Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit

Candace S. Britton1,2, Trevor R. Sorrells1,2, Alexander D. Johnson1

Changes in both the coding sequence of transcriptional regulators and in the cis-regulatory sequences recognized by these regulators have been implicated in the evolution of transcriptional circuits. However, little is known about how they evolved in concert. We describe an evolutionary pathway in fungi where a new transcriptional circuit (α-specific gene repression by the homeodomain protein Matα2) evolved by coding changes in this ancient regulator, followed millions of years later by cis-regulatory sequence changes in the genes of its future regulon. By analyzing a group of species that has acquired the coding changes but not the cis-regulatory sites, we show that the coding changes became necessary for the regulator’s deeply conserved function, thereby poised the regulator to jump-start formation of the new circuit.

Protein-coding changes over evolutionary time are an important source of organismal novelty. Such circuits are typically composed of one or more transcriptional regulators (sequence-specific DNA binding proteins) and their direct target genes, which contain cis-regulatory sequences recognized by the regulators. Although changes in cis-regulatory sequences are often stressed as sources of novelty that avoid extensive pleiotropy, it is clear that coding changes in the transcriptional regulatory proteins are also of key importance (1–6). Some well-documented changes in transcriptional circuitry require concerted changes in both elements (7, 8). Although such concerted changes are likely to be widespread, we know little about how they occur.

In this work, we study a case in the fungal lineage where gains in cis-regulatory sequences and coding changes in the transcriptional regulator were both required for a new circuit to have evolved. Specifically, we addressed which came first: the changes in the regulatory protein or the changes in the cis-regulatory sequences of its 5 to 10 target genes. The system we analyzed consists of an ancient regulator, the homeodomain protein Matα2, and the changes—both in the protein itself and in the regulatory regions of the genes it controls—that occurred across the Saccharomycotina clade of fungi, which spans roughly 300 million years. (In terms of protein diversity, this represents roughly the range between humans and sea sponges (9)). Throughout this time, Matα2 has maintained its ancient function: It binds cooperatively to DNA with a second homeodomain protein, Mata1, to repress a group of genes called the haploid-specific genes (Fig. 1). More recently, Matα2 formed an additional circuit, which is present in only a subset of the Saccharomycotina: It binds DNA cooperatively with the MADS box protein Mcm1 to repress the α-specific genes (Fig. 1). Before this time, the α-specific genes were

![Diagram of protein-coding changes preceded cis-regulatory gains](image-url)

**Fig. 1.** Cell type–specific gene expression in the Saccharomycotina yeast. **(A)** Across the Saccharomycotina clade, α and α cells each express a set of genes specific to that cell type (α- and α-specific genes, or asgs and αsbs, respectively), as well as a shared set of haploid-specific genes (hsbs). α and α cells can mate to form α/α cells, which do not express the α-, α-, or haploid-specific genes (22). Wavy arrows represent active transcription. **(B)** The mechanism underlying the expression of α-specific genes is different among species. In the last common ancestor of the Saccharomycotina yeast (see circled A in the figure), transcription of the α-specific genes was activated by Matα2, a protein produced only in cells, which binds directly to the regulatory region of each α-specific gene (10, 23). Much later in evolutionary time (see circled E in the figure), repression of the α-specific genes by direct binding by Matα2 evolved. Still later, the Matα2-positive form of control was lost in some species (including *S. cerevisiae*), leaving only the Matα2-negative form. mya, million years ago.
is required for a cells to mate (see supplementary text for details). (D) mRNA sequencing (mRNA-seq) (tpm, transcripts per million) of wild-type W. anomalus a cells (MATa) compared with a cells with MATa deleted (MATa mata2-Δ), a-specific genes STE2, AXL1, ASG7, BARI, STE6, and MATa are shown in green. Expression of MATa and the marker used to delete it (Nat) are shown in pink and opaque black, respectively. Data from independent replicates are given in fig. S3. (E) a-specific gene expression levels in a wild-type W. anomalus a cells (MATa) compared with a cells with MATa deleted (MATa mata2-Δ), measured by the NanoString nCounter system (24). For comparison, expression levels of the a-specific gene STE3 and the haploid-specific gene STE4 are also given. Means and SDs of two cultures per genotype, grown and tested in parallel, are shown.

regulated by a different mechanism—positive control by the HMG-domain protein Mata2 (10, 11).

The switch between the two mechanisms of controlling the a-specific genes occurred sometime before the divergence of Saccharomyces cerevisiae and Kluyveromyces lactis (formally known as the Saccharomyctaceae, here called the S. cerevisiae clade) but after the divergence of this clade and that containing Candida albicans and Pichia membrifacia (formally known as the Pichiaeae and Debaryomycetae, here called the C. albicans clade) (Fig. 1B). Three events must have occurred for the newer (represion) scheme to have evolved: (i) Mata2 acquired the ability to contact the Tup1-Ssn6 co-repressor, bringing it to DNA to carry out the repression function; (ii) Mata2 acquired the ability to bind to DNA cooperatively (through a direct protein-protein contact) with Mcm1; and (iii) the a-specific genes (numbering between 5 and 10, depending on the species) each acquired a new cis-regulatory site for the Mata2-Mcm1 combination (Fig. 1B).

To determine the order of these events, we studied Mata2 and the regulation of the a-specific genes in a clade that branched from the ancestor before the occurrences of all three of these events. We reasoned that this group of species might have acquired some, but not all, of the changes needed to form the new circuit, and it therefore might provide clues to the evolutionary history. This approach was made possible by the genome sequencing of a monophyletic group of species that branches before the last common ancestor of the S. cerevisiae clade (formally known as the Phaffomyctaceae) (Fig. 1B) (12, 13). We chose the species Wickerhamomyces anomalus, and we were able to optimize relatively simple procedures to alter it genetically (14).

We examined the W. anomalus Mata2 protein sequence to determine whether it is more similar to the ancestral (represented by C. albicans) or the derived (represented by S. cerevisiae) form of Mata2. Alignment of the Mata2 coding sequences across many species indicated that, of the five functional regions described for the S. cerevisiae protein (Fig. 2A and fig. S1), the W. anomalus protein shares all of them. In particular, it has a similar Tup1-interacting region (region 1, Fig. 2A) and Mcm1-interacting region (region 3, Fig. 2A); these regions are missing in outgroup proteins and are needed to repress the a-specific genes in S. cerevisiae (11, 15). By swapping these W. anomalus regions into the S. cerevisiae protein, we confirmed that they are functional in repressing the a-specific genes (Fig. 2B). In the course of these experiments, we found that the homeodomain of the W. anomalus protein contained mutations that prevented its binding to the a-specific gene cis-regulatory sequence in S. cerevisiae, a derived change within this clade alone (Fig. 2B and fig. S1). Similar results were obtained with the Mata2 protein from two additional species that branch with W. anomalus, indicating that these two conclusions—that W. anomalus clade Mata2 bears functional protein-protein interactions but cannot bind
the \textit{S. cerevisiae} a-specific genes—are characteristic of the \textit{W. anomalus} clade rather than of a single species (fig. S1D).

The observation that the \textit{W. anomalus} Mata2 protein acquired the necessary coding changes to interact with Tup1 and Mcm1 but could not bind to the \textit{S. cerevisiae} a-specific gene control region raised the question of whether it has any role in regulating the a-specific genes in \textit{W. anomalus}. A series of otherwise-isogenic strains was constructed with Mata2 (and Mata2) deleted, and the results show that, in this species, Mata2 does not regulate the a-specific genes; they are instead regulated by Mata2 (Fig. 2, C to E, and fig. S8). Thus, despite the changes in Mata2, \textit{W. anomalus} retains the ancestral form of a-specific gene regulation and activation by Mata2. This conclusion is supported by a bioinformatic analysis showing that the a-specific genes possess Mata2-Mcm1, but not Mata2-Mcm1 cis-regulatory sequences (fig. S4B). These results argue against the possibility that direct, a-specific gene repression by Mata2 existed in an ancestor of \textit{W. anomalus} but was subsequently lost, as this would have required the independent loss of Mata2 binding sites from all of the a-specific genes across numerous species.

Our experiments up to this point demonstrate that Mata2 had acquired the coding changes needed to repress the a-specific genes millions of years before its cis-regulatory sequences appeared in the a-specific genes. We next addressed how these changes in the Mata2 protein could have been maintained in the absence of their usefulness in repressing the a-specific genes. One hypothesis focuses on Mata2’s ancient function—repressing the haploid-specific genes with Mata1—and holds that the Mata2 coding changes became required for this function only in the \textit{W. anomalus} clade. To test this idea, we analyzed the requirements for haploid-specific gene repression in \textit{W. anomalus}. We deleted \textit{MATa}2 and \textit{MATa}1 in a/a cells and found that they are both necessary for haploid-specific gene repression, a conclusion confirmed by chromatin immunoprecipitation (Fig. 3A and figs. S5 and S6C). However, unlike in species outside the \textit{W. anomalus} clade, the Tup1-interaction region and the Mcm1-interaction region of Mata2 are necessary for repression of the haploid-specific genes within the clade (Fig. 2A and fig. S6B). Finally, an Mcm1 cis-regulatory site is also required for the repression of the \textit{W. anomalus} haploid-specific gene \textit{RME1} (Fig. 3C and fig. S6). Taken together, these experiments show that Mata2, Mata1, and Mcm1 are all required for haploid-specific gene repression in \textit{W. anomalus}, and that the portions of Mata2 that interact with Mcm1 and Tup1 are
also required. This three-part recognition of the haploid-specific genes in the W. anomalus clade was not anticipated from studies of other species. Even in the S. cerevisiae clade, where Mcm1 and Mat2 are known to interact, this interaction is not required for haploid-specific gene repression (11). These results explain the observation that the key changes in Mato2 needed for the new a-specific gene circuit were already in place in the last common ancestor of S. cerevisiae and W. anomalus, long before the circuit came into play (Fig. 4). An alternative scenario—in which the Mato2 protein gained the Mcm1-interaction region twice, once in the S. cerevisiae clade and once in the W. anomalus clade—is unlikely because the same seven amino acids would have had to be gained in exactly the same position in the protein (fig. S1).

This study helps to illuminate several long-standing issues. First, how is pleiotropy avoided when transcriptional regulators acquire new functions? The modular structure of Mato2 is evident from the protein domain swap experiments (Fig. 2B and fig. S6B), showing that the derived regions of the protein (Tup1- and Mcm1-interaction regions) can be transplanted to a variety of outgroup Mato2 proteins and that they endow the ancestral proteins with the new functions without compromising the existing functions (11). However, there is a second, more subtle way that extensive pleiotropy was avoided in the case studied in this work. In the shift between the different ways of controlling the haploid-specific genes, pleiotropy was avoided automatically; even before the new a-specific gene circuit was formed, the Mato2-Mcm1 combination (which forms the basis of the new circuit) had been “vetted” for millions of years as being compatible with the ancestral function of Mato2.

Second, is the evolutionary pathway we describe in this paper compatible with the concept of constructive neutral evolution, or the idea that new functions can evolve through evolutionary transitions of approximately equal fitness (16–18)? Before the results presented here were obtained, it was difficult to understand how the derived circuit represented by S. cerevisiae (repression of the a-specific genes by Mato2 in α cells) could have evolved because it required changes in both the Mato2 coding region and in the cis-regulatory sequences controlling the 5 to 10 a-specific genes. We propose that the prior changes to Mato2 represent an example of constructive neutral evolution, in the sense that the neutral sampling of different ways to repress the haploid-specific genes over evolutionary time led to changes in Mato2 that, millions of years later through exaptation, formed the basis of the new circuit. Although we cannot rule out the possibility that the differences in the way that the haploid-specific genes were repressed were somehow adaptive, it seems more likely that they occurred neutrally—an explanation consistent with a wide variety of theoretical work (16–19). In any case, there is no obvious adaptive explanation, and neutral evolution is an appropriate default hypothesis.

Third, is there an inherent logic to the mechanisms underlying a given transcription circuit? In this paper, we show that some clades regulate the haploid-specific genes with a combination of three proteins, whereas others use only two of the proteins, even though the third is present. Nonetheless, the overall pattern of haploid-specific gene expression is the same. If there is any overriding design logic to the different mechanisms of regulating these genes, it is difficult to discern (20). More broadly, the work presented here illustrates that a given transcription circuit is best understood as one of several possible interchangeable, mechanistic solutions rather than as a finished, optimized design (21).
Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit

Candace S. Britton, Trevor R. Sorrells and Alexander D. Johnson

Science 367 (6473), 96-100.
DOI: 10.1126/science.aax5217

Generating a new transcriptional network

Organismal novelties result from changes in transcriptional circuits. But what comes first, changes in regulatory protein or changes in cis-regulatory sequences? Britton et al. examine the Matα2 protein in a Saccharomycotina clade of fungi. They show that a newly evolved transcription circuit involving repression of the a-specific genes by the ancient homeodomain protein Matα2 occurred in two stages separated by millions of years. In the first stage, Matα2 acquired several coding changes followed by changes in cis-regulatory sequences. This clade-specific requirement explains how the coding changes of Matα2 were in place long before the new a-specific gene repression circuit arose.

Science, this issue p. 96

ARTICLE TOOLS
http://science.sciencemag.org/content/367/6473/96

SUPPLEMENTARY MATERIALS
http://science.sciencemag.org/content/suppl/2019/12/30/367.6473.96.DC1

REFERENCES
This article cites 44 articles, 11 of which you can access for free
http://science.sciencemag.org/content/367/6473/96#BIBL

PERMISSIONS
http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title Science is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works