How Transcription Networks Evolve and Produce Biological Novelty

ISABEL NOCEDAL AND ALEXANDER D. JOHNSON

Departments of Microbiology and Immunology and of Biochemistry and Biophysics,
University of California, San Francisco, California 94158

Correspondence: ajohnson@cgl.ucsf.edu

The rewiring of gene regulatory networks over evolutionary timescales produces changes in the patterns of gene expression and is a major source of diversity among species. Yet the molecular mechanisms underlying evolutionary rewiring are only beginning to be understood. Here, we discuss recent analyses in ascomycete yeasts that have revealed several general principles of network rewiring. Specifically, we discuss how transcription networks can maintain a functional output despite changes in mechanism, how specific types of constraints alter available evolutionary trajectories, and how regulatory rewiring can ultimately lead to phenotypic novelty. We also argue that the structure and “logic” of extant gene regulatory networks can largely be accounted for by constraints that shape their evolutionary trajectories.

It has long been appreciated that evolutionary changes in gene expression patterns underlie much of the diversity of life. A change even in the regulation of a single gene can have important consequences for modern species (Gompel et al. 2005; Prud’homme et al. 2007; Wray 2007; Mclean et al. 2011; Shim et al. 2012). For example, the ability of human populations to digest lactose as adults is caused by increased expression of the enzyme that breaks down lactose (Lewinsky et al. 2005; Tishkoff et al. 2007). However, we know that most biological processes require the coordinated expression of many genes rather than a single gene. Often hundreds of genes must be expressed in a specific spatial or temporal pattern to produce a useful phenotype. These concerted changes are often achieved in eukaryotes through the actions of transcription regulators—proteins that bind to specific cis-regulatory DNA sequences to activate or repress the transcription of nearby genes. By binding to a specific cis-regulatory sequence at many locations in the genome, a single transcription regulator can control the expression of large numbers of genes. Some of these target genes may themselves be transcription regulators, leading to further downstream gene regulation. In this way, large numbers of genes can be linked together in regulatory networks, where they are coregulated by one or more transcription regulators. This organization makes it possible, through only a few point mutations, to up-regulate or down-regulate whole sets of genes over evolutionary time by accumulating mutations that alter the expression pattern or protein sequence of a key transcriptional regulator. For example, in stickleback fish, small changes in the expression pattern of a single transcription regulator are linked to reductions in bone structure (Shapiro et al. 2004; Chan et al. 2010).

For truly novel transcription networks to form, however, many ancestral connections between transcription regulators and target genes must be broken and new ones formed over evolutionary timescales. In this article, we will review work from our laboratory and others that attempts to understand how these large-scale changes happen and how they can lead to new phenotypes. This approach requires a detailed understanding of the connections between transcription regulators and target genes in a network, and how these connections break and form over evolutionary time. For several reasons, the ascomycete yeasts have proven a powerful model system for this work. First, there are a large number of ascomycete yeast species with sequenced genomes. The most highly diverged ascomycete species discussed in this article last shared a common ancestor approximately 300 million years ago (Taylor and Berbee 2006); based on the rates of change of deeply conserved proteins, this evolutionary distance corresponds roughly to the sequence divergence between humans and fish (Dujon 2006). Thus, these yeast species represent a sizeable amount of evolutionary history. Second, many of these yeast species are genetically tractable, meaning that regulatory networks can be experimentally mapped. For example, genomic locations bound by a given transcription regulator can be determined by full genome chromatin immunoprecipitation (ChIP-seq), and the effects of knocking out a transcription regulator can be assayed by sequencing mRNA (RNA-seq). Third, because yeasts are single-celled organisms, the regulatory networks, although complex by any reasonable metric, are often simpler than those in multicellular organisms.

Analysis of yeast regulatory networks has clearly shown that transcription network rewiring—the breaking of old connections and formation of new ones—occurs at surprisingly high rates over evolutionary timescales (Tsong et al. 2006; Borneman et al. 2007; Martchenko et al. 2007; Tuch et al. 2008a; Lavoie et al. 2009). Because transcription regulators are deeply conserved (often retaining the same DNA-binding specificity for hundreds

Copyright © 2015 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/sqb.2015.80.027557
of millions of years) and most “target” genes (i.e., genes whose expression is controlled by transcription regulators) are also conserved, it was reasonable to expect that the connections between regulators and target genes would also be conserved. Although this is sometimes the case, there are so many counterexamples to this expectation that one is tempted to view rewiring as the norm with deeply preserved connections as the exception. For example, the transcription regulator Mcm1 regulates many hundreds of genes in *Saccharomyces cerevisiae* (Messenguy and Dubois 2003). It is sufficiently conserved that a human homolog can substitute for it in vivo (Primig et al. 1991); the human and yeast proteins have retained a nearly identical DNA-binding specificity (Hayes et al. 1988). However, only 15% of the connections between this regulator and its deeply conserved target genes are observed in two other ascomycete species, *Candida albicans* and *Kluyveromyces lactis* (Tuch et al. 2008a). Many of these changes are due to the losses and gains of cis-regulatory sequences (e.g., through mutational inactivation of ancestral sequences and de novo formation of derived sequences), but some are also due to the breakdown and formation of protein–protein interactions between Mcm1 and its partner transcription regulators (Tuch et al. 2008a; Askew et al. 2010). In another study of two regulators of pseudohyphal growth, <25% of the genes bound in *S. cerevisiae* were also bound in the closely related yeasts *S. mikatae* and *S. bayanus*, although these genes are present in all three species (Borneman et al. 2007). A third case—that of the ribosomal protein genes—shows that high levels of transcription rewiring occur even across a very deeply conserved target gene set (Tanay et al. 2005; Hogues et al. 2008; Tuch et al. 2008a; Lavoie et al. 2010). For example, the two key transcription regulators of ribosomal protein gene expression in *S. cerevisiae* (Rap1 and Hmo1) have no role regulating these genes in *C. albicans*; instead two other regulators (Tbf1 and Cbf1), which are also present in *S. cerevisiae* (but have no role in ribosomal regulation in that species), serve this role in *C. albicans* (Lavoie et al. 2010). These studies show that, even for processes fundamental to the cell that require many highly conserved proteins to be coexpressed in the appropriate stoichiometry, transcription rewiring nevertheless occurs at high rates.

These observations raise an apparent paradox: If sets of genes are deeply conserved across species and if the regulatory proteins that control their expression are also deeply conserved, why are the ancestral regulatory connections not preserved? In other words, why is network rewiring not selected against? In the next section, we will review several examples of regulatory rewiring that we believe provide insight into this fundamental question.

**NETWORKS CAN REWIRE IN WAYS THAT PRESERVE OUTPUT**

Several features of transcription networks have been identified that—in principle—contribute to their ability to change rapidly over evolutionary time (for reviews, see Carroll 2000; Wray et al. 2003; Tuch et al. 2008b). Transcription regulators bind to short, degenerate DNA sequences—usually fewer than 10 nucleotide pairs—that can be positioned at variable locations within the control region of a gene. Functional cis-regulatory sequences are therefore relatively easy to acquire (and lose) through random mutation, resulting in changes in connections between regulators and target genes. For example, in any stretch of 1000 nucleotide pairs, there are probably several “near-miss” sequences for any given transcription regulator; by a single point mutation, they can be converted to a functional cis-regulatory sequence. In addition, transcription regulators often bind to gene control regions cooperatively with other sequence-specific transcription regulators. The protein–protein interactions between regulators that can produce the cooperativity are often weak, meaning that a few amino acid changes may be sufficient to gain or lose an interaction with another regulator (Mead et al. 1996; Arnone and Davidson 1997; Lynch and Wagner 2008; Baker et al. 2013; Sorrells and Johnson 2015). Finally, transcription regulators have been shown to be modular, with DNA-binding specificity typically uncoupled from the ability to interact with other regulators or with the general transcription machinery. Thus, it is possible for a small number of amino acid changes to alter one of these three functions without affecting the other two. All of these features allow new patterns of gene expression to arise from a small number of mutations. However, these considerations by themselves cannot explain the high rates of transcription rewiring; if interactions between a transcription regulator and a DNA sequence or another protein provide a fitness advantage, they should be preserved by purifying selection. Thus, despite the fact that transcription circuits are inherently malleable by mutation, one might have predicted that purifying selection would prevent these changes from accumulating.

The evolution of mating-type regulation in the ascomycete yeasts has provided a clear understanding of how circuit rewiring can occur despite constraints imposed by purifying selection. In the ascomycete yeast species, cells exist in one of three cell types: a, α, and a/α cells (Hesklovitz 1989; Hull et al. 2000). The mating competent forms, a cells and α cells, can mate with each other to form a/α cells. In order to ensure that only a cells and α cells mate, each of the three cell types express a distinct set of cell type-specific genes. In a cells, these genes are known as the a-specific genes, and they encode mating pheromone, pheromone receptor, agglutinins, and a pheromone exporter, each required for a cells to mate efficiently with α cells. These genes are controlled by transcription regulators encoded by the mating-type locus, which ensure that they are expressed in a cells but not in α cells or a/α cells.

The mechanism through which the a-specific genes are regulated differs between the modern species *S. cerevisiae* and *C. albicans* (Tsong et al. 2003, 2006; Baker et al. 2012). In the last shared ancestor of these two species, it has been deduced (based on a variety of approaches), that cell type specification was performed by a transcriptional activator (Mata2, hereafter referred to as α2), encoded by the mating-type locus (Tsong et al. 2006). In α cells and
a/α cells, the a-specific genes are off by default, because these cell types do not make a2. This form of positive regulation is preserved in the extant species of *C. albicans* (Tsong et al. 2003). In this species, a2 binds to specific DNA sequences cooperatively with the transcription regulator, Mcm1 (which is expressed in all three cell types) and thereby activates the a-specific genes. However, on the lineage leading to *S. cerevisiae*, there was a major rewiring of this network (Tsong et al. 2006; Baker et al. 2012). First, a product of the α cell mating-type locus, (Mato2, hereafter referred to as α2), acquired the ability to repress the a-specific genes in α cells. This occurred through the gain of a cooperative binding interaction with Mcm1, coupled with the gain of α2 cis-regulatory sequences upstream of the a-specific genes. This acquisition resulted in dual-control regulation of the a-specific genes: In a cells they are activated by a2 and Mcm1, whereas in α cells and a/α cells they are repressed by α2 and Mcm1. This dual form of regulation is conserved in the extant clade represented by the species *L. kluyveri*. However, this dual system of regulation was not maintained in all descendant species. On the lineage leading to *K. lactis*, the α2 repression mode was lost, reverting to the ancestral form of a2 activation-only regulation. On the lineage leading to *S. cerevisiae*, the opposite shift occurred: The ancestral form of regulation (activation by a2) was lost and α2 repression-only regulation was retained (Fig. 1). We know that the *S. cerevisiae* lineage is now locked in the negative-only form of regulation because the gene encoding a2 was lost, making this change effectively irreversible. Thus, in modern species, at least three different regulatory schemes (positive regulation, negative regulation, and dual-control regulation) are used to ensure a-specific gene expression (Baker et al. 2012).

Perhaps the most important conclusion of this study is that the overall regulation of the a-specific genes (expression in a cells, no expression in α and a/α cells) is preserved throughout these evolutionary changes; that is, the output of the circuit is never broken despite the major changes in the molecular mechanism. It is not known whether any of these changes were adaptive; indeed, it is possible that these multiple rewiring events could have occurred simply by nonadaptive drift. The DNA sequences recognized by a2 and α2 are very similar (Smith and Johnson 1992; Tsong et al. 2006), despite the fact that these proteins are members of different transcription factor families, the former being an HMG (high mobility group)-domain protein and the latter a homeodomain protein. This means that very few mutations were required to convert an existing a2 site into an α2 site. In addition, weak interactions with Mcm1 can stabilize α2 binding in the absence of its cis-regulatory site (Baker et al. 2012), suggesting that the protein may have been able to bind and weakly repress the a-specific genes even before strong α2 cis-regulatory sites emerged. Thus, it is possible that only a few mutations were required for α2 to initially gain the ability to repress the a-specific genes, a change that would occur without compromising ancestral regulation by a2. Because selection acts only on the output of this system—the ability of a cells to mate with α cells—the network architecture may have been free to change (e.g., from positive control to dual control to neg-

**Figure 1.** Phylogenetic tree showing the evolution of a-specific gene regulation. In the ancestor of *Saccharomyces cerevisiae* and *Candida albicans*, the regulator a2 activated a-specific gene expression in a cells, whereas these genes are not expressed in α cells because these cells lack a2. This positive regulation mode is conserved in *C. albicans*. On the lineage leading to *S. cerevisiae*, dual-control regulation emerged when a2 gained the ability to repress a-specific genes in α cells. This dual-control regulation was resolved in three different ways in extant species: in *Klyuymeroncyes lactis*, the new α2 repression mode was lost, reverting to the ancestral form of positive regulation; in *Lanchnacea klyuyveri* the dual-control regulation was maintained; and in *S. cerevisiae* the positive form was discarded leaving only the negative form of regulation. In all cases, a2 and α2 regulate expression cooperatively with the regulator Mcm1 (not shown), which is expressed in both cell types. Correct regulation of the a-specific genes (on in a cells, off in α cells) is maintained through these transitions and exists in all extant species. (Adapted from Baker et al. 2012.)
ative control) simply because it retained the appropriate output. We suspect such dual-control regulatory networks serve as common evolutionary intermediates in many regulatory rewiring events. In the same way that duplicated genes can diversify, we predict that dual-control circuits often lead to different regulatory schemes in extant species. This idea can easily explain how transcription circuits can become rewired over an evolutionary timescale without a loss in fitness (Li and Johnson 2010; Weirauch and Hughes 2010).

**PATHS OF REGULATORY NETWORK REWIRING ARE CONSTRAINED**

For any regulatory network there exists, in theory, a large number of possible trajectories for network rewiring. However, many of these trajectories—for example, those that destroy an important connection—are unlikely to be realized as they would have a detrimental effect on the output of the network. This imposes a set of severe constraints on available paths of evolutionary rewiring. Understanding the constraints imposed on any particular network helps reveal the ways in which networks can be rewired, and, ultimately, may explain why networks in modern cells have particular types of configurations.

**Constraints on Intersecting Transcription Networks**

Most transcription networks do not exist in isolation, but instead overlap with other regulatory networks. For example, the activation of a single gene may be the downstream response for more than one regulatory network, each activated by different set of environmental cues. As a result, networks are often subject to constraint imposed by selection on the output of multiple networks. In other words, a given network cannot change in such a way as to disrupt the output of that network or any other intersecting networks, providing these networks are under purifying selection. Because of this, certain pathways through which a network could—in principle—change are not permissible because the change would disrupt an intersecting network. An example of this principle is seen with the network that controls the pheromone response in the ascomycete yeast species, which intersects with the previously described a-specific gene network. Here, a previously inaccessible evolutionary path of rewiring of the pheromone-response network was made accessible by the rewiring of the a-specific gene network.

In both a and α cells, a set of pheromone-responsive genes are activated when pheromone is sensed. Some of these genes are cell type specific, including the a-specific genes, and some are common to both a and α cell types. In the presence of pheromone, the transcription regulator Ste12 is activated by phosphorylation and up-regulates the pheromone-responsive genes (Herskowitz 1995; Coria et al. 2006). The a-specific genes, therefore, are a part of two different, intersecting transcription networks: the a-specific gene network with the master transcriptional regulators a2 and α2 (as described above), and the pheromone-response network with the master regulator Ste12. The way in which Ste12 regulates the a-specific genes varies among the ascomycete yeast species (Fig. 2A). In the *C. albicans* and *K. lactis* clades, Ste12 is brought to the a-specific genes indirectly, through a protein–protein interaction with a2 (Sorrells et al. 2015). This indirect recruitment occurs despite the fact that Ste12 is capable of binding directly to specific DNA sequences to activate gene expression; the general pheromone-responsive genes are activated in this manner (Dolan et al. 1989). In the *S. cerevisiae* clade, the binding of Ste12 to the a-specific genes occurs through direct DNA binding—the a-specific genes in this clade contain optimal Ste12 cis-regulatory sequences (Fig. 2A).

Experiments performed in *K. lactis* (a proxy for the ancestral form of regulation) show that the gain of direct Ste12 cis-regulatory sequences in the *K. lactis* a-specific genes compromises cell type regulation of the a-specific genes, causing their ectopic expression in α cells (Sorrells et al. 2015). However, this gain of cis-regulatory sites could occur without compromising a-specific gene regulation once the a-specific network had shifted from positive regulation (by a2) to negative regulation (by α2), as described in the preceding section. In this way, a change in the regulation of the a-specific genes allowed for a change in Ste12 regulation of the pheromone-response genes without compromising the outputs of either network (Fig. 2B).

This study highlights the importance of fully understanding available evolutionary pathways to truly understand the “logic” of extant transcription networks. In comparing the pheromone regulation of the a-specific genes between *S. cerevisiae* and *K. lactis*, one might have assumed that there must be an adaptive explanation for Ste12 binding directly to these genes in *S. cerevisiae* but indirectly in *K. lactis*. We suggest an alternative explanation: the direct Ste12 binding pathway was not a viable trajectory from the ancestral mode of a-specific gene regulation and therefore could not occur in *K. lactis* without a detrimental effect (Fig. 2B). The switch from the indirect to direct mode of Ste12 recruitment required a prior change in a-specific gene regulation, which occurred well before the Ste12 shift. This case clearly demonstrates that new trajectories of gene regulation depend on prior changes, a situation analogous to sign epistasis in protein evolution where the effect of a given amino acid change critically depends on prior changes in the protein (Mohrig et al. 1995; Bloom et al. 2006; Ortlund et al. 2007; Harms and Thornton 2014).

**Constraints on Duplicated Transcription Regulators**

Gene duplication has long been recognized as an important source for producing new transcription networks (Ohno 1970; Conant and Wolfe 2008; Innan and Kondrashov 2010). When a transcription regulator gene is duplicated, for example, its functions can be partitioned
Figure 2. Evolutionary trajectory of rewiring of Ste12 regulation of the α-specific genes. (A) In the ancestor of *S. cerevisiae* and *C. albicans*, Ste12 activated α-specific gene expression in α cells indirectly, by binding to the regulator α2. After the dual-control regulatory state of α-specific gene regulation emerged, Ste12 acquired the ability—in *S. cerevisiae*—to bind these genes directly. (B) The order of these changes was constrained. The cube shows all possible paths of rewiring of the α-specific genes, with each axis representing one of the three required changes. Dark lines with arrows represent the most likely evolutionary path, as other paths (shown in dashed lines) lead to misregulation of the α-specific genes. In all cases, α2 and α2 regulate expression cooperatively with the regulator Mcm1 (not shown), which is expressed in both cell types. (Adapted from New and Lehner 2015; Sorrells et al. 2015.)
between the two paralogs, a process known as subfunctionalization, which results in both genes being retained by selection (Lynch and Force 2000; Lynch 2007). Less well appreciated, however, are constraints that may occur in the subsequent diversification of these duplicated genes. For example, certain evolutionary changes in one transcription regulator paralog could interfere with the function of the other and therefore be removed by selection. An example of the constraints imposed by such “paralog interference” has been documented for the duplication of the transcription regulator Mcm1 (Baker et al. 2013).

As previously discussed, Mcm1 is required in ascomycetes for proper regulation of the α-specific genes. It is also required for proper regulation of the α-specific genes, where it binds cooperatively with α1, a product of the α mating-type locus (Bender and Sprague 1987). In addition to the cell type–specific genes, Mcm1 also regulates the genes required for arginine metabolism, where it binds cooperatively with the transcriptional regulator Arg81 (Messenguy and Dubois 1993). In the ancestor of S. cerevisiae and C. albicans, Mcm1 binds to both the α-specific genes and the arginine metabolism genes as a homodimer. In both cases, the Mcm1 cis-regulatory site is adjacent to the binding site of the cofactor, allowing cooperative binding with either α1 or Arg81 (Fig. 3; Tuch et al. 2008a).

Mcm1 underwent a segmental duplication event on the lineage leading to S. cerevisiae. One of the paralogs, called Mcm1 in S. cerevisiae, lost the ability to interact with Arg81. The other paralog, called Arg80 in S. cerevisiae, lost the ability to interact with α1 (Baker et al. 2013). The result is a classic case of subfunctionalization: in S. cerevisiae Mcm1 controls α-specific genes, and Arg80 controls arginine metabolism genes (Fig. 3; Messenguy and Dubois 2003). However, while this appears to be a neat partitioning of ancestral function between the two paralogs, paralog interference necessitated additional changes in the regulatory mechanism.

Paralog interference occurs because transcription regulators are modular, with different parts of the protein responsible for different physical interactions. Mutating a region in one paralog can therefore result in it acting as a “dominant negative” on the other. Reconstruction experiments indicate that this took place when Mcm1 and Arg80 underwent a subfunctionalization (Baker et al. 2013). In particular, the reconstruction experiments show that an ancestral form of Arg80 interfered with Mcm1 by binding the cis-regulatory sequences at the α-specific genes but failing to interact productively with α1. This interference was relieved (or more likely, circumvented) by a reduction in the ability of Arg80 to bind DNA, allowing Mcm1 to outcompete Arg80 for binding at the α-specific genes. The order in which these mutations occurred is not known, but the end result is that S. cerevisiae Arg80 has lost the ability to interact with α1 and has a reduced DNA-binding affinity (Baker et al. 2013). This pathway accounts for the previously puzzling fact that in S. cerevisiae Arg80 binds to the arginine genes as a heterodimer with Mcm1: The distal cis-regulatory site is bound by an Mcm1 monomer because of its increased affinity for DNA, whereas the proximal site is bound by an Arg80 monomer because of its cooperative interaction with Arg81 (Fig. 3). This evolutionary origin for the Arg80-Mcm1 heterodimer has much more explanatory power than a simple assertion that, for S. cerevisiae,

![Diagram](https://example.com/diagram.png)

**Figure 3.** Phylogenetic tree showing evolution of Mcm1 regulation of α-specific genes and arginine metabolism genes (ARG genes). In the ancestor of S. cerevisiae and C. albicans, a homodimer of Mcm1 regulated the α-specific genes by binding to DNA with α1, and a homodimer of Mcm1 regulated the ARG genes by binding to DNA with Arg81. This ancestral regulatory state is maintained in K. lactis and C. albicans. On the lineage leading to S. cerevisiae, Mcm1 underwent a segmental duplication resulting in the paralogs Mcm1 and Arg80. The resulting paralogs subfunctionalized: in S. cerevisiae, a homodimer of Mcm1 regulates α-specific genes, whereas a heterodimer of Mcm1 and Arg80 regulates ARG genes. The regulatory scheme of L. kluveri has not been determined but is presumed to be the ancestral form. (Redrawn from Baker et al. 2013.)
the heterodimer is somehow an “improvement” over the ancestral homodimer.

TRANSCRIPTION NETWORK REWIRING CAN GENERATE BIOLOGICAL NOVELTY

The examples of transcription network rewiring discussed so far are all neutral in the sense that the overall regulatory output has been preserved despite the changes in molecular mechanism: The α-specific genes are still expressed specifically in α cells, they are induced by pheromone, and the arginine metabolism genes are still regulated in response to arginine. Although the quantitative aspects of these regulons may have been altered by the changes in regulation (and some of these could be adaptive), the overall gene expression patterns have not changed. The same forces and constraints that produce these neutral changes in molecular wiring can also produce new transcription circuits that produce novel phenotypes, a topic we now discuss.

Generation of Novel Network Behavior by Intercalation of a Regulator

One way an existing network can gain a novel behavior is by integrating an existing regulator into the network. In this way, a whole set of previously linked genes can become responsive to a new condition (Gehring and Ikeo 1999). This integration of a novel regulator into an existing circuit has been demonstrated for a set of yeast cell type-specific genes, the haploid-specific genes (Booth et al. 2010). These genes are expressed in α and α cells, but not in a/α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989). In S. cerevisiae and C. albicans, these genes are on by default in α and α cells, and they code for proteins that both α and α cells use to transduce the pheromone signal and activate Ste12 (Herskowitz 1989).

As discussed above, gene duplication is widely regarded as one of the most important sources of evolutionary novelty (Ohno 1970; Conant and Wolfe 2008; Innan and Kondrashov 2010). Unlike the partitioning of an ancestral network between two paralogs (as occurred in the case of Mcm1) gene duplication of a transcription regulator can also lead to a large network expansion. A striking example occurred relatively recently in the C. albicans clade, where successive duplications of a deeply conserved transcriptional regulator, Lys14, led to the formation of new regulatory networks important for the ability of C. albicans to colonize and cause disease in mammalian hosts (Pérez et al. 2013, 2014). C. albicans is one of only a few ascomycete yeast species that can colonize

Figure 4. Phylogenetic tree showing the evolution of regulation of the core haploid-specific genes (core hsgs). In the ancestor of S. cerevisiae and C. albicans, a heterodimer of α1 and α2 directly repressed the core hsgs in a/α cells. This direct regulation is retained in S. cerevisiae and C. albicans. In K. lactis, however, α1-α2 now represses the core hsgs indirectly by repressing the regulator Rme1, which activates the core hsgs. As Rme1 expression is activated by starvation, the core hsgs now require both starvation and the absence of α1/α2 to be fully expressed. (Adapted from Booth et al. 2010.)
humans (Butler et al. 2009), and the formation of these new regulatory networks appears to be a crucial adaptation for the tight association of this species with its human host.

In *S. cerevisiae*, the transcription regulator Lys 14 binds to a specific *cis*-regulatory sequence and thereby regulates expression of the genes required for lysine biosynthesis, hence its name (Ramos et al. 1988; Feller et al. 1994). On the lineage leading to *C. albicans*, this gene underwent several successive duplications to produce four paralogs in *C. albicans* (Fig. 5). None of the four paralogs regulate the lysine biosynthesis genes in this species (Homann et al. 2009); instead, they have roles in other processes crucial for the biology of *C. albicans*. Two of the paralogs, Lys14 and Lys144 (originally named with the prefix “Lys” because of sequence similarity with the *S. cerevisiae* Lys14 gene rather than function) have been shown to have roles in host–pathogen interactions. Lys14 is required for *C. albicans* to cause a bloodstream infection, whereas Lys144 is required for colonization of the intestinal tract (Pérez et al. 2013). One of the other paralogs, Lys143, has been shown to affect the ability of cells to switch between white and opaque types (Pérez et al. 2014), two distinct heritable cell types, a behavior that arose on the lineage to *C. albicans* and does not exist in *S. cerevisiae* (Soll 2004; Lohse and Johnson 2009). The biological role of the fourth paralog is not currently known. In each case, only a single paralog is required for the phenotype in question, demonstrating the nearly complete diversification of these four transcription regulators subsequent to the gene duplications. This phenotypic diversification was confirmed by analyzing the direct gene targets of each of the four paralogs. Although some connections are shared between them, a large fraction of the target genes are bound by one paralog but not by any of the other three paralogs (Pérez et al. 2014).

This diversification in DNA-binding specificity across the four paralogs is due to a combination of three different mechanisms (Pérez et al. 2014). The first is a change in the monomer DNA-binding specificity. Each paralog binds DNA as a homodimer, and the optimal DNA sequence recognized by a given monomer is slightly different than that of the other paralogs. More importantly, the orientation and spacing of the monomer binding have di-
REFERENCES


Coria R, Kawasaki L, Torres-Quiroz AF, Ongay-Larios L, Arnone MI, Davidson EH. 1997. The hardwiring of development networks change and form new networks. These evolutionary pathways typically do not correspond to the way a scientist or engineer would change a network. Rather, severe constraints on available evolutionary pathways arise because networks cannot pass through “broken” intermediates. These constraints, in turn, force networks to evolve in specific ways. A corollary of this idea is that the structure and logic of modern transcriptional networks are dictated by their evolutionary histories. We suggest that any attempt to rationalize network structures in the absence of evolutionary considerations will often lead to mistaken conclusions regarding the advantage to the cell of one type of network architecture over another (Sorrells and Johnson 2015).

Many networks are much larger than those discussed here, sometimes encompassing hundreds or thousands of genes. Although there is no direct evidence available, we believe that the evolution of these large networks is shaped, at least in part, by the same types of constraints we have described. However, we do not yet understand how a large network, composed of many transcription regulators and hundreds or thousands of target genes, forms in the first place. Although a few general principles underlying network evolution have now been deduced, we simply do not know how many additional principles await discovery.

ACKNOWLEDGMENTS

We thank T. Sorrells, C. Britton, S. Singh-Babak, K. Fowler, N. Ziv, and L. Noiman for comments on the manuscript and N. Hartooni for advice on figures. We are also grateful to the past and present members of the laboratory and our collaborators who contributed to the published experiments cited in this review. These include C. Baker, L. Booth, P. Fordyce, D. Galgozcy, V. Hanson-Smith, H. Li, M. Lohse, J.C. Perez, T. Sorrells, A. Tsong, and B. Tuch. Work from the authors’ laboratory was supported by grant R01 GM037049 from the National Institutes of Health.


Homann OR, Dea J, Noble SM, Johnson AD. 2009. A pheno-


Lynch M, Force A. 2000. The probability of duplicate gene pres-

Lynch VJ, Wagner GP. 2008. Resurrecting the role of transcrip-


Perez JC, Kumamoto CA, Johnson AD. 2013. Candida albicans commensalism and pathogenicity are intertwined traits direct-


Primig M, Winkler H, Ammerger G. 1991. The DNA binding and oligomerization domain of MCM1 is sufficient for its inter-
action with other regulatory proteins. EMBO J 10: 4209–4218.

Prud’homme B, Gompel N, Carroll SB. 2007. Emerging prin-


Shim S, Kwan KY, Li M, Lefebvre V, Séstan N. 2012. Cis-


Sorrells TR, Booth LN, Tuch BB, Johnson AD. 2015. Intersect-
ing transcription networks constrain gene regulatory evolu-


Tseng AE, Tuch BB, Li H, Johnson A. 2006. Evolution of al-


Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rock-
How Transcription Networks Evolve and Produce Biological Novelty

Isabel Nocedal and Alexander D. Johnson

Cold Spring Harb Symp Quant Biol 2015 80: 265-274 originally published online December 9, 2015
Access the most recent version at doi:10.1101/sqb.2015.80.027557

References
This article cites 68 articles, 22 of which can be accessed free at:
http://symposium.cshlp.org/content/80/265.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to Cold Spring Harbor Symposia on Quantitative Biology go to:
http://symposium.cshlp.org/subscriptions

Copyright © 2015 Cold Spring Harbor Laboratory Press; all rights reserved