Title
Evolutionary rewiring of the gene regulatory network controlled by Ndt80

Author
Nocedal, Isabel

Publication Date
2015

Peer reviewed|Thesis/dissertation

Evolutionary rewiring of the gene regulatory network controlled by Ndt80

by

Isabel Nocedal

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements

This dissertation would not have been possible without the incredible support I received from colleagues, friends, and family. My advisor, Sandy Johnson, was instrumental in developing and shaping this project, as well as training me to become a rigorous and inquisitive scientist. Sandy taught me how to formulate scientific questions, how to view your research within the context of a much larger scientific community, and how to successfully present your research to others. I feel that most of what I know about what it means to be a scientist I learned from Sandy, and for that I will be forever grateful.

In addition to being an excellent advisor, Sandy also fosters a wonderful, collaborative community in his lab, and I was incredibly lucky to work with so many amazing people in my time in the Johnson Lab. I can't imagine having finished this project without their invaluable scientific feedback, as well as their support when I was feeling demoralized. In particular, it was a pleasure going through the whole process with Trevor Sorrells, who joined the lab at the same time as me. Trevor's enthusiasm for science is infectious, and I hope we continue to be friends and scientific collaborators for many years to come. Lauren Booth, Chris Baker, Quinn Mitrovich, and Clarissa Nobile were all really important, supportive mentors for me when I joined the lab and provided incredibly helpful scientific advice as well as general guidance and support when I was a young, clueless, intimidated graduate student. Chiraj Dalal, Sheena Singh-Babak, Eugenio Mancera, and Naomi Ziv have all been awesome mentors, as well as friends, who have been a pleasure to work with over the years. They are all going to make excellent faculty members one day. The other graduate students that have passed through the lab in my time there—Emily Fox, Liron Noiman, Candace Britton, Kyle Fowler—are all awesome, smart, interesting people, and they made getting through grad school a little less miserable. Also I

would be remiss not that thank Matt Lohse, a fixture in the Johnson lab the whole time I was there, without whom the lab would cease to function, and who gave countless technical help during the course of this project, and who's bizarre sense of humor made every day an adventure in the Johnson Lab.

I am also incredibly grateful for my fantastic classmates who made graduate school so much more fun. I didn't expect to meet such a fun group of smart, interesting people when I started graduate school, and its been a pleasure going through this process with them. In particular Alison Leaf always provided a much needed laugh when I was down, Lindsey Pack provided crucial sports-watching companionship as well as much-appreciated sarcasm, Sam Liang was always there for a fancy Starbucks drink when we were both bored in lab, Coral Zhou was always ready for random, spontaneous activities, and Kelly Nissen was always there to vent about science and entertain my ramblings about politics. It was a pleasure getting to know them, and the rest of my classmates, and I can't wait to see all the cool things they are going to accomplish in their careers.

Finally, I have to thank family and friends who provided general support and encouragement through this whole process. All my Bay Area friends were incredibly patient with my strange work hours and seemingly-interminable graduate program, and I feel so lucky that I got to go on so many fun outdoor adventures with them over the years. My parents and my brother have always been incredibly encouraging of my career, and have never been impatient with how long it seemed to be taking for me to get my PhD. And of course I have to thank my husband, Max, who put up with all kinds of nonsense over the years that were required for this dissertation. I won't say I couldn't have done it without him, but completing graduate school certainly would have been a lot less pleasant without him.

Acknowledgement of previously published work

Chapter 1 was published in the following journal:

Nocedal I., Johnson A.D. How Transcription Networks Evolve and Produce Biological Novelty. Cold Spring Harb Symp Quant Biol. 2015 Dec 9. pii: 027557

Evolutionary Rewiring of the Gene Regulatory Network Controlled by Ndt80

Isabel Nocedal

Gene regulatory networks—composed of transcription regulators and their downstream "target" genes—have been shown to control many biological processes, and changes to either the regulators, the target genes, or the connections between them can lead to alterations in the level, timing, or patterning of gene products. However, it remains poorly understood how a complex gene regulatory network, controlling a novel phenotype, arises over evolutionary time. Here we demonstrate that the fungal transcription regulator Ndt80 has undergone a dramatic evolutionary switch in regulatory function—from an ancestral role in regulating sporulation and meiosis to a derived role in regulating biofilm formation. This switch in function, which took place along the lineage leading to the *Candida* clade, corresponded to a large-scale change in the genes regulated by Ndt80, as it was incorporated into a novel gene regulatory network. However, we demonstrate, through experiments in multiple extant yeast lineages, that the connections between Ndt80 and its target genes were undergoing rapid rewiring prior to the switch in Ndt80's regulatory function. We propose that drift in the Ndt80 regulon, which occurs along all lineages examined, facilitated its dramatic switch in function along one lineage, allowing it to become integrated into a novel gene regulatory network. This idea that the promiscuity of the Ndt80 regulon could be exploited in the evolution of a novel phenotype is analogous in many ways to studies of protein evolution that have demonstrated that ancestral promiscuity in protein function can be exploited in the evolution of novel protein functions. We believe this work provides the first example of the exploitation of regulatory network promiscuity in the evolution of a novel regulatory network, and emphasizes the importance of regulatory network drift in the exploration of new network configurations and the evolution of novel phenotypes.

Table of Contents

List of Tables

Chapter 2

List of Figures

Chapter 1

Chapter 2

CHAPTER 1

INTRODUCTION

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 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 co-regulated 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 regulation 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 paper, we will review work from our laboratory and others that

attempts to understand how this happens and how it 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 paper 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). And third, because yeast 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 demonstrated that transcription network rewiring—breaking old connections and forming 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 of millions of years) and most "target" genes (that is, genes whose expression is controlled by transcription regulators) are deeply conserved, it was reasonable to assume that the connections between regulators and target

genes would also be conserved. Although this is sometimes the case, there are so many counter examples to this expectation that one is tempted to view rewiring as the norm with preserved connections as the exception. For example, the transcription regulator Mcm1 regulates many hundreds of genes in *S. 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, *C. albicans* and *K. lactis* (Tuch et al. 2008a). Many of these changes are due to the losses and gains of *cis*-regulatory sequences (for example, through mutational inactivation of ancestral sequences and *de novo* formation of derived sequences), but some are also due to the breakage and formation of protein-proteins 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, fewer than 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; Lavoie et al. 2010). 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 demonstrate that, even for processes fundamental to the cell that require many highly conserved proteins to be co-

expressed 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 aren't the ancestral regulatory connections preserved? In other words, why isn't network rewiring 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 (reviewed by Carroll 2000; Wray et al. 2003; Tuch et al. 2008b). Transcription regulators bind to short, degenerate DNA sequences—usually fewer than 10 base 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 nucleotides, there are probably several "near-miss" sequences for any given transcription regulator that can, by a single point mutation, 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 (Arnone and Davidson 1997; Lynch and Wagner 2008; 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 (Herskowitz 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 exporters, 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 carried out by a transcriptional activator (Mata2, hereafter

referred to as **a**2), 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 **a**2. This form of positive regulation is preserved in the extant species of *C. albicans* (Tsong et al. 2003). In this species, **a**2 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 (Baker et al. 2012). First, a product of the α cell mating-type locus, (Matα2, hereafter referred to as α2), acquired the ability to repress the **a**-specific genes in α cells. This occurred through a gain of cooperative binding interaction with Mcm1, coupled with the gain of α2 *cis*-regulatory sequences upstream of the **a**-specific genes. This acquisition resulted in a dual-control regulation of the **a**-specific genes: in **a** cells they are activated by **a**2 and Mcm1, while 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 **a**2 activation-only regulation. On the lineage leading to *S. cerevisiae*, the opposite shift occurred: the ancestral form of regulation (activation by **a**2) was lost and α2 repression-only regulation was retained (Figure 1). We know that the *S. cerevisiae* lineage is now locked in the negative-only form of regulation because the gene encoding **a**2 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 non-adaptive drift. The DNA sequences recognized by **a**2 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-domain protein and the latter a Homeodomain protein. This means that very few mutations were required to convert an existing **a**2 site into an α2 site. In addition, even 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 **a**2. 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 (for example from positive control to dual-control to negative control) simply because it retained the appropriate output. We suspect such dual-control regulatory networks may 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 (Weirauch and Hughes 2010; Li and Johnson 2010).

Paths of regulatory network rewiring are constrained

For any regulatory network there exists, in theory, a vast 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 of the output of the network. This imposes a set of 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 discussed **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, when pheromone is sensed, a set of pheromone-responsive genes are activated. 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 **a**2 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 (Figure 2A). In the *C. albicans* and *K. lactis* clades, Ste12 is brought to the **a**-specific genes indirectly, through a protein-protein interaction with **a**2 (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 (Figure 2A).

Experiments performed in *K. lactis* (a proxy for the ancestral form of regulation) show that the gain of 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 **a**2) 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 pheremone response genes without compromising the outputs of either network (Figure 2B).

This study highlights the importance of fully understanding available evolutionary pathways in order 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 (Figure 2B). The switch from the indirect to direct mode of Ste12 recruitment required the shift in **a**-specific gene regulation, and occurred well after the shift happened. This case clearly demonstrates that new trajectories of gene regulation depend on prior changes, a situation analogous to sign epistasis in protein evolution (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 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 **a**-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 Jr. 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 (Figure 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 (Figure 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 due to its increased affinity for DNA, while the proximal site is bound by an Arg80 monomer due to its cooperative interaction with Arg81 (Figure 3). This evolutionary origin for the Arg80-Mcm1 heterodimer has much more explanatory power than a simple assertion that, for *S. cerevisiae*, 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 **a**-specific genes are still expressed specifically in **a** 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 **a** and α cells, but not in **a**/α cells, and they code for proteins that both **a** 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 **a** and α cells, and repressed in **a**/α cells by a heterodimer of the regulators **a**1 and α2 (Figure 4) (Miller and Johnson 2002; Galgoczy et al. 2004). In the species *K. lactis*, however, a new tier of regulation has entered this network (Booth et al. 2010). Here, the **a**1/α2 heterodimer represses another regulator, Rme1, and Rme1 in turn activates the haploid-specific genes in haploids. The overall effect of **a**1 and α2 on haploid-specific gene expression is therefore preserved: the heterodimer represses the haploid-specific genes by turning off their activator. However the molecular complexity of the circuit has increased, as an additional transcription regulator has been intercalated into the ancestral circuit (Figure 4). Rme1 is present in all ascomycetes (where it is often activated by starvation to induce meiosis), but only in the *K. lactis* clade has it entered the haploid-specific gene network.

This change in regulation also results in a change in the logic of the circuit. Because Rme1 expression itself is regulated by starvation (Covitz and Mitchell 1993), expression of the haploid-specific genes in *K. lactis* now becomes dependent on starvation. It has been known for many years that *K. lactis*, unlike *S. cerevisiae* and *C. albicans*, requires starvation to mate (Herman 1970), and this novel behavior can be attributed to the intercalation of Rme1 into the haploid-specific network. Although the overall cell-type output of the network has been

conserved (the haploid-specific genes are off in **a**/α cells), the intercalation of Rme1 into the ancestral network has generated a novel cell behavior: the stimulation of *K. lactis* mating in response to nutrient deprivation (Booth et al. 2010).

Generation of novel networks through gene duplication

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 duplication 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; Perez et al. 2014). *C. albicans* is one of only a few ascomycete yeast species that can colonize humans (Butler et al. 2009), and the formation of these new regulatory networks is believed 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* (Figure 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 important in the biology of *C. albicans*. Two of the paralogs, Lys14 and Lys144 (originally named with the prefix "Lys" due to homology

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, while 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 (Perez 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* and related species (Soll 2004; Lohse and Johnson 2009). The biological role of the fourth paralog is not 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 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 (Perez et al. 2014).

This diversification in DNA binding specificity across the four paralogs is due to a combination of three different mechanisms (Perez 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 has diverged significantly, indicating key differences in the way that each paralog forms homodimers. Finally, one of the regulators, Lys144, has also acquired the ability to bind DNA cooperatively with Mcm1. Together, these changes allow the four paralogs to bind to different sets of genes, and therefore control different networks (Figure 5). While it is not yet known when, with respect to the successive gene duplications, these new networks formed, the end result is four new

transcription networks that evolved relatively recently and almost certainly strengthened the links between *C. albicans* and its mammalian hosts.

Conclusions

The study of regulatory networks across the ascomycete yeasts has provided several important insights into the molecular mechanisms through which ancestral transcription 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.

Acknowledgements

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

References

Arnone MI, Davidson EH. 1997. The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**: 1851–1864.

Askew C, Sellam A, Epp E, Mallick J, Hogues H, Mullick A, Nantel A, Whiteway M. 2010. The zinc cluster transcription factor Ahr1p directs Mcm1p regulation of Candida albicans adhesion. *Mol Microbiol* **79**: 940–953.

Baker C, Booth L, Sorrells T, Johnson A. 2012. Protein Modularity, Cooperative Binding, and Hybrid Regulatory States Underlie Transcriptional Network Diversification. *Cell* **151**: 80–95.

Baker C, Hanson-Smith V, Johnson A. 2013. Following Gene Duplication, Paralog Interference Constrains Transcriptional Circuit Evolution. *Science* **342**: 104–108.

Bender A, Sprague Jr. GF. 1987. MATα1 protein, a yeast transcription activator, binds synergistically with a second protein to a set of cell-type-specific genes. *Cell* **50**: 681–691.

Blake VM, Barolo S. 2014. Genome Evolution: How Sister Genes Grow Apart. *Curr Biol* **24**: R695–R697.

Bloom JD, Labthavikul ST, Otey CR, Arnold FH. 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci* **103**: 5869–5874.

Booth LN, Tuch BB, Johnson A. 2010. Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature* **468**: 959–963.

Borneman A, Gianoulis T, Zhang Z, Yu H, Rozowsky J, Seringhaus M, Wang L, Gerstein M, Snyder M. 2007. Divergence of Transcription Factor Binding Sites Across Related Yeast Species. *Science* **317**: 815–819.

Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy JL, et al. 2009. Evolution of pathogenicity and sexual reproduction in eight Candida genomes. *Nature* **459**: 657–662.

Carroll SB. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**: 577–580.

Chan Y, Marks M, Jones F, Villarreal G, Shapiro M, Brady S, Southwick A, Absher D, 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.

Conant G, Wolfe KH. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet* **9**: 938–950.

Coria R, Kawasaki L, Torres-Quiroz F, Ongay-Larios L, Sánchez-Paredes E, Velázquez-Zavala N, Navarro-Olmos R, Rodríguez-González M, Aguilar-Corachán R, Coello G. 2006. The pheromone response pathway of Kluyveromyces lactis. *FEMS Yeast Res* **6**: 336–344.

Covitz P a., Mitchell a. P. 1993. Repression by the yeast meiotic inhibitor RME1. *Genes Dev* **7**: 1598–1608.

Dolan JW, Kirkman C, Fields S. 1989. The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. *Proc Natl Acad Sci U S A* **86**: 5703–5707.

Dujon B. 2006. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet* **22**: 375–387.

Feller a, Dubois E, Ramos F, Piérard A. 1994. Repression of the genes for lysine biosynthesis in Saccharomyces cerevisiae is caused by limitation of Lys14-dependent transcriptional activation. *Mol Cell Biol* **14**: 6411–6418.

Galgoczy DJ, Cassidy-Stone A, Llinás M, O'Rourke SM, Herskowitz I, Derisi JL, Johnson A. 2004. Genomic dissection of the cell-type-specification circuit in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **101**: 18069–18074.

Gehring WJ, Ikeo K. 1999. Pax 6: mastering eye morphogenesis and eye evolution. *Trends Genet* **15**: 371–377.

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

Harms MJ, Thornton JW. 2014. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* **512**: 203–207.

Hayes TE, Sengupta P, Cochran BH. 1988. The human c-fos serum response factor and the yeast factors GRM/PRTF have related DNA-binding specificities. *Genes Dev* **2**: 1713–1722.

Herman AI. 1970. Interspecies sex-specific growth responses in Kluyveromyces. *Antonie Van Leeuwenhoek* **36**: 421–425.

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

Herskowitz I. 1995. MAP kinase pathways in yeast: for mating and more. *Cell* **80**: 187–197.

Hogues H, Lavoie H, Sellam A, Mangos M, Roemer T, Purisima E, Nantel A, Whiteway M. 2008. Transcription factor substitution during the evolution of fungal ribosome regulation. *Mol Cell* **29**: 552–562.

Homann OR, Dea J, Noble S, Johnson A. 2009. A phenotypic profile of the Candida albicans regulatory network. *PLoS Genet* **5**: e1000783.

Hull C, Raisner R, AD J. 2000. Evidence for Mating of the "Asexual" Yeast Candida albicans in a Mammalian Host. *Science* **289**: 307–310.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* **11**: 97–108.

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

Lavoie H, Hogues H, Whiteway M. 2009. Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. *Curr Opin Microbiol* **12**: 655–663.

Lewinsky RH. 2005. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Hum Mol Genet* **14**: 3945–3953.

Li H, Johnson A. 2010. Evolution of transcription networks - lessons from yeasts. *Curr Biol* **20**: R746–53.

Lohse M, Johnson A. 2009. White-opaque switching in Candida albicans. *Curr Opin Microbiol* **12**: 650–654.

Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA* **104 Suppl**: 8597–8604.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459–473.

Lynch VJ, Wagner GP. 2008. Resurrecting the role of transcription factor change in developmental evolution. *Evolution* **62**: 2131–2154.

Martchenko M, Levitin A, Hogues H, Nantel A, Whiteway M. 2007. Transcriptional rewiring of fungal galactose-metabolism circuitry. *Curr Biol* **17**: 1007–1013.

Mclean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**: 216–219.

Messenguy F, Dubois E. 1993. Genetic evidence for a role for MCM1 in the regulation of arginine metabolism in Saccharomyces cerevisiae. *Mol Cell Biol* **13**: 2586–2592.

Messenguy F, Dubois E. 2003. Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* **316**: 1–21.

Miller MG, Johnson AD. 2002. White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* **110**: 293–302.

Mohrig, Moerke KA, Cloutier DL, Lane BD, Person EC, Onasch TB. 1995. Importance of historical contingency in the stereochemistry of hydratase-dehydratase enzymes. *Science* **269**: 527–529.

New AM, Lehner B. 2015. Systems biology: Network evolution hinges on history. *Nature* **523**: 297–298.

Ohno S. 1970. *Evolution by gene duplication*. Springer-Verlag, New York, NY.

Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science* **317**: 1544–1548.

Pérez J, Kumamoto C, Johnson A. 2013. Candida albicans Commensalism and Pathogenicity Are Intertwined Traits Directed by a Tightly Knit Transcriptional Regulatory Circuit. *PLoS Biol* **11**: e1001510.

Perez JC, Fordyce PM, Lohse MB, Hanson-Smith V, Derisi JL, Johnson AD. 2014. How duplicated transcription regulators can diversify to govern the expression of nonoverlapping sets of genes. *Genes Dev* **28**: 1272–1277.

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

Prud'homme B, Gompel N, Carroll SB. 2007. Emerging principles of regulatory evolution. *Proc Natl Acad Sci U S A* **104 Suppl**: 8605–8612.

Ramos F, Dubois E, Piérard A. 1988. Control of enzyme synthesis in the lysine biosynthetic pathway of Saccharomyces cerevisiae. Evidence for a regulatory role of gene LYS14. *Eur J Biochem* **171**: 171–176.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**: 717–723.

Shim S, Kwan KY, Li M, Lefebvre V, Šestan N. 2012. Cis-regulatory control of corticospinal system development and evolution. *Nature* **486**: 74–79.

Smith DL, Johnson AD. 1992. A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an alpha 2 dimer. *Cell* **68**: 133–142.

Soll DR. 2004. Mating-type locus homozygosis, phenotypic switching and mating: A unique sequence of dependencies in Candida albicans. *BioEssays* **26**: 10–20.

Sorrells TR, Booth LN, Tuch BB, Johnson AD. 2015. Intersecting transcription networks constrain gene regulatory evolution. *Nature* **523**: 361–365.

Sorrells TR, Johnson AD. 2015. Making Sense of Transcription Networks. *Cell* **161**: 714–723.

Tanay A, Regev A, Shamir R. 2005. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc Natl Acad Sci U S A* **102**: 7203–7208.

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

Tishkoff S, Reed F, Ranciaro A, Voight B, Babbitt C, Silverman J, Powell K, Mortensen H, Hirbo J, Osman M, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* **39**: 31–40.

Tsong AE, Miller MG, Raisner RM, Johnson A. 2003. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* **115**: 389–399.

Tsong AE, Tuch BB, Li H, Johnson A. 2006. Evolution of alternative transcriptional circuits with identical logic. *Nature* **443**: 415–420.

Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson A. 2008a. The evolution of combinatorial gene regulation in fungi. *PLoS Biol* **6**: e38.

Tuch BB, Li H, Johnson A. 2008b. Evolution of eukaryotic transcription circuits. *Science* **319**: 1797–1799.

Weirauch MT, Hughes TR. 2010. Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet* **26**: 66–74.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**: 206–216.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman M V, Romano LA. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**: 1377–1419.

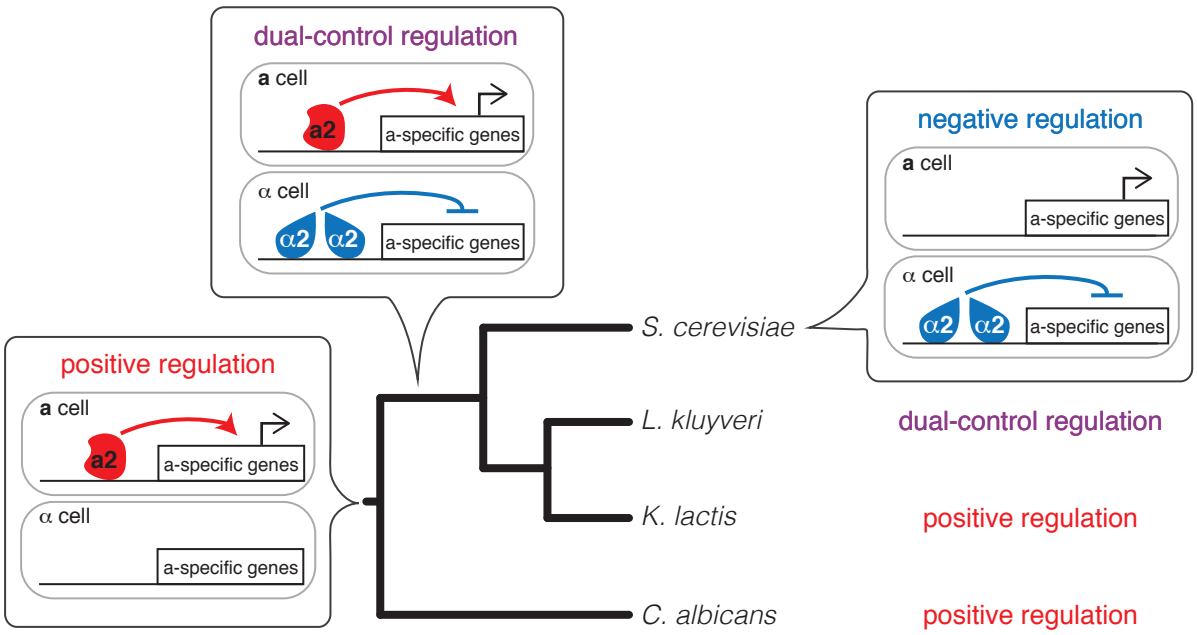


Figure 1: Phylogenetic tree showing the evolution of a-specific gene regulation. In the ancestor of S. cerevisiae and C. albicans, the regulator a2 activates a-specific gene expression in a cells while these genes are not expressed in alpha cells because these cells lack a2. This positive regulation mode is conserved in C. albicans. On the lineage leading to S. cerevisiae, dual-control regulation emerged when alpha2 gained the ability to repress a-specific genes in alpha cells. This dual-control regulation was resolved in three different ways in extant species: in K. lactis, the new alpha2 repression mode was lost, reverting to the ancestral form of positive regulation; in L. kluyveri 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 alpha2 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 alpha cells) is maintained through all of these transitions. (Adapted from Baker et al. 2012)

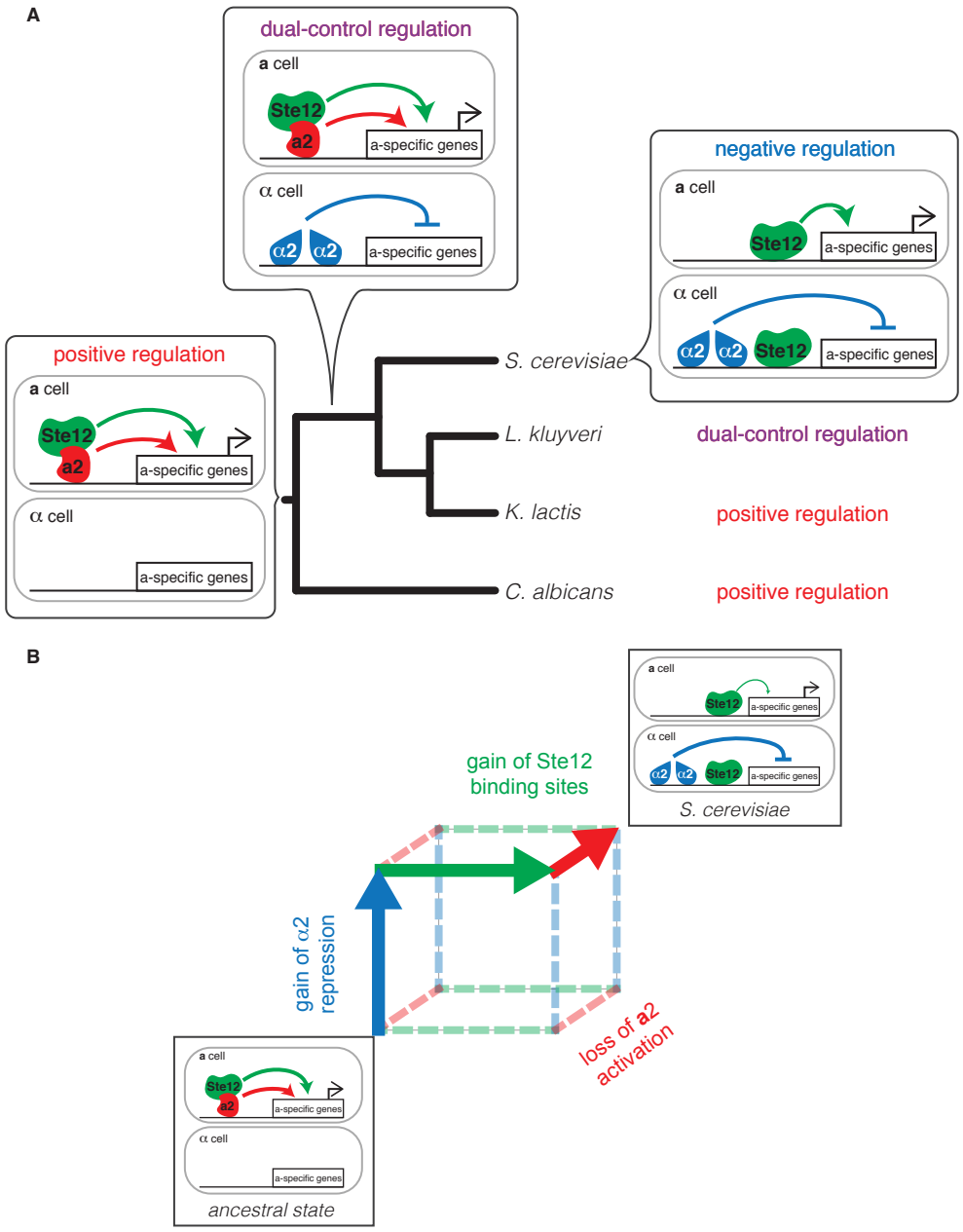


Figure 2: Evolutionary trajectory of rewiring of Ste12 regulation of the a-specific genes. (A) In the ancestor of S. cerevisiae and C. albicans, Ste12 activates a-specific gene expression in a cells indirectly, by binding to the regulator a2. After the dual-control regulatory state of a-specific gene regulation emerged, Ste12 acquired the ability to bind these genes directly to activate their expression. (B) The order of these changes was constrained. The cube shows all possible paths of rewiring of the a-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 mis-regulation of the a-specific genes. In all cases, a2 and alpha2 regulate expression cooperatively with the regulator Mcm1 (not shown), which is expressed in both cell types. (Adapted from Sorrells et al. 2015; New and Lehner 2015)

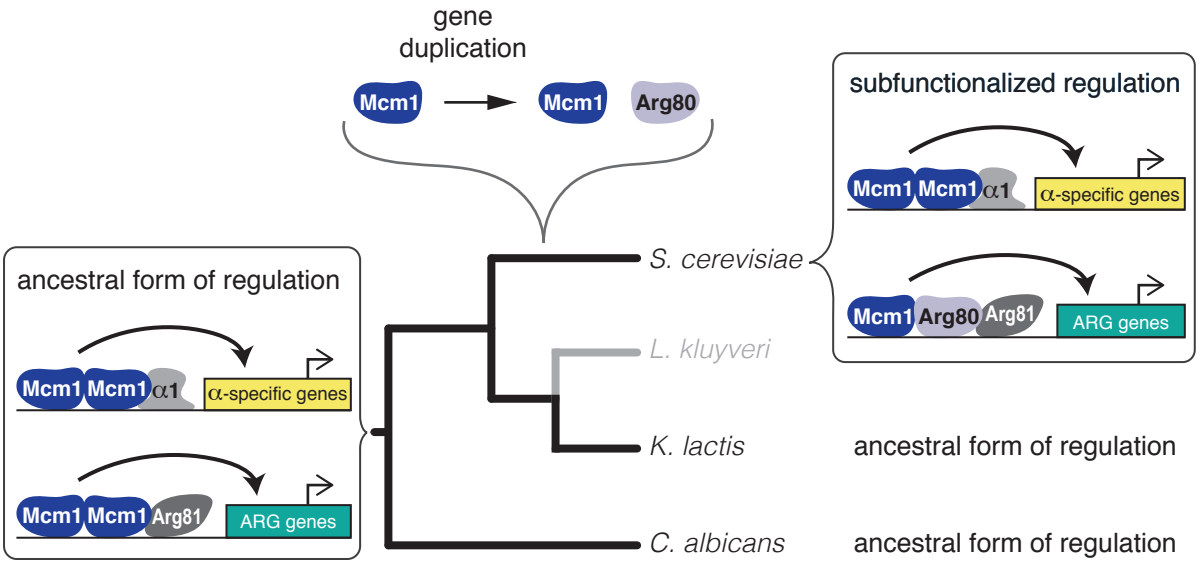


Figure 3: Phylogenetic tree showing evolution of Mcm1 regulation of alpha-specific genes and arginine metabolism genes (ARG genes). In the ancestor of S. cerevisiae and C. albicans, a homodimer of Mcm1 regulates the alpha-specific genes by binding to DNA with alpha1, and a homodimer of Mcm1 regulates 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 alpha-specific genes, while a heterodimer of Mcm1 and Arg80 regulates ARG genes. The regulatory scheme of L. kluyveri has not been determined, but is presumed to be the ancestral form. (Redrawn from Baker et al. 2013)

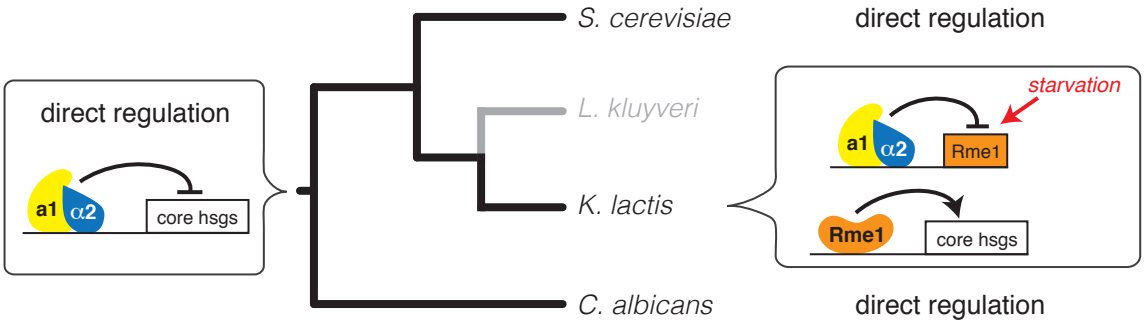


Figure 4: Phylogenetic tree showing the evolution of regulation of the core haploid-specific genes (core hsgs). In the ancestor of S. cerevisiae and C. albicans, a heterodimer of a1 and alpha2 directly repress the core haploid-specific genes in a/alpha cells. This direct regulation is retained in S. cerevisiae and C. albicans. In K. lactis, however, a1-alpha2 now represses the core hsgs indirectly by repressing the regulator Rme1, which activates the core haploid-specific genes. As Rme1 expression is activated by starvation, the core haploid-specific genes now require both starvation and the absence of a1/alpha2 to be fully expressed. (Adapted from Booth et al. 2010)

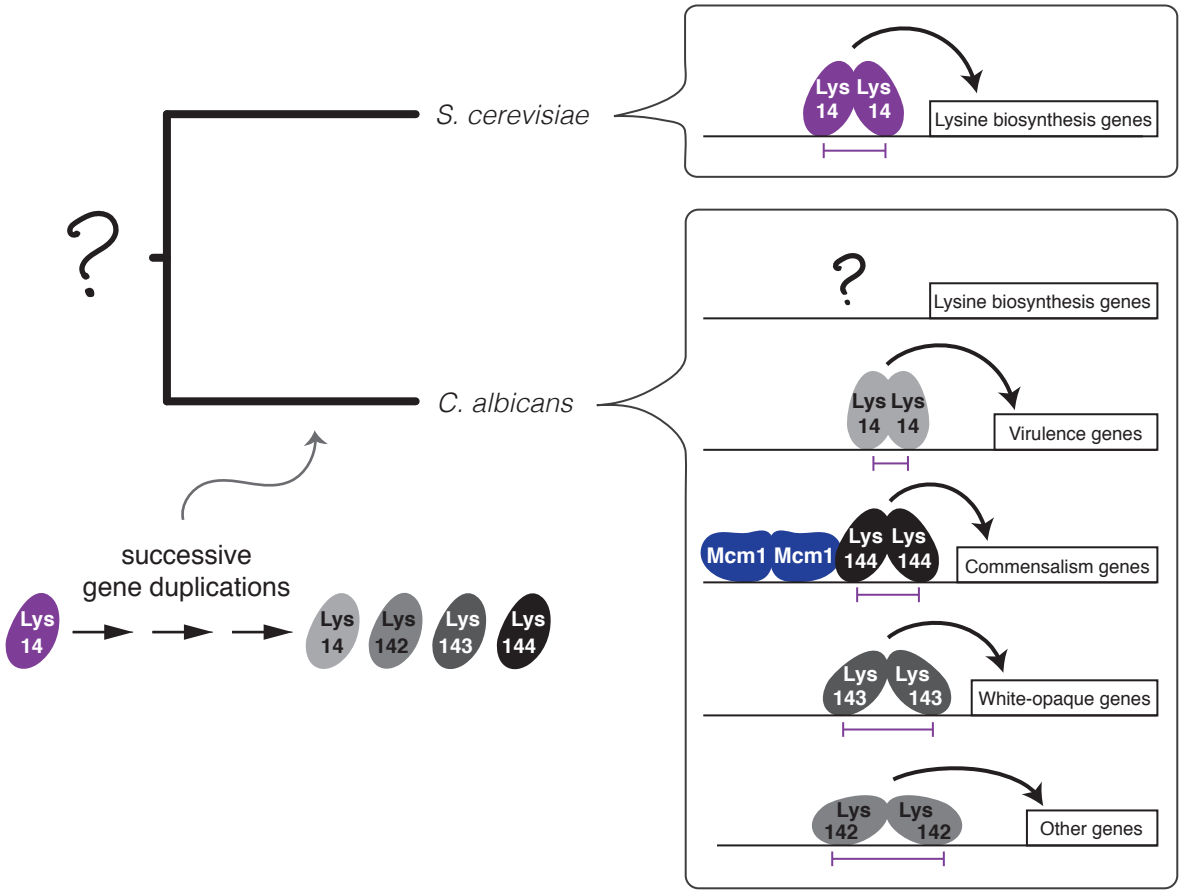


Figure 5: Phylogenetic tree showing evolution of the Lys14 family of regulators. In S. cerevisiae, a Lys14 homodimer regulates lysine biosynthesis genes. On the lineage leading to C. albicans, Lys14 underwent three successive duplications resulting in four paralogs. In C. albicans, none of these four genes regulates lysine biosynthesis. Each paralog recognizes a different DNA sequence with different spacing between monomer binding sites to regulate a unique gene set. In addition, Lys144 now binds cooperatively with a homodimer of Mcm1. The regulation in the ancestor is unknown. (Adapted from Perez et al. 2014; Blake and Barolo 2014)

CHAPTER 2

EXTREME GENE REGULATORY NETWORK PLASTICITY FACILITATES A SWITCH IN FUNCTION OF A CONSERVED TRANSCRIPTION REGULATOR

Introduction

The emergence of novel traits has long fascinated evolutionary biologists, with many intriguing examples existing across the tree of life (*1*-*4*). However, we still lack a detailed understanding of the molecular changes that underlie the evolution of truly novel phenotypes. The rewiring of gene regulatory networks over time is a key source of variation that facilitates the modification of complex phenotypes (*5*-*8*), however it has not been shown how complex gene regulatory networks actually form over evolutionary time. Typically, gene regulatory networks are composed of an interlocking group of "master" transcription regulators and many downstream "target" genes, whose expression is controlled by these regulators in response to environmental or developmental signals. It has been proposed that new gene regulatory networks are a composite of both *de novo* genes and pre-existing genes that are coopted into the newly forming network through a change in regulation (*2*). Yet it remains unclear how such combinations of genes become incorporated into an emerging regulatory network. In particular, an important open question is how an ancient transcription regulator can form the basis of a newly evolving regulatory network.

Here, we reveal the evolutionary history of Ndt80, a sequence-specific transcriptional regulator that has lost its ancestral role to become one of the master regulators in a newly evolved gene regulatory network in the human fungal pathogen *Candida albicans*. In *C. albicans*, Ndt80 is a master regulator of the gene regulatory network that controls the formation of biofilms, multicellular communities of surface-associated cells, which enable persistent human infection on mucosal surfaces of implanted medical devices (Fig. 1A, (*9*-*11*)). The biofilm gene regulatory network in *C. albicans* is both complex—Ndt80 alone controls the

expression of hundreds of genes involved in biofilm formation—and also appears to have evolved relatively recently (*9*). However, Ndt80 regulates a very different process in *Saccharomyces cerevisiae*, another fungal species that last shared a common ancestor with *C. albicans* around 300 million years ago (*12*). In *S. cerevisiae*, Ndt80 regulates the expression of hundreds of genes involved in sporulation, the process in which a diploid cell undergoes meiosis to form four haploid spores (Fig. 1A, (*13-18*)).

In this study, we determine that Ndt80's ancestral role was in the regulation of sporulation. This ancestral role was lost coincident with Ndt80's switch into the network controlling biofilm formation, between approximately 20 and 100 million years ago (*12*). Ndt80 regulates hundreds of genes in *S. cerevisiae* and *C. albicans*, and we show there is little overlap between these gene sets. Thus, the shift in Ndt80 function involved the breaking and forming of hundreds of target gene connections. We can rigorously exclude a model in which the Ndt80 network remained static until a sudden change occurred in the *Candida* clade when its overall function changed from sporulation to biofilm formation. Instead, we found that the Ndt80 regulon was undergoing significant rewiring prior to the switch in Ndt80 function. Specifically, we found extensive rewiring in the target genes of Ndt80 in clades where Ndt80's role in sporulation was conserved. We propose that the inherent, continuous flexibility of the Ndt80 regulon facilitated the exploration of new regulatory networks, leading to the evolution of a novel phenotype. This work is analogous to studies of protein evolution that have demonstrated that promiscuity in protein function, such as the ability of an enzyme to weakly catalyze an alternative reaction, can evolve with no apparent detriment to the function of the protein, but can later be exploited in the evolution of a novel function (*19-23*). This study provides the first account of a similar exploitation of regulatory network promiscuity in the evolution of novel

transcription regulator function. We believe such network promiscuity may be a general feature of regulatory networks and provides an important mechanism to facilitate the emergence of novel phenotypes.

Results

As described above, Ndt80 is required for two very different processes – sporulation/meiosis in *S. cerevisiae* and biofilm formation in *C. albicans* (Fig. 1A). In principle, the difference in phenotype could be due to a difference in the identity of the genes regulated by Ndt80 between these species; alternatively, Ndt80 could regulate the same genes in both species and the difference in phenotype could be due to something else, for example, Ndt80 target genes acquiring new enzymatic functions.

To distinguish between these models, we identified the genes directly regulated by Ndt80 in both *S. cerevisiae* and *C. albicans* using chromatin immunoprecipitation of epitope-tagged Ndt80 followed by high throughput sequencing (ChIP-Seq). To minimize the effect of species-specific differences in Ndt80 expression (and to capture as much of the network as possible) the promoters of Ndt80 in both species were replaced with high-expression constitutive promoters (see Supplemental Materials and Methods). It has been well established that, even using proper controls, chromatin immunoprecipitation produces false positives in addition to valid instances of binding (*24*, *25*). To eliminate spurious signals and identify bona fide instances of Ndt80 binding, we employed four different criteria to identify and filter Ndt80 targets (Fig. 1C). First, we identified genes with significant peaks of ChIP enrichment in the intergenic region upstream of that gene as our least stringent criteria. Second, we filtered this set to include only those ChIP

enrichment peaks that contained an Ndt80 *cis*-regulatory motif in the intergenic region. Third, of those ChIP peaks that contained an Ndt80 motif, we further refined the set of Ndt80 targets by requiring that the motif also be present in the orthologous intergenic region of two very closely related species. (For *S. cerevisiae*, we used *S. mikatae* and *S. kudriavzevii*, for *C. albicans* we used *C. tropicalis* and *C. dubliniensis* (*26*, *27*)). Lastly, we used gene expression data to restrict Ndt80 targets to ChIP peaks with an Ndt80 motif where the adjacent gene also exhibited a significant change in gene expression when Ndt80 was deleted (*9*, *14*).

Each of the four sets of criteria were used to identify and compare the genes regulated by Ndt80 between *S. cerevisiae* and *C. albicans*. While one could argue *a priori* of the superiority of one set of criteria over another, we note that the major conclusions of this analysis were independent of the approach used. Thus, while the number of genes identified as Ndt80 targets differs depending on the criteria used, all four methods produce the same overall result: Ndt80 regulates many genes, the majority of which differ between *S. cerevisiae* and *C. albicans* (Fig. 1D). By all methods, fewer than 13% of the targets of Ndt80 in these two species are shared; with the most stringent criteria this drops to 2% (Fig. 1D). To ensure that constitutive overexpression of Ndt80 was not responsible for this large difference in Ndt80 targets, we also performed ChIP-Seq on Ndt80 in both species under the control of the endogenous promoter. While many fewer regions are bound than with constitutive Ndt80 expression, we similarly find very little overlap in the Ndt80 targets in *S. cerevisiae* and *C. albicans*.

These results all support the idea that Ndt80 regulates different sets of genes in *S. cerevisiae* and *C. albicans*. Most of these genes have 1:1 orthologs between the two species, however, there are a smaller number of species-specific genes. If we exclude these genes from our analysis, however, we still find very little overlap in Ndt80 targets (2-20% targets shared,

depending on target identification method), suggesting that gene gain/loss alone cannot account for the change in Ndt80 targets between *S. cerevisiae* and *C. albicans*. Despite the significant differences in Ndt80 targets, however, the *cis*-regulatory sequence bound by Ndt80 in both species is highly conserved ((*9*, *28*), Fig. 1B). This observation indicates that the Ndt80 regulon has been significantly rewired between *S. cerevisiae* and *C. albicans* without a change in the DNA-binding specificity of Ndt80.

While the Ndt80 genes in *S. cerevisiae* and *C. albicans* tested above are unambiguous orthologs, in *C. albicans* there is a second copy of Ndt80, resulting from a recent gene duplication (Fig. 2A, (*29*)). The other paralog, which we refer to as Ndt80A, shares synteny with the pre-duplication Ndt80 (*26*), indicating that it derives from the original gene. Ndt80B—the paralog required for biofilm formation whose targets we describe above—derives from the duplicated copy. The regulatory function of Ndt80A is unknown, but deletion of Ndt80A has no effect on biofilm formation ((*9*), Fig. 4H). To ensure that the difference in targets between Ndt80 in *S. cerevisiae* and Ndt80B in *C. albicans* is not simply a consequence of the gene duplication, we identified the regulatory targets of Ndt80A in *C. albicans* by ChIP-Seq under control of the same constitutive promoter used for Ndt80B. We found that the targets regulated by Ndt80A represent a small subset of the targets of Ndt80B; that is, Ndt80A binds to many fewer genomic regions, but all of these regions are also bound by Ndt80B. Notably, if we compare the targets of Ndt80 in *S. cerevisiae* to the targets of Ndt80A in *C. albicans*, we find that less than 3% of the overall targets are shared, demonstrating that Ndt80A's targets are no more similar to the targets of Ndt80 in *S. cerevisiae* than that of Ndt80B. These results show that the rewiring of Ndt80 targets that occurred since *S. cerevisiae* and *C. albicans* diverged was not

the result of a diversification in Ndt80 paralogs following the gene duplication in the *Candida* clade.

To reconstruct a timeline of the evolution of the Ndt80 gene regulatory network, we conducted ChIP-Seq experiments in three additional species that branch from the lineages to *S. cerevisiae* and *C. albicans* at highly informative points: *K. lactis, P. pastoris* and *P. stipitis* (Fig. 2A). In each case, we used the strategy of expressing Ndt80 under the control of a strong constitutive promoter to minimize expression differences between species. In *C. albicans* and *P. stipitis* there are two paralogs of Ndt80 and so targets were identified for each paralog separately and the union was taken to represent all Ndt80 targets in that species.

We first investigated whether the regulatory targets of Ndt80 in each of these species more closely resemble the targets in *S. cerevisiae* or the targets in *C. albicans*. Ndt80 targets were identified in all species using ChIP enrichment plus Ndt80 motif presence (Fig. 1C), and compared to the targets of Ndt80 in *S. cerevisiae* and *C. albicans*. We found, counter to our expectation, that a very large number of Ndt80 targets in each species are not targets of Ndt80 in either *S. cerevisiae* or *C. albicans* (50% of all targets in *K. lactis*, 70% in *P. pastoris*, and 48% in *P. stipitis,* Fig. 2B). Thus, the Ndt80 regulons in these three species do not closely resemble either those of *S. cerevisiae* or *C. albicans*; instead, each have acquired a distinctive set of Ndt80 target genes. If we repeat this analysis using alternative criteria for identifying Ndt80 targets (Fig. 1C), the largest category of targets in each species is still genes that are targets in neither *S. cerevisiae* or *C. albicans*, indicating that this conclusion is robust to the stringency of Ndt80 target identification.

To further analyze the similarity in Ndt80 targets across all five species tested, we identified shared Ndt80 targets for every two-species comparisons. For this analysis, only genes

with identifiable 1:1 orthologs in both species were considered, to correct for changes due to gains and losses of target genes themselves and focus instead on changes in Ndt80 regulation of conserved genes. This analysis shows very little overlap in Ndt80 targets between any two species tested (Fig. 2C). *P. stipitis* and *C. albicans* share the largest overlap, with 25% of the targets in either species bound in both, likely due to the fact that these species are the most closely related of the five tested (Fig. 2A). The remaining species, however, share less than 15% of their targets, with no discernable pattern predicted by the phylogenetic relationships between the species. These results demonstrate that Ndt80 has undergone extensive rewiring of its target genes along several independent phylogenetic branches since the last common ancestor of *S. cerevisiae* and *C. albicans*. These changes have resulted in very different Ndt80 regulons in each of the five extant species tested. Consistent with this model, only 10 genes are targets of Ndt80 in all five species tested (Table S1). While this is more than expected strictly by chance considering the number of regions bound in each species ($p < 10^{-6}$), this number pales in comparison to the total of 3,261 genes that are Ndt80 targets in just one of the five species. In short, the Ndt80 regulon exhibits a high degree of rewiring in each branch leading to these five extant species.

Because the Ndt80-target gene connections differ significantly between species, we considered whether these differences were due primarily to changes in the Ndt80 protein itself, or to changes in the distribution of its *cis*-regulatory motif across the different genomes. Although the *cis*-regulatory sequence recognized by Ndt80 is conserved across these species, changes in the Ndt80 protein itself that alter, for example, protein-protein interactions with other transcription regulators, could account for some of the differences between species. The sequence of the Ndt80 protein does differ considerably across the species tested, with only 55%

similarity in the DNA-binding domains of Ndt80 in *S. cerevisiae* and *C. albicans* (*29*), with the remainder of the protein showing much less similarity.

To test whether these changes in protein sequence were primarily responsible for the observed differences in Ndt80 target genes between species, we moved the sequence for *P. pastoris* Ndt80 into *S. cerevisiae* under the control of the same constitutive promoter used for ChIP-Seq of the endogenous Ndt80 in *S. cerevisiae*, and carried out ChIP-Seq on the *P. pastoris* Ndt80. If we compare Ndt80 ChIP enrichment genome-wide in this experiment to Ndt80 enrichment for the native Ndt80s in each species, we find that many more of the targets of the ectopic *P. pastoris* Ndt80 are shared with the targets of the native Ndt80 in *S. cerevisiae* (Fig. 3A and B) than the targets of the Ndt80 in *P. pastoris* (Fig. 3C and D). These results demonstrate that changes in Ndt80 protein sequence across species have had little effect on the Ndt80-target gene connections; instead, the connections are predominantly determined by the distribution of *cis*-regulatory sequences across the different genomes. This result does not mean that changes in the protein are unimportant; indeed, the ectopic *P. pastoris* Ndt80 does not complement an Ndt80 deletion in *S. cerevisiae* (data not shown). Our results do show, however, that the ectopic *P. pastoris* protein binds to the *S. cerevisiae*-specific positions of the genome. Thus, the majority of difference in the Ndt80 regulons across species are due to gains and losses of the conserved Ndt80 DNA-binding motif.

To determine how the rewiring of the Ndt80 gene regulatory network corresponds to changes in the overall function of Ndt80, we determined the phenotype of an Ndt80 deletion in *K. lactis*, *P. pastoris,* and *C. lusitaniae* (Fig. 2A). We first tested whether deletions had an effect on sporulation. In *K. lactis* and *P. pastoris* the Ndt80 deletion mutant is completely deficient in its ability to sporulate (Fig. 4A and B). C. *lusitaniae*, like *C. albicans*, has two paralogs of

Ndt80. Deletion of one paralog (Ndt80A) results in a complete deficiency in sporulation, while deletion of the other paralog (Ndt80B) results in a dramatic reduction in sporulation efficiency (Fig. 4C). In contrast to these four species, *C. albicans* has never been observed to undergo sporulation and meiosis, relying instead on an alternate parasexual cycle (*30*, *31*). Given that Ndt80 is required for sporulation in *S. cerevisiae* (*13*, *18*), *K. lactis, P. pastoris*, and *C. lusitaniae*, and given the phylogenetic relationship between these species (Fig. 2A), the most parsimonious model is that, in the ancestor of these species, Ndt80 regulated sporulation.

We next investigated the requirement for Ndt80 in biofilm formation in these species. In *C. albicans*, biofilms consist of thick structures of different cell types (yeast-form and hyphae) that can form both *in vitro* and *in vivo* (on implanted catheters in a mouse model, (*32*)). Deletion of Ndt80B in *C. albicans* completely eliminates surface adherence and biofilm formation (Fig. 4H). In contrast, in *S. cerevisiae*, *K. lactis*, *P. pastoris*, and *C. lusitaniae*, the Ndt80 mutants and wild-type are comparable in their ability to adhere to a solid surface (Fig. 4D - G). It is worth noting, however, that these species form only a relatively thin layer of yeast-form cells on the solid surface rather than the type of thick biofilms characteristic of those formed in *C. albicans*, highlighting the pivotal role likely played by Ndt80 in the evolution of the complex biofilm phenotype.

Taken together, these results indicate that Ndt80 regulated sporulation in the shared ancestor of *S. cerevisiae* and *C. albicans*, and that this role was lost and the role in biofilm formation gained along the *C. albicans* lineage, after it branched off from *C. lusitaniae* (Fig. 5A). We believe this scenario is more likely than that of the next most parsimonious model, which holds that Ndt80 regulated both sporulation and biofilm formation in the shared ancestor, requiring four independent losses (three losses of biofilm regulation, one loss of sporulation

regulation). Given that Ndt80 regulates hundreds of genes in each species, a single loss and a single gain of Ndt80 function seem more plausible than four independent losses.

Combining this phenotypic data with the Ndt80 target identification previously discussed, we can begin to understand the importance of individual genes in the overall switch in Ndt80 function. Only 10 genes are targets of Ndt80 in all five species tested (*S. cerevisiae*, *K. lactis, P. pastoris*, *P. stipitis*, and *C. albicans*) (Figure 2E). This suggests that a very limited set of genes were repurposed from sporulation regulation to biofilm regulation. If we examine these genes and compare them to their known functions in *S. cerevisiae*, we find three genes involved in cell cycle regulation (Clg1, Cln2, Yox1), four metabolic enzymes (Gdh2, Gln1, Rki1, Sga1), two proteins associated with the ribosome (Rps7a, Tma10), and one involved in regulation of cell wall biosynthesis (Usv1) (Figure 2E), suggesting that these may be important links between the phenotypes of biofilm formation and sporulation. The gene Usv1 is particularly intriguing, as its ortholog in *C. albicans*, Bcr1, is one of the other master regulators of biofilm formation, along with Ndt80 (*9*, *33*). This suggests that two of the master regulators of biofilm formation, Ndt80 and Bcr1, may have had an ancestral regulatory relationship that was co-opted in the evolution of the biofilm regulatory network.

Because the targets of Ndt80 appear to have rewiring significantly during the switch from sporulation regulation to biofilm regulation, we next investigated the significance of this rewiring by comparing the overall phenotype of Ndt80 in each species to its underlying regulatory network. In particular, we tested whether Ndt80 targets are more similar among species with a conserved overall Ndt80 function than among species with a diverged Ndt80 function. If so, this would suggest a dramatic rewiring of targets associated with the switch in Ndt80 regulatory function. To test this idea, we focused on four species in which we know both

the Ndt80 phenotype and the binding targets: *S. cerevisiae*, *K. lactis*, *P. pastoris*, and *C. albicans*. We then compared between these species, determining the conservation in Ndt80 targets normalized to the overall sequence conservation between species (for all six two-species comparisons) using ChIP Peaks with Ndt80 motifs (Fig. 1C). In contrast to our simple expectation, we found that conservation of phenotype does not correlate with lower rates of divergence in Ndt80 targets (Fig. 5B). In some cases, comparisons between species in which Ndt80 has a conserved role in sporulation regulation show higher rates of Ndt80 target divergence than comparisons between species in which Ndt80 regulates sporulation in one species and biofilm formation in the other (Fig. 5B). This pattern also holds if we use the more stringent criteria to identify Ndt80 targets in each species. These results show that the targets of Ndt80 change rapidly, even while Ndt80's conserved role in sporulation regulation is maintained. While we cannot entirely rule out the possibility that Ndt80 has another, unknown regulatory function in one or more of these species that accounts for the significant difference in Ndt80 targets, in *S. cerevisiae* Ndt80 is not produced in mitotically dividing cells, and is produced specifically during sporulation (*13*). Similarly, in *K. lactis* and *P. pastoris* Ndt80 is not expressed in rich growth media (data not shown), consistent with the idea that Ndt80 is only present during sporulation and is unlikely to have another more general regulatory role.

The dramatic rewiring of Ndt80 targets that appears to have taken place even while the overall phenotype of Ndt80 was conserved suggests that only a small number of the Ndt80 targets must be conserved in order for the sporulation network to continue to be functional. There are only 20 genes that are targets of Ndt80 in all three sporulation species (*S. cerevisiae*, *K. lactis*, and *P. pastoris,* Table S2). If we examine the known functions of these genes in *S. cerevisiae*, we find no obvious pattern in the types of genes conserved compared to those rewired

between the species. For example, there are no gene ontology terms significantly enriched in either the genes that are conserved, or those that are not conserved, amongst the sporulation species. One possible explanation for this flexibility is that there are sets of genes for which some, but not all, of the genes must be activated by Ndt80 in order for sporulation to proceed. For example, there are 62 genes bound by Ndt80 in one of the three sporulation species that are involved in cell wall assembly (GO annotation: cell wall organization or biogenesis). There are 57 bound in *S. cerevisiae*, 26 bound in *K. lactis*, and 32 bound in *P. pastoris*. However, only one of these genes, Skg1, is bound in all three species (Table S2). As sporulation requires the formation of ascospores and the generation of new cell wall, it is likely that these cell wall assembly genes play an important role in sporulation. However, it is possible that only a subset of these genes need to be activated by Ndt80 in order for cell wall assembly to function properly in sporulation. This would allow Ndt80 to drift between different configurations of regulation of these genes, while its overall role activating cell wall assembly during sporulation was conserved.

In support of this idea that only a small number of individual genes must be regulated by Ndt80 for the sporulation network to be functional, we find that two genes that are particularly important in sporulation in *S. cerevisiae* are shared Ndt80 targets across the three species: Ime2 and Cdc5. Ime2 is a serine/threonine kinase which has been shown to be required for sporulation in *S. cerevisiae* (*34*, *35*), and which is required for the full activity of Ndt80. Intriguingly, the regulatory relationship between Ime2 and Ndt80 appears to be quite ancient, as experiments in the distantly related species *Neurospora crassa* and *Aspergillus nidulans* demonstrate that homologs of Ime2 negatively regulate the expression of Ndt80 homologs (*36*, *37*). Thus, the ability of Ndt80 to regulate the expression of Ime2 may be a particularly important regulatory

connection, explaining why it is one of the few conserved amongst all extant sporulation species tested. Similarly, the polo-like kinase Cdc5, is a conserved Ndt80 target in all three sporulation species. In *S. cerevisiae*, Cdc5 has been shown to be a particularly important target of Ndt80, as its ectopic expression in an Ndt80 deletion allows cells to bypass the meiotic block observed in the absence of Ndt80 (*38*, *39*). This suggests that induction of Cdc5 is one of the more crucial functions of Ndt80 in *S. cerevisiae*, and it is therefore not surprising that this relationship is conserved among other sporulation species. Interestingly, Cdc5 is also a target of Ndt80 in *C. albicans*, suggesting this kinase may have acquired a new role in biofilm regulation in this species. Taken together, these results suggest that only a small number of Ndt80 regulatory connections must be conserved in order for Ndt80 to maintain to regulate sporulation, and that significant drift in the other connections is possible without compromising the function of the overall regulatory network.

One explanation for how Ndt80 maintained a role in sporulation, despite the near complete turnover of its target genes, is that the genes directly involved in sporulation themselves are changing across species. This idea would account, at least in part, for how the regulatory targets of Ndt80 can differ so significantly while Ndt80 maintains a conserved overall function. To test this idea, we measured global gene expression during sporulation and compared it to mitotic growth conditions in *K. lactis* and *P. pastoris* (see Supplemental Materials and Methods). We then compared the genes specifically up-regulated during sporulation in these species to those upregulated during sporulation in *S. cerevisiae* (*14*). We found extensive differences in the genes activated during sporulation across these species. Only 23 genes show sporulation-activation in *S. cerevisiae*, *K. lactis*, and *P. pastoris*. In contrast, hundreds of genes are activated during sporulation uniquely in each of the three species. We also measured global

gene expression in an Ndt80 deletion in both *K. lactis* and *P. pastoris* under sporulation conditions, and found that genes whose expression depends on Ndt80 during sporulation also differ significantly across species . These results show that the genes induced during sporulation differ significantly between *S. cerevisiae*, *K. lactis*, and *P. pastoris*, and that this can explain, at least in part, the large differences in Ndt80 targets seen across these species, despite the overall conservation of Ndt80 function in sporulation regulation. This analysis also reveals that there is no single "solution" to sporulation and meiosis and that widely divergent networks can nonetheless orchestrate this conserved outcome.

Discussion

We have combined ChIP-Seq, RNA-Seq, molecular genetic analysis, and bioinformatics to identify the function and gene regulatory network of Ndt80 in a variety of extant fungal species, encompassing approximately 300 million years of divergence time. We have deduced that the function of Ndt80 in the most recent shared common ancestor of *S. cerevisiae* and *C. albicans* was likely the regulation of sporulation, and that a switch in overall function to the regulation of biofilm formation occurred in a recent ancestor of *C. albicans*. While a few conserved transcription regulators have previously been shown to regulate different genes in different extant species, we believe this is the most dramatic example yet described of a transcription regulator switching its regulatory role, in this case from sporulation to the seemingly unrelated process of biofilm formation. We have shown that this switch was not accompanied by a change in the DNA-binding specificity of Ndt80 (Fig. 1B) or another major change to the Ndt80 protein itself (Fig. 3). Rather, a change in the distribution of the conserved

Ndt80 *cis*-regulatory sequence across hundreds of genes in the genome accounts for the different functions of Ndt80 in *S. cerevisiae* and *C. albicans*.

Prior to these experiments, we considered two likely explanations for the change in Ndt80 function. First, we hypothesized that the duplication of Ndt80 could have allowed the ancestral function to be conserved in one paralog while the other paralog acquired a novel role in biofilm regulation. However, the similarity in function of Ndt80 paralogs in *C. lusitaniae* (Fig. 4C and F) as well as the similarity in targets of Ndt80 paralogs in *C. albicans* rule out this model. A second plausible model was that a sudden change in Ndt80-target gene connections occurred along the *C. albicans* lineage, perhaps triggered by the loss of meiosis in this clade, relaxing constraints on Ndt80. However, we also ruled out this model by showing that similarly high rates of Ndt80 rewiring occurred in all phylogenetic branches examined, even those where the function of Ndt80 remains conserved in sporulation (Fig. 5B). For example, we find that only ~15% of the hundreds of Ndt80 targets are shared between any of the two species in which Ndt80 regulates sporulation. Rather than a sudden shift in regulation, our results indicate significant flexibility in the regulon of Ndt80; the Ndt80 regulatory connections appear able to move through many configurations even while retaining a conserved output in the regulation of sporulation. In support of this idea, we showed that genes up-regulated during sporulation (by Ndt80 and other sporulation regulators) vary considerably across different extant species. In addition, it has previously been noted that several key genes required for meiosis and sporulation in *S. cerevisiae* (Ime1, Zip2, Spo13) are missing from the genomes of species such as *C. lusitaniae* that undergo a similar pathway of meiosis and sporulation (*40*). All of these observations indicate that there is no "universal" solution to meiosis and sporulation in terms of both the genes required for this process and the sporulation network itself.

This work illustrates how a transcription regulator with hundreds of connections can nonetheless shift its role in the cell. We propose that the continuous exploration of regulatory connections available to Ndt80—while still maintaining its role in sporulation—facilitated its relatively recent change to a regulator of biofilm formation. Had Ndt80 been more heavily constrained in its target gene connections, it seems unlikely that it could have made this dramatic switch in function. A corollary of this conclusion is that high rates of rewiring per se are not predictive of a change in function; direct experimentation was required to determine the biological function of Ndt80 across species.

Although other transcription regulators may be subject to more severe evolutionary constraints than Ndt80, there is ample evidence that many transcription networks can drift through new configurations while preserving output (*41-45*). For example, the rapid reconfiguration of *cis*-regulatory sequences in *Drosophila* enhancers, despite a constant output, provide a different but nonetheless striking demonstration of the flexibility of transcription networks (*46-48*). While much regulatory rewiring likely occurs through drift with no obvious shift in phenotype, we propose that this rewiring can also predispose a regulator to a dramatic shift in function. In this way, a newly evolved phenotype—complex biofilm formation in *C. albicans*—can come to be controlled by a deeply conserved transcription regulator that has an entirely different role in other species. Such extreme, inherent flexibility in gene regulatory networks, such as that exhibited by Ndt80, may prove to be an important contributor to the formation of new networks underlying novel phenotypes.

Finally, we note that the behavior observed for Ndt80—rapid explorations of new circuit configurations that nonetheless support sporulation and meiosis—strongly supports computational models of gene network drift(*49*). For example, Wagner and colleagues have

argued that metabolic pathways, protein sequences, and gene regulatory networks are all inherently flexible, allowing an exploration of many alternatives that still maintain function (*22*). This exploration, in turn, allows evolving networks to sample many points in "network space". Some of these points may be only a few changes away from generating a network with novel function, but without the ability to drift through many configurations that maintain function, their novel configurations could not be reached without destroying the original network. We believe that our work with Ndt80 represents a tangible example of this idea: it is only through the inherent flexibility of the Ndt80 connections that the network—while still maintaining sporulation and meiosis—could reach a position where the transition to regulating biofilm formation was a relatively smooth one. This conclusion is analogous in many ways to studies of "molecular exploitation" that have demonstrated how drift in protein function can facilitate the evolution of novel functions without compromising the ancestral function of the protein(*20*, *23*, *50*). We propose that a similar inherent promiscuity in the Ndt80 regulatory network was crucial for the repurposing of Ndt80 from regulating sporulation and meiosis to regulating biofilm formation.

Materials and Methods

Media

All strains were grown in YPD at 30°C unless otherwise noted. For sporulation, *S. cerevisiae* strains were grown in liquid YPA (2% Peptone, 2% Potassium Acetate, 1% Yeast Extract) and incubated at room temperature for 20-30 hours. For *K. lactis*, saturated liquid cultures in YPD were spotted onto SPO plates (1% potassium acetate, 2% agar + amino acids) and incubated at room temperature for 3 days. For *P. pastoris*, cells growing on YPD plates were patched onto PpSPO plates (0.5% Sodium acetate, 1% potassium chloride, 1% glucose, 2% agar) and incubated at room temperature for 3 days. For *C. lusitaniae*, cells growing on YPD plates were patched onto PDA plates (0.37% Potato Dextrose + 1.45% Agar) and incubated at room temperature for 4 days. To quantify sporulation efficiency, three technical replicates were performed and 200 cells were counted for each sample by DIC microscopy. Nuclear staining using DAPI (*K. lactis*, *P. pastoris*) and Hoecsht (*C. lusitaniae*) was used to verify spore formation.

Strain construction

Strains used in this study are listed in Table 1 and primers used for strain construction are listed in Table 2. Gene disruption cassettes for Ndt80 deletions were constructed by fusion PCR. For *K. lactis*, two cassettes were constructed, one with a Ura3 marker, one with a KanMX marker, each with 700 bp homology flanking the markers on either side. These constructs were transformed, sequentially, into a wild-type diploid strain. For *P. pastoris* and *C. lusitaniae*, a

split marker approach was used and two constructs were generated for each gene disruption. For *P. pastoris*, a Hygromycin resistance marker and 700 bp flanking homology were used, and the split markers were transformed into two haploid strains of complementary mating types. These strains were then mated together to form a diploid Ndt80 deletion (ploidy was verified by FACS). For *C. lusitaniae*, a NAT resistance marker was used along with 1 kb of flanking homology. The split markers were transformed into a wild-type α-cell, that was then mated with a wild-type **a**-cell, sporulated, and Ndt80 deletion **a**-cells were picked and verified by PCR. These were then mated with the Ndt80 deletion α-cell to generate a diploid Ndt80 deletion strain.

Tagged strains for ChIP were generated using a 13x c-terminal Myc tag(*51*) inserted into the genome at the c-terminus of the endogenous Ndt80 sequence. For *S. cerevisiae*, long primers were used to amplify the Myc tag fused to a KAN marker from pFA6a-13Myc- kanMX6 (*51*) and this construct was integrated into a wild-type W303 strain. For *K. lactis*, split marker constructs were generated with 700 bp flanking homology fused to the Myc-KAN cassette and transformed into a wild-type diploid strain. For *P. pastoris*, split marker constructs were generated with 700 bp flanking homology fused to the Myc-NAT construct amplified from pADH34 (*52*). In *P. stipitis*, homologous recombination was found to be very inefficient, and thus constructs were randomly integrated into the genome. Constructs containing a NAT marker upstream of an *A. gossypii* Tef1 promoter upstream of *P. stipitis* Ndt80 fused to a 13x c-terminal Myc tag with a Sat1 terminator were generated by PCR for both Ndt80A and Ndt80B and transformed into a wild-type strain. Two independent isolates were generated for each paralog and tested independently by ChIP-Seq to verify that the location of integration did not affect

Ndt80 function or genomic binding. To test binding of *P. pastoris* Ndt80 in *S. cerevisiae*, fusion PCR was used to generate a construct with a Hygromycin resistance marker fused to the Gal1 promoter from *S. cerevisiae* upstream of *P. pastoris* Ndt80 fused to a 13x c-terminal Myc tag with a Sat1 terminator. This was fused to homology to Ura3 and integrated in an *S. cerevisiae* wild-type at the Ura3 locus.

To generate strains with highly expressed Ndt80, different constitutive promoters were used, and constructs generated by fusion PCR to integrate these upstream of Ndt80. In *S. cerevisiae*, pGal1 from *S. cerevisiae* was used; in *K. lactis* pGal1 from *K. lactis* was used; in *P. pastoris* and *P. stipitis* pTef1 from A. gossypii was used; in *C. albicans* pTDH3 from *C. albicans* was used.

S. cerevisiae and *C. albicans* were transformed using standard lithium acetate protocols. Electroporation protocols were used for *K. lactis* (*53*), *P. pastoris* (*54*), and *P. stipitis*. The protocol for *P. stipitis* was adapted from (*54*), with cells harvested at an OD of 1.3-1.5.

Chromatin immunoprecipitation and high throughput sequencing

For all ChIP experiments except the mid-sporulation ChIP, samples were isolated from log-phase cultures in YPD and chromatin immunoprecipitation was performed as previously described (*52*). Libraries were prepared using NEBNext Multiplex Kit for Illumina as previously described (*55*) and sequenced on an Illumina HiSeq 4000. For mid-sporulation ChIP in *S. cerevisiae*, cells were induced to sporulate as previously described (*13*, *15*), and samples were isolated after 22 hours in sporulation media.

mRNA expression experiments

RNA expression was measured in *P. pastoris* using RNA-Seq. Samples were isolated from log-phase cultures in YPD or from mid-sporulation cultures after 30 hours. Total RNA was isolated using the Ambion RiboPure kit, and mRNA was isolated using the Oligotex mRNA Mini kit. Purified mRNA was then concentrated using the Zymo RNA Clean and Concentrator kit and sequenced on an Illumina HiSeq 4000.

RNA expression was measured in *K. lactis* by whole-genome tiling microarrays as previously described (*9*).

Biofilm experiments

Biofilms were grown and imaged by confocal scanning laser microscopy (CSLM) similar to previously described(*56*). In brief, silicon squares were pre-incubated overnight at 30^{o}C in bovine serum and after washing with phosphate-buffered saline (PBS) they were inoculated in Spider media containing 1% glucose instead of mannitol with cells coming from an overnight YEPD 30^{o}C culture to an OD_{600} of 0.5. The squares were then incubated for 90 minutes at 30^{o}C shaking at 200 rpm for cell adherence. After adherence the squares were washed with PBS and transferred to fresh Spider 1% glucose to be incubated for 48 hours at 30^{o}C shaking at 200 rpm. For CSLM, the biofilms grown on the silicon squares were stained with concanavalin A Alexa Fluor 594 conjugate (50 μg/ml) and visualized using a Nikon Eclipse C1si upright spectral imaging confocal microscope and a 40x/0.80W Nikon objective.

Identifying regions of Ndt80 binding

Sequencing reads were aligned to the genome using Bowtie (*57*). Peaks and fold enrichment values were generated using MACS (*58*) with peak shift sizes generated using SPP (*59*). Peaks were mapped to intergenic regions with any overlap with the peak and any neighboring genes assigned to that peak using MochiView (*60*). A gene was considered to have an Ndt80 motif if a peak was present in its intergenic for both replicates. Fold enrichment values were generated for every base in the genome and a maximum was taken for every intergenic region for the *P. pastoris* swap experiment.

DNA sequence motif generation discovery and scoring

Ndt80 motifs were generated *de novo* from regions of binding using MEME (*61*). The union of peaks in two replicates were trimmed to 200 bp around the midpoint (*60*) and submitted to MEME-ChIP using default parameters. Ndt80 consensus sequence of 'CACAAA' was used for filtering of peaks, scoring was performed in Mochiview (*60*), with each intergenic region scored for presence/absence of the consensus motif. For identifying genes with Ndt80 motif present in three closely related species, orthologs were identified (discussed below) and orthologous intergenic sequence was scored for consensus 'CACAAA' sequence using MochiView. For *S. cerevisiae*, S. mikatae and S. kudriavzevii were used as a comparison; for *K. lactis*, *E. cymbalariae and E. gossypii* were used; and for *C. albicans, C. dubliniensis* and *C. tropicalis* were used.

Mapping gene orthologs across species

Orthologs were identified in *S. cerevisiae, S. mikatae, S. kudriavzevii, K. lactis, E. cymbalariae, E. gossypii, P. stipitis, C. dubliniensis, C. tropicalis,* and *C. albicans* using synteny-based orthology databases YGOB (*26*) and CGOB (*62*). *P. pastoris* orthologs were identified using YGAP (*26*).

RNA-Seq analysis

RNA-Seq analysis was performed using the TopHat and Cufflinks suites as previously described (*63*).

Phylogenetic tree building

A phylogenetic tree of relevant species was constructed as previously described (*64*). Protein sequences for 73 orthologs present in a single copy in all species were concatenated and aligned using Clustal (*65*) and a tree was constructed using FastTree (*66*) Two additional outgroup species, *N. crassa* and *A. nidulans* were used in the building of the tree to improve root placement, but were omitted from the figure for simplicity.

Estimating species divergence

To determine overall sequence conservation for two-species comparisons for Fig. 5B, branch lengths from phylogenetic tree were used. The fraction of targets conserved in any two-species comparison were multiplied by the branch lengths of the two relevant species from their most recent common ancestor to normalize target conservation to overall species conservation.

References

1. M. Pigliucci, What, if Anything, Is an Evolutionary Novelty? *Philosophy of Science*. **75**, 887–898 (2012).

2. G. P. Wagner, V. J. Lynch, Evolutionary novelties. *Curr Biol*. **20**, R48–52 (2010).

3. O. Rieppel, Turtles as hopeful monsters. *BioEssays*. **23**, 987–991 (2001).

4. N. Shubin, C. Tabin, S. Carroll, Deep homology and the origins of evolutionary novelty. *Nature*. **457**, 818–823 (2009).

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

6. G. A. Wray, The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet*. **8**, 206–216 (2007).

7. E. H. Davidson, D. H. Erwin, Gene regulatory networks and the evolution of animal body plans. *Science*. **311**, 796–800 (2006).

8. H. Li, A. D. Johnson, Evolution of transcription networks - lessons from yeasts. *Curr Biol*. **20**, R746–53 (2010).

9. C. J. Nobile *et al.*, A Recently Evolved Transcriptional Network Controls Biofilm Development in Candida albicans. *Cell*. **148**, 126–138 (2012).

10. J. Bonhomme, C. D'enfert, Candida albicans biofilms: building a heterogeneous, drug-tolerant environment. *Curr Opin Microbiol*. **16**, 398–403 (2013).

11. E. M. Kojic, R. O. Darouiche, Candida Infections of Medical Devices. *Clinical Microbiology Reviews*. **17**, 255–267 (2004).

12. J. W. Taylor, M. L. Berbee, Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia*. **98**, 838–849 (2006).

13. S. Chu, I. Herskowitz, Gametogenesis in yeast is regulated by a transcriptional cascade dependent on Ndt80. *Mol Cell*. **1**, 685–696 (1998).

14. S. Chu *et al.*, The transcriptional program of sporulation in budding yeast. *Science*. **282**, 699–705 (1998).

15. S. R. Hepworth, H. Friesen, J. Segall, NDT80 and the meiotic recombination checkpoint regulate expression of middle sporulation-specific genes in Saccharomyces cerevisiae. *Molecular and Cellular Biology*. **18**, 5750–5761 (1998).

16. M. Pierce *et al.*, Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. *Molecular and Cellular Biology*. **23**, 4814–4825 (2003).

17. A. K. Vershon, M. Pierce, Transcriptional regulation of meiosis in yeast. *Curr Opin Cell Biol*. **12**, 334–339 (2000).

18. L. Xu, M. Ajimura, R. Padmore, C. Klein, N. Kleckner, NDT80, a meiosis-specific gene required for exit from pachytene in Saccharomyces cerevisiae. *Molecular and Cellular Biology*. **15**, 6572–6581 (1995).

19. A. Aharoni *et al.*, The "evolvability" of promiscuous protein functions. *Nat Genet*, 1–4 (2004).

20. J. T. Bridgham, S. M. Carroll, J. W. Thornton, Evolution of hormone-receptor complexity by molecular exploitation. *Science*. **312**, 97–101 (2006).

21. O. Khersonsky, D. S. Tawfik, Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem*. **79**, 471–505 (2010).

22. A. Wagner, Robustness, evolvability, and neutrality. *FEBS Letters*. **579**, 1772–1778 (2005).

23. S. M. Coyle, J. Flores, W. A. Lim, Exploitation of Latent Allostery Enables the Evolution of New Modes of MAP Kinase Regulation. *Cell*. **154**, 875–887 (2013).

24. L. Teytelman, D. M. Thurtle, J. Rine, A. Van Oudenaarden, Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci USA*. **110**, 18602–18607 (2013).

25. X. Fan, K. Struhl, Where Does Mediator Bind In Vivo? *PLoS ONE*. **4**, e5029–6 (2009).

26. K. P. Byrne, K. H. Wolfe, The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*. **15**, 1456–1461 (2005).

27. M. B. Lohse *et al.*, Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains. *Proc Natl Acad Sci USA*. **110**, 7660–7665 (2013).

28. E. R. Jolly, C.-S. Chin, I. Herskowitz, H. Li, Genome-wide identification of the regulatory targets of a transcription factor using biochemical characterization and computational genomic analysis. *BMC Bioinformatics*. **6**, 275 (2005).

29. A. Sellam *et al.*, Role of transcription factor CaNdt80p in cell separation, hyphal growth, and virulence in Candida albicans. *Eukaryotic Cell*. **9**, 634–644 (2010).

30. R. J. Bennett, A. D. Johnson, Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains. *The EMBO Journal*. **22**, 2505–2515 (2003).

31. G. Butler *et al.*, Evolution of pathogenicity and sexual reproduction in eight Candida

genomes. *Nature*. **459**, 657–662 (2009).

32. D. Andes *et al.*, Development and Characterization of an In Vivo Central Venous Catheter Candida albicans Biofilm Model. *Infection and Immunity*. **72**, 6023–6031 (2004).

33. C. J. Nobile, A. P. Mitchell, Regulation of cell-surface genes and biofilm formation by the C. albicans transcription factor Bcr1p. *Curr Biol*. **15**, 1150–1155 (2005).

34. R. Sopko, S. Raithatha, D. Stuart, Phosphorylation and maximal activity of Saccharomyces cerevisiae meiosis-specific transcription factor Ndt80 is dependent on Ime2. *Molecular and Cellular Biology*. **22**, 7024–7040 (2002).

35. N. Guttmann-Raviv, S. Martin, Y. Kassir, Ime2, a meiosis-specific kinase in yeast, is required for destabilization of its transcriptional activator, Ime1. *Molecular and Cellular Biology*. **22**, 2047–2056 (2002).

36. E. A. Hutchison, N. L. Glass, Meiotic Regulators Ndt80 and Ime2 Have Different Roles in Saccharomyces and Neurospora. *Genetics*. **185**, 1271–1282 (2010).

37. M. E. Katz, S. Cooper, Extreme Diversity in the Regulation of Ndt80-Like Transcription Factors in Fungi. *G3 (Bethesda)*. **5**, 2783–2792 (2015).

38. I. Acosta, D. Ontoso, P. A. San-Segundo, The budding yeast polo-like kinase Cdc5 regulates the Ndt80 branch of the meiotic recombination checkpoint pathway. *Mol Biol Cell*. **22**, 3478–3490 (2011).

39. A. Sourirajan, M. Lichten, Polo-like kinase Cdc5 drives exit from pachytene during budding yeast meiosis. *Genes Dev*. **22**, 2627–2632 (2008).

40. R. K. Sherwood, R. J. Bennett, Fungal meiosis and parasexual reproduction—lessons from pathogenic yeast. *Curr Opin Microbiol*. **12**, 599–607 (2009).

41. D. Villar *et al.*, Enhancer Evolution across 20 Mammalian Species. *Cell*. **160**, 554–566 (2015).

42. A. E. Tsong, B. B. Tuch, H. Li, A. D. Johnson, Evolution of alternative transcriptional circuits with identical logic. *Nature*. **443**, 415–420 (2006).

43. H. Lavoie *et al.*, Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol*. **8**, e1000329 (2010).

44. C. R. Baker, L. N. Booth, T. R. Sorrells, A. D. Johnson, Protein Modularity, Cooperative Binding, and Hybrid Regulatory States Underlie Transcriptional Network Diversification. *Cell*. **151**, 80–95 (2012).

45. A. Tanay, A. Regev, R. Shamir, Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc Natl Acad Sci USA*. **102**, 7203–7208 (2005).

46. C. I. Swanson, D. B. Schwimmer, S. Barolo, Rapid Evolutionary Rewiring of a Structurally Constrained Eye Enhancer. *Current Biology*. **21**, 1186–1196 (2011).

47. E. E. Hare, B. K. Peterson, V. N. Iyer, R. Meier, M. B. Eisen, Sepsid even-skipped Enhancers Are Functionally Conserved in Drosophila Despite Lack of Sequence Conservation. *PLoS Genet*. **4**, e1000106–13 (2008).

48. A. M. Moses *et al.*, Large-Scale Turnover of Functional Transcription Factor Binding Sites in Drosophila. *PLoS Comput Biol*. **2**, e130–13 (2006).

49. A. Barve, A. Wagner, A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*. **500**, 203–206 (2013).

50. P. J. O'Brien, D. Herschlag, Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105 (1999).

51. M. S. Longtine *et al.*, Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast*. **14**, 953–961 (1998).

52. A. D. Hernday, S. M. Noble, Q. M. Mitrovich, A. D. Johnson, Methods in Enzymology. **470**, 737–758 (2010).

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

54. J. M. Cregg *et al.*, Expression in the yeast Pichia pastoris. *Meth Enzymol*. **463**, 169–189 (2009).

55. T. R. Sorrells, L. N. Booth, B. B. Tuch, A. D. Johnson, Intersecting transcription networks constrain gene regulatory evolution. *Nature*. **523**, 361–365 (2015).

56. E. Mancera, A. M. Porman, C. A. Cuomo, R. J. Bennett, A. D. Johnson, Finding a Missing Gene: EFG1 Regulates Morphogenesis in Candida tropicalis. *G3 (Bethesda)*. **5**, 849–856 (2015).

57. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

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

59. P. V. Kharchenko, M. Y. Tolstorukov, P. J. Park, Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology*. **26**, 1351–1359 (2008).

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

61. T. L. Bailey *et al.*, MEME SUITE: tools for motif discovery and searching. *Nucleic Acids*

Research. **37**, W202–W208 (2009).

62. S. L. Maguire *et al.*, Comparative Genome Analysis and Gene Finding in Candida Species Using CGOB. *Mol Biol Evol*. **30**, 1281–1291 (2013).

63. P. A. Dumesic *et al.*, Product Binding Enforces the Genomic Specificity of a Yeast Polycomb Repressive Complex. *Cell*. **160**, 204–218 (2015).

64. M. B. Lohse *et al.*, Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains. *Proc Natl Acad Sci USA*. **110**, 7660–7665 (2013).

65. J. D. Thompson, D. G. Higgins, T. J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. **22**, 4673–4680 (1994).

66. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol Biol Evol*. **26**, 1641–1650 (2009).

Table 1

Name	Species	Description	Previous Name	Genotype
yIN198	*S. cerevisiae*	Wild-type (SK1 background)	NKY611[1]	ho::LYS2/ho::LYS2 lys2/lys2 ura3/ura3 leu2::hisG/leu2::hisG
yIN181	*S. cerevisiae*	Ndt80 deletion (SK1 background)	NKY2296[1]	ho::LYS2/ho::LYS2 lys2/lys2 ura3/ura3 leu2::hisG/leu2::hisG ndt80::LEU2/ndt80::LEU2
YTS89	*S. cerevisiae*	Wild-type (W303 background)	yTS89[2]	MAT**a** leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15
yIN209	*S. cerevisiae*	Ndt80-Myc (W303 background)		ho::LYS2/ho::LYS2 lys2/lys2 ura3/ura3 leu2::hisG/leu2::hisG Ndt80-Myc-Kan/Ndt80
yIN219	*S. cerevisiae*	Constitutive promoter-Ndt80-Myc (W303 background)		unknown background, pNdt80:pNdt80-pGal1-NAT, Ndt80-Myc-KAN
yIN065	*K. lactis*	Wild-type	yLB5[3]	ade1 trp1 leu2 metA1 uraA1 nej1Δ::LEU2 lysA1 trp1 leu2 metA1 uraA1 nej1Δ::LEU2
yIN071	*K. lactis*	Ndt80 deletion		ade1 trp1 leu2 metA1 uraA1 nej1Δ::LEU2 ndt80::Ura3 lysA1 trp1 leu2 metA1 uraA1 nej1Δ::LEU2 ndt80::Kan
yIN223	*K. lactis*	Constitutive promoter-Ndt80-Myc		pNdt80-Ndt80:pGal1-Ndt80-Myc-KAN-NAT ade1/ade1 trp1/trip1 leu2/leu2 metA1 uraA1 nej1Δ::LEU2 lysA1/lysA1
yIN091	*P.pastoris*	Wild-type haploid	JC304[4]	ade1 his4
yIN092	*P. pastoris*	Wild-type haploid	JC306[4]	arg4 ura3
yIN104	*P. pastoris*	Wild-type diploid		arg4/ARG4 ade1/ADE1 his4/HIS4 ura3/URA3
yIN110	*P. pastoris*	Ndt80 deletion		arg4/ARG4 ade1/ADE1 his4/HIS4 ura3/URA3 Ndt80::Hph/Ndt80::Hph
yIN183	*P. pastoris*	Constitutive promoter-Ndt80-Myc		arg4/ARG4 ade1/ADE1 his4/HIS4 ura3/URA3 pTEF1-truncated Ndt80-Myc/NDT80
yIN161	*P. stipitis*	Wild-type		Prototrophic
yIN194	*P. stipitis*	Ndt80A deletion		pTEF1-Ndt80A-Myc-NAT Randomly integrated
yIN185	*P. stipitis*	Ndt80B deletion		pTEF1-Ndt80B-Myc-NAT Randomly integrated

yIN194	*P. stipitis*	Constitutive promoter-Ndt80A-Myc		pTEF1-Ndt80A-Myc-NAT Randomly integrated
yIN185	*P. stipitis*	Constitutive promoter-Ndt80B-Myc		pTEF1-Ndt80B-Myc-NAT Randomly integrated
INy081	*C. lusitaniae*	Wild-type a-cell	RSY284[5]	ura3
INy082	*C. lusitaniae*	Wild-type alpha-cell	RSY411[5]	arg ade CycR (Cycloheximide Resistance)
yIN138	*C. lusitaniae*	Wild-type diploid		Ura3 +/- Arg4 +/- Ade2 +/- CycR +/-
yIN141	*C. lusitaniae*	Ndt80A deletion		Ura3 +/- Arg4 +/- Ade2 +/- CycR
yIN148	*C. lusitaniae*	Ndt80B deletion		Ura3 +/- Arg4 +/- ade2- CycR
yIN124	*C. albicans*	Wild-type	SN250[6]	ura3D-iro1D::imm^{434}/URA3-IRO1, his1D/his1D, arg4D/arg4D, leu2D/leu2D
TF178	*C. albicans*	Ndt80A deletion	TF178[7]	arg4D/arg4D, leu2D/leu2D, his1D/his1D, URA3/ura3D, IRO1/iro1D , Ndt80B:LEU2/Ndt80B:HIS1
TF095	*C. albicans*	Ndt80B deletion	TF095[7]	arg4D/arg4D, leu2D/leu2D, his1D/his1D, URA3/ura3D, IRO1/iro1D , Ndt80B:LEU2/Ndt80B:HIS1
yIN123	*C. albicans*	Ndt80B-Myc	CJN1748[7]	ura3D-iro1D::imm^{434}/URA3-IRO1, his1D/his1D, arg4D/arg4D, leu2:cdHIS1/leu2:cmLEU2, orf19.2119-Myc/Orf19.2119
yIN118	*C. albicans*	Constitutive promoter-Ndt80A-Myc		ura3D-iro1D::imm434/URA3-IRO1, his1D/his1D, arg4D/arg4D, leu2:cdHIS1/leu2:cmLEU2, orf19.513::orf19.513-Myc/orf19.513 pOrf19.513::pTDH3-Nat/pOrf19.513
yIN120	*C. albicans*	Constitutive promoter-Ndt80B-Myc		ura3D-iro1D::imm434/URA3-IRO1, his1D/his1D, arg4D/arg4D, leu2:cdHIS1/leu2:cmLEU2, orf19.2119::orf19.2119-Myc/orf19.2119 pOrf19.513::pTDH3-Nat/pOrf19.513

yIN218	*S. cerevisiae*	*P. pastoris* Ndt80-Myc expressed in *S. cerevisiae*		MATa leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 ura3:pGal1-PpNdt80-Myc-Hyg

1. Xu, L., Ajimura, M., Padmore, R., Klein, C. & Kleckner, N. NDT80, a meiosis-specific gene required for exit from pachytene in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **15,** 6572–6581 (1995).

2. Sorrells, T. R., Booth, L. N., Tuch, B. B. & Johnson, A. D. Intersecting transcription networks constrain gene regulatory evolution. *Nature* **523,** 361–365 (2015).

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

4. Lin Cereghino, G. P. *et al.* New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of Pichia pastoris. *Gene* **263,** 159–169 (2001).

5. Sherwood, R. K., Scaduto, C. M., Torres, S. E. & Bennett, R. J. Convergent evolution of a fused sexual cycle promotes the haploid lifestyle. *Nature* **506,** 387–390 (2014).

6. Noble, S. M. & Johnson, A. D. Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen Candida albicans. *Eukaryotic Cell* **4,** 298–309 (2005).

7. Nobile, C. J. *et al.* A Recently Evolved Transcriptional Network Controls Biofilm Development in Candida albicans. *Cell* **148,** 126–138 (2012).

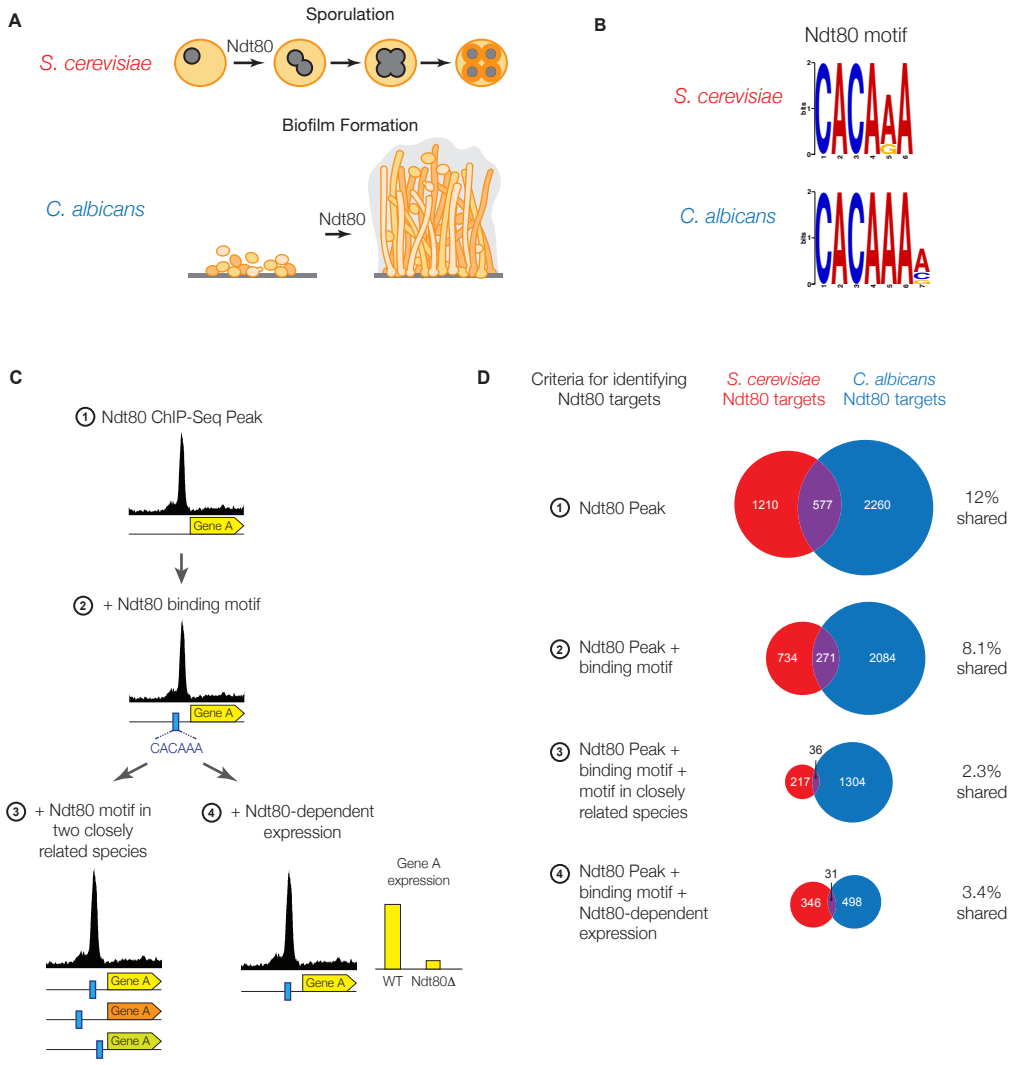
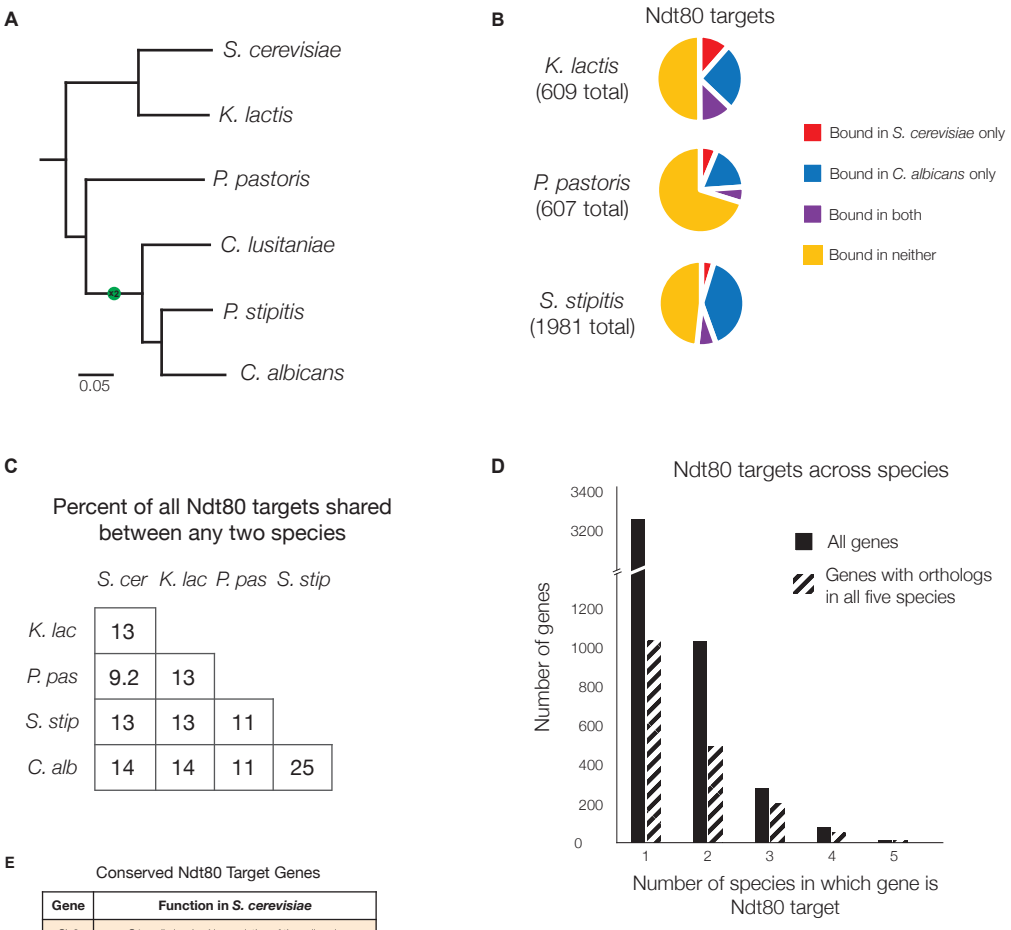


Figure 1: Ndt80 – target gene connections differ between S. cerevisiae and C. albicans. (A) Diagram of Ndt80 phenotypes in the regulation of sporulation in S. cerevisiae and biofilm formation in C. albicans. (B) The cis-regulatory motif most highly enriched at locations of Ndt80 ChIP binding in S. cerevisiae and C. albicans. Motifs were generated independently for each species. The S. cerevisiae Ndt80 motif in this study closely matches that determined previously for Ndt80 (references). (C) Diagram of the four criteria used to identify Ndt80 regulatory targets. Criteria1: ChIP-Seq enrichment in the intergenic upstream of a gene relative to untagged control experiments. Criteria 2: ChIP-Seq enrichment and the presence of anNdt80 motif in the intergenic. Criteria 3: ChIP-Seq enrichment with the Ndt80 motif present in the intergenic region and also in orthologous intergenic regions of at least two very closely related species (indicates the motif has been maintained by selection). Criteria 4: ChIP-Seq enrichment with the Ndt80 motif present in the intergenic region and Ndt80-dependent expression of the nearby gene (indicates expression of the gene is in fact under Ndt80 control). (D) Overlap in targets of S. cerevisiae Ndt80 (red) and C. albicans Ndt80B (blue), using different criteria of identifying targets. Venn diagrams are roughly area-proportional.



E

Conserved Ndt80 Target Genes

Gene	Function in *S. cerevisiae*
Cln2	G1 cyclin involved in regulation of the cell cycle
Clg1	Cyclin-like protein
Yox1	Transcriptional repressor; binds promoters of cell cycle-regulated genes
Rki	Ribose-5-phosphate ketol-isomerase
Gln1	Glutamine synthetase
Sga1	Intracellular sporulation-specific glucoamylase
Gdh2	NAD(+)-dependent glutamate dehydrogenase
Rps7a	Protein component of the small ribosomal subunit
Tma10	Protein of unknown function that associates with ribosomes
Usv1	Putative transcription factor; affects regulation of genes involved in cell wall biosynthesis

Figure 2: Ndt80 targets differ across S. cerevisiae, K. lactis, P. pastoris, and P. stipitis
(A) Phylogenetic tree of species investigated, with the scale representing the number of nucleotide substitutions per site. The likely timing of the Ndt80 gene duplication is indicated with an arrow. (B) For each species, Ndt80 targets were identified by Criteria 2 (Fig. 1A) and categorized according to whether they are shared with Ndt80 targets in S. cerevisiae, C. albicans, or both. (C) Percent of all Ndt80-bound genes shared between any two species tested using Criteria 2; for this analysis, only 1:1 orthologs in each two-species comparison were considered. (D) Histogram of all Ndt80 targets (using Criteria 2) in the five species tested (S. cerevisiae, K. lactis, P. pastoris, P. stipitis, and C. albicans) according to the number of species in which that gene is a target. All genes (black) and genes with 1:1 orthologs across all five species (dashed) are shown. (E) List of genes bound by Ndt80 in S. cerevisiae, K. lactis, P. pastoris, P. stipitis, and C. albicans with annotations from S. cerevisiae (SGD).

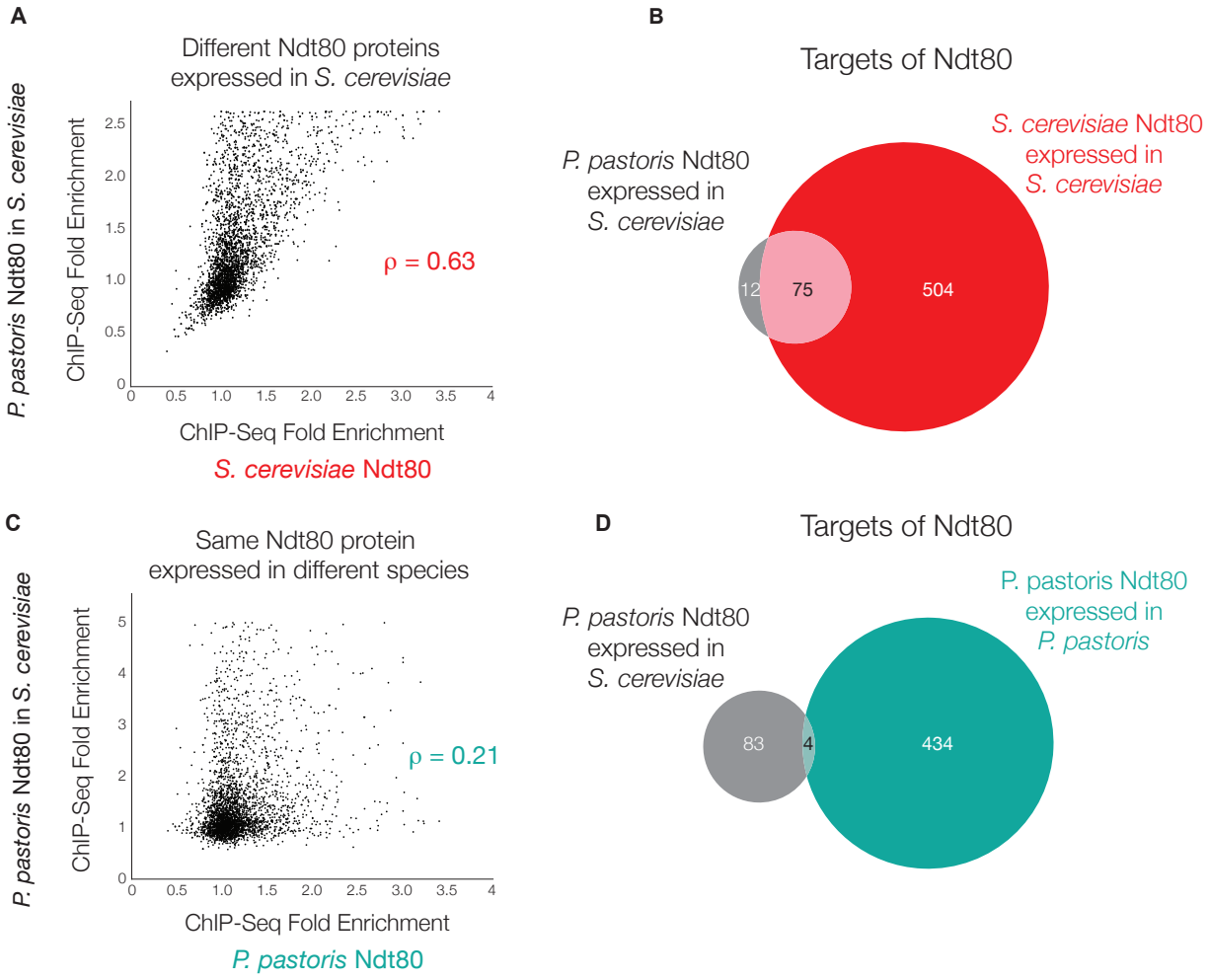


Figure 3: P. pastoris Ndt80 binds to S. cerevisiae Ndt80 targets when ectopically expressed in S. cerevisiae

(A and C) ChIP-Seq fold enrichment for all genes with 1:1 orthologs in S. cerevisiae and P. pastoris with correlation coefficient from Spearman's rank test. (A) S. cerevisiae Ndt80 vs. P. pastoris Ndt80 expressed in S. cerevisiae. (C) P. pastoris Ndt80 vs. P. pastoris Ndt80 expressed in S. cerevisiae. (B and D) Ndt80 targets, defined by ChIP-Seq enrichment alone (Criteria 1), for P. pastoris Ndt80 when expressed in S. cerevisiae, compared to targets of Ndt80 in S. cerevisiae (C) and P. pastoris (D). The genes that are Ndt80 targets in both S. cerevisiae and P. pastoris were omitted from analysis, as they are not informative in determining Ndt80 binding specificity.

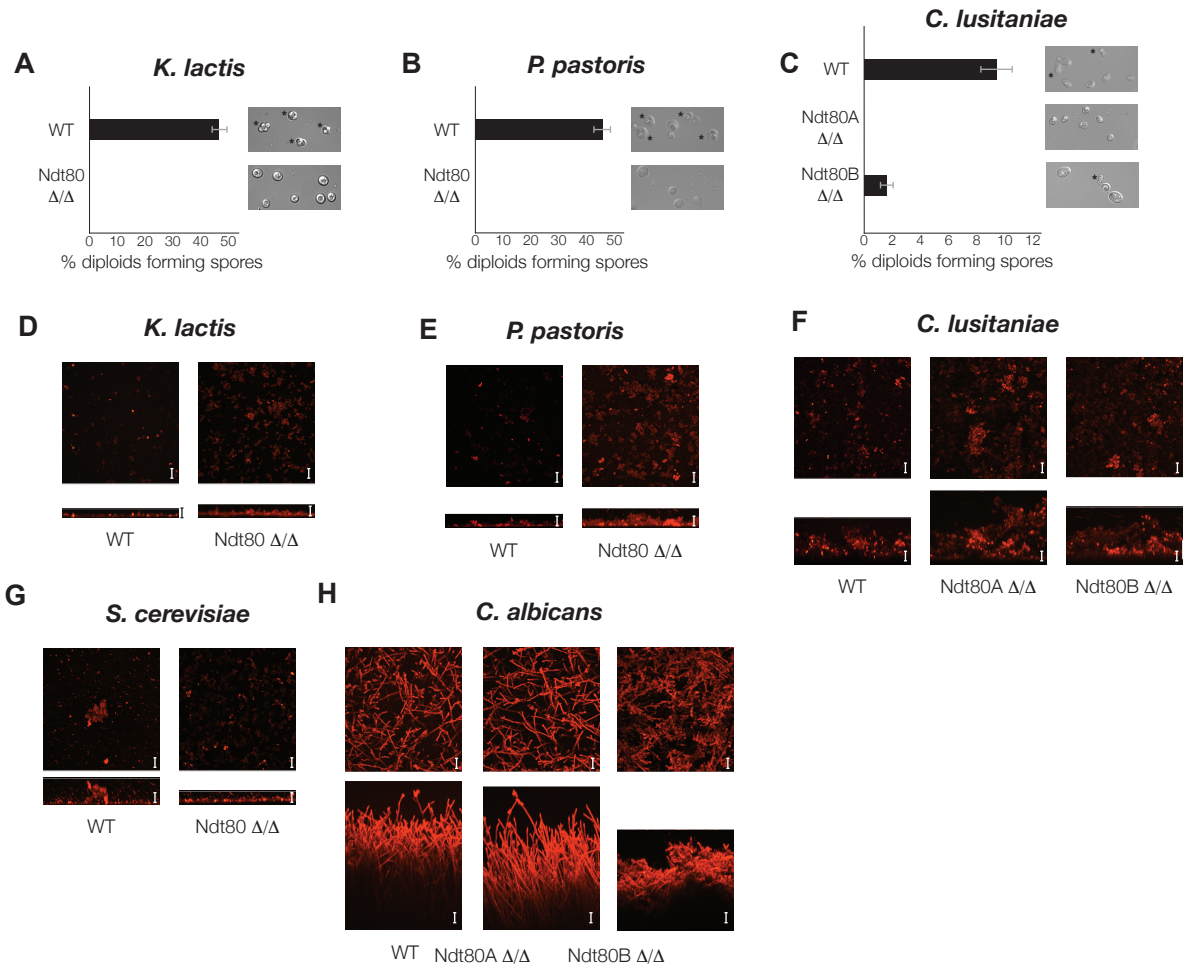


Figure 4: Ndt80 required for sporulation, dispensible for biofilm formation, in K. lactis, P. pastoris, and C. lusitaniae

(A-C) Light microscope images of genetically matched wild-type and Ndt80 deletion strains (Stars indicate diploids that have undergone sporulation) and quantification of the percent of cells exhibiting spores, as measured by microscopy (200 cells counted for each strain). (D-H) Confocal scanning laser microscopy images of biofilm formation for genetically matched wild-type and Ndt80 deletion strains. Top view of biofilm shown above side view for each, with scale bar representing 25 um.

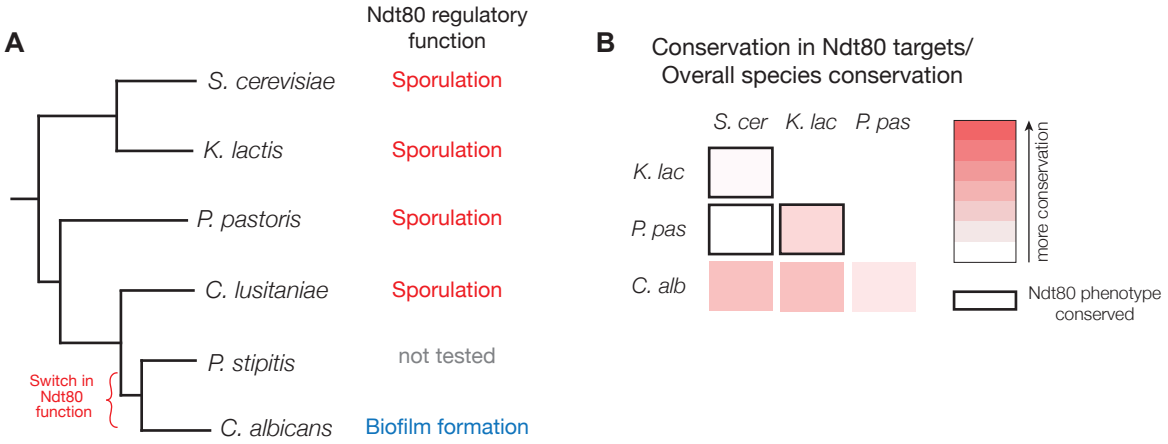


Figure 5: The targets of Ndt80 differ among species with conserved Ndt80 phenotype
(A) Phylogenetic tree of species investigated, with the Ndt80 deletion phenotype indicated. The red bracket indicates the switch in Ndt80 function from sporulation to biofilm regulation. (B) Conservation in Ndt80 targets normalized to overall conservation of genomic DNA, as measured by the nucleotide substitution rate (see Supplemental Materials and Methods). Species comparisons between two species with a conserved Ndt80 phenotype shown with black outline.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

8/11/16

Date