cerevisiae* genome using a standard lithium-acetate transformation. Yeast were grown on non-selective media for 24h then replica plated onto plates containing at least 100µg/mL Hygromycin B. For the spacing and weak Mcm1 reporter series, the reporters are enumerated in the plasmid list but not the strain list. This is because they were transformed into *K. lactis*, tested, then discarded due to their large numbers. For each reporter, 4 independent isolates were measured, but isolates where the full reporter had not integrated were discarded, resulting in 3 or 4 replicates per construct.

The *HIS3* reporters were generated by performing PCR on the equivalent GFP reporters to generate wild-type, Rap1^{AS}, and scrambled versions of the *RPS23* fragment containing Rap1 and Mcm1 binding sites with NcoI and SacII restriction sites on the ends. These fragments were cloned in a 3:1 ratio into similarly digested *UAS*_{Rap1^{WT}}-*HIS3* reporter plasmid (Garbett et al., 2007; Mencía et al., 2002). These reporters were digested

with SpeI and SalI and integrated into the *S. cerevisiae* genome via lithium-acetate transformation. Transformants were selected on media lacking *TRP1* and integration at the correct locus was confirmed via PCR.

GFP Reporter Assays

K. lactis and *S. cerevisiae* reporters were grown overnight in 1mL cultures in 96 well plates in synthetic complete media. The next day, cells were diluted into synthetic complete media to $OD_{600} = 0.025-0.05$ and grown for 3h. Cells were measured by flow cytometry on a BD LSR II between 3h and 4h after dilution. A total of 10,000 cells per strain were recorded. Cells were gated to exclude debris, and the mean fluorescence for each strain was used for comparing among different strains. For each reporter, three to four independent isolates were checked. In the case that one of the isolates anomalously showed expression equivalent to background, while the other isolates showed similar but detectable fluorescence, the anomalous isolate was excluded from analysis. Experiments were performed a minimum of two times on different days.

HIS3 Reporter Assays

HIS3 reporter expression in *S. cerevisiae* was scored by growth assays performed on three independent biological replicates. In these assays, *S. cerevisiae* were grown overnight to saturation and serially diluted 1:4 in sterile water in 96-well plates. These dilution series were spotted using a pinning tool (Sigma) onto non-selective media (SC+ His) and selective media plates (SC-His + 3-AT). Plate images were acquired using the

ChemiDoc MP Imager (Bio-Rad) and processed using ImageLab software (Bio-Rad) after 2 days of growth at 30°C.

Interspecies hybrids

To construct the interspecies hybrids for allele-specific expression measurements, multiple isolates of different species were mated together. Strains with complementary markers were grown on YEPD plates for 2 days, then mixed together in patches on 5% malt extract plates for 2-4 days. These patches were then observed under the microscope to check for zygotes and streaked out onto plates that select for mating products. Hybrids were tested by PCR for products that were species, and mating-type specific. *Kl. lactis*-*Kl. dobzhanskii* matings were attempted, but were unsuccessful, perhaps because the *Kl. dobzhanskii* isolate used was an \mathbf{a}/α strain. *Kl. lactis*-*L. kluyveri* zygotes were observed when the two species were mixed with *Kl. lactis* alpha-factor, but no mating products were obtained. One cross of *Kl. lactis* and *Kl. aestuarii* produced zygotes and mating products, but the *Kl. aestuarii* isolate turned out to be an isolate of *Kl. wickerhamii* instead (discovered upon genome-sequencing). In all, three *Kl. lactis*-*Kl. wickerhamii* matings, and one *Kl. lactis*-*Kl. marxianus* mating—each with three isolates—were obtained and carried forward for analysis. *Kl. lactis*-*Kl. wickerhamii* hybrids are yTS347 and yTS349 (two matings between yLB13a and yLB122), and yTS353 (a mating between yLB72 and yLB66c). The *Kl. lactis*-*Kl. marxianus* mating was yTS352 (a mating between yLB72 and CB63).

Both mRNA and genomic DNA were sequenced from the hybrids. Genomic DNA of one isolate of each of the *Kl. lactis*-*Kl. wickerhamii* hybrids (yTS347, yTS349 and

yTS353) was sequenced, along with all three isolates of the *Kl. lactis*-*Kl. marxianus* hybrid, and one isolate of each of the parental strains. Cells were grown in 5mL cultures overnight in YEPD. Genomic DNA was prepared using a standard “smash and grab” protocol, where cells are lysed with phenol/choloroform/isoamyl alcohol and glass beads. DNA was precipitated twice and treated with RNase A, then sheared on a Diagenode Bioruptor for 2 x 10 min (30 s on 1 min off) on medium intensity. Genomic DNA was prepared for sequencing using the NEBNext Ultra DNA Prep Kit for Illumina E7370 (New England BioLabs).

All three isolates of each of the four hybrids were prepared for mRNA sequencing. Cells were grown overnight, then diluted back into YEPD to $OD_{600}=0.2$ and grown for 4-8h until they reached $OD_{600}=0.7-1.0$. The growth rate of the *Kl. lactis*-*Kl. wickerhamii* hybrids was slower and more variable between isolates. At this point, cells were pelleted and frozen in liquid nitrogen. mRNA was extracted using the RiboPure kit AM1926 (Applied Biosystems). Polyadenylated RNAs were selected using two rounds of the Oligotex mRNA Kit 70022 (Qiagen). The samples were then concentrated using the RNA Clean & Concentrator-5 (Zymo Research).

Libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina E7420 (New England BioLabs). mRNA and library quality were assessed using a Bioanalyzer 2100 (Agilent). Sequencing was performed at the University of California, San Francisco Center for Advanced Technology on an Illumina HiSeq 4000.

Allele-specific expression analysis

Raw sequencing data was checked for quality control using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Next, each of the genomic DNA isolates was aligned separately to each of four genomes: *Kl. lactis*, *Kl. marxianus*, *Kl. wickerhamii*, and *Kl. aestuarii*. In each case, reads uniquely mapped to the expected genome(s) and few reads mapped to other genomes. The strain yTS349 was previously thought to be a *Kl. lactis*-*Kl. aestuarii* hybrid, but sequencing revealed it was in fact a *Kl. lactis*-*Kl. wickerhamii* hybrid, and was treated as such for the analyses.

The genomes for *Kl. lactis*, *Kl. marxianus*, and *Kl. wickerhamii* were annotated using the yeast genome annotation pipeline from YGOB to standardize the gene annotation and ortholog assignment. *Kl. lactis* is already included in YGOB. These files were converted to gff format using `convert_YGAP_GFF.py` and genes were extracted using `pull_genes.py`. Next, hybrid genomes were created *in silico* by concatenating fasta sequences of the genes of each species. mRNA and gDNA reads were aligned to the hybrid genomes on a computer cluster using the script `aln_reads.py`, which calls Bowtie 2 (Langmead and Salzberg, 2012). Default parameters were used, which allow mismatches in Bowtie 2. However, reads that mapped equally well to multiple locations in the genome were removed after alignment using `discard_multimapping.py`. Because ribosomal proteins are highly conserved, they contain stretches of more than 50bp that are identical between the orthologs belonging to each species in the hybrid. Thus, this filtering step is necessary to assure reads map uniquely to the ortholog from one species or the other.

To quantify allele-specific expression, the reads aligning to each gene were counted using `ASE_server.py`. mRNA counts of the three *Kl. lactis*-*Kl. marxianus*

replicates and seven of the nine *Kl. lactis*-*Kl. wickerhamii* hybrids that were sequenced were highly similar. The two other isolates showed that one of the genomes in the hybrid was present at a lower level than the other suggesting it had been lost from some of the cells (although it was a different parental genome in each of the two isolates). The seven reproducible isolates were then treated as replicates for the rest of the analysis. Genome sequencing also revealed that there were two copies of the *Kl. marxianus* genome in each of the *lactis-marxianus* hybrids, suggesting that our parental *Kl. marxianus* strain was a diploid. The mRNA read counts for each gene in each replicate were then normalized to the total reads in the experiment. Second, they were divided by the gDNA read counts from each gene, thus controlling for the effect of two *Kl. marxianus* genomes in the *lactis-marxianus* hybrids. (The gDNA read counts per gene were averaged across replicates for each hybrid, so all the replicates were divided by the same gDNA count). Finally, the *Kl. lactis* ortholog read count was divided by either the *Kl. marxianus* or the *Kl. wickerhamii* ortholog to get an allele-specific expression value for each ortholog pair in each replicate.

To calculate significance of allele-specific expression, a two-sided one-sample t-test was used on the $\log_2(\text{allele-specific expression})$, across replicates (three replicates in the case of the *Kl. lactis*-*Kl. marxianus* hybrid and seven in the case of the *Kl. lactis*-*Kl. wickerhamii* hybrid). The significance of each gene was calculated using the Benjamini-Hochberg procedure to control the false-discovery rate at 0.05. To test for concerted allele-specific expression in the ribosomal proteins, as well as across all gene ontology (GO) terms, the geometric mean of each group of genes was calculated, then tested by the hypergeometric test to see if they were enriched for genes at least 1.1-fold up or down in

Kl. lactis. Altering this fold cutoff had little effect on the results. These tests were carried out using the ASE_local.py script.

Protein expression and purification

His₆-Rap1 (*S. cerevisiae*) used in main text gel shift experiments was expressed using a previously generated pET28a-Rap1 expression vector in Rosetta II DE3 *E. coli* and purified as described (Johnson and Weil, 2017). In brief, after inducing expression in *E. coli* cells grown to an OD₆₀₀ of 0.5-1 for 4 h with 1 mM IPTG at 37°C, cell pellets from 500mL of culture were resuspended in 20 ml of Rap1 Lysis/Wash buffer (25 mM HEPES-NaOH (pH 7.6), 10% v/v glycerol, 300 mM NaCl, 0.01% v/v Nonidet P-40, 1 mM Benzamidine, 0.2 mM PMSF) and lysed via treatment with 1 mg/mL lysozyme and sonication. Lysate cleared via centrifugation was incubated with 2.5 mL Ni-NTA agarose (Qiagen) equilibrated with Rap1 Lysis/Wash buffer for 3 hours at 4°C to allow for His₆-Rap1 protein binding. Following 3 X washes with Rap1 Lysis/Wash buffer, Ni-NTA agarose-bound proteins were transferred to a disposable column and eluted using Rap1 Lysis/Wash buffer containing 200 mM Imidazole.

To generate *S. cerevisiae* Mcm1 for the gel shift and Far Western protein-protein binding assays presented in the main text, *S. cerevisiae* *MCM1* was cloned into a vector that would allow its expression and purification with an N-terminal MBP tag and PreScission protease cleavage site. Specifically, *S. cerevisiae* *MCM1* was amplified from *S. cerevisiae* genomic DNA and cloned into a p425 GAL1 expression vector (Mumberg et al., 1994) containing an MBP-3C tag (Feigerle and Weil, 2016) using the SpeI and XhoI restriction enzymes. This p425 GAL1 MBP-3C-Mcm1 vector was expressed in

yeast grown to an OD₆₀₀ of ~3 in 1% w/v raffinose via induction with 2% w/v galactose for 3 hours at 30°C. Yeast cell pellets obtained from 1L of culture were resuspended in 4mL Mcm1 Lysis Buffer (20 mM HEPES-KOH (pH = 7.6), 500 mM potassium acetate, 10% v/v glycerol, 0.5% v/v Nonidet P-40, 1mM DTT + 1X protease inhibitors (0.1mM PMSF, 1mM Benzamidine HCl, 2.5µg/mL aprotinin, 2.5µg/mL leupeptin, 1µg/mL pepstatin A)). Cells were lysed via glass bead lysis. Soluble cell extract was obtained via centrifugation and mixed with 2 mL DE-52 resin pre-equilibrated with Mcm1 Lysis Buffer for 5 min at 4°C. Flowthrough from the DE-52 purification was collected and diluted with 20 mM HEPES-KOH, 10% v/v glycerol, and protease inhibitors to reduce the concentration of potassium acetate and Nonidet P-40 to 200 mM and 0.2% v/v, respectively. Binding to 600 µl amylose resin was performed in batch for 2 hr at 4°C. Amylose resin-bound proteins were transferred to a disposable column, washed with 10 column volumes of Mcm1 Wash Buffer (20 mM HEPES-KOH, 200mM potassium acetate, 10% v/v glycerol, and protease inhibitors) and eluted with 10 column buffers of Mcm1 Wash Buffer containing 10mM maltose. For gel shift reactions, the N-terminal MBP tag was removed by incubating 12pmol MBP-Mcm1 in multiple 50µl reactions each containing 48pmol lab-generated 3C protease for 30 min at 4°C.

Taf1-TAP TFIID used for Far Western protein-protein binding analyses was purified via a modified tandem affinity protocol as previously described (Feigerle and Weil, 2016). GST-, His₆-Taf3, His₆-Taf4 and GST-Taf4 were expressed in *E. coli* and purified via chromatographic methods that varied upon each protein (Layer et al., 2010).

To generate material for the gel shift experiments in the supplement, the full-length Rap1-His₆ protein from *Kl. lactis* was purified from *E. coli*. Rap1 was amplified

from genomic DNA and cloned into the pLIC-H3 expression vector using XmaI and XhoI to make pTS207. This protein was purified using a His₆ tag as previously described (Lohse et al., 2013). The full-length Mcm1-HA protein from *Kl. lactis* and *S. cerevisiae* was purified from *S. cerevisiae*. The genes were amplified from genomic DNA and cloned into p426 Gal1P-MCS (ATCC 87331) using BamHI and HindIII to make pTS226 (*S. cerevisiae* Mcm1) and pTS227 (*Kl. lactis* Mcm1). These plasmids were transformed into *S. cerevisiae* W303 for expression. Cells were grown 12h in SD –Leu at 30°C, then in YPGL (1% yeast extract, 2% peptone, 3% glycerol, 2% lactate) for 10-12h. Then cells were diluted into 1L YPGL to OD₆₀₀=0.3 and grown for 4-5h until OD₆₀₀=0.6. Cells were induced with 2% galactose (added from a 40% galactose stock) for 1h. Then cells were pelleted, resuspended in an equal volume of lysis buffer, and pipetted into liquid nitrogen to make pellets. Cells were lysed in a Cryomill (Retsch) 6 times for 3 min at 30Hz, refreezing in liquid nitrogen between cycles.

Cell powder was thawed on ice and diluted to 4mL/g frozen pellet with lysis buffer (100mM Tris pH 8.0, 1mM EDTA, 10mM fresh β-mercaptoethanol, 10% glycerol, 200mM NaCl). Lysate was resuspended by pipetting, then cleared by centrifugation for 2h at 200,000 x g. Protein was bound to 250μl of HA-7-agarose (Sigma) slurry per liter yeast culture for 1h at 4°C. Lysate was applied to a Polyprep mini disposable gravity column (Biorad) and washed 4 times with 10mL lysis buffer. The protein was eluted 4 times with 1 bed volume of 1mg/mL HA peptide in lysis buffer after incubating 30 minutes on a tilt board at room temperature. The protein was stored in aliquots at -80°C.

Extract for gel shifts was prepared as previously described (Baker et al., 2011). 25mL cultures of cells were grown to OD = 0.8-1.0, and frozen at -80°C. Cells were lysed

in 300 μ l extract gel shift buffer (100mM Tris pH 8.0, 200mM NaCl, 1mM EDTA, 10mM MgCl₂, 10mM β -mercaptoethanol, 20% glycerol, EDTA-free protease inhibitor cocktail (Roche)) with 200 μ l glass beads on a vortexer for 30 min. Lysate was cleared by centrifugation at 18,000 x g for 20 minutes and diluted for experiments.

Gel-mobility shift assays

For the main text experiments, purified *S. cerevisiae* His₆-Rap1 and Mcm1 were incubated either individually or in combination in increasing amounts as indicated in the figure legend. All binding reactions were performed using 10 fmol (7,000 cpm) of a 79-bp ³²P-labeled fragment of the *Kl. lactis RPS23* promoter containing the Rap1 and Mcm1 binding sites generated via PCR, EcoRI restriction enzyme digestion, and native PAGE purification. Binding reactions were performed in binding buffer (20mM HEPES-KOH (pH 7.6), 10% v/v glycerol, 100mM KCl, 0.1mM EDTA, 1mM DTT, 25 μ g/ μ l BSA, 25 μ g/ μ l Poly(dG-dC) (double-stranded, alternating copolymer) in a final volume of 20 μ l. For competition reactions, binding was performed in the presence of 100-fold mole excess of cold Rap1 WT (R_{wt}) or Rap1 scrambled (R_{sc}) sequences and/or the Mcm1 WT (M_{wt}) or Mcm1 scrambled (M_{sc}). Reactions were allowed to proceed for 20 min at room temperature before loading onto 0.5X TBE-buffered (44.5 mM Tris, 44.5 mM Boric acid, 1mM EDTA (pH 8.0)) 5% polyacrylamide gels and electrophoresed for 1 hr at 200V at room temperature. Gels were vacuumed dried and ³²P -DNA signals detected via K-screen imaging using a BioRad Pharos FX imager.

Gel shift experiments in the supplemental material were performed as previously described (Lohse et al., 2010). Binding reactions were carried out in 20mM Tris pH 8.0 50mM potassium acetate, 5% glycerol, 5mM MgCl₂, 1mM DTT, 0.5μg/μl BSA, 25μg/μl poly(dI-dC) (Sigma).

Far Western Assay

Purified proteins tested for direct interaction with Mcm1 in Far western protein-protein binding assays were separated on parallel 4-12% NuPAGE Bis-Tris Polyacrylamide gels (Life Technologies). 0.5pmol Taf1-TAP TFIID, 1pmol His₆-Taf3, and 1pmol His₆-Taf4 were used in the assay to test for direct Mcm1 interaction with TFIID subunits. In the assay used to identify the Taf4 Mcm1 binding domain, ~ 0.4pmol of each Taf4 form, 0.8pmol His₆-Taf3, and 0.8pmol GST were used. For each experiment, one gel was stained using Sypro Ruby Protein stain (Invitrogen) to monitor protein integrity and amount. The other gel(s) were transferred to PVDF membranes pre-equilibrated in transfer buffer. Following transfer, PVDF membranes were incubated with renaturation buffer (20mM HEPES-KOH, pH 7.6, 75mM KCl, 25mM MgCl₂, 0.25mM EDTA, 0.05v/v% Triton X-100 and 1mM DTT freshly added) for 90min at 4°C on a tiltboard. The PVDF membrane was then blocked using 5% non-fat milk in renaturation buffer for 30min at room temperature on a tiltboard. The overlays were performed overnight with 10nM MBP or 10nM MBP-Mcm1 with 1% BSA (Roche) as a nonspecific competitor in renaturation buffer. Bound MBP- or MBP-Mcm1 was detected using a standard immunoblotting protocol (primary antibody MBP (NEB Catalog#E8032), used at a dilution of 1:5,000), secondary antibody horse anti-mouse

IgG, HRP-linked (Cell Signalling Catalog#7076), used at a dilution of 1:5,000).

Detection of bound proteins was achieved via incubation with ECL (GE) and X-ray film.

Figures

Figure 1

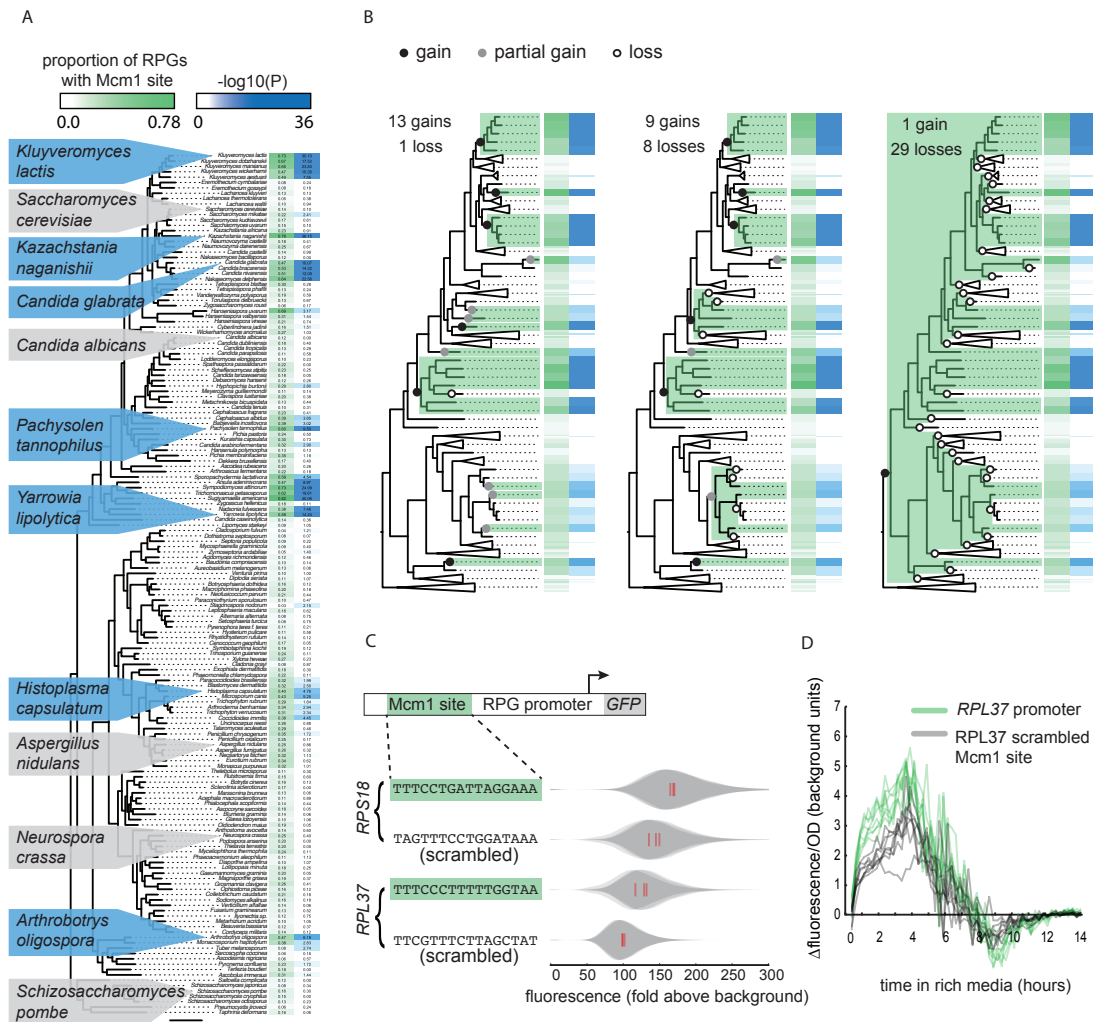


Figure 1. Repeated evolution of Mcm1 *cis*-regulatory sites at RPGs

(A) The Mcm1 sites are found upstream of the ribosomal protein genes (RPGs) in several different clades in the Ascomycete fungi. The first column, colored in green, shows the proportion of ribosomal proteins in each species that contains at least one Mcm1 site at a cutoff of ~50% the maximum log likelihood score using a position weight matrix. The second column, colored in blue, shows the $-\log_{10}(P)$ for the enrichment of these Mcm1 binding sites relative to upstream regulatory regions genome-wide, as expected under the hypergeometric distribution. Dis-enrichment values are not highlighted. The phylogeny is a maximum likelihood tree based on the protein sequences of 79 genes found in single copy in most species. Key species discussed in this paper and previous literature are highlighted with blue background indicating enrichment for Mcm1 binding sites or gray indicating no enrichment. (B) Three scenarios for how the Mcm1 sites could have arisen. Ancestors containing Mcm1 sites are shown highlighted in green. Full gains ($-\log_{10}(P) > 6$) are shown as black circles, partial gains ($6 > -\log_{10}(P) > 3$) and losses are shown as open circles. Clades lacking Mcm1 sites are condensed for simplicity. (C) Intergenic regions upstream of two ribosomal proteins in the species *Kluyveromyces lactis* were positioned upstream of a GFP reporter. The Mcm1 *cis*-regulatory sites were scrambled and the wild-type and mutant reporters were integrated into the *Kl. lactis* genome. Cells were grown for 6 hours in rich media and expression was measured by flow cytometry. Shown is the single-cell fluorescence distribution for three independent genetic isolates and the median (red bar), normalized by forward-scatter values. The values were divided by the average fluorescence for a cell lacking a GFP reporter (fold above background). (D) The *RPL37* reporter strains were diluted into rich media and fluorescence and optical cell density (OD_{600}) were measured every 15 minutes in a plate reader. Shown is the

change in fluorescence between consecutive time points divided by the OD_{600} for eight technical replicates comprised of the three independent genetic isolates of each strain.

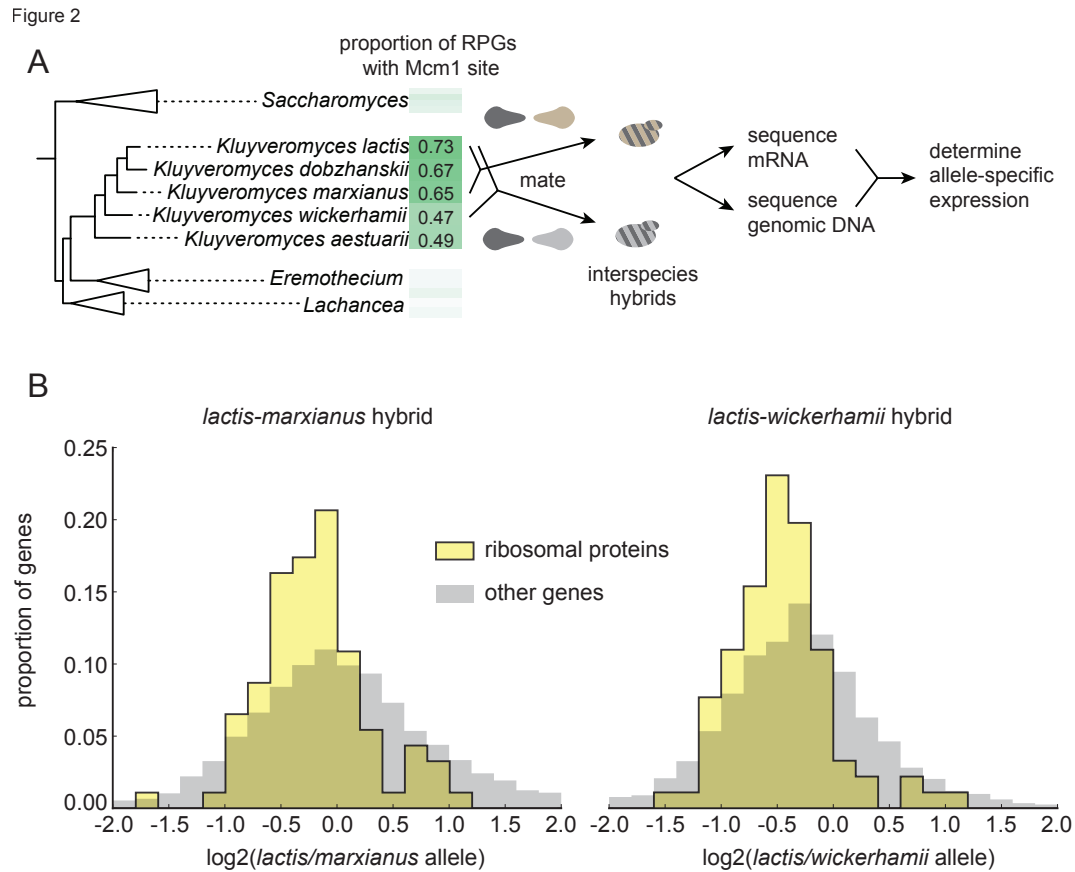


Figure 2. Selection on RPG expression level in *Kluyveromyces*

(A) Schematic of the experimental approach to measure differences in gene expression between *Kluyveromyces* yeast species. Two interspecies hybrids were constructed through mating and mRNA and genomic DNA were sequenced. The allele-specific expression is the ratio of the number of reads mapping to the coding sequence of one gene vs. its ortholog in the genome of the other species. The mRNA reads for each gene were normalized to the total reads and to the genomic DNA reads mapping to the same region to control for biases introduced in the sequencing and analysis process. (B) The

log₂-ratio of allele-specific expression with the *lactis* allele in the numerator is shown for (left) the *lactis-marxianus* hybrid and (right) the *lactis-wickerhamii* hybrid. Shown are histograms for ribosomal protein genes and the rest of the identified orthologs in the genome.

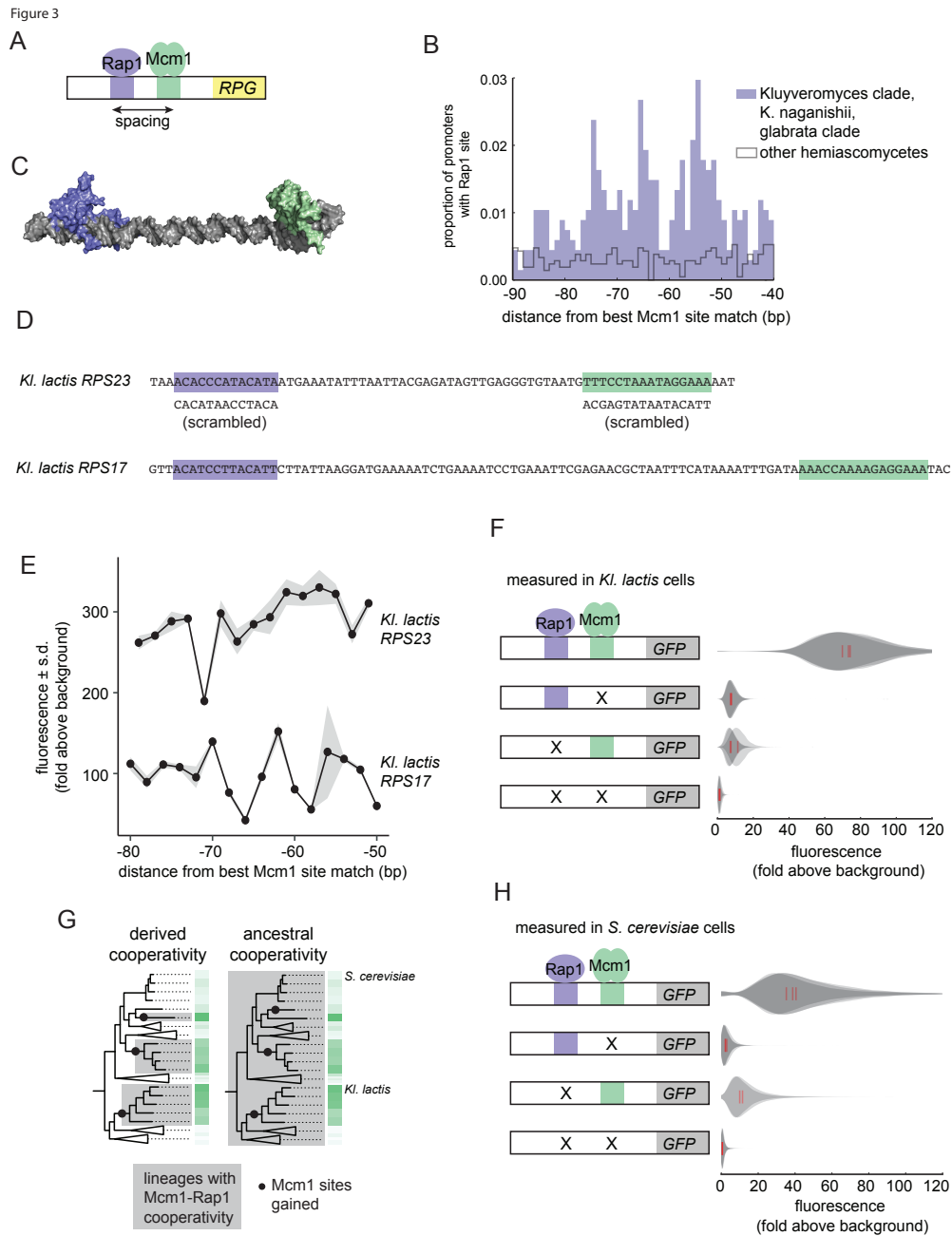


Figure 3. Ancestral cooperativity of Mcm1 with Rap1

(A) Schematic of Mcm1 and Rap1 *cis*-regulatory sites at the RPGs. (B) RPG promoters were aligned at the strongest hit to the Mcm1 position weight matrix and the relative location of Rap1 *cis*-regulatory sites was plotted. Sites for Rap1 with log-likelihood > 6.0 are shown at 1bp resolution. Hemiascomycete yeast (the 29 species at the top of the tree in Figure 1A) are divided into those with large numbers of Mcm1 sites at the ribosomal protein genes (purple shading) and those without (gray line). (C) Schematic using published structures of Mcm1 and Rap1 DNA-binding domains (PDB IDs: 1MNM and 3UKG) bound to DNA connected by a DNA linker corresponding to 55bp spacing between their *cis*-regulatory sites. Rap1 is shown in purple and Mcm1 is shown in green. (D-G) The ability of Mcm1 to work with the ribosomal protein gene regulator Rap1 was tested using a GFP reporter. (D) The Rap1 and Mcm1 *cis*-regulatory sites from *Kl. lactis* *RPS23* and *RPS17* were placed in a reporter containing the *S. cerevisiae* *CYC1* basal promoter. Reporter variants were generated by altering the spacing between these sites and by mutating the sites individually and in combination. (E, F) Reporter variants were integrated into the genome of *Kl. lactis*. Cells were grown for 4 hours in rich media and expression was measured by flow cytometry. (E) Shown is the mean fluorescence for at least three independent genetic isolates. The values were divided by the average fluorescence for a cell lacking a GFP reporter (fold above background). The measurements for *RPS23* and *RPS17* were collected on separate days and are shown on the same axes for clarity. (F) Shown is the single-cell fluorescence distribution for three independent genetic isolates and the median (red bar), normalized by side-scatter values. (G) Diagrams showing the phylogenetic distribution of ancestral or derived cooperativity.

(H) The *RPS23* reporter variants were integrated into the genome of *S. cerevisiae*, a species that lacks *Mcm1* *cis*-regulatory sequences at the ribosomal protein genes. Cells were grown and measured as described in (F). (For the third construct, one isolate had multiple reporter insertions and was not included.)

Figure 4

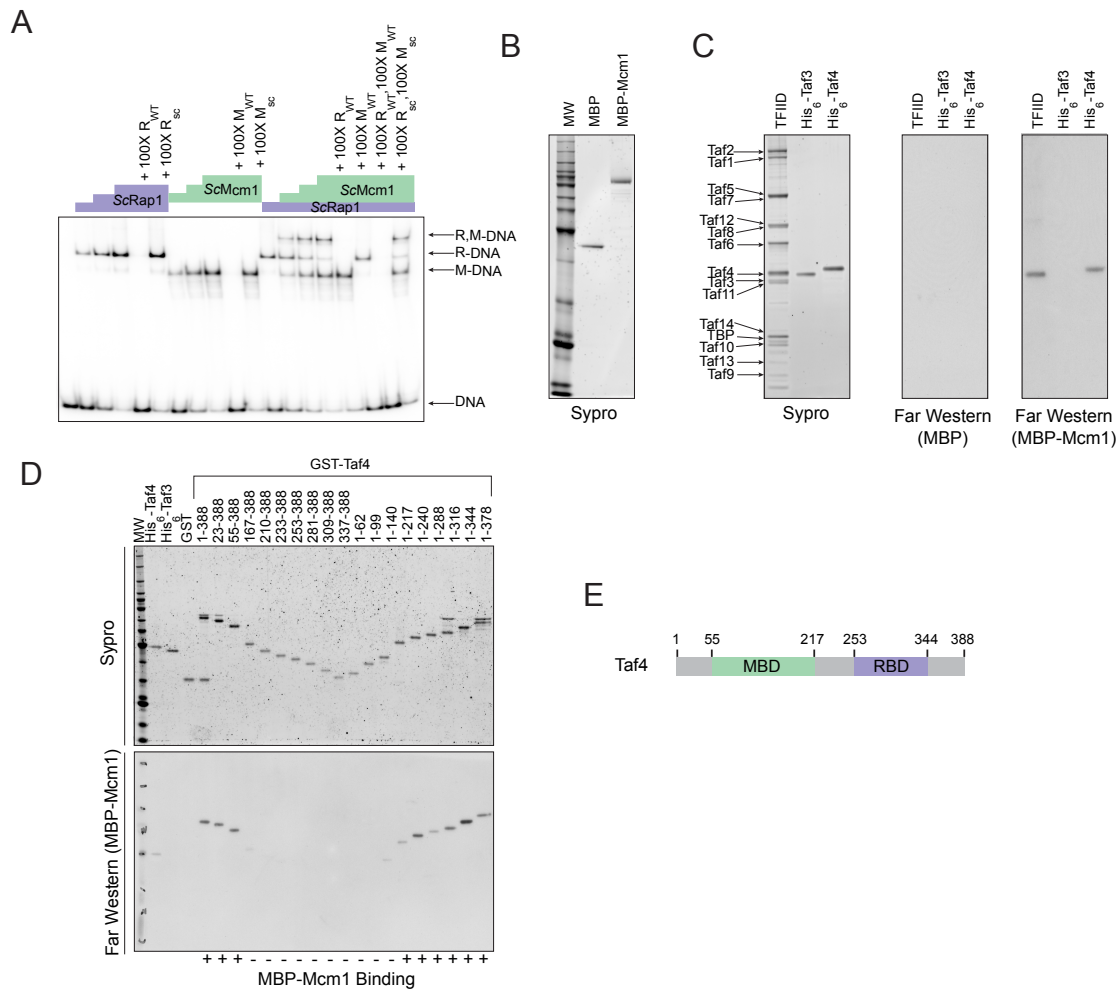


Figure 4. Mcm1 interacts with TFIID

Experiments were performed in *S. cerevisiae* or using *S. cerevisiae* proteins to test the mechanism of Rap1-Mcm1 cooperativity. (A) Gel shift DNA binding assays were performed to test possible cooperative DNA binding between purified *ScMcm1* and

ScRap1. Gel shift reactions were performed by incubating 10 fmol (~7,000 cpm) of a 79-bp ³²P-labeled fragment of the *Kl. lactis RPS23* promoter containing the Rap1 and Mcm1 binding sites (see Figure 3D) with either no protein, 2.5fmol Rap1, 5fmol Rap1, 10fmol Rap1, 10fmol Mcm1, 20fmol Mcm1, 30fmol Mcm1, or 2.5fmol Rap1 with 10, 20, or 30fmol Mcm1. Reactions also included either no cold competitor DNA or a 100-fold mole excess of cold ~20bp DNA containing either the Rap1 WT (R_{wt}) or Rap1 scrambled (R_{sc}) sequences and/or the Mcm1 WT (M_{wt}) or Mcm1 scrambled (M_{sc}) sequences in a final volume of 20μl. Reactions were fractionated on non-denaturing polyacrylamide gels, vacuum dried, and imaged using a Bio-Rad Pharos FX imager. Radiolabeled species are indicated on the left (R,M-DNA = Rap1-Mcm1-DNA, R-DNA = Rap1-DNA, M-DNA = Mcm1-DNA) (B) Sypro Ruby stain of SDS-PAGE fractionated MBP (2.4pmol) and MBP-Mcm1 (1.3pmol) probe proteins used for Far Western protein-protein binding analyses. (C) Far Western protein-protein binding analysis of Mcm1 binding to TFIID. Purified TFIID, His₆-Taf3, and His₆-Taf4 were separated on two SDS-PAGE gels. One gel was stained with Sypro Ruby for total protein visualization (left panel). The other was electrotransferred to a membrane for protein-protein binding analysis (middle and right panels). Membranes were probed with either control MBP (middle panel) or MBP-Mcm1 (right panel). Binding of probe proteins to Tafs was detecting using an anti-MBP antibody. (D) Mapping the Mcm1 Binding Domain (MBD) of Taf4. Roughly equal molar amounts of His₆-Taf4, His₆-Taf3, GST, GST-Taf4, and GST-Taf4 deletion variants were fractionated on two SDS-PAGE gels. One gel was stained with Sypro Ruby for total protein visualization. The other gel was electrotransferred to a membrane and Mcm1-Taf protein-protein binding was assayed as described in (C) using the MBP-Mcm1 as the

overlay protein. (E) Taf4 protein map indicating the location of the Taf4 MBD mapped in this study (green) as well as the Rap1 Binding Domain (RBD, purple) mapped in a previous study (Layer, et al. (2010)).

Figure 5

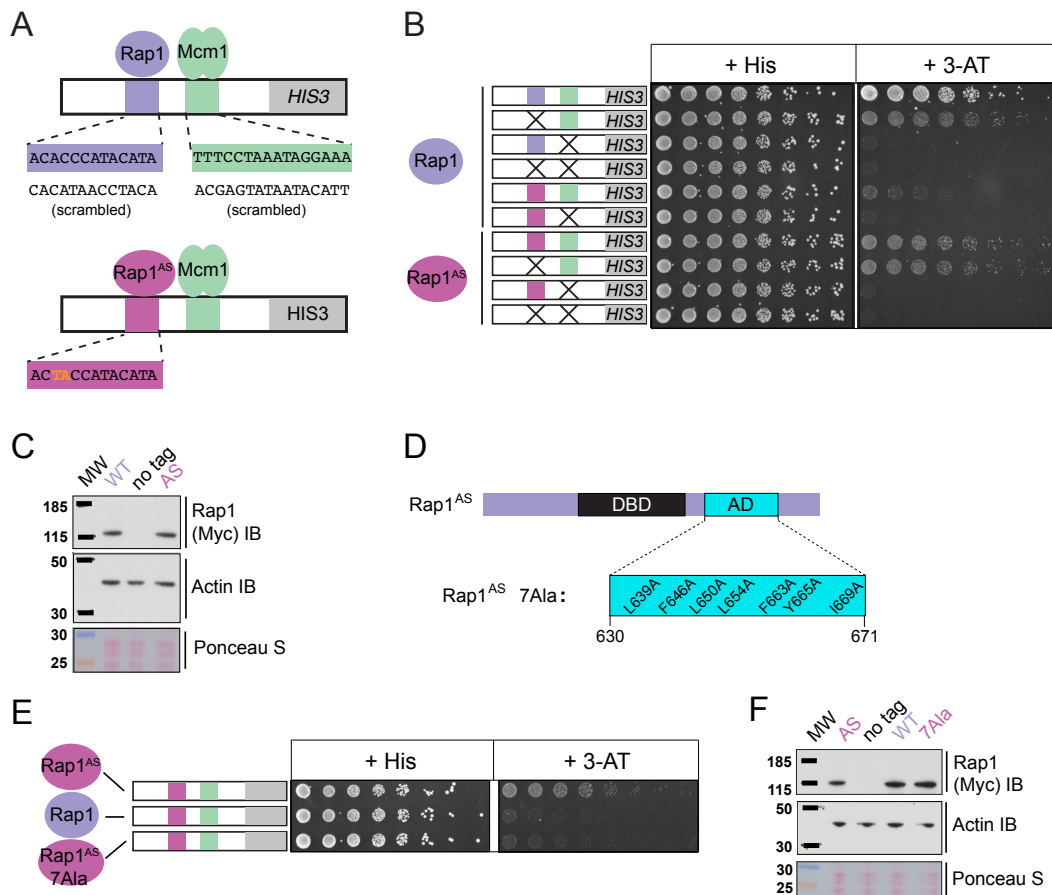


Figure 5. Mcm1-Rap1 cooperative activation requires Rap1-TFIID contacts.

(A) A series of reporter constructs were designed to test the mechanism of Rap1-Mcm1 cooperative activation. Experiments were performed in *S. cerevisiae* using *S. cerevisiae* proteins. (B) Growth analysis of yeast strains carrying the $UAS_{Rap1-Mcm1}$ reporter (containing a fragment of the *Kl. lactis* *RPS23* promoter) indicated in the diagram and either an altered

DNA-binding specificity Rap1 variant (Rap1^{AS}, magenta) or a second copy of Rap1^{WT} (purple). To perform these analyses, yeast were grown overnight to saturation, serially diluted 1:4 and spotted using a pinning tool onto either non-selective media plates (+ His) or plates containing 3-Aminotriazole (+ 3-AT), which selects for expression of the *HIS3* reporter gene. Plates were incubated for 3 days at 30° C and imaged using a Bio-Rad ChemiDoc MP imager. (C) Immunoblot analyses of the expression levels of Myc-tagged Rap1^{WT} and Myc-tagged Rap1^{AS} (Myc IB) compared to actin (Actin IB) and total protein (Ponceau S) loading controls. (D). Rap1 protein map indicating the *Sc*Rap1 AD mapped to a location C-terminal of the Rap1 DBD and the 7 key AD amino acids. These amino acids were mutated to alanine to inactivate the *Sc*Rap1 AD and create the Rap1^{AS} 7Ala mutant variant. (E) Growth analyses performed using yeast carrying the *UAS_{Rap1AS-Mem1}-HIS3* reporter and either Rap1^{AS}, a second copy of Rap1^{WT}, or Rap1^{AS7Ala} performed as described in “B.” (F) Immunoblot analysis of the Rap1 forms tested in (E) performed as described in (C).

Figure 6

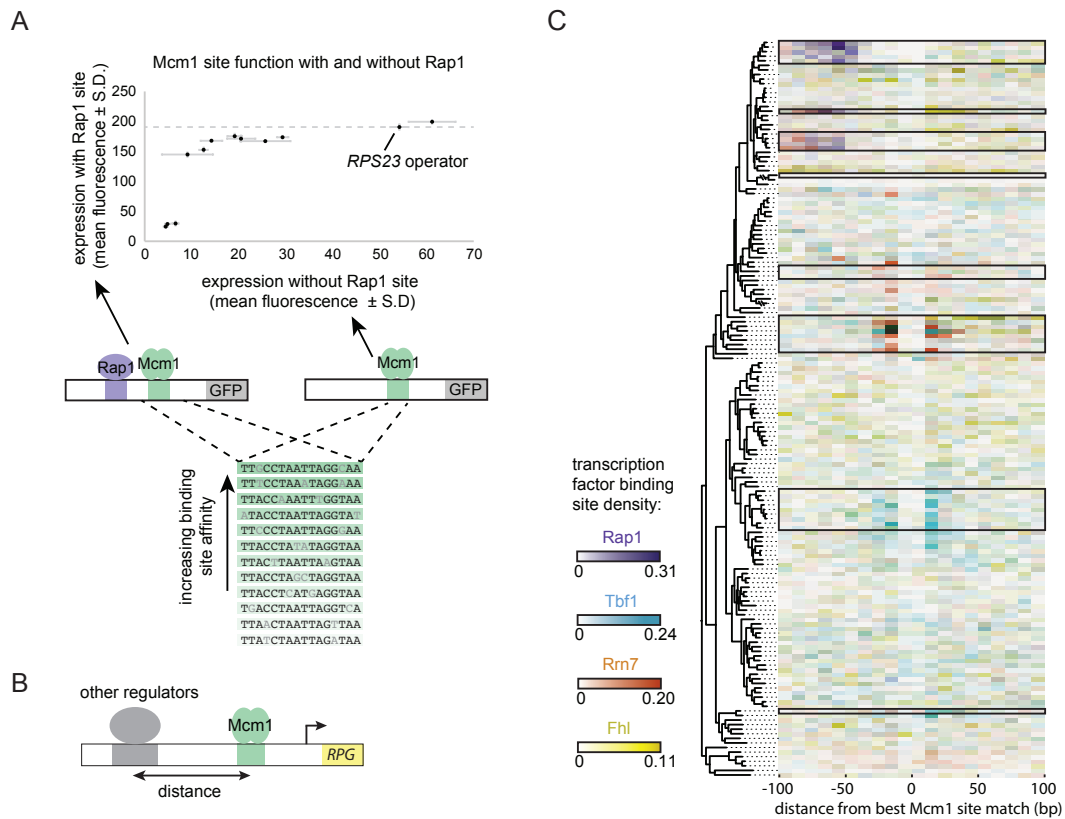


Figure 6. Evolutionary implications of intrinsic cooperative activation

(A) A series of reporters were designed to test the transcriptional activation of a weak Mcm1 site in the presence and absence of a Rap1 site. A series of Mcm1 *cis*-regulatory sites were chosen with a range of affinities that correlate with transcription rate (Acton et al., 1997). The order of the sequences shown corresponds to their expression level on the x-axis. These sites were introduced to the *S. cerevisiae* *CYC1* reporter and tested with (y-axis) and without (x-axis) an upstream Rap1 binding site. Cells were grown and measured as described in Figure 3E. The expression level of the WT *RPS23* operator is shown as a dotted line. (B) A computational analysis was designed to detect evolution of Mcm1 sites at fixed distances for other ribosomal protein regulators. Ribosomal protein

gene promoters were aligned at the strongest hit to the Mcm1 position weight matrix and the relative location of Rap1 *cis*-regulatory sites for other transcription regulators was plotted. (C) The shading in each rectangle represents the proportion of ribosomal protein gene promoters in that species that have the given *cis*-regulatory site in that 10bp interval. The clades with a large number of Mcm1 *cis*-regulatory sequences are shown with black boxes.

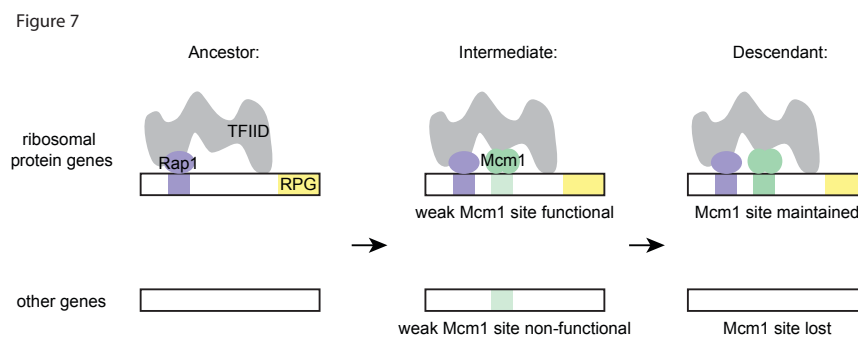


Figure 7 Model for evolution of *cis*-regulatory sites through intrinsic cooperativity

Multiple gains of new Mcm1 sites occur in the ribosomal protein genes because Rap1 and Mcm1 both bind to TFIID, a general transcription factor. Due to the intrinsic cooperativity of Rap1 and Mcm1 (which is ancestral to the gains of Mcm1 *cis*-regulatory sequences) the evolution of even a weak Mcm1 site near an existing Rap1 site would produce an effect on transcription. Because they are more likely to be functional, weak Mcm1 *cis*-regulatory sequences are preferentially retained in the population if they arise at a specified distance (as determined by the shape of TFIID) from Rap1 *cis*-regulatory sequences.

References

Acton, T.B., Zhong, H., and Vershon, A.K. (1997). 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.

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

Baker, C.R., Tuch, B.B., and Johnson, A.D. (2011). Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc Natl Acad Sci USA* *108*, 7493–7498.

Bellott, D.W., Skaletsky, H., Pyntikova, T., Mardis, E.R., Graves, T., Kremitzki, C., Brown, L.G., Rozen, S., Warren, W.C., Wilson, R.K., et al. (2010). Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* *466*, 612–616.

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. *Genome Biol* *5*, R62.

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D.S. (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* *444*, 929–932.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. *Proc Natl Acad Sci USA* *103*, 5869–5874.

Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012). Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* *489*, 513–518.

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

Borneman, A.R., Gianoulis, T.A., Zhang, Z.D., Yu, H., Rozowsky, J., Sringhaus, M.R., Wang, L.Y., Gerstein, M., and Snyder, M. (2007). Divergence of transcription factor binding sites across related yeast species. *Science* *317*, 815–819.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* *18*, 1752–1762.

Britten, R.J., and Davidson, E.H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol* *46*, 111–138.

Bullard, J.H., Mostovoy, Y., Dudoit, S., and Brem, R.B. (2010). Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proceedings of the National Academy of Sciences* *107*, 5058–5063.

Camacho, C., Coulouris, G., Avagyan, V., and Ma, N. (2009). BLAST+: architecture and applications | *BMC Bioinformatics* | Full Text. *Bmc*

Carroll, S.B. (2005). Evolution at two levels: on genes and form. *PLoS Biol* *3*, e245.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* *327*, 302–305.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* *351*, 1083–1087.

Clowers, K.J., Will, J.L., and Gasch, A.P. (2015). A unique ecological niche fosters hybridization of oak-tree and vineyard isolates of *Saccharomyces cerevisiae*. *Mol Ecol*.

Coleman, R.A., Qiao, Z., Singh, S.K., Peng, C.S., Cianfrocco, M., Zhang, Z., Song, L., Piasecka, A., Aldeborgh, H., Basishvili, G., et al. (2016). p53 dynamically directs TFIID assembly on target gene promoters. *bioRxiv* 083014.

Darwin, C. (1883). *On the Origin of Species*.

Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., et al. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* *345*, 1181–1184.

Dermitzakis, E.T., and Clark, A.G. (2002). Evolution of Transcription Factor Binding Sites in Mammalian Gene Regulatory Regions: Conservation and Turnover. *Mol Biol Evol* *19*, 1114–1121.

ENCODE Project Consortium, Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74.

Feigerle, J.T., and Weil, P.A. (2016). The C Terminus of the RNA Polymerase II Transcription Factor IID (TFIID) Subunit Taf2 Mediates Stable Association of Subunit Taf14 into the Yeast TFIID Complex. *J. Biol. Chem.* *291*, 22721–22740.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* *461*, 186–192.

- Gallant, J.R., Traeger, L.L., Volkening, J.D., Moffett, H., Chen, P.-H., Novina, C.D., Phillips, G.N., Anand, R., Wells, G.B., Pinch, M., et al. (2014). Nonhuman genetics. Genomic basis for the convergent evolution of electric organs. *Science* *344*, 1522–1525.
- Garbett, K.A., Tripathi, M.K., Cencki, B., Layer, J.H., and Weil, P.A. (2007). Yeast TFIID serves as a coactivator for Rap1p by direct protein-protein interaction. *Mol Cell Biol* *27*, 297–311.
- Gasch, A.P., Moses, A.M., Chiang, D.Y., Fraser, H.B., Berardini, M., and Eisen, M.B. (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biol* *2*, e398.
- Gavin, A.-C., Bösch, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* *415*, 141–147.
- Gordon, J.L., Armisén, D., Proux-Wéra, E., ÓhÉigeartaigh, S.S., Byrne, K.P., and Wolfe, K.H. (2011). Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proc Natl Acad Sci USA* *108*, 20024–20029.
- Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otilar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* *42*, D699–D704.
- Guarente, L., and Ptashne, M. (1981). Fusion of *Escherichia coli* lacZ to the cytochrome c gene of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* *78*, 2199–2203.
- Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yanai, I., McClellan, M., Berman, J., and Barkai, N. (2005). Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science* *309*, 938–940.
- Johnson, A.N., and Weil, P.A. (2017). Identification of a transcriptional activation domain in yeast repressor activator protein 1 (Rap1) using an altered DNA-binding specificity variant. *J. Biol. Chem.* *292*, 5705–5723.
- Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E., et al. (2010). Variation in Transcription Factor Binding Among Humans. *Science* *328*, 232–235.
- Knutson, B.A., Luo, J., Ranish, J., and Hahn, S. (2014). Architecture of the *Saccharomyces cerevisiae* RNA polymerase I Core Factor complex. *Nat Struct Mol Biol* *21*, 810–816.

- Koerber, R.T., Rhee, H.S., Jiang, C., and Pugh, B.F. (2009). Interaction of transcriptional regulators with specific nucleosomes across the *Saccharomyces* genome. *Mol Cell* 35, 889–902.
- Kohler, A., Kuo, A., Nagy, L.G., Morin, E., Barry, K.W., Buscot, F., Canbäck, B., Choi, C., Cichocki, N., Clum, A., et al. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet* 47, 410–415.
- Kooistra, R., Hooykaas, P.J.J., and Steensma, H.Y. (2004). Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* 21, 781–792.
- Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.-L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. *Mol Cell* 60, 422–434.
- Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 42, 631–634.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359.
- Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., and Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol* 8, e1000329.
- Layer, J.H., Miller, S.G., and Weil, P.A. (2010). Direct transactivator-transcription factor IID (TFIID) contacts drive yeast ribosomal protein gene transcription. *285*, 15489–15499.
- Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* 36, 900–905.
- Lee, H.N., Mostovoy, Y., Hsu, T.Y., Chang, A.H., and Brem, R.B. (2013). Divergence of iron metabolism in wild Malaysian yeast. *G3 (Bethesda)* 3, 2187–2194.
- Lickwar, C.R., Mueller, F., Hanlon, S.E., McNally, J.G., and Lieb, J.D. (2012). Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* 484, 251–255.
- Lin, Y.S., Carey, M., Ptashne, M., and Green, M.R. (1990). How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature* 345, 359–361.
- Lohse, M.B., Hernday, A.D., Fordyce, P.M., Noiman, L., Sorrells, T.R., Hanson-Smith, V., Nobile, C.J., DeRisi, J.L., and Johnson, A.D. (2013). Identification and characterization

of a previously undescribed family of sequence-specific DNA-binding domains. *Proc Natl Acad Sci USA* *110*, 7660–7665.

Lohse, M.B., Zordan, R.E., Cain, C.W., and Johnson, A.D. (2010). Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in *Candida albicans*. *107*, 14105–14110.

Losos, J.B. (2011). Convergence, adaptation, and constraint. *Evolution* *65*, 1827–1840.

Lowe, C.B., Kellis, M., Siepel, A., Raney, B.J., Clamp, M., Salama, S.R., Kingsley, D.M., Lindblad-Toh, K., and Haussler, D. (2011). Three periods of regulatory innovation during vertebrate evolution. *Science* *333*, 1019–1024.

Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* *43*, 1154–1159.

Lynch, V.J., Tanzer, A., Wang, Y., Leung, F.C., Gellersen, B., Emera, D., and Wagner, G.P. (2008). Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proc Natl Acad Sci USA* *105*, 14928–14933.

Mallick, J., and Whiteway, M. (2013). The evolutionary rewiring of the ribosomal protein transcription pathway modifies the interaction of transcription factor heteromer Interacts with Fork-head 1-Fork-head Like 1 (Ifh1-Fhl1) with the DNA-binding specificity element. *288*, 17508–17519.

Marcet-Houben, M., and Gabaldón, T. (2015). Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLoS Biol* *13*, e1002220.

Martin, H.C., Roop, J.I., Schraiber, J.G., Hsu, T.Y., and Brem, R.B. (2012). Evolution of a membrane protein regulon in *Saccharomyces*. *Mol Biol Evol*.

Marvig, R.L., Sommer, L.M., Molin, S., and Johansen, H.K. (2015). Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat Genet* *47*, 57–64.

McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2009). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proceedings of the National Academy of Sciences* *106*, 15394–15399.

McGregor, A.P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D.G., Payre, F., and Stern, D.L. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* *448*, 587–590.

Mencía, M., Moqtaderi, Z., Geisberg, J.V., Kuras, L., and Struhl, K. (2002). Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Mol Cell* 9, 823–833.

Mumberg, D., Müller, R., and Funk, M. (1994). Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res* 22, 5767–5768.

Nagy, L.G., Ohm, R.A., Kovács, G.M., Floudas, D., Riley, R., Gácsér, A., Sipiczki, M., Davis, J.M., Doty, S.L., de Hoog, G.S., et al. (2014). Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat Commun* 5, 4471.

Nogales, E., Fang, J., and Louder, R.K. (2016). Structural dynamics and DNA interaction of human TFIID. *Transcription* 8, 55–60.

Ortlund, E.A., Bridgham, J.T., Redinbo, M.R., and Thornton, J.W. (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317, 1544–1548.

Oughtred, R., Chatr-aryamontri, A., Breitkreutz, B.-J., Chang, C.S., Rust, J.M., Theesfeld, C.L., Heinicke, S., Breitkreutz, A., Chen, D., Hirschman, J., et al. (2016). Use of the BioGRID Database for Analysis of Yeast Protein and Genetic Interactions. *Cold Spring Harb Protoc* 2016, pdb.prot088880.

Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., and Schultz, P. (2010). TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. *Nature* 465, 956–960.

Papai, G., Weil, P.A., and Schultz, P. (2011). New insights into the function of transcription factor TFIID from recent structural studies. *Curr Opin Genet Dev* 21, 219–224.

Paris, M., Kaplan, T., Li, X.-Y., Villalta, J.E., Lott, S.E., and Eisen, M.B. (2013). Extensive divergence of transcription factor binding in *Drosophila* embryos with highly conserved gene expression. *PLoS Genet* 9, e1003748.

Perez, J.C., Fordyce, P.M., Lohse, M.B., Hanson-Smith, V., DeRisi, J.L., and Johnson, A.D. (2014). How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes. *Genes Dev* 28, 1272–1277.

Pfenning, A.R., Hara, E., Whitney, O., Rivas, M.V., Wang, R., Roulhac, P.L., Howard, J.T., Wirthlin, M., Lovell, P.V., Ganapathy, G., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346, 1256846.

- Piasecki, B.P., Burghoorn, J., and Swoboda, P. (2010). Regulatory Factor X (RFX)-mediated transcriptional rewiring of ciliary genes in animals. *Proceedings of the National Academy of Sciences* *107*, 12969–12974.
- Platt, J.M., Ryvkin, P., Wanat, J.J., Donahue, G., Ricketts, M.D., Barrett, S.P., Waters, H.J., Song, S., Chavez, A., Abdallah, K.O., et al. (2013). Rap1 relocalization contributes to the chromatin-mediated gene expression profile and pace of cell senescence. *Genes Dev* *27*, 1406–1420.
- Rebeiz, M., Pool, J.E., Kassner, V.A., Aquadro, C.F., and Carroll, S.B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* *326*, 1663–1667.
- Reed, R.D., Papa, R., Martin, A., Hines, H.M., Counterman, B.A., Pardo-Diaz, C., Jiggins, C.D., Chamberlain, N.L., Kronforst, M.R., Chen, R., et al. (2011). optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* *333*, 1137–1141.
- Reja, R., Vinayachandran, V., Ghosh, S., and Pugh, B.F. (2015). Molecular mechanisms of ribosomal protein gene coregulation. *Genes Dev* *29*, 1942–1954.
- Rossetti, L., Cacchione, S., De Menna, A., Chapman, L., Rhodes, D., and Savino, M. (2001). Specific interactions of the telomeric protein rap1p with nucleosomal binding sites. *J Mol Biol* *306*, 903–913.
- Sanders, S.L., Garbett, K.A., and Weil, P.A. (2002). Molecular characterization of *Saccharomyces cerevisiae* TFIID. *Mol Cell Biol* *22*, 6000–6013.
- Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* *148*, 335–348.
- Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* *328*, 1036–1040.
- Sherwood, R.K., Scaduto, C.M., Torres, S.E., and Bennett, R.J. (2014). Convergent evolution of a fused sexual cycle promotes the haploid lifestyle. *Nature* *506*, 387–390.
- Soria-Carrasco, V., Gompert, Z., Comeault, A.A., and Farkas, T.E. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*.
- Sorrells, T.R., Booth, L.N., Tuch, B.B., and Johnson, A.D. (2015). Intersecting transcription networks constrain gene regulatory evolution. *Nature* *523*, 361–365.

- Spivak, A.T., and Stormo, G.D. (2012). ScerTF: a comprehensive database of benchmarked position weight matrices for *Saccharomyces* species. *Nucleic Acids Res* *40*, D162–D168.
- Starr, T.N., Picton, L.K., and Thornton, J.W. (2017). Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* *5*, e16965.
- Stern, D.L. (2013). The genetic causes of convergent evolution. *Nat Rev Genet* *14*, 751–764.
- Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution* *62*, 2155–2177.
- Stone, J.R., and Wray, G.A. (2001). Rapid evolution of cis-regulatory sequences via local point mutations. *Mol Biol Evol* *18*, 1764–1770.
- Sudarsanam, P., Iyer, V.R., Brown, P.O., and Winston, F. (2000). Whole-genome expression analysis of *snf/swi* mutants of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* *97*, 3364–3369.
- Tanay, A. (2004). A Global View of the Selection Forces in the Evolution of Yeast Cis-Regulation. *Genome Res* *14*, 829–834.
- Tanay, A., Regev, A., and Shamir, R. (2005). Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc Natl Acad Sci USA* *102*, 7203–7208.
- Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. *Nature* *443*, 415–420.
- Tuch, B.B., Galgoczy, D.J., Hernday, A.D., Li, H., and Johnson, A.D. (2008a). The evolution of combinatorial gene regulation in fungi. *PLoS Biol* *6*, e38.
- Tuch, B.B., Li, H., and Johnson, A.D. (2008b). Evolution of eukaryotic transcription circuits. *Science* *319*, 1797–1799.
- Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* *24*, 437–440.
- Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature* *430*, 85–88.
- Woolford, J.L., and Baserga, S.J. (2013). Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* *195*, 643–681.
- Wu, S.Y., and Chiang, C.M. (2001). TATA-binding protein-associated factors enhance the recruitment of RNA polymerase II by transcriptional activators. *276*, 34235–34243.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 25, 2227–2241.

Zeevi, D., Lubliner, S., Lotan-Pompan, M., Hodis, E., Vesterman, R., Weinberger, A., and Segal, E. (2014). Molecular dissection of the genetic mechanisms that underlie expression conservation in orthologous yeast ribosomal promoters. *Genome Res* 24, 1991–1999.

Zeevi, D., Sharon, E., Lotan-Pompan, M., Lubling, Y., Shipony, Z., Raveh-Sadka, T., Keren, L., Levo, M., Weinberger, A., and Segal, E. (2011). Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. *Genome Res* 21, 2114–2128.

Zhong, P., and Melcher, K. (2010). Identification and characterization of the activation domain of Ifh1, an activator of model TATA-less genes. *Biochem Biophys Res Commun* 392, 77–82.