Title
Evolution of a combinatorial transcriptional circuit: A case study in yeasts

Author
Tsong, Annie E

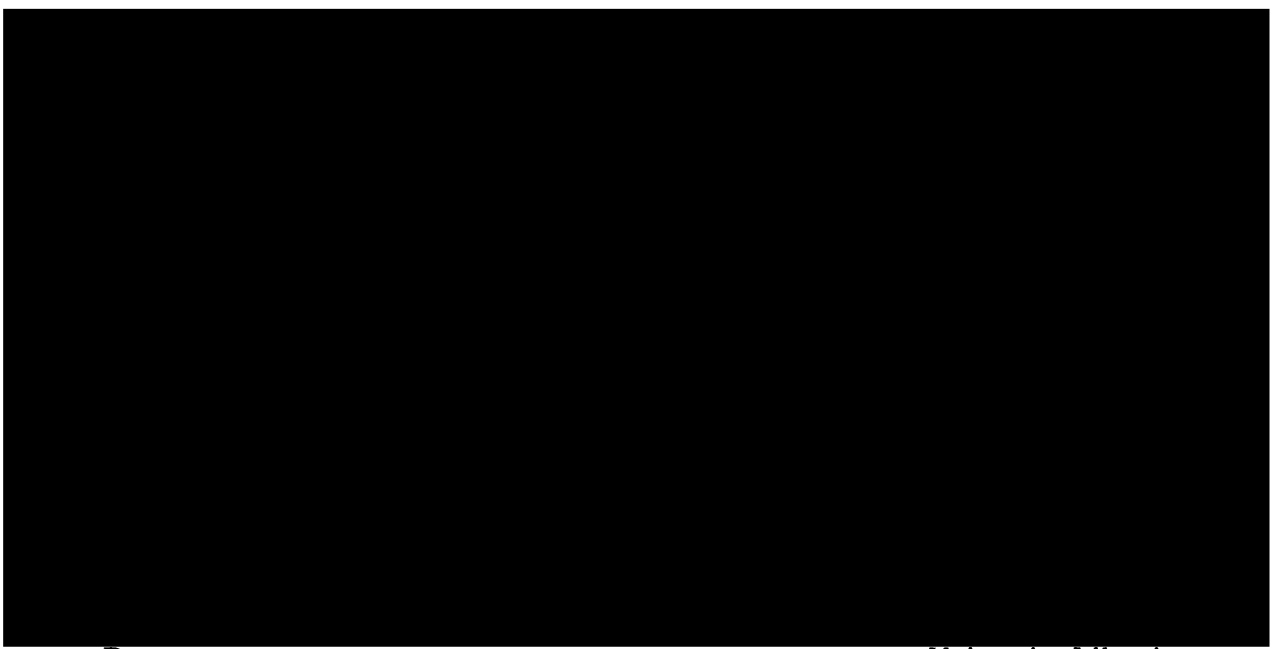
Publication Date
2005

Peer reviewed|Thesis/dissertation

Evolution of a combinatorial transcriptional circuit:
a case study in yeasts

Annie E. Tsong

DISSERTATION
Submitted in partial satisfaction of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
GENETICS
in the
GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA SAN FRANCISCO



Date University Librarian

Degree conferred:

This thesis is dedicated to my parents, Tian Yow Tsong and Liyun Wu Tsong. My father, Tian Yow Tsong, has always taken great pride in the fact that I enjoyed categorizing animals and plants as a toddler. Because scientists love to classify things, he took this as an encouraging sign that I would follow in his footsteps as a scientist. It has been a joy to be able to communicate my work to my father, and as I have gained knowledge, to be able to understand just a little bit of what he does. My father's scientific work in multiple fields in biophysics inspires me in more ways than I can count. Original, groundbreaking, iconoclast, and ahead-of-its-time are just a few adjectives that come to mind. I have learned from my father not to settle for "cookbook science," and I hope this thesis reflects my father's influence on me. My mother, Liyun Tsong, has kept me going through grad school, which has been quite a feat. Although this thesis presents a tidy story, the reality was messy, filled with dead-ends and seemingly wasted years gone down the drain from chasing red herrings. Many times I have gone home in distress, but have come back to San Francisco refreshed and ready to work because visiting my mother let me regain my footing and strength. My mother, a piano teacher, has also given me the gift of music. Sitting in front of the familiar keys of a piano is a trick that always transports me to a safe place. I would like to dedicate this thesis to my parents to thank them for all of their support over these many years.

Acknowledgments

Like many pursuits, success in science requires a unique combination of timing, hard work, thoroughness, and pure luck. In this case, it also required a huge team of enthusiastic and collaborative labmates and colleagues, and an unfailingly supportive thesis advisor who allowed me to pursue whatever questions crossed my mind.

It is no secret to my family and friends that I was prepared to quit graduate school several years into my degree. At a time when most thesis advisors would have kicked a student out of the graduate program, Sandy patiently waited for my mind to "come back" to science. I am grateful to have been in an environment where at the end of my graduate career, I feel even more excited about science than when I started, even now in the throes of thesis-writing. I am certain that I owe Sandy my career.

I also owe my thesis committee, Hiten Madhani and Dave Morgan, a great deal. Their enthusiasm and joy in a good scientific story made me feel that my work was worthwhile, even when I myself was having doubts, and their thoughts and advice always made me want to grab a pipetteman and a petri dish and start testing some new ideas.

Two faculty members outside of UCSF, Ken Wolfe and Stefan Ästrøm, must be acknowledged for their enormous but largely unwritten role in this thesis. I have never actually met either of them, and we have exchanged only a few e-mails, but their work on evolution of the mating-type locus, and their willingness to share unpublished results, strains, and ideas has been absolutely critical to my project.

I must acknowledge many people in the Johnson Lab. First and foremost, I would like to acknowledge Ann Cassidy-Stone. If I owe Sandy my career, I can say honestly that I owe Ann the happiness in my life. I owe an absolutely enormous THANK YOU to

Ryan Raisner. Not only do I owe Ryan the project that is the subject of this thesis, but Ryan has also been the truest and most loyal of friends over these many years. Brian Tuch has been an absolutely outstanding partner in research. He was able to take my inchoate, vague ideas and pitiable excuses of informatics and use them as a launching point for developing a credible and fascinating view of transcriptional evolution.

Andrew Uhl's knowledge of the field of pathogenesis is staggering, and his generosity inside and outside the lab makes me have more faith in people. Joanne Penko was a wonderful, irreverent, hilarious, and understanding baymate for many years, and has continued to be a very close friend. She is really going to make a difference in this world, to a lot of people. Matt Miller was an absolute inspiration in the lab, not to mention a great classmate. His brand of science is something that I hope to emulate one day. Suzanne Noble has been like a big sister to me. Her advice on all things science and otherwise has been enormously valuable and comforting to me over the years, and I look forward to receiving it for years to come. Richard Bennett has a heart of gold, and has been an amazing friend and labmate for years. Sudarsi Desta has made this thesis possible with her good humor, her gentle cheer, and literally tons of top-notch media and solutions. From Bethann Hromatka's hands sprang a thousand microarrays, of which I used hundreds. It was such a pleasure to work with Bethann, a model of efficiency, thoughtfulness, and intelligence. Quinn Mitrovich provided a show-tunes soundtrack to life in the lab, not to mention many question during group meeting that cut to the heart of my project. Diane Inglis spearheaded the microarray effort in the Johnson lab, and her can-do, no nonsense enthusiasm for the project made my own project possible. Matt Lohse and Becky Zordon provided a great deal of valuable feedback on this project

during our numerous subgroup meetings. Matt's absurdist humor and Becky's willingness to tackle difficult problems have added much to the lab. Christina Hull is really the mother of the field of *C. albicans* mating. Although we only overlapped in the Johnson Lab for a few months, her influence on my science has been incalculable. Many other members of the Johnson lab have contributed to my happy times, both scientifically and otherwise, in the lab. Sarah Elson, Josie Haduong, Brad Green, Shannon Stroschein, David Kadosh, Sarah Green, Hien Tran, David Galgoczy, Lisa Kohn, Burk Braun, and Oliver Homann are all part of a fantastic tradition of great science and great times, and supplying the lab with some much-needed indie cred (Thanks Brad!).

Though not in the Johnson lab, Corey Adams has been a major influence on my work and my life. Not many people study molecular evolution at UCSF right now, so having that scientific kinship was a vital part of my graduate work. He has been a great friend in so many ways. There are many more people I would like to acknowledge, but I have to start this thesis somewhere. So here we go.

Chapter 2 is reprinted with minor changes from Cell, 115, Annie E. Tsong, Mathew G. Miller, Ryan M. Raisner, and Alexander D. Johnson, *Evolution of a Combinatorial Transcriptional Circuit: A Case Study in Yeasts*, 389-399, Copyright 2003, with permission from Elsevier.

From: H.Nixon@elsevier.com
Subject: RE: Obtain Permission
Date: December 9, 2005 4:16:44 AM PST
To: atsong@itsa.ucsf.edu

We hereby grant you permission to reprint the material mentioned below at no charge in your thesis subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgment to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier".

3. Reproduction of this material is confined to the purpose for which permission is hereby given.

4. This permission is granted for non-exclusive world English rights only. For other languages please reapply separately for each one required. Permission excludes use in an electronic form. Should you have a specific electronic project in mind please reapply for permission.

5. This includes permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

Yours sincerely
Harry Nixon
Rights Assistant

Global Rights Department
Phone - 0044 (0) 1865 843517
Fax - 0044 (0) 1865 853333
Email - H.Nixon@Elsevier.com

-----Original Message-----
From: atsong@itsa.ucsf.edu [mailto:atsong@itsa.ucsf.edu]
Sent: 08 December 2005 20:19
To: permissions@elsevier.com
Subject: Obtain Permission

This Email was sent from the Elsevier Corporate Web Site and is related to Obtain Permission form:

Product: Customer Support
Component: Obtain Permission
Web server: http://www.elsevier.com
IP address: 10.10.24.149
Client: Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/416.12 (KHTML, like Gecko) Safari/416.13
Invoked from:
http://www.elsevier.com/wps/find/obtainpermissionform.cws_home?isSubmitted=yes&navigateXmlFileName=/store/prod_webcache_act/framework_support/obtainpermission.xml
Request From:
Student Annie E. Tsong
UCSF
600 16th Street 2200
94143-2200
San Francisco
United States
Contact Details:
Telephone: (415) 502-0859
Fax: (415) 502-4315
Email Address: atsong@itsa.ucsf.edu
To use the following material:
ISSN/ISBN:
Title: Cell
Author(s): Tsong AE, Miller MG, Raisner RM, Johnson AD
Volume: 115
Issue: November 14, 2003
Year: 2003
Pages: 389 - 399
Article title: Evolution of a Transcriptional Circuit
How much of the requested material is to be used:
The entire article will be used
Are you the author: Yes
Author at institute: Yes
How/where will the requested material be used: [how_used]

Details:
I would like permission to include in my dissertation a copy of the paper cited, which was published in Cell. The dissertation will be microfiled by University Microfilms Incorporated, and they request permission to provide single copies on demand. Thank you.

Abstract

Developing new regulation of existing genes is a major source of evolutionary novelty. In this thesis, I examine evolution of transcriptional regulation by dissecting a combinatorial regulatory circuit that governs mating-type in the ascomycete branch of the fungal lineage. I first determine the means by which four conserved transcription factors regulate mating-type in the pathogenic yeast *C. albicans*. I then closely compare the *C. albicans* circuit to that of *S. cerevisiae*, identifying several classes of changes that have arisen since their divergence from a common ancestor. Among the many changes identified, I focus on a group of orthologs, the **a**-specific genes, that is positively regulated in *C. albicans*, but is negatively regulated in *S. cerevisiae*. I demonstrate that positive regulation represents the ancestral form, and that the *S. cerevisiae* mode of negative regulation is a recently derived innovation. By examining the regulation of asgs in a group of 16 modern yeasts that diverged at successively later times from a common ancestor, I deduce specific, sequential changes in both *cis*- and *trans*-regulatory elements that constitute the transition from positive to negative regulation.

TABLE OF CONTENTS

List of tables

Chapter 2:

Chapter 3:

Chapter 4:

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

Evolution

Life on earth is astonishing in its variety of form, function, and habitat. Over one million species have been identified thus far, and it is estimated that as many as 10 to 20 million more have yet to be discovered (Futuyma, 1997). Only one condition - the presence of liquid water - appears to consistently constrain life's appearance on our planet. Aside from this requirement, unique adaptations allow organisms to thrive in vastly disparate niches which range the extremes of temperature, atmospheric pressure, osmolarity, and radiation. Single-celled extremophiles such as the archae *Pyrolobus fumarii*, for example, are capable of growing at temperatures up to 113°C, while at the opposite temperature extreme, representatives of all taxa thrive in the Antarctic ocean which is perpetually at or below freezing (Rothschild and Mancinelli, 2001). Despite this massive diversity of life, however, all organisms share a common origin.

Charles Darwin's theory of evolution by natural selection explains how such diverse modern species emerged from a common ancestor (Darwin, 1859). His theory of evolution includes several fundamental concepts:

- Organisms generate offspring that resemble themselves.
- Organisms exist as part of a larger population, within which there is some variation in characteristics.
- Variations in characteristics allow individual organisms to reproduce at different rates, changing the proportion of characteristics within a population over time.
- Whether a specific characteristic enhances reproduction depends on the interplay of environmental pressures with an organism. Over time, this selection

results in adaptation of a population to a particular environment. Isolation of subgroups within the larger population can result in generation of new species.

When Charles Darwin began formulating his theory of evolution in the 1830s, he based his work largely on fossil records, pigeon and livestock breeding, and observations of closely-related finches in the Galapagos islands. Considering the diversity of life, the vague contemporary understanding of heredity, and the black box which comprised variation at the time, it is all the more extraordinary that Darwin was sufficiently unrelenting in his logic to extrapolate his theory to all observed life on earth.

In the years since Darwin formulated his theory, the twin concepts of common ancestry and of evolution by natural selection have been validated thousands of times over. The discovery of the principles of heredity revealed the basis for similarity between parents and offspring and accounted for the distribution of variation in a population. The decoding of DNA, unimaginable in Darwin's time, has confirmed his theory. It has uncovered the physical basis of heredity, revealed the sources of variation in population, provided nucleotide-by-nucleotide evidence for random mutation and environmental selection, and revealed astounding underlying similarities between all organisms, for all their exquisite specialization.

Variation is the raw material honed by natural selection to generate adaptive changes to an environment. The work in this thesis concerns a mechanism which underlies one kind of variation: specifically, I have explored evolution of transcriptional regulation, a mode of generating variation which is increasingly thought to account for much of the diversity of form in multicellular organisms.

Evolution of coding sequence

Evolutionary innovation comes from many categories of changes in the genome. Among them are changes in protein coding sequence, and changes in regulatory mechanisms ranging from alternative splicing to post-translational modifications to transcription.

Changes in coding sequence are central to optimization of molecular functions. Archae that live at high temperatures, for instance, have evolved exceptionally thermostable versions of conserved proteins; this is generally accomplished by optimization of specific amino acid side-chain interaction in the protein core (Scandurra et al., 2000). These principles have even been harnessed to computationally engineer artificially thermostable proteins (Korkegian et al., 2005).

Existing protein folds can also be adapted or co-opted for new functions. The soil bacteria *Sphingomonas chlorophenolica*, for example, is capable of degrading the anthropogenic pesticide pentachlorophenol. To accomplish this task, enzymes have been recruited from two other pathways; one of these enzymes, which converts pentachlorophenol to tetrachlorohydroquinone in a reductive dehalogenation, is derived from the enzyme maleylacetoacetate isomerase, which has a native function in metabolism of phenylalanine and tyrosine (Copley, 2000).

Coding sequence evolution encompasses not only adaptation of existing genes, but also the birth of entirely new genes. In one spectacular documented example, it was demonstrated that the functional properties of the antifreeze glycoprotein of the Antarctic fish *Dissostichus mawsoni* are derived from what was originally a non-coding region in

an ancestor (Chen et al., 1997). This region was iteratively duplicated to give 41 repeated segments, and the resulting threonine-alanine-alanine repeats act to prevent ice-crystal formation in the fish. The scaffold of the protein is related to the enzyme pancreatic trypsinogen, which retains a signal peptide that guides proper secretion of the antifreeze protein.

Evolution of transcriptional regulation

Although vivid examples of coding sequence evolution pepper the literature, recent work in comparative genomics has revealed that increases in organismal complexity do not scale with increases in gene number. Despite the complexity of the human organism, which comprise on the order of tens of trillions of cells acting in concert, the human genome is estimated to encode only 25,000 genes (Lander et al., 2001; Venter et al., 2001); in comparison, the genome of the diminutive 959-cell worm *C. elegans* encodes an estimated 19,000 genes (C. elegans sequencing Consortium, 1998). Moreover, the increase in human gene number over that of *C. elegans* is due almost entirely to duplication of existing genes, rather than innovation of new ones.

In response to this apparent contradiction, researchers have suggested that complexity must arise in part from increasing the complexity of gene expression patterns throughout development, rather than simply increasing the number of genes (Carroll, 2001; Davidson, 2001; Gerhart and Kirschner, 1997). The proposed mechanism for evolutionary innovation by changes in transcriptional regulation is appealing on multiple levels, outlined below:

1) Reconciling organismal complexity and gene number

As mentioned above, evolution of transcriptional regulation partially accounts for the discontinuity between organismal complexity and gene number in wide evolutionary windows. Organismal development unfolds in a series of precise temporal and spatial transformations. One way of achieving this level of precison could be to use different gene products for each stage in development, each cell-type, or each anatomical trait; in reality, however, such precision appears to be achieved by evolving progressively more complex regulation of existing genes rather than increasing bulk gene number. More complex regulation also means that a single gene product can be used and re-used in multiple developmental contexts and stages. For instance, the *toll* and *dorsal* gene products of *Drosophila melanogaster* are required for dorsal-ventral patterning in the early embryo, as well as the innate immune response into adulthood (Belvin and Anderson, 1996).

The concept of evolving transcriptional regulation also has considerable explanatory power in narrower evolutionary windows. Chimpanzees and humans, for instance, share more than 98.5% of their genome sequence (Chimpanzee Sequencing and Analysis Consortium, 2005). Long before the whole genome sequencing of chimpanzee, King and Wilson suggested that differences in chimp and human coding sequence were insufficient to account for differences in social behavior and anatomy, and that the differences between humans and chimps must involve evolution of gene regulation (King and Wilson, 1975). Recently, Wray and colleagues identified what may be the first specific example of *cis*- regulatory evolution between humans and chimps, demonstrating that the opioid receptor prodynorphin is differentially regulated in humans and chimps

(Rockman et al., 2005). Although the biological consequences of this differential regulation have yet to be determined, prodynorphin has demonstrated roles in perception, behavior, and memory. Interestingly, a survey of selective pressures on protein-coding genes in the human genome showed that as a class, transcription factors are among the most rapidly evolving genes in the human genome when compared to the chimp genome (Bustamante et al., 2005).

2) Resolution of the "Hox Paradox"

Evolution of transcriptional regulation accounts for the so-called "Hox paradox." The Hox paradox arose from the stunning discovery that organisms as distinct as humans, flies, and sea urchins share a core set of transcription factors (the homeobox, or Hox factors) that govern basic aspects of morphogenesis - and perhaps even more unexpectedly, that they direct development of analogous structures even when specific developmental pathways have diverged to the point of being unrecognizable.

One oft-cited example of the paradox is that of Pax6, a key regulator of eye development in vertebrates and invertebrates alike. Pax6 itself is a paired domain DNA binding protein (not a homeodomain protein); in organisms as different as *Drosophila* and mice, molluscs and jellyfish, Pax6 directs eye development (Kozmik, 2005). Yet eyes in these organisms have strikingly little in common in terms of anatomical structure, morphogenesis, and even molecular components. For instance, Pax6 directs formation of the lens by regulating crystallin production in vertebrate eyes. Not only do invertebrate eyes lack lenses altogether, but crystallin itself comprises a diverse family of proteins lacking functional homologs in invertebrates (Davidson, 2001).

A possible explanation for the involvement of Pax6 in eye development lies in the observation that in all cases, a Pax6 homolog regulates rhodopsin and other conserved photoreceptor genes. It has been proposed that Pax6 was initially assigned to photoreceptor regulation, and morphogenenic regulatory genes subsequently came under control of Pax6 through changes in transcriptional regulation (Davidson, 2001). Changes in which genes are regulated by specific transcription factors must be invoked to explain how orthologous transcription factors direct such vastly different developmental programs.

3) Eukaryotic promoter architecture

The mechanisms of transcriptional regulation itself are thought to be particularly amenable to evolution on multiple levels. Non-coding regions are not constrained by reading frame, and are also tolerant to changes in orientation and spacing, rendering them especially plastic with regards to sequence requirements. This is evident in comparing sequence of closely related organisms. For example, *S. cerevisiae* and its close relative *S. paradoxus* are 90% identical in genic regions, but 80% identical in non-coding regions (Kellis et al., 2003).

Furthermore, many eukaryotic promoters are modular in form; that is, individual regions within the promoters are sufficient to direct a discrete fraction of the expression program. Combined, these regions coordinate the complete expression program, directing a gene to be expressed in multiple spatial regions in a developing organism, or at multiple times during development. Due to the modular format of eukaryotic promoters, individual elements act - and can therefore evolve - independently of other

elements. This feature of eukaryotic promoters is proposed to be advantageous because it decreases the likelihood of pleiotropic effects of mutation (Carroll, 2005); however, whether minimization of pleiotropy is actually a principle of evolution remains under debate (Hansen, 2003).

The expression of *even-skipped (eve)* in *Drosophila* is a prime example of modular promoter architecture (Goto et al., 1989; Harding et al., 1989; Stanojevic et al., 1991). The regulatory region of *eve* spans some 20,000 bp, and consists of five modules, each of which specifies one to two of the seven stripes of *eve* expressed in the developing larva. The second of these stripes, *eve* stripe 2, has been the focus of extensive functional and evolutionary studies, and is completely specified by a 480 nucleotide region of the *eve* promoter (Small et al., 1992). Thus, a change in this 480 nucleotide region specifically affects the expression of *eve* stripe 2, while the remaining 6 stripes maintain wildtype expression.

The concept of pleiotropy-avoidance in evolutionary innovation has gained some experimental support in recent years as the sequencing of genomes in related species has allowed direct measurement of evolutionary rate. Many investigators have attempted to correlate evolutionary rate and selective pressure of coding sequence with potential metrics of pleiotropy such as the effects of protein deletion on growth rate, protein network connectivity, expression levels, and recent gene duplication, to varying degrees of success and controversy (Bloom and Adami, 2003; Fraser, 2005; Fraser and Hirsh, 2004; Fraser et al., 2002; Fraser et al., 2003; Gu et al., 2005; Gu et al., 2002; Hirsh and Fraser, 2001; Jordan et al., 2003; Wall et al., 2005).

It should be noted that other strategies can also buffer the pleiotropy of mutations.

Gene duplication and alternative splicing are classic examples of evolution that are thought to be advantageous precisely for their ability to minimize pleiotropy. In the case of duplication, a duplicate gene can evolve rapidly in coding sequence or transcriptional regulation while the original gene retains its original function (Gu et al., 2005; Katju and Lynch, 2003; Lynch and Katju, 2004); a similar effect is achieved in alternative splicing, where a given exon may evolve rapidly, but another spliced variant remains unchanged (Cusack and Wolfe, 2005).

4) Mechanistic consequences of combinatorial interactions

Finally, transcriptional control in eukaryotes is largely combinatorial - that is, multiple transcription factors physically cooperate to control gene expression. The molecular mechanism underlying combinatorial interactions has major ramifications in evolution. While activation or repression of gene expression has been demonstrated to occur via allosteric control of RNA polymerase in a few instances in prokaryotes (Benoff et al., 2002; Lawson et al., 2004; Niu et al., 1996; Toulokhonov et al., 2001), it is thought that simple recruitment of either RNA polymerase or repressive machinery is the primary mechanism of transcriptional control in eukaryotes (Ptashne and Gann, 1998). Unlike allosteric interactions which work by inducing precise conformational changes within RNA polymerase, recruitment can be achieved by simply changing the number or size of surfaces available to interact with RNA polymerase or repressive machinery at a given gene. Such changes in protein interaction are commonly found in evolution (Ptashne and Gann, 1998); moreover, in contrast to allosteric control, transcriptional control using weak interactions is in principle compatible with unlimited combinatorial

integration of inputs. Recently, it was demonstrated that a mechanism of transcriptional regulation via weak interactions is sufficient to generate all possible complex logical computations (Buchler et al., 2003), and that the *in silico* produced *cis*-regulatory regions resemble natural promoters in their modularity and use of combinatorial interactions.

Biologically speaking, the use of combinatorial interactions in transcriptional control allows a high degree of specificity to be achieved via integration of multiple interactions, any of which individually may have low affinity or specificity. This observation has basis in many developmental studies, such as those of *eve* stripe 2, which show that a small number of transcription factors can yield a highly specific output (Howard and Davidson, 2004). In the case of *eve* stripe 2 expression, four transcription factors, which vary in concentration across the embryo, act in concert to specify a narow band of *even-skipped.*

Because low-specificity protein-protein and protein-DNA contacts are at the heart of combinatorial interactions, small changes in regulatory proteins can generate major changes in regulation. For instance, in this thesis, I demonstrate that the interaction between *S. cerevisiae* α2 and Mcm1, conferred by a short linker region of just eight amino acids in α2 (Mead et al., 1996), arose recently in evolution; as a consequence of this new interaction, an entire group of genes is under negative control of the Mcm1-α2 heterotetramer.

A further consequence of combinatorial control is that a small number of transcription factors can be re-used in multiple contexts to yield multiple outcomes. Mcm1, for instance, functions in at least nine different complexes in *S. cerevisiae*, each of which controls a different group of genes (Herskowitz, 1992; Kumar et al., 2000;

Kunoh et al., 2000; Pramila et al., 2002; Turner et al., 2002). These groups of genes range in function from arginine metabolism to cell-cycle control to mating-type determination.

Specificity from combinatorial control is achieved using a combination of many short interactions rather than a small number of highly specific interactions. Reflecting this, nearly all eukaryotic transcription factors have short, degenerate recognition sequences. This feature of eukaryotic transcription also has consequences for evolution, as degeneracy in principle allows rapid turnover and generation of new binding sites on evolutionary time-scales (Wray et al., 2003). The *eve* stripe 2 module provides a fine example of rapid turnover: binding site turnover in closely related species of *Drosophila* has rendered the module unalignable by sequence, yet each species still achieves equivalent expression of *eve* stripe 2 using the same transcription factors (Ludwig and Kreitman, 1995; Ludwig et al., 2005; Ludwig et al., 1998). In this example, binding site turnover is accompanied by identical biological output; whether the degeneracy of transcription factors is advantageous in conferring new biological outputs remains to be seen. Notably, all identified examples of binding-site turnover thus far involve redundant regulation or combinatorial interactions (Wray et al., 2003).

Difficulties in studying evolution of transcriptional regulation

Evolution of transcriptional regulation can occur by many mechanisms. Existing genes may come under new control of an existing transcriptional regulator through mutation of promoter sequences, or by recombination with a region already containing a regulatory site. Conversely, changes in a transcriptional regulator may alter its binding or

activity, bringing new genes under its control or changing the nature of its control over an existing regulon.

Recently, many remarkable examples of transcriptional circuit evolution have been identified (Carroll, 2001; Davidson, 2001; Doebley and Lukens, 1998; Duboule and Wilkins, 1998; Gasch et al., 2004; Landry et al., 2005; Tanay et al., 2005). Changes in both *cis*- and *trans*- elements have been shown to have a major impact on animal morphology or pigmentation (Galant and Carroll, 2002; Gompel et al., 2005; Ronshaugen et al., 2002). For instance, a transcriptional repression domain in the homeodomain protein Ultrabithorax (Ubx) is key in reducing the number of limbs in insects to three thoracic pairs, while crustaceans with Ubx homologs that lack the repression function retain multiple limbs (Galant and Carroll, 2002; Ronshaugen et al., 2002). Due to limitations in model organisms, however, detailed analysis is often limited in scope to either the *cis*- or the *trans*-elements. In many cases, the corresponding *trans*- or *cis*-elements are unknown, or are difficult to evaluate at a comparable level of detail. Moreover, due to complexity in developmental programs, it is difficult if not impossible to understand how a change in a particular regulatory event has become integrated into the existing circuit.

While the examples of morphological or pigmental changes via evolution of transcription are undeniably dramatic, the inability to analyze the changes in greater detail can be a substantial obstacle towards understanding the mechanisms which underlie transcriptional regulatory evolution. For example, the principle observation behind resolution of the hox paradox is that orthologous transcription factors control analogous processes via different target genes. However, converting such an observation

into a mechanistic understanding of *how* different target genes come under control of a specific transcription factor – for instance, by recombination of promoter sequences, or by binding site turnover – requires identification of those targets, their *cis*-elements, and all associated co-activators or co-repressors. Yet limitations in multicellular model systems hamper their identification. Similarly, the evolution of Ubx (as referenced above) is associated with repression of limb formation in hexapods. Without knowing the exact genes regulated by Ubx in both crustaceans and hexapods, or the identities or sequences of co-regulators involved in Ubx regulation, it is impossible to know how much of the effect is attributable to the identified change in Ubx activity, or how much is due to combined effect of changes in Ubx, its co-factors, and its target genes.

Studying evolution of transcriptional regulation in yeasts

In this thesis, I have chosen to study evolution of trancriptional regulation primarily in the fungi *Saccharomyces cerevisiae* and *Candida albicans,* which diverged between 200 and 800 million years ago (Calderone, 2002; Hedges, 2002). I also extend these studies to include a group of 14 related ascomycetes, the most distant of which diverged from *S. cerevisiae* ~1 billion years ago (Hedges, 2002).

The yeast model system provides considerable advantages over other major systems used to study evolution of transcription, such as fruit flies and sea urchins. For instance, the genomes of multiple yeasts have been sequenced and annotated (Cliften et al., 2003; Dujon et al., 2004; Kellis et al., 2004; Kellis et al., 2003; Souciet et al., 2000). The facility of genetic and reverse genetic methods, combined with whole genome transcriptional analysis made possible by genome sequence, allows comprehensive

mapping of transcription factors to the gene products they regulate. Furthermore, many of the sequenced yeasts are closely related in evolution, allowing bioinformatic identification of important non-coding sequences using phylogenetic footprinting (Cliften et al., 2003; Kellis et al., 2003). A myriad of yeasts in the ascomycete lineage spanning over a billion years of evolution have now been sequenced; for these and other reasons, yeasts have been a primary subject of many recent molecular evolution studies, ranging from ribosome regulation to evolution of the mating-type locus (Butler et al., 2004; Fabre et al., 2005; Gasch et al., 2004; Ihmels et al., 2005; Tanay et al., 2005).

Mating type regulation in *S. cerevisiae* and *C. albicans*

The mating-type system in *S. cerevisiae* is particularly well-studied (Herskowitz, 1992; Johnson, 1995). All known transcription factors involved in the circuit have been identified; the targets of their regulation have not only been identified, but their biological roles characterized as well. The mechanisms by which the transcription factors activate and repress gene expression has also been analyzed in extensive mutational and structural studies (Acton et al., 2000; Carr et al., 2004; Goutte and Johnson, 1988; Goutte and Johnson, 1993; Goutte and Johnson, 1994; Keleher et al., 1988; Li et al., 1998; Li et al., 1995; Mead et al., 2002; Mead et al., 1996; Phillips et al., 1994; Redd et al., 1996; Stark et al., 1999; Stark and Johnson, 1994; Tan and Richmond, 1998; Vershon and Johnson, 1993; Zhong et al., 1999).

To summarize, mating type in *S. cerevisiae* is controlled by a single locus in the genome called the MAT locus. The MAT locus exists as one of two alleles, "**a**" or "α." Cells which express only the **a** allele ("**a**-type") can mate with those that express only the

α allele ("α-type"), resulting in formation of a non-mating cell which expresses both alleles ("**a**/α-type"). The MAT locus encodes three transcription factors: a1 (encoded at MAT**a**), α1, and α2 (encoded at MATα). These transcription factors specify mating behavior by turning on or off specific batteries of genes, including genes involved with pheromone production and pheromone sensing (Chapter 3, figure 1, page 108).

At the outset of this project, mating-type regulation in *C. albicans* had just been discovered. In contrast to *S. cerevisiae* whose main modern environmental repository is in vineyards and breweries, *C. albicans* is a commensal organism in the digestive tracts of mammals. It is also a major pathogen of humans, particularly in immunocompromised patients. *C. albicans* causes both superficial and systemic infections, and in the latter case is capable of efficiently colonizing every human organ. Many features of *C. albicans* are thought to contribute to its pathogenicity, including the ability to grow filamentously, and the ability to respond swiftly to environmental signals present in the mammalian host such as serum, high temperature, and changes in pH.

Until recently, *C. albicans* was classified as an asexual yeast. However, a locus with remarkable similarity to the mating type locus of *S. cerevisiae* was discovered in 1999, and was named the *MTL* locus, for Mating Type Like locus (Hull and Johnson, 1999). As in *S. cerevisiae*, this locus exists as one of two alleles, **a** or α. Each allele encodes transcription factors with homology to those encoded at the *S. cerevisiae MAT* locus; in addition, these factors have conserved configuration within the locus, direction of transcription, and the location of introns (see Chapter 2, figure 1, page 50). Soon after discovery of the *MTL* locus, targeted deletion of either the **a** or the α allele led to the discovery that *C. albicans* is capable of mating (Hull et al., 2000; Magee and Magee,

2000). As in *S. cerevisiae*, *C. albicans* cells which express only the **a** allele can mate with those that express only the α allele, while cells expressing both alleles do not mate.

A key difference between *S. cerevisiae* and *C. albicans* biology was discovered and characterized in 2002, when Miller and colleagues demonstrated that in order to mate, *C. albicans* cells must first undergo a massive phenotypic transition to a mating-competent form (Miller and Johnson, 2002). This transition, called the "white-opaque switch," is repressed in cells where both a1 and α2 are expressed; moreover, the switch only occurs under very specific environmental conditions even in cells that are "switch-competent." This level of mating regulation has no correlate in *S. cerevisiae*.

Despite this major difference, mating in *S. cerevisiae* and *C. albicans* share several basic features. First, **a** cells mate can mate with α cells, producing **a**/α cells which do not mate. Second, the combined activity of a1/α2 represses mating in both yeasts. In *S. cerevisiae,* this is due to repression of the haploid-specific genes; in *C. albicans*, this is due to the repression of the white-to-opaque transition.

Findings of this thesis

In this work, I investigate evolution of transcriptional regulation by dissecting the mating-type regulatory circuit in *C. albicans*, then closely comparing it to that of *S. cerevisiae*.

In Chapter 2, I describe experiments designed to comprehensively determine how the *MTL* transcription factors specify mating-type in *C. albicans*. My colleagues and I use targeted gene deletion strategies to construct strains bearing all possible combinations of the *MTL* transcription factors. We then examine the mating behaviors of these strains

and their transcriptional profiles to determine both the overall role of each transcription factor in specifying mating type, as well as their specific regulatory targets. We reveal that the *C. albicans MTL* encodes four transcription factors that regulate approximately 500 genes total; in contrast, the *S. cerevisiae MAT* locus uses three transcription factors to regulate approximately 40 genes. Several classes of changes in mating-type regulation have occurred since the divergence of *C. albicans* and *S. cerevisiae*: differences in the regulation of individual target genes, the requirement for α-factor induction in asg expression in *C. albicans* but not *S. cerevisiae*, replacement of positive regulation of the asgs by negative regulation in *S. cerevisiae*, and incorporation of an additional level mating control in the form of the white-opaque switch. These changes are discussed in detail in Chapter 2.

We then focus on one specific change, the finding that a positive regulator of asg expression, a2, was retained from an ancestor in *C. albicans* but was lost in *S. cerevisiae*. Instead, *S. cerevisiae* uses a modified transcription factor, α2, to repress asgs in the opposite mating type. This addition conserves the logical output of the circuit. To learn more about the ancestral state of positive regulation, we first identify the promoter element responsible for asg activation in *C. albicans*. Based on our findings in *C. albicans*, we deduce that Mcm1 was involved in asg regulation in the common ancestor of *C. albicans* and *S. cerevisiae*; we also find that α2 and a2 share similar binding specificity. Combined, these two findings suggest an evolutionary scheme whereby Mcm1 facilitated the co-option of an a2 binding site by α2. The mechanistic and evolutionary implications of this scheme are far-reaching, potentially representing a general principle of combinatorial circuit evolution. These ideas are discussed in detail in

Chapters 3 and 4.

By extending our studies to a series of 16 ascomycetes that are related to varying extents, we find evidence in both *cis*- and *trans*-regulatory elements that the repression of asgs by α2 evolved prior to the loss of asgs by a2. Our observations are consistent with a model in which a dual regulatory strategy was employed as an intermediate during the transition from positive to negative regulation.

Summary

The lack of genome fossils is a major hurdle that prevents scientists from directly observing molecular evolution over large time-scales. However, I demonstrate in this thesis that it is nevertheless possible to identify highly detailed instances of transcriptional evolution, including the transcription factors, the target genes, and the target *cis*- elements involved in evolution of a given combinatorial transcriptional circuit. Moreover, I determine the relevant changes in both the *cis*- and *trans*- elements, which together define the overall evolutionary transition. These changes have major, experimentally testable implications for the mechanism underlying the transition between two vastly different forms of transcriptional regulation. Finally, I show that show that the genomes of extant organisms can be used to infer the order in which a series of evolutionary changes occurred. Importantly, the order of changes determines the specific fitness barriers that were faced during evolution of the circuit, tying our observations of evolution of transcription to natural selection.

References

Acton, T. B., Mead, J., Steiner, A. M., and Vershon, A. K. (2000). Scanning mutagenesis of Mcm1: residues required for DNA binding, DNA bending, and transcriptional activation by a MADS-box protein. Mol Cell Biol *20*, 1-11.

Belvin, M. P., and Anderson, K. V. (1996). A conserved signaling pathway: the Drosophila toll-dorsal pathway. Annu Rev Cell Dev Biol *12*, 393-416.

Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y. W., Berman, H. M., and Ebright, R. H. (2002). Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. Science *297*, 1562-1566.

Bloom, J. D., and Adami, C. (2003). Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets. BMC Evol Biol *3*, 21.

Buchler, N. E., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. Proc Natl Acad Sci U S A *100*, 5136-5141.

Bustamante, C. D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M. T., Glanowski, S., Tanenbaum, D. M., White, T. J., Sninsky, J. J., Hernandez, R. D., *et al.* (2005). Natural selection on protein-coding genes in the human genome. Nature *437*, 1153-1157.

Butler, G., Kenny, C., Fagan, A., Kurischko, C., Gaillardin, C., and Wolfe, K. H. (2004). Evolution of the MAT locus and its Ho endonuclease in yeast species. Proc Natl Acad Sci U S A *101*, 1632-1637.

Calderone, R. A. (2002). Candida and Candidiasis (Washington, D.C.: ASM Press).

Carr, E. A., Mead, J., and Vershon, A. K. (2004). Alpha1-induced DNA bending is required for transcriptional activation by the Mcm1-alpha1 complex. Nucleic Acids Res *32*, 2298-2305.

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

Carroll, S. B., Grenier J. K., Weatherbee, S. D. (2001). From DNA to Diversity (Malden, Massachusetts: Blackwell Science, Inc.).

Chen, L., DeVries, A. L., and Cheng, C. H. (1997). Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc Natl Acad Sci U S A *94*, 3811-3816.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A., and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science *301*, 71-76.

Consortium, *C. e.* S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium. Science *282*, 2012-2018.

Consortium, C. S. a. A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature *437*, 69-87.

Copley, S. D. (2000). Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. Trends Biochem Sci *25*, 261-265.

Cusack, B. P., and Wolfe, K. H. (2005). Changes in alternative splicing of human and mouse genes are accompanied by faster evolution of constitutive exons. Mol Biol Evol *22*, 2198-2208.

Darwin, C. R. (1859). The Origin of Species (New York, New York: Gramercy).

Davidson, E. H. (2001). Genomic Regulatory Systems (San Diego, CA: Academic Press).

Doebley, J., and Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. Plant Cell *10*, 1075-1082.

Duboule, D., and Wilkins, A. S. (1998). The evolution of 'bricolage'. Trends Genet *14*, 54-59.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., *et al.* (2004). Genome evolution in yeasts. Nature *430*, 35-44.

Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B., and Fairhead, C. (2005). Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. Mol Biol Evol *22*, 856-873.

Fraser, H. B. (2005). Modularity and evolutionary constraint on proteins. Nat Genet *37*, 351-352.

Fraser, H. B., and Hirsh, A. E. (2004). Evolutionary rate depends on number of protein-protein interactions independently of gene expression level. BMC Evol Biol *4*, 13.

Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C., and Feldman, M. W. (2002). Evolutionary rate in the protein interaction network. Science *296*, 750-752.

Fraser, H. B., Wall, D. P., and Hirsh, A. E. (2003). A simple dependence between protein evolution rate and the number of protein-protein interactions. BMC Evol Biol *3*, 11.

Futuyma, D. J. (1997). Evolutionary Biology (Sunderland, Massachusetts: Sinauer).

Galant, R., and Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. Nature *415*, 910-913.

Gasch, A. P., Moses, A. M., Chiang, D. Y., Fraser, H. B., Berardini, M., and Eisen, M. B. (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol *2*, e398.

Gerhart, J., and Kirschner, M. (1997). Cells, Embryos, and Evolution (Malden, Massachusetts: Blackwell Science, Inc.).

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., and Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature *433*, 481-487.

Goto, T., Macdonald, P., and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell *57*, 413-422.

Goutte, C., and Johnson, A. D. (1988). a1 protein alters the DNA binding specificity of alpha 2 repressor. Cell *52*, 875-882.

Goutte, C., and Johnson, A. D. (1993). Yeast a1 and alpha 2 homeodomain proteins form a DNA-binding activity with properties distinct from those of either protein. J Mol Biol *233*, 359-371.

Goutte, C., and Johnson, A. D. (1994). Recognition of a DNA operator by a dimer composed of two different homeodomain proteins. Embo J *13*, 1434-1442.

Gu, X., Zhang, Z., and Huang, W. (2005). Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proc Natl Acad Sci U S A *102*, 707-712.

Gu, Z., Nicolae, D., Lu, H. H., and Li, W. H. (2002). Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet *18*, 609-613.

Hansen, T. F. (2003). Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. Biosystems *69*, 83-94.

Harding, K., Hoey, T., Warrior, R., and Levine, M. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. Embo J *8*, 1205-1212.

Hedges, S. B. (2002). The origin and evolution of model organisms. Nat Rev Genet *3*, 838-849.

Herskowitz, I., Rine, J., and Strathern, J. (1992). Mating-type determination and mating-type interconversion in *Saccharomyces cerevisiae*., In The Molceular Cellular Biology fo

the Yeast *Saccharomyces*, J. R. P. E.W. Jones, J.R. Broach, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 583-657.

Hirsh, A. E., and Fraser, H. B. (2001). Protein dispensability and rate of evolution. Nature *411*, 1046-1049.

Howard, M. L., and Davidson, E. H. (2004). cis-Regulatory control circuits in development. Dev Biol *271*, 109-118.

Hull, C. M., and Johnson, A. D. (1999). Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans. Science *285*, 1271-1275.

Hull, C. M., Raisner, R. M., and Johnson, A. D. (2000). Evidence for mating of the "asexual" yeast Candida albicans in a mammalian host. Science *289*, 307-310.

Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yanai, I., McClellan, M., Berman, J., and Barkai, N. (2005). Rewiring of the yeast transcriptional network through the evolution of motif usage. Science *309*, 938-940.

Johnson, A. D. (1995). Molecular mechanisms of cell-type determination in budding yeast. Curr Opin Genet Dev *5*, 552-558.

Jordan, I. K., Wolf, Y. I., and Koonin, E. V. (2003). No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol *3*, 1.

Katju, V., and Lynch, M. (2003). The structure and early evolution of recently arisen gene duplicates in the Caenorhabditis elegans genome. Genetics *165*, 1793-1803.

Keleher, C. A., Goutte, C., and Johnson, A. D. (1988). The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. Cell *53*, 927-936.

Kellis, M., Birren, B. W., and Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature *428*, 617-624.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241-254.

King, M. C., and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107-116.

Korkegian, A., Black, M. E., Baker, D., and Stoddard, B. L. (2005). Computational thermostabilization of an enzyme. Science *308*, 857-860.

Kozmik, Z. (2005). Pax genes in eye development and evolution. Curr Opin Genet Dev *15*, 430-438.

Kumar, R., Reynolds, D. M., Shevchenko, A., Shevchenko, A., Goldstone, S. D., and Dalton, S. (2000). Forkhead transcription factors, Fkh1p and Fkh2p, collaborate with Mcm1p to control transcription required for M-phase. Curr Biol *10*, 896-906.

Kunoh, T., Kaneko, Y., and Harashima, S. (2000). Positive regulation of transcription of homeoprotein-encoding YHP1 by the two-component regulator Sln1 in Saccharomyces cerevisiae. Biochem Biophys Res Commun *278*, 344-348.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., and Hartl, D. L. (2005). Compensatory cis-trans Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of Drosophila. Genetics.

Lawson, C. L., Swigon, D., Murakami, K. S., Darst, S. A., Berman, H. M., and Ebright, R. H. (2004). Catabolite activator protein: DNA binding and transcription activation. Curr Opin Struct Biol *14*, 10-20.

Li, T., Jin, Y., Vershon, A. K., and Wolberger, C. (1998). Crystal structure of the MATa1/MATalpha2 homeodomain heterodimer in complex with DNA containing an A-tract. Nucleic Acids Res *26*, 5707-5718.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995). Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. Science *270*, 262-269.

Ludwig, M. Z., and Kreitman, M. (1995). Evolutionary dynamics of the enhancer region of even-skipped in Drosophila. Mol Biol Evol *12*, 1002-1011.

Ludwig, M. Z., Palsson, A., Alekseeva, E., Bergman, C. M., Nathan, J., and Kreitman, M. (2005). Functional evolution of a cis-regulatory module. PLoS Biol *3*, e93.

Ludwig, M. Z., Patel, N. H., and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development *125*, 949-958.

Lynch, M., and Katju, V. (2004). The altered evolutionary trajectories of gene duplicates. Trends Genet *20*, 544-549.

Magee, B. B., and Magee, P. T. (2000). Induction of mating in Candida albicans by construction of MTLa and MTLalpha strains. Science *289*, 310-313.

Mead, J., Bruning, A. R., Gill, M. K., Steiner, A. M., Acton, T. B., and Vershon, A. K. (2002). Interactions of the Mcm1 MADS box protein with cofactors that regulate mating in yeast. Mol Cell Biol *22*, 4607-4621.

Mead, J., Zhong, H., Acton, T. B., and Vershon, A. K. (1996). The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. Mol Cell Biol *16*, 2135-2143.

Miller, M. G., and Johnson, A. D. (2002). White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. Cell *110*, 293-302.

Niu, W., Kim, Y., Tau, G., Heyduk, T., and Ebright, R. H. (1996). Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. Cell *87*, 1123-1134.

Phillips, C. L., Stark, M. R., Johnson, A. D., and Dahlquist, F. W. (1994). Heterodimerization of the yeast homeodomain transcriptional regulators alpha 2 and a1 induces an interfacial helix in alpha 2. Biochemistry *33*, 9294-9302.

Pramila, T., Miles, S., GuhaThakurta, D., Jemiolo, D., and Breeden, L. L. (2002). Conserved homeodomain proteins interact with MADS box protein Mcm1 to restrict ECB-dependent transcription to the M/G1 phase of the cell cycle. Genes Dev *16*, 3034-3045.

Ptashne, M., and Gann, A. (1998). Imposing specificity by localization: mechanism and evolvability. Curr Biol *8*, R897.

Redd, M. J., Stark, M. R., and Johnson, A. D. (1996). Accessibility of alpha 2-repressed promoters to the activator Gal4. Mol Cell Biol *16*, 2865-2869.

Rockman, M. V., Hahn, M. W., Soranzo, N., Zimprich, F., Goldstein, D. B., and Wray, G. A. (2005). Ancient and Recent Positive Selection Transformed Opioid cis-Regulation in Humans. PLoS Biol *3*, e387.

Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. Nature *415*, 914-917.

Rothschild, L. J., and Mancinelli, R. L. (2001). Life in extreme environments. Nature *409*, 1092-1101.

Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L., and Engel, P. C. (2000). Protein stability in extremophilic archaea. Front Biosci *5*, D787-795.

Small, S., Blair, A., and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. Embo J *11*, 4047-4057.

Souciet, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de Montigny, J., Dujon, B., *et al.* (2000). Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett *487*, 3-12.

Stanojevic, D., Small, S., and Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. Science *254*, 1385-1387.

Stark, M. R., Escher, D., and Johnson, A. D. (1999). A trans-acting peptide activates the yeast a1 repressor by raising its DNA-binding affinity. Embo J *18*, 1621-1629.

Stark, M. R., and Johnson, A. D. (1994). Interaction between two homeodomain proteins is specified by a short C-terminal tail. Nature *371*, 429-432.

Tan, S., and Richmond, T. J. (1998). Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature *391*, 660-666.

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

Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001). Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. Science *292*, 730-733.

Turner, S. D., Ricci, A. R., Petropoulos, H., Genereaux, J., Skerjanc, I. S., and Brandl, C. J. (2002). The E2 ubiquitin conjugase Rad6 is required for the ArgR/Mcm1 repression of ARG1 transcription. Mol Cell Biol *22*, 4011-4019.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Vershon, A. K., and Johnson, A. D. (1993). A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. Cell *72*, 105-112.

Wall, D. P., Hirsh, A. E., Fraser, H. B., Kumm, J., Giaever, G., Eisen, M. B., and Feldman, M. W. (2005). Functional genomic analysis of the rates of protein evolution. Proc Natl Acad Sci U S A *102*, 5483-5488.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. (2003). The Evolution of Transcriptional Regulation in Eukaryotes. Mol Biol Evol.

Zhong, H., McCord, R., and Vershon, A. K. (1999). Identification of target sites of the alpha2-Mcm1 repressor complex in the yeast genome. Genome Res *9*, 1040-1047.

Chapter 2

Evolution of a Combinatorial Transcriptional Circuit

This version of Chapter 2 is reprinted with minor changes from Cell, 115, Annie E. Tsong, Mathew G. Miller, Ryan M. Raisner, and Alexander D. Johnson, *Evolution of a Combinatorial Transcriptional Circuit: A Case Study in Yeasts*, 389-399, Copyright 2003, with permission from Elsevier.

The work in this chapter was carried out by myself, Mathew Miller, and Ryan Raisner. I contributed most of the strains eventually used in the paper, as well as most of the microarray experiments and all of the quantitative mating experiments. Matt contributed several microarray experiments involving comparisons between white and opaque strains. Ryan did all of the initial experiments that lay the groundwork for this project. Matt and Ryan also constructed strains used in this project prior to the discovery of a2 (See Appendix 2).

ABSTRACT

Developing new regulation of existing genes is likely a key mechanism by which organismal complexity arises in evolution. To examine plasticity of gene regulation over evolutionary timescales, we have determined the transcriptional circuit regulating mating-type in the human fungal pathogen *Candida albicans*, and compared it to that of *Saccharomyces cerevisiae*. Since the two yeasts last shared an ancestor 100-800 million years ago, several major differences in circuitry have arisen. For example, a positive regulator of mating-type was retained in *C. albicans* but lost in *S. cerevisiae*; this circuit branch was replaced by the modification of an existing negative regulator, thereby conserving the circuit output. We also characterize a tier of mating-type transcriptional regulation that is present only in *C. albicans*, and likely results from the vastly different environmental selections imposed on the two yeasts – in this case, the pressure on *C. albicans* to survive in a mammalian host.

INTRODUCTION

A mechanism by which evolutionary novelty is thought to arise involves rewiring transcriptional circuitry to allow new regulatory patterns of existing gene products (Carroll, 2001; Davidson, 2001; Gerhart and Kirschner,1997). The importance of this general strategy is becoming evident as more genomes are sequenced. For example, despite the complexity of the human organism (composed of trillions of cells of diverse cell types) in comparison to the nematode *C. elegans* (composed of less than 1000 cells and only a subset of the cell types present in humans), it has been estimated that our genomes contains only 1/3 more genes than *C. elegans* (*C. elegans* Sequencing Consortium, 1998; Lander et al., 2001). One view reconciling the small increase in gene number with the large increase in organismal complexity is that complexity arises principally from increasing the number of different gene expression patterns throughout development (for examples, see Carroll, 2000; Doebley and Lukens, 1998; Duboule and Wilkins, 1998; Levine and Tjian, 2003).

New transcription patterns are thought to arise through many mechanisms. Individual genes may come under new control of a transcriptional regulator through mutation of promoter sequences, or through recombination of promoter sequences with those of a region already containing a regulatory site. Conversely, changes in a transcriptional regulator may alter its binding specificity or activity, causing new genes to come under its control, or changing the nature of its control over an existing regulon. Relatively simple molecular changes in transcriptional circuitry can result in large morphological changes in multicellular organisms (Averof and Patel, 1997; Carroll, 2001;

Sharkey et al., 1997). For example, a transcriptional repression domain in the homeodomain protein Ultrabithorax (Ubx) is key in reducing the number of limbs in insects to three thoracic pairs, while crustaceans with Ubx homologs that lack the repression function retain multiple limbs (Galant and Carroll, 2002; Ronshaugen et al., 2002).

In metazoans, pathways regulating sex determination appear unusually plastic, being less conserved across phyla than other developmental circuits. The specific forces of selection driving rapid divergence of sex determination remain to be understood in these cases (de Bono and Hodgkin, 1996; Haag et al., 2002; O'Neil and Belote, 1992; Whitfield et al., 1993). In this paper, we describe the transcriptional circuit that regulates sex determination (termed mating-type in fungi) in a single-celled eukaryote, the human fungal pathogen *Candida albicans*. To do this, we systematically deleted each of the four transcriptional regulators encoded at the *C. albicans* mating-type locus, singly and in all possible combinations. Two of these transcriptional regulators are homeodomain proteins, and the other two are members of the α-domain- and HMG-box-containing families, respectively (Coppin et al., 1997; Gehring et al., 1994; Grosschedl et al., 1994; Kronstad and Staben, 1997; Scott et al., 1989). By analyzing both the mating behavior and transcriptional profiles of these strains, we determined how the transcriptional regulators control mating-type, and which target genes are responsible for the output of the circuit.

We then compared the circuit to that of *Saccharomyces cerevisiae,* which last shared a common ancestor with *C. albicans* some 100 to 800 million years ago (Calderone, 2002; Hedges, 2002). The comparison between sex determining circuits of *S. cerevisiae* and *C. albicans* revealed several major differences, as well as many minor

differences. In the first major example of divergence between *C. albicans* and *S. cerevisiae* mating-type regulation, we find that *C. albicans* has retained a positive regulator of **a**-type mating from a common ancestor, while *S. cerevisiae* has lost this regulator. We propose that in response to this evolutionary contingency, a new scheme of negative regulation has been added in *S. cerevisiae*, preserving the overall output of the transcriptional circuit while altering the specific roles of the regulatory proteins.

A second major example of divergence is the interposition of an additional layer of transcriptional control in the *C. albicans* mating-type circuit. It has previously been shown that *C. albicans* **a** and α cells must undergo a "phenotypic switch" from the white phase to the opaque phase before they are competent to mate (Miller and Johnson, 2002). This switch is governed by the mating type locus: two homeodomain proteins (a1 and α2)—one from the **a** locus and one from the α locus—cooperate to repress the switching, thereby assuring that **a** and α, but not **a**/α cells, can do it (Lockhart et al., 2002; Miller and Johnson, 2002). In this paper, we show that the *C. albicans* a1 and α2 proteins repress a handful of genes directly, but control many more indirectly by governing white-opaque switching. This indirect regulation of mating-competency by a1-α2 in *C. albicans* constitutes an additional, large-scale layer of transcriptional regulation — absent in *S. cerevisiae* — that ensures that mating only occurs in specific environments; we believe that *C. albicans* has developed this additional layer to more closely control mating, a likely necessity in the hostile environment of a mammalian host.

RESULTS

The *C. albicans* mating type locus encodes an extra transcriptional regulator relative to that of *S. cerevisiae*

The *C. albicans* mating type locus (called the *MTL* locus) closely resembles the *S. cerevisiae MAT* locus (Figure 1, Hull and Johnson, 1999). In both organisms, this locus can exist as one of two alleles, either **a** or α. Cells carrying only the **a**-allele mate as **a** cells, cells carrying only α-allele mate as α cells, and cells carrying both alleles (**a**/α cells) do not mate. The mating type locus encodes transcriptional regulators that control cell-type by turning on and off appropriate sets of mating genes. Three of these regulators (α1, α2, and a1) have close relatives in both *S. cerevisiae* and *C. albicans*; moreover, the gene order, the direction of transcription, and the location of introns within them are preserved. Each of these regulatory proteins belongs to a conserved family found in many eukaryotic organisms. α2 and a1 are homeodomain proteins, and α1 is a so-called "α domain" protein (see Introduction for references). All three regulate transcription by binding to specific sequences of DNA, usually in combination with other sequence-specific DNA binding proteins (for reviews, see Herskowitz, 1992; Johnson, 1995). In this paper, we will use the following conventions: the gene encoding each transcriptional regulator will be referred to by its complete name (e.g. *MTLa1* or *MTLα1*), and the gene product will be abbreviated (e.g. a1 or α1).

In laboratory strains of *S. cerevisiae*, **a** and α cells are generally haploid, and hence contain only one of the two *MAT* loci, while **a**/α cells are generally diploid. Higher ploidy states also exist in nature. In *C. albicans*, the three cell-types exist only in the diploid state (Hull et al., 2000; Lockhart et al., 2002; Magee and Magee, 2000). Naturally occurring diploid **a** and α cells are homozygous at their *MTL* loci (Lockhart et al., 2002),

and it is thought that *C. albicans* undergoes a diploid-tetraploid parasexual cycle (Bennett et al., 2003). To preserve isogenicity, the strains in this paper are diploid and heterozygous at the *MTL* locus.

In the course of disrupting the *MTLa1* gene of *C. albicans*, we unintentionally created a functional knock-out of an adjacent, previously undescribed ORF at the *MTLa* locus of *C. albicans* (Figure 1). This strain exhibited a mating behavior distinct from that of a "clean" *MTLa1* deletion. The existence of this ORF was generously brought to our attention by Kenneth Wolfe, Smurfit Institute, University of Dublin, Ireland. This ORF (which we have named *MTLa2*) encodes a putative DNA binding protein (a2) of 201 amino acids that shows significant sequence similarity to HMG-box proteins present in mating type loci of other fungi, including *C. parasitica*, *K. lactis*, *N. crassa*, and *P. anserina* (Astrom et al., 2000; Coppin et al., 1997). The HMG-box family of DNA binding proteins constitute a large class of well-characterized transcriptional regulators found in many eukaryotes (Grosschedl et al., 1994). The closest known relative of the *C. albicans* a2 protein is the MAT1-1-3 protein from *Cryphonectria parasitica*, which shows 24% sequence identity across the whole of the protein and 33% across the HMG-box (McGuire et al., 2001). The *S. cerevisiae MAT* locus lacks a homolog of *C. albicans MTLa2*, although it clearly shares homologs with the three other transcriptional regulators encoded by the *C. albicans MTL*, a1, α1, and α2.

a-type mating is positively regulated in *C. albicans*

In order to determine the specific roles of the *MTL*-encoded regulators in specifying mating-type, we tested mating behavior in a collection of diploid isogenic

strains bearing all 16 possible combinations of the four genes (Table 1). Each ORF was knocked out precisely, leaving DNA sequences up- and down-stream of the coding region intact. All strains were constructed as ura-ADE+ auxotrophs and were tested for mating against two URA+ade- tester strains: an α strain, and an **a** strain (see Experimental procedures for details of strain construction).

In order to mate efficiently, *C. albicans* must undergo a "phenotypic switch" from the white phase, the prevalent form of *C. albicans*, to the opaque phase, which is stable at 23°C, but not 30°C. As previously described, this transition is blocked by the combination of a1 and α2 proteins encoded at the *MTL* locus (Miller and Johnson, 2002). We first tested the 16 strains for their ability to undergo white-to-opaque switching. All but four strains could efficiently switch (Table 1). These four strains are the only ones that retain both a1 and α2 (lines 1, 3, 9, and 11), confirming the previous conclusion that a1 and α2 control white-opaque switching. This experiment also shows that the newly described a2 is not involved in this control.

Of the twelve strains that could undergo white-opaque switching, some could mate and others could not, revealing a complex pattern of mating behavior that is governed by the *MTL* locus (Table 1). These mating experiments were carried out in two ways. In one set of experiments, the experimental strain in the white phase was mated to tester strains in the opaque phase, and mating was quantified using a filter assay described by Miller and Johnson (2002). In these experiments, the mating frequency presumably reflects the ability of the unknown strain to undergo the white-to-opaque transition and to subsequently mate. In the second set of experiment, both the experimental and the tester strains were in the opaque form; these frequencies reflect only mating ability (Table 1).

The mating table appears complex, but is entirely self-consistent. Moreover, the systematic experimental design provided significantly more information than is minimally necessary to construct the genetic circuit. Thus, each of the following conclusions can be independently derived from different subsets of the data: efficient α-type mating (defined as the ability to mate with the **a**-type mating-type tester) requires only α1 (Table 1, compare lines 10, 12, and 14). Likewise, efficient **a**-type mating requires only a2 (Table 1, compare lines 7, 8, and 15). When both α1 and a2 are present, the strain mates inefficiently with both tester strains (Table 1, lines 2, 5, and 6); that is, it is a weak bi-mater. In contrast, a1 and α2 are dispensable for mating in these 12 strains. Based on these results, we propose a genetic circuit governing mating-type in *C. albicans* (Figure 2). Each of the two mating-type alleles (**a** and α) contributes a positive regulator of its respective mating-type. Additionally, each allele contributes one half of a heterodimer that negatively regulates mating-competency. A critical feature of this *C. albicans* model is that each half of the heterodimer—on its own—has no regulatory activity.

This regulation differs significantly from that of *S. cerevisiae*, which lacks an HMG-box a2. In *C. albicans*, **a**-type mating requires positive regulation by a2, whereas in *S. cerevisiae*, **a**-type mating requires no input from the *MAT* locus. In order to prevent **a**-type mating in α cells, the *MATa* locus represses **a**-type mating by means of α2; this branch of the pathway is absent in *C. albicans* (Table 1, compare lines 4 and 8).

Microarray analysis recapitulates mating behavior and reveals different classes of MTL-regulated genes

To identify specific genes regulated by the *C. albicans* a1, a2, α1, and α2 proteins, we performed microarray analysis on strains carrying the 16 combinations of *MTL* genes described above. We analyzed all 16 strains in the white phase; for the 12 strains competent to switch from white to opaque, we also analyzed the transcriptional profiles of the opaque forms. Multiple isolates of each strain in both ura- and URA+ forms were tested; in addition, several different growth conditions were analyzed to eliminate non-*MTL* controlled transcriptional effects (see Experimental procedures). In all experiments, a 2-fold change was considered significant when it reproduced in all trials.

First we compared the 16 white strains of each *MTL* configuration under several conditions. Although there were variations from experiment to experiment, we found 7 genes that were reproducibly repressed 2- to 8-fold by a1 and α2 working together in the white phase. These results were reproduced in 8 independent trials using multiple isolates and conditions, 4 of which are shown in Figure 3a. This gene cluster is comprised of *CAG1* (homologous to *S. cerevisiae GPA1*), *FUS3*, *FAR1*, *STE2*, *YEL003w*, and two ORFS with no homology to any known genes. Because this control is observed in white cells, these 7 genes are controlled by the a1-α2 heterodimer independent of the white-to-opaque transition. We saw no additional transcripts with *MTL*-configuration-dependent expression in the white phase. In other words, we observed no significant differences between the expression patterns of **a** and α white cells.

We next examined the interplay between genes regulated by the *MTL* locus and genes regulated in the white-to-opaque transition. White and opaque versions of strains carrying the 12 *MTL* configurations permissive for white-opaque switching were compared to a white control strain carrying an intact *MTL*. We identified 237 genes that

are upregulated in the opaque phase, and 197 genes that are downregulated in the opaque phase, irrespective of the *MTL* configuration (providing, of course, it was one that permitted the white-to-opaque transition, Figure 3b). These genes, listed in the supplementary table, were regulated from 2- to 200-fold across five independent experiments. Transcripts upregulated in the opaque phase included some likely to be involved in mating in both **a** and α cells, such as *STE4* and *FUS3* (see Discussion).

The set of white and opaque-specific transcripts observed here overlaps with genes identified previously in a comparison of the white and opaque phases of strain WO-1 (Lan et al., 2002). WO-1, which has a substantially different genetic background than the strains used in this study, is the "classic" white-opaque switching strain, hence its name (Slutsky et al., 1987); it is now known to be homozygous for *MTLα*. As described by Lan et al. (2000), opaque-specific genes include large numbers of genes involving metabolic specialization, cell adhesion, cell surface composition, and virulence.

a- and α-specific genes

By comparing strains in white and opaque phases bearing all possible combinations of *MTL* transcriptional regulators, we were able to identify **a**- and α-specific genes. We found that *STE3* and *MFα1* are highly induced in opaque strains relative to white strains (300- and 1000-fold, respectively), but only in those that carry an intact *MTLα1* gene – that is, those able to mate as α-cells (Table 1 and Figure 2c). Thus, we consider these two genes to be α-specific. In *S. cerevisiae*, *STE3* encodes a cell surface receptor protein that is required for α cells to respond to **a**-factor, a mating pheromone produced by **a** cells. It presumably has this same role in *C. albicans*. In both

S. cerevisiae and *C. albicans*, *MFα1* encodes α-factor, the mating pheromone produced by α cells (Bennett et al., 2003; Sprague and Thorner, 1992). Thus, the activation of these two genes by the *C. albicans* α1 makes conceptual sense, as both genes encode α-specific functions.

Initially, we were unable to find equivalent **a**-specific genes in opaque-phase **a**-type cells (data not shown). However, it was recently reported that α-factor strongly induces expression of a number of genes in **a** opaque cells (Bennett et al., 2003). Thus, it was possible that **a**-specific genes are detectably upregulated only in response to pheromone. To test this idea, we treated four *MTL* mutant strains with 10 μg/ml α-factor: one strain had both a1 and a2, two had either a1 or a2, and one strain had neither (Figure 3d). We found that the two strains with a2 formed characteristic mating projections in response to α-factor (data not shown, but see Bennett et al., 2003 for examples of mating projections), while the two strains missing a2 did not. Transcriptional profiling of the strains' response to α-factor recapitulated these results (Figure 3d), demonstrating that a2 is the sole *MTL*-encoded regulator necessary for the morphological and transcriptional response to α-factor.

Although we have identified a set of genes whose induction is dependent on the presence of a2 and α-factor, some of these genes may be induced in α-type maters exposed to **a**-factor (as of yet unidentified in *C. albicans*), making them pheromone-induced genes in both **a** and α cells. However, some of the α-factor induced genes, such as *STE6* or *RAM2*, have specific roles in processing and exporting a-factor in *S. cerevisiae*; indeed, *STE6* is required for *C. albicans* **a** cells (but not α cells) to mate (Magee et al., 2002). Thus, we provisionally include *STE6* and *RAM2* as **a**-specific genes

with the understanding that some of the other α-factor induced genes may also turn out to be **a**-specific.

DISCUSSION

The *C. albicans* mating-type circuitry

In this paper, we show that the four transcriptional regulators encoded at the *C. albicans* mating type locus (a1, a2, α1, and α2) control the expression of several hundred genes and thereby specify the different *C. albicans* cell types (Figure 4). a2 and α1 are activators of the **a**-specific and α-specific genes, respectively. The other two *MTL* regulators, the homeodomain proteins a1 and α2, have no apparent regulatory activities on their own, but together they shut down mating by turning off many genes. A few of these genes are turned off directly by a1-α2 in white cells, while a much larger set (the opaque-specific genes) is turned off indirectly by virtue of the fact that a1-α2 represses white to opaque switching.

Differences between *C. albicans* and *S. cerevisiae* Mating-Type Circuitry

We discuss four types of differences in the circuits regulating mating-type in *S. cerevisiae* and *C. albicans*, and speculate on the evolutionary developments that may have been responsible for them: (1) individual genes that are under direct mating type locus control in one organism but not the other, (2) a branch of the circuit where gene expression requires an external signal in one organism but not the other, (3) the replacement of a positively regulated branch of the circuit in *C. albicans* with a negatively regulated branch in *S. cerevisiae* to produce the same logical output, and (4)

the presence of an additional layer of mating-type regulation in *C. albicans* that likely arose in response to pressure to thrive in a mammalian host.

1) Changes in the Regulation of Individual Target Genes

We have observed many instances in which genes directly controlled by the mating type locus in one organism are out from under this control in the other. Some of these changes preserve the overall output of the circuit, while some change it. An example of the former is the regulation of the *STE4* gene, which encodes the β-subunit of the trimeric G-protein used for pheromone signaling in *S. cerevisiae*. In both organisms, *STE4* is turned off in the presence of a1-α2. In *S. cerevisiae*, the control is direct, whereas in *C. albicans*, it is indirect: *STE4* is an opaque-specific gene, and the white-to-opaque switch is blocked by a1-α2. Although the basic output of the circuit is conserved (*STE4* can be expressed in **a** and α cells but not in **a**/α cells), there has been a subtle change in *C. albicans*: *STE4* is now upregulated in opaque **a** and α cells (the mating competent form) relative to the white cells (the non-competent form). An example of a gene whose output is altered is illustrated by *YEL007w*, a gene of unknown function in both organisms. In *C. albicans*, this gene is expressed in **a** and α white cells but repressed by a1-α2 in **a**/α white cells; in *S. cerevisiae*, it is expressed at equal levels in **a**, α, and **a**/α cells and is therefore not regulated by the mating type locus (Galitski et al., 1999). Conversely, the *BUD31* gene, involved in bud-site selection, is regulated by the mating type locus in *S. cerevisiae* but not in *C. albicans*. There are many additional examples of this type of difference.

Two simple scenarios could account for such differences. According to one, the common ancestor of *C. albicans* and *S. cerevisiae* had the gene in question under mating-

type control, and the control sequences became mutated or otherwise lost in one lineage. Alternatively, the gene was not under mating-type control in the common ancestor and acquired regulation in one lineage by simple mutation or gene conversion of its transcriptional control sequence.

2) Changes in the Requirement of an External Signal for Expression of One Branch of the Circuit

In *C. albicans*, **a**-type maters in the opaque phase do not detectably express **a**-specific genes at levels above those of other cell-types. However, when α-factor (the mating pheromone from α cells) is added, the **a**-specific genes are transcriptionally activated (Bennett et al., 2003). This situation differs from that of *S. cerevisiae*, where naïve (pheromone-unstimulated) **a**-type maters clearly express the **a**-specific genes. Again, the change that could lead to this difference is simple to imagine, since in *S. cerevisiae* most **a**-specific genes are further upregulated when **a**-cells are exposed to α-factor. Thus, the main difference between the two organisms in this regard appears to be in the basal (e.g. pheromone unstimulated) level at which **a**-specific genes are expressed. Such a difference could be achieved by modulating the basal activity or abundance of a transcriptional activator required for **a**-specific gene expression.

3) Replacement of a Positive Branch of the Circuit in One Organism by a Negative Branch in the Other

One of the most intriguing changes between the mating circuitry of *C. albicans* and *S. cerevisiae* is the replacement of a positive branch of regulation by a negative branch, a change that nonetheless conserves the overall output of the circuit (figures 2 and 4). In *C. albicans*, **a**-type mating requires the positive regulatory activity of a2. In

contrast, **a**-type mating in *S. cerevisiae* is the "default" behavior, and requires no input from the *MAT* locus. However, to prevent **a**-specific genes from erroneous activation, an additional branch of control is needed in *S. cerevisiae*: α2 now represses **a**-specific genes. How might this circuit change have evolved? More specifically, did *C. albicans* gain a2 or did *S. cerevisiae* lose it?

The *C. albicans* a2 protein is an HMG-box protein that is highly conserved among ascomycetes (Figure 5). Fungi as distant as *S. pombe* carry an HMG-box gene at their mating type loci, where it positively regulates mating behavior (Kjaerulff et al., 1997). An exception, however, is the evolutionary lineage carrying *S. cerevisiae*, which lacks this mating-type regulator. Given the broad phylogenetic presence of a2-like regulators of mating, the most economical model is that the a2 gene was simply lost in the *S. cerevisiae* lineage. This scenario raises a number of interesting questions. First, how might HMG-box a2 have been lost from the *S. cerevisiae* lineage? Second, how did **a**-specific genes subsequently bypass the need for this positive regulator? Third, what change in α2 allowed it to function as a repressor in the absence of its normal heterodimeric partner a1?

Obviously, we cannot answer these questions definitively, but there are some tantalizing clues from a variety of studies. For example, there is a precedent for deleting genes from a mating type locus; indeed, deletion of the HMG-box a2-type gene appears to be part of the natural life cycle in some species of the filamentous ascomycete *Ceratocystis* (Harrington and McNew, 1997; Witthuhn et al., 2000). A similar event (or perhaps an error in the gene conversion event underlying mating-type interconversion) could explain why a2 is missing in *S. cerevisiae*.

The loss of HMG-box a2 from *S. cerevisiae* has apparently been compensated for by α2 taking on a new function since *C. albicans* and *S. cerevisiae* diverged. In addition to working with a1 (as it does in *C. albicans*), α2 acts independently of a1 to negatively regulate **a**-specific genes. Previous work suggests that α2's ability to act with and without a1 evolved separately, as these two activities are dependent on different portions of the α2 protein (Mak and Johnson, 1993; Mead et al., 1996; Vershon and Johnson, 1993). Based on the results from this paper as well as existing information about regulation of **a**-specific genes in *S. cerevisiae*, we propose a hypothetical series of events accounting for how an ancestral a2-controlled positive circuit branch was replaced by an α2-controlled negative circuit branch in *S. cerevisiae* (figure 6).

4) Incorporation of an Additional Level of Control of Mating

The most visually dramatic difference between the mating-circuitry of *S. cerevisiae* and *C. albicans*, apparent even to the naked eye, is the white-opaque phenotypic transition that precedes mating in *C. albicans* (for reviews, see Johnson, 2003; Soll et al., 2003). This transition was previously shown to be under mating-type control, being repressed by a1-α2 (Miller and Johnson, 2002). In this paper, we have examined genes repressed by a1-α2 in the white phase, as well as those indirectly regulated via the white-opaque transition.

In exponentially dividing *S. cerevisiae* cells, a1-α2 represses approximately 20-30 genes (Galitski et al., 1999). In contrast, only 7 genes are downregulated by a1-α2 in exponentially growing *C. albicans* white cells. In *S. cerevisiae*, genes repressed by a1-α2 include genes involved in mating as well as genes specific to the haploid lifestyle, e.g. budding pattern determinants and DNA repair enzymes (Galitski et al., 1999; Valencia et

al., 2001). Homologs of some of these haploid-lifestyle genes are not repressed by a1-α2 in *C. albicans*, potentially indicating the lack of haploid-diploid life-cycle differences.

While direct a1-α2 repression in *C. albicans* white cells is sparse, indirect regulation via white-to-opaque switching is extensive, with the total number of regulated genes (approximately 197 genes down-regulated and 237 up-regulated) greatly exceeding that regulated by a1-α2 in growing *S. cerevisiae* cells. A few genes involved in mating are among those indirectly repressed by a1-α2; however, most opaque-specific genes appear to have no role in mating. As discussed in Lan et al. (2002), this set of genes concerns many aspects of cell biology, including metabolic specialization, cell surface properties, and virulence.

We do not know whether white-opaque switching developed independently of the *MTL* locus and later came to be under its control, or whether it developed under *MTL* regulation in the first place. However, phenotypic switching (of which white-opaque switching is an example) is a common phenomenon in many mammalian pathogens, and it has been suggested that *C. albicans* phenotypic variants are able to infect vastly different host microenvironments (Soll, 1992). Consistent with this view, opaque cells are much less efficient at host colonization than their white counterparts in a mouse tail-vein model of infection, yet are more robust in a mouse skin model of infection (Kvaal et al., 1999; Kvaal et al., 1997). Host microenvironments also likely affect mating; given the complex and hostile mammalian host response, it has been proposed that some aspects of mating, such as pheromone signaling and membrane fusion, may be incompatible with life in some microenvironments but compatible with others (Gow, 2002). Consistent with this idea, skin surfaces appear to be amenable to mating (Lachke, 2003). It seems likely

that white-opaque switching is a relatively recent evolutionary novelty, probably brought about by the pressure of life in a mammalian host. However, white-opaque switching is under the control of an ancient regulatory circuit, and this coupling ensures that mating is closely integrated with host environments.

Summary

By dissecting the inner workings of the *C. albicans* mating-type regulatory circuit and comparing it to that of *S. cerevisiae*, we have been able to document many intriguing and specific changes in two circuits regulating mating-type that diverged more than 100 million years ago. Because each of the regulators encoded at the mating type loci represents a highly conserved DNA-binding motif, we believe that these specific circuit divergences illustrate several general mechanisms by which eukaryotic regulatory circuits can change over evolutionary timescales.

EXPERIMENTAL PROCEDURES

Strain construction

All strains were derived from CAI4 (Δura3::imm434/Δura3::imm434) (Fonzi and Irwin, 1993). In different isolates, the α1 and α2 genes were deleted either using strategies and constructs outlined in Hull et al (2000), or a PCR method (Wilson et al., 2000). The a1 and a2 genes were knocked out using strategies outlined Wilson et al., 2000. The following PCR primers were used to knock out

a1:

5'AATATTGATTAGAGGCACAAAATAAAAATCACCTTCAACCTCCTCGTTTTTT

CCAAAGCAGTTTTCCCAGTCACGACGTT3' and

5'ATACCACTTACTATTACTACTACTTACCACCACTTATTTCCAGTGGAGTTATT

GTGGCTATGTGGAATTGTGAGCGGATA3'.

a2:

5'GATGCATTTTTTAACACGTTCTCTCTTTTTGATTAGTTTTTAAGACAAATACC

CAATAATGTTTTCCCAGTCACGACGTT3' and

5'GCCTCTAATCAATATTTTTAGAGGGAAAAAAAGTTATTTATCACAATACGCA

GAAAAAAATGTGGAATTGTGAGCGGATA3'.

α1:

5'CTCTAATTTTTTAACAATTCCGTTCGACTATGTGTAACAGAAGATACCACTT

CAACAGATGTTTTCCCAGTCACGACGTT3' and

5'ATATATATTTAATGTTATCAACTTGCTGTATATTTTACTTCATTATGTAAACA

TCCTCAATGTGGAATTGTGAGCGGATA3'.

α2:

5'ATGAATTCACATCTGGAGGCACTCTTTGAAATAAATCAACGCTTGCTGTCTC

ACATAAGAGTTTTCCCAGTCACGACGTT3' and

5'CCTATCAATATGTTAGTAGGGTTACAAAGAATGAAAATTAATTAGTTATTTG

ATACATACTGTGGAATTGTGAGCGGATA3'.

Strains were switched to the opaque phase as previously described (Miller and Johnson, 2002).

Quantitative mating analysis

Quantitative mating analysis was previously described (Miller and Johnson, 2002).

Preparation of cultures and cDNA for microarray experiments

For white and opaque cultures, 1 ml cultures were grown overnight at 23°C in SC+100μg/ml uridine + 55 μg/ml adenine, and used to inoculate 100ml cultures of either SC+100μg/ml uridine + 55 μg/ml adenine, or YEPD + 55 μg/ml adenine at various temperatures. Cultures were harvested at OD 1.0 by vacuum filtration, and stored at -80°C. For α-factor induction experiments, cultures were grown as described above to OD 1.0 in SC+100μg/ml uridine + 55 μg/ml adenine at 23°C. An aliquot was taken as a zero time-point. The culture was split into two. To one half, α-factor dissolved in 10% DMSO was added to the culture to a final concentration of 10 μg/ml. To the other half, an equal quantity of 10% DMSO was added. The cultures were grown an additional 4 hours at 23°, and were then harvested by vacuum filtration. cDNA was prepared as previously described (Bennett et al., 2003). Construction and analysis of *C. albicans* microarrays was also previously described (Bennett et al., 2003)

FIGURE 1

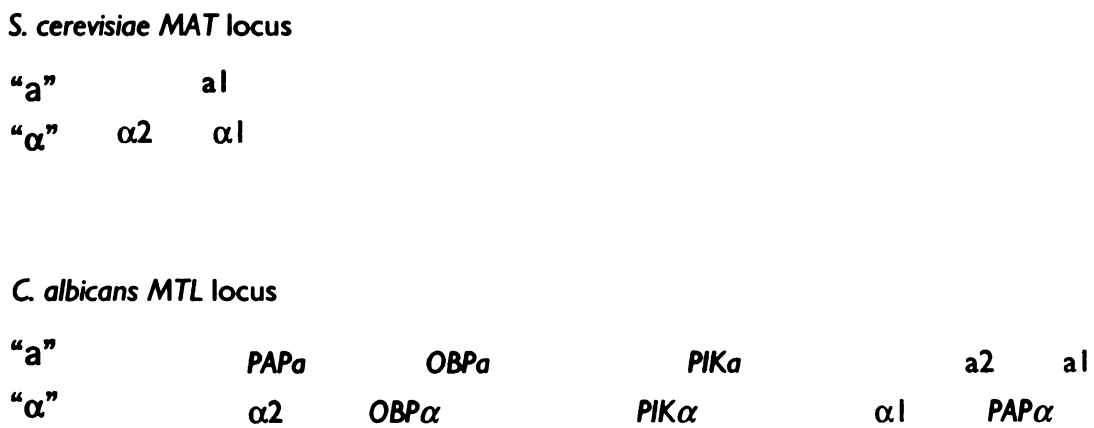


Figure 1. The *C. albicans* and *S. cerevisiae* mating type loci

The *S. cerevisiae MAT* locus has two alleles, *MAT***a** and *MATα*. Cells carrying only *MAT**a*** mate as **a** cells, cells carrying only *MATα* mate as α cells, and cells carrying both (**a**/α cells) do not mate. Together, *MAT**a*** and *MATα* encode three transcriptional regulators: a1, a1, and a2. The *MTL* of *C. albicans* also consists of two alleles. As in *S. cerevisiae*, cells carrying only *MTL**a*** are **a**-maters, cells carrying only *MTLα* are α-maters, and cells carrying both are **a**/α non-maters. *MTL**a*** encodes the transcriptional regulators a1 and a2, and *MTLα* encodes the transcriptional regulators α1 and α2. In addition to these regulators, each *MTL* allele encodes a poly(A) polymerase, a phosphatidyl inositol kinase, and an oxysterol binding protein. The *MAT* locus of *S. cerevisiae* and the *MTL* locus of *C. albicans* resemble each other in the arrangement, direction of transcription, and intron positions of their a1, α1, and α2 homologs.

Line	a2 / α2	a1 / α1	MTL configuration of experimental strain a1	a2	α1	α2	Ability to W-O switch Yes/No	Frequency of mating with opaque testers, where experimental strain is white x α-type tester	x a-type tester	Frequency of mating with opaque testers, where experimental strain is opaque x α-type tester	x a-type tester
1			+	+	+	+	No	-	-	n/a	n/a
2		Δ	-	+	+	+	Yes	9.1 x 10-8	3.3 x 10-7	1.1 x 10-6	3.1 x 10-4
3		Δ	+	+	-	+	No	-	-	n/a	n/a
4		ΔΔ	-	+	-	+	Yes	7.8 x 10-5	-	2.8 x 10-3	-
5	Δ		+	+	+	-	Yes	9.5 x 10-8	7.5 x 10-7	8.3 x 10-6	5.0 x 10-5
6	Δ	Δ	-	+	+	-	Yes	6.7 x 10-8	8.3 x 10-8	5.3 x 10-6	9.5 x 10-5
7	Δ	Δ	+	+	-	-	Yes	6.3 x 10-5	-	6.3 x 10-3	-
8	Δ	ΔΔ	-	+	-	-	Yes	3.3 x 10-5	-	3.3 x 10-3	-
9	Δ		+	-	+	+	No	-	-	n/a	n/a
10	Δ	Δ	-	-	+	+	Yes	-	2.7 x 10-6	-	4.8 x 10-3
11	Δ	Δ	+	-	-	+	No	-	-	n/a	n/a
12	Δ	ΔΔ	-	-	-	+	Yes	-	-	-	-
13	ΔΔ		+	-	+	-	Yes	-	1.3 x 10-5	-	1.7 x 10-3
14	ΔΔ	Δ	-	-	+	-	Yes	-	1.2 x 10-6	-	1.6 x 10-3
15	ΔΔ	Δ	+	-	-	-	Yes	-	-	-	-
16	ΔΔ	ΔΔ	-	-	-	-	Yes	-	-	-	-

where - is < 1.0x10-8

Table 1. Mating behavior of strains bearing all possible combinations of *MTL* transcriptional regulators

Strains bearing all possible combinations of *MTL*-encoded transcriptional regulators were allowed to mate with **a** and α-type tester strains. From left to right, the table shows the *MTL* configuration of each strain, its ability to undergo white-to-opaque switching, and its mating efficiency in two types of test crosses against **a**- and α-type tester strains. In the first test cross, the experimental strain was in the white phase, and both **a**- and α-type tester strains were in the opaque phase. The strains were mated for 4 days at 23°. In the second test cross, both the experimental strain and the tester strains were in the opaque phase (those unable to switch to the opaque phase are marked "n/a" in the results columns). The strains were mated for 2 days at 23°.

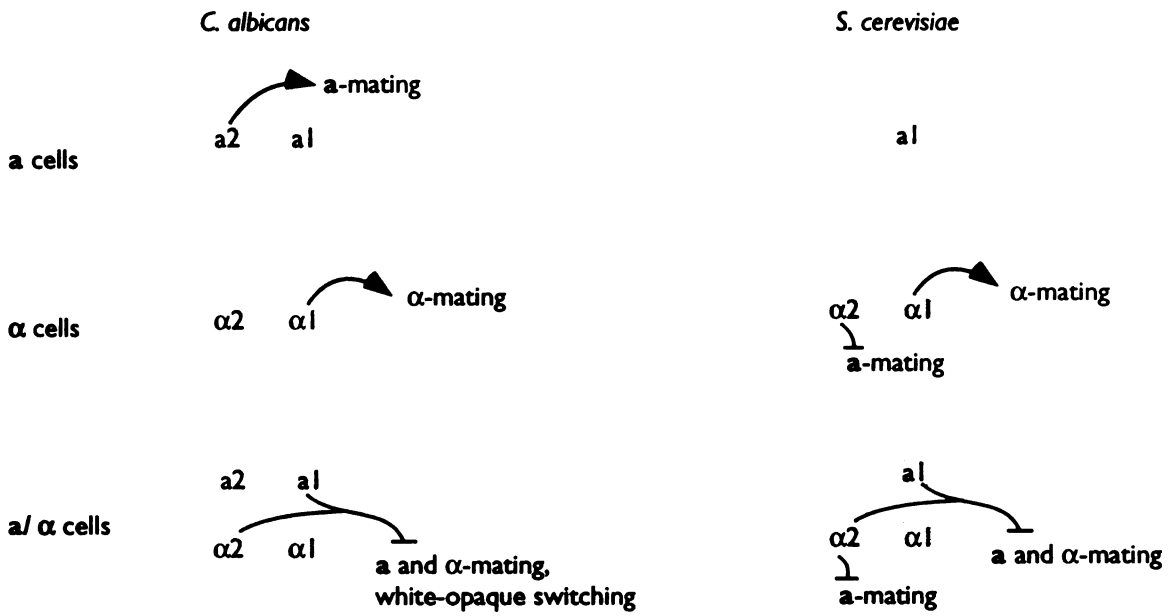


Figure 2. A genetic circuit of mating-type determination in *C. albicans*

In *C. albicans,* a2 encoded by *MTLa* positively regulates **a**-type mating, and α1 encoded by *MTLα* positively regulates α-type mating. In an **a**/α cell which carries both *MTL* alleles, a1 and α2 work together to repress the white-opaque switch and mating. This circuit differs from that of *S. cerevisiae*, where **a**-type mating requires no input from the *MAT* locus. Instead, **a**-type mating is regulated negatively by α2.

Figure 3

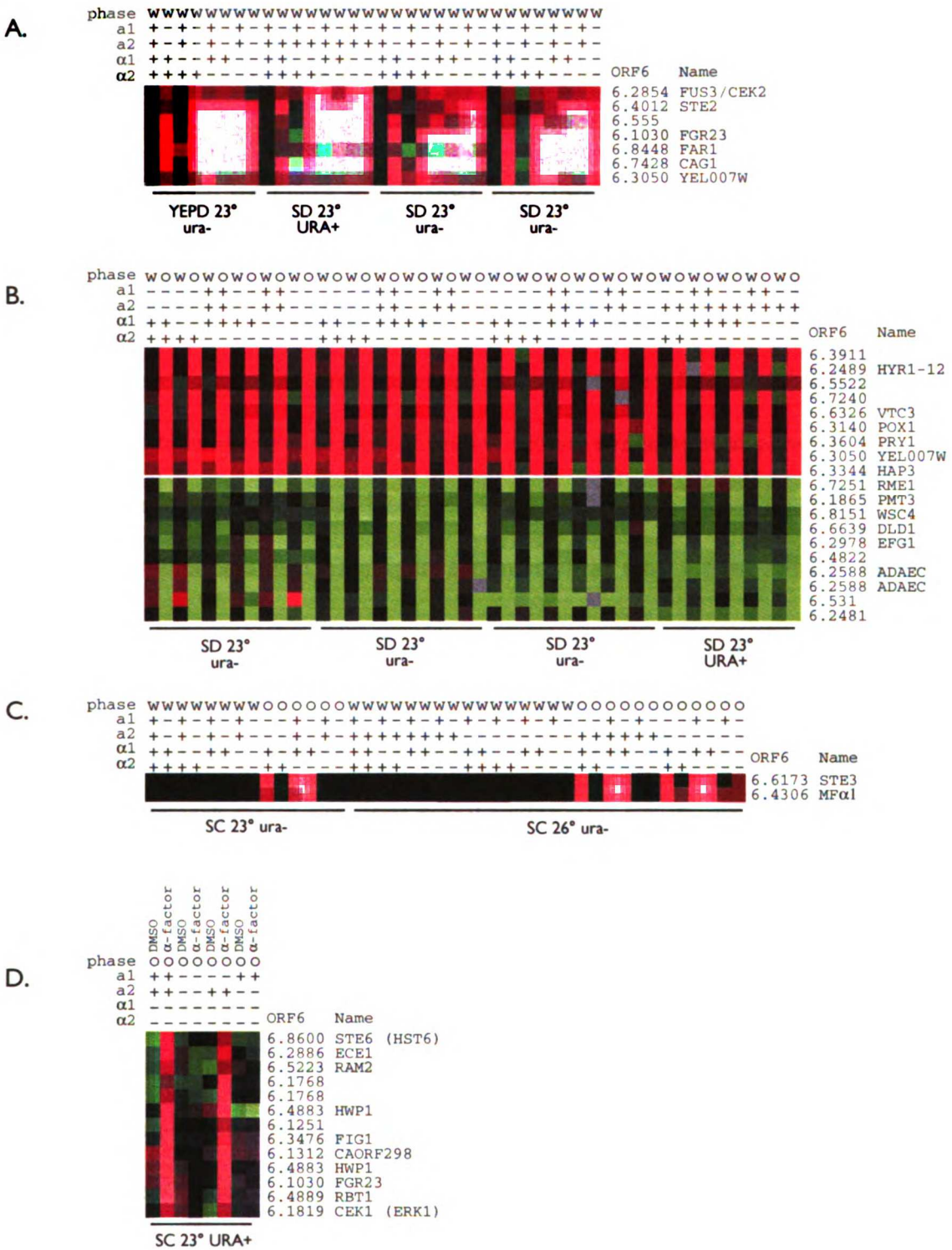


Figure 3. Transcriptional profiling of strains bearing all combinations of *MTL* transcriptional regulators, in both white and opaque phases

mRNA was prepared from each of the strain summarized in Table I (in both white and opaque phases) and hybridized to DNA microarrays containing ~10,000 PCR products and representing ~6,300 ORFs, with many ORFs represented by more than one spot. Each experiment was replicated several times, in most cases using two independently constructed strains of the same genotype. 2-fold and greater changes were considered significant when consistent across all experimental replicates. Each spot has been given a standard ORF6 number (sequences for each ORF6 are available at http://www-sequence.stanford.edu/group/candida), and if available, a name based on homology to a gene from another organism. This figure summarizes the most salient points of the analyses.

(A) Genes repressed by a1-α2 in white cells

White strains, carrying multiple *MTL* configurations, were grown at 23° in YEPD or SC as indicated. In each experiment, data for all strains were transformed to the strain carrying intact *MTLa* and *MTLα* alleles (that is, an a/α strain). Shown are genes that are derepressed in all strains lacking a1, α2, or both. These results were obtained in four additional trials using both URA+ and ura- cells (data not shown).

(B) Opaque-specific genes and white-specific genes

URA+ and ura- strains carrying multiple *MTL* configurations were switched to the opaque phase. Opaque and white cells from the same colony were subcultured and mRNA was prepared for analysis. In each experiment, the data were transformed to a white-phase control strain carrying intact *MTLa* and *MTLα* alleles. A subset of the

opaque-specific and white-specific gene clusters are shown. In all, 237 genes are upregulated in the opaque phase, and 197 genes are downregulated (see Supplementary tables). Note that these white- and opaque-specific genes are regulated independently of the *MTL* configuration, aside from the requirement that the strain lack a1-α2 repression.

(C) α-specific genes

White and opaque strains carrying all *MTL* configurations were transformed to a white-phase control strain carrying intact *MTLa* and *MTLα* alleles. As shown, expression of *MFα1* (the α-factor structural gene) and *STE3* (a-factor receptor) is dependent on two criteria: the strain must carry α1, and it must be in the opaque phase. The genes are upregulated ~300-fold and ~1000-fold, respectively, compared to the **a**/α strain.

d. **a**-specific genes

Δ*MTLa1*Δ*MTLa2* opaque strains carrying a1 and a2, a1 alone, a2 alone, or neither, were treated with 10 μg/ml α-factor in DMSO or DMSO alone for 4 hours. Data were transformed to the respective 0-hour time-point sample (see Experimental Procedures). As shown, the transcriptional response to α-factor is dependent on a2, but not a1.

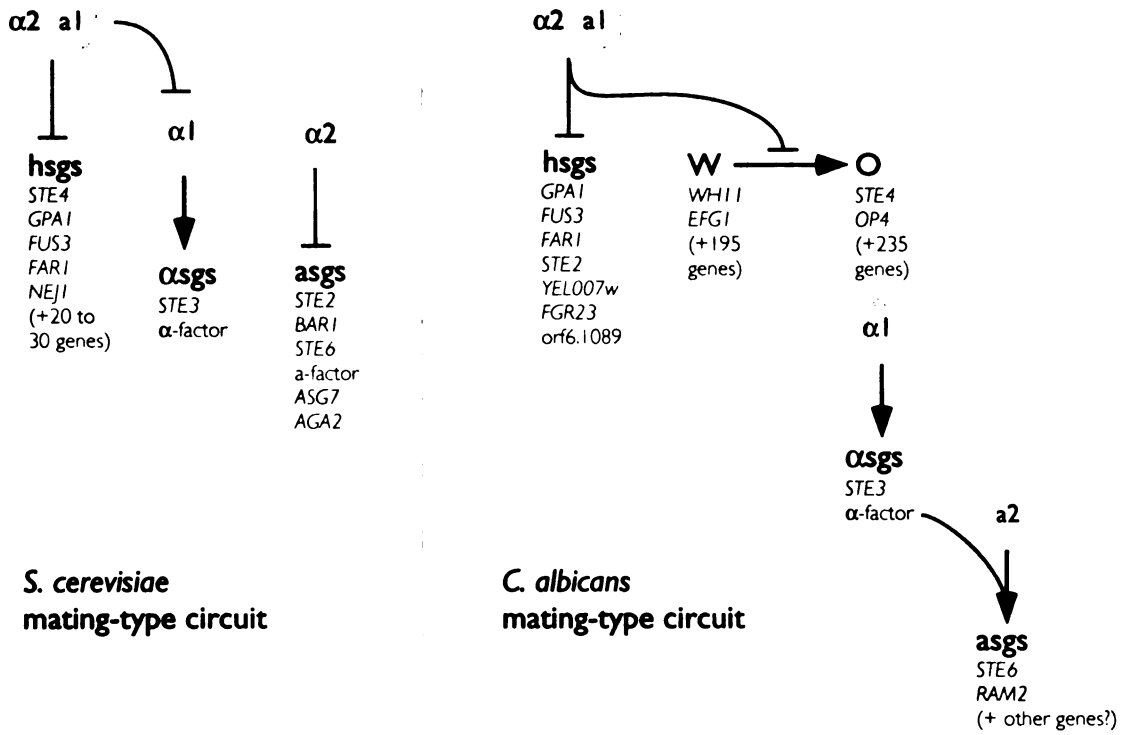


Figure 4. Transcriptional circuit regulating mating-type in *C. albicans* and *S. cerevisiae*

In *C. albicans* white phase cells, the homeodomain proteins a1 and α2 repress a small number of genes directly, some of which share homologs with the so-called "haploid specific genes" (hsgs) in *S. cerevisiae*. Although the nomenclature is not relevant to *C. albicans*, which has no known haploid phase, we will nonetheless refer to genes as "haploid-specific" to preserve the analogy with *S. cerevisiae*. a1 and α2 are also required to repress the white-to-opaque switch in *C. albicans*. In the absence of either component of this heterodimer, the cell can switch to the opaque phase, which is accompanied by the up- and down-regulation of more than 400 genes. α-specific genes (αsgs) are turned on

in α-cells by α1, but only if the cells are in the opaque phase. While naïve opaque **a**-cells do not detectably express **a**-specific transcripts, they show upregulation of 65 genes upon exposure to α-factor, some of which have **a**-specific homologs in *S. cerevisiae*. We have provisionally included these genes as **a**-specific genes (**a**sgs). There are two major differences between the *S. cerevisiae* and *C. albicans* mating circuitry. First, in contrast to *S. cerevisiae*, relief of a1-α2 repression is necessary but not sufficient for mating-competency in *C. albicans*. *C. albicans* must further undergo the environment-dependent white-to-opaque transition to achieve mating competency, constituting a layer of transcriptional control over mating that is not present in *S. cerevisiae*. Second, individual components of the mating-type circuit have assumed different roles in the two yeasts; while **a**-type mating is positively regulated by a2 in *C. albicans*, it is negatively regulated in *S. cerevisiae* by α2. Despite these changes, the overall output of the circuit is the same in both organisms: **a**-cells mate with α-cells and form a non-mating conjugant.

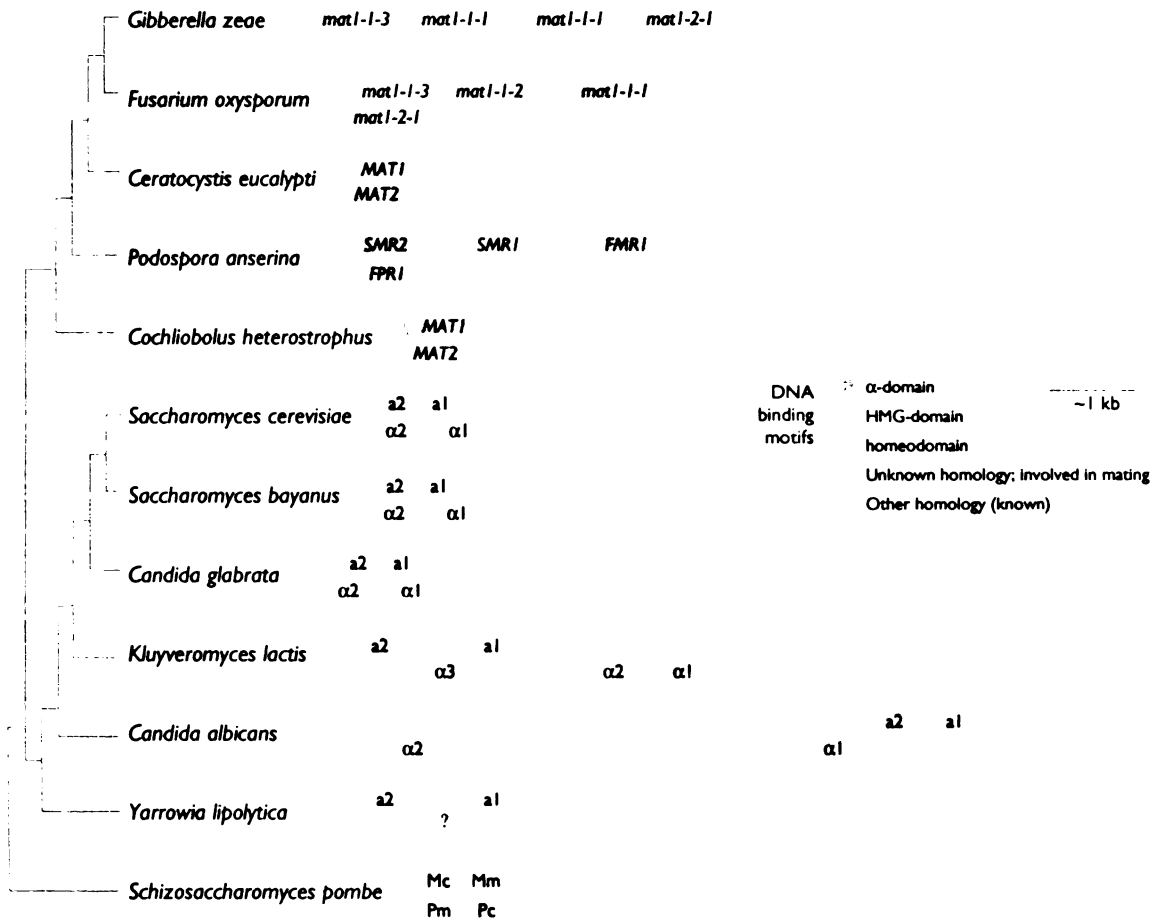


Figure 5. Family tree of fungal mating type loci

Shown are diagrams of mating type loci from various ascomycetes, arranged on a tree (based on ribosomal RNA subunit sequences) indicating their phylogenetic relationships. An HMG-box homolog at the *MAT* locus is present in many modern species ranging from the filamentous ascomycete *C. heterostrophus* to *S. pombe*. The *S. cerevisiae* lineage lacks this regulator, suggesting that the lineage lost this regulator from its *MAT* locus in relatively recent evolutionary time, sometime after the divergence of *S. cerevisiae* and *K. lactis*. (This figure is based on Astrom et al., 2000; Casselton and Olesnicky, 1998;

Cliften et al., 2003; Coppin et al., 1997; Debuchy et al., 1993; Harrington and McNew, 1997; Herskowitz, 1992; Hull and Johnson, 1999; Kellis et al., 2003; Kelly et al., 1988; Keogh et al., 1998; Kronstad and Staben, 1997; Kurischko et al., 1999; Picard et al., 1991; Souciet et al., 2000; Srikantha et al., 2003; Turgeon et al., 1993; Witthuhn et al., 2000; Wong et al., 2002; Wong et al., 2003; Yun et al., 2000).

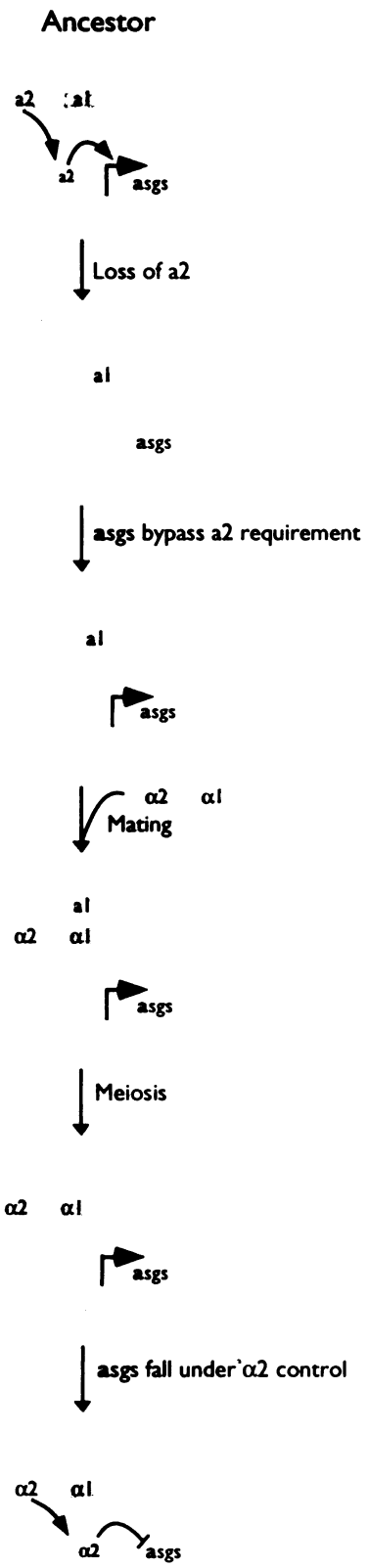


Figure 6. Model for the evolutionary change in a-specific gene regulation

Shown is a plausible scheme by which a positive regulator of **a**-specific genes in an ancestor of *C. albicans* and *S. cerevisiae* could have been replaced by a negative regulator in modern *S. cerevisiae*. First, a2 was lost (see Discussion). Subsequently, a minimal set of **a**-specific genes would have to bypass the a2 requirement for activation in order for the strain to re-enter the mating population. This might occur by co-opting an abundant and/or highly adaptable regulator of transcription; in *S. cerevisiae*, Mcm1 has fulfilled this role (for review, see Johnson, 1995). Upon mating, **a**-specific genes would be misexpressed in the resulting **a**/α strain. Meiotic progeny derived from this strain would include genotypically α-cells that misexpressed **a**-specific genes; this type of cell can mate, but does so inefficiently (Table I; see also Siliciano and Tatchell, 1986). Efficient α-type mating would be restored when **a**-specific genes came under negative control. Although many mechanisms of repressing

a-specific genes are plausible, an activity encoded at *MAT*α such asα2, would be highly advantageous since it would continue to confer α-mating efficiency in α-cells following the independent assortment of alleles associated with meiotic cycling. In *S. cerevisiae*, α2 works with Mcm1 to repress **a**-specific genes (Keleher et al., 1989). The use of Mcm1 and α2 provides a simple and efficient solution to the requirement for both activation and regulated repression of **a**-specific genes following loss of a2.

ACKNOWLEDGEMENTS

We are very grateful to Kenneth Wolfe for generously bringing the existence of the *MTLa2* ORF to our attention, and to Mike Lorentz and Gerald Fink for the collaboration that produced the DNA microarrays used in this paper. We are also grateful to the Stanford Genome Technology Center (http://www-sequence.stanford.edu/group/candida) for providing sequence data for *C. albicans*. Sequencing of the *C. albicans* genome was supported by NIDR and the Burroughs Wellcome Fund. We thank Diane Inglis, Andrew Uhl, and Burk Braun for their extensive efforts in genome annotation, Bethann Hromatka for overseeing microarray printing, and Richard Bennett and members of the Johnson lab for many fruitful discussions. This work was supported by grants from the Burroughs Wellcome Fund (993218) and the NIH (R01 AI49187) to A.D.J.. A.E.T. and M.G.M. are Howard Hughes Medical Institute Predoctoral Fellows.

REFERENCES

Astrom, S. U., Kegel, A., Sjostrand, J. O., and Rine, J. (2000). Kluyveromyces lactis Sir2p regulates cation sensitivity and maintains a specialized chromatin structure at the cryptic alpha-locus. Genetics *156*, 81-91.

Averof, M., and Patel, N. H. (1997). Crustacean appendage evolution associated with changes in Hox gene expression. Nature *388*, 682-686.

Bennett, R. J., Uhl, M. A., Miller, M. G., and Johnson, A. D. (2003). Identification and characterization of a mating pheromone in *C. albicans*. Mol Cell Biol *In press*.

Calderone, R. A. (2002). Candida and Candidiasis (Washington, D.C., ASM Press).

Carroll, S. B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. Cell *101*, 577-580.

Carroll, S. B., Grenier J. K., Weatherbee, S. D. (2001). From DNA to Diversity (Malden, Massachusetts, Blackwell Science, Inc.).

Casselton, L. A., and Olesnicky, N. S. (1998). Molecular genetics of mating recognition in basidiomycete fungi. Microbiol Mol Biol Rev *62*, 55-70.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A., and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science *301*, 71-76.

Consortium, *C. e. S.* (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium. Science *282*, 2012-2018.

Coppin, E., Debuchy, R., Arnaise, S., and Picard, M. (1997). Mating types and sexual development in filamentous ascomycetes. Microbiol Mol Biol Rev *61*, 411-428.

Davidson, E. H. (2001). Genomic Regulatory Systems (San Diego, CA, Academic Press).

de Bono, M., and Hodgkin, J. (1996). Evolution of sex determination in caenorhabditis: unusually high divergence of tra-1 and its functional consequences. Genetics *144*, 587-595.

Debuchy, R., Arnaise, S., and Lecellier, G. (1993). The mat- allele of Podospora anserina contains three regulatory genes required for the development of fertilized female organs. Mol Gen Genet *241*, 667-673.

Doebley, J., and Lukens, L. (1998). Transcriptional regulators and the evolution of plant form. Plant Cell *10*, 1075-1082.

Duboule, D., and Wilkins, A. S. (1998). The evolution of 'bricolage'. Trends Genet *14*, 54-59.

Fonzi, W. A., and Irwin, M. Y. (1993). Isogenic strain construction and gene mapping in Candida albicans. Genetics *134*, 717-728.

Galant, R., and Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. Nature *415*, 910-913.

Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999). Ploidy regulation of gene expression. Science *285*, 251-254.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994). Homeodomain-DNA recognition. Cell *78*, 211-223.

Gerhart, J., and Kirschner, M. (1997). Cells, Embryos, and Evolution (Malden, Massachusetts, Blackwell Science, Inc.).

Gow, N. A. (2002). Candida albicans switches mates. Mol Cell *10*, 217-218.

Grosschedl, R., Giese, K., and Pagel, J. (1994). HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. Trends Genet *10*, 94-100.

Haag, E. S., Wang, S., and Kimble, J. (2002). Rapid coevolution of the nematode sex-determining genes fem-3 and tra-2. Curr Biol *12*, 2035-2041.

Harrington, T. C., and McNew, D. L. (1997). Self-fertility and uni-directional mating-type switching in Ceratocystis coerulescens, a filamentous ascomycete. Curr Genet *32*, 52-59.

Hedges, S. B. (2002). The origin and evolution of model organisms. Nat Rev Genet *3*, 838-849.

Herskowitz, I., Rine, J., and Strathern, J. (1992). Mating-type determination and mating-type interconversion in *Saccharomyces cerevisiae*. In The Molecular Cellular Biology of the Yeast *Saccharomyces*, J. R. P. E.W. Jones, J.R. Broach, ed. (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press), pp. 583-657.

Hull, C. M., and Johnson, A. D. (1999). Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans. Science *285*, 1271-1275.

Hull, C. M., Raisner, R. M., and Johnson, A. D. (2000). Evidence for mating of the "asexual" yeast Candida albicans in a mammalian host. Science *289*, 307-310.

Johnson, A. D. (1995). Molecular mechanisms of cell-type determination in budding yeast. Curr Opin Genet Dev *5*, 552-558.

Johnson, A. D. (2003). The biology of mating in *Candida albicans*. Nat Rev Microbiol, in press.

Keleher, C. A., Passmore, S., and Johnson, A. D. (1989). Yeast repressor alpha 2 binds to its operator cooperatively with yeast protein Mcm1. Mol Cell Biol *9*, 5228-5230.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241-254.

Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988). Four mating-type genes control sexual differentiation in the fission yeast. Embo J *7*, 1537-1547.

Keogh, R. S., Seoighe, C., and Wolfe, K. H. (1998). Evolution of gene order and chromosome number in Saccharomyces, Kluyveromyces and related fungi. Yeast *14*, 443-457.

Kjaerulff, S., Dooijes, D., Clevers, H., and Nielsen, O. (1997). Cell differentiation by interaction of two HMG-box proteins: Matl-Mc activates M cell-specific genes in S. pombe by recruiting the ubiquitous transcription factor Stel1 to weak binding sites. Embo J *16*, 4021-4033.

Kronstad, J. W., and Staben, C. (1997). Mating type in filamentous fungi. Annu Rev Genet *31*, 245-276.

Kurischko, C., Schilhabel, M. B., Kunze, I., and Franzl, E. (1999). The MATA locus of the dimorphic yeast Yarrowia lipolytica consists of two divergently oriented genes. Mol Gen Genet *262*, 180-188.

Kvaal, C., Lachke, S. A., Srikantha, T., Daniels, K., McCoy, J., and Soll, D. R. (1999). Misexpression of the opaque-phase-specific gene PEP1 (SAP1) in the white phase of Candida albicans confers increased virulence in a mouse model of cutaneous infection. Infect Immun *67*, 6652-6662.

Kvaal, C. A., Srikantha, T., and Soll, D. R. (1997). Misexpression of the white-phase-specific gene WH11 in the opaque phase of Candida albicans affects switching and virulence. Infect Immun *65*, 4468-4475.

Lachke, S. A., Lockhart, S.R., Daniels, K.J., and Soll, D.R. (2003). Skin facilitates *Candida albicans* mating. Infect Immun *71*, 4970-4976.

Lan, C. Y., Newport, G., Murillo, L. A., Jones, T., Scherer, S., Davis, R. W., and Agabian, N. (2002). Metabolic specialization associated with phenotypic switching in Candidaalbicans. Proc Natl Acad Sci U S A *99*, 14907-14912.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.

Lockhart, S. R., Pujol, C., Daniels, K. J., Miller, M. G., Johnson, A. D., Pfaller, M. A., and Soll, D. R. (2002). In Candida albicans, white-opaque switchers are homozygous for mating type. Genetics *162*, 737-745.

Magee, B. B., Legrand, M., Alarco, A. M., Raymond, M., and Magee, P. T. (2002). Many of the genes required for mating in Saccharomyces cerevisiae are also required for mating in Candida albicans. Mol Microbiol *46*, 1345-1351.

Magee, B. B., and Magee, P. T. (2000). Induction of mating in Candida albicans by construction of MTLa and MTLalpha strains. Science *289*, 310-313.

Mak, A., and Johnson, A. D. (1993). The carboxy-terminal tail of the homeo domain protein alpha 2 is required for function with a second homeo domain protein. Genes Dev *7*, 1862-1870.

McGuire, I. C., Marra, R. E., Turgeon, B. G., and Milgroom, M. G. (2001). Analysis of mating-type genes in the chestnut blight fungus, Cryphonectria parasitica. Fungal Genet Biol *34*, 131-144.

Mead, J., Zhong, H., Acton, T. B., and Vershon, A. K. (1996). The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. Mol Cell Biol *16*, 2135-2143.

Miller, M. G., and Johnson, A. D. (2002). White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. Cell *110*, 293-302.

O'Neil, M. T., and Belote, J. M. (1992). Interspecific comparison of the transformer gene of Drosophila reveals an unusually high degree of evolutionary divergence. Genetics *131*, 113-128.

Picard, M., Debuchy, R., and Coppin, E. (1991). Cloning the mating types of the heterothallic fungus Podospora anserina: developmental features of haploid transformants carrying both mating types. Genetics *128*, 539-547.

Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. Nature *415*, 914-917.

Scott, M. P., Tamkun, J. W., and Hartzell, G. W., 3rd (1989). The structure and function of the homeodomain. Biochim Biophys Acta *989*, 25-48.

Sharkey, M., Graba, Y., and Scott, M. P. (1997). Hox genes in evolution: protein surfaces and paralog groups. Trends Genet *13*, 145-151.

Slutsky, B., Staebell, M., Anderson, J., Risen, L., Pfaller, M., and Soll, D. R. (1987). "White-opaque transition": a second high-frequency switching system in Candida albicans. J Bacteriol *169*, 189-197.

Soll, D. R. (1992). High-frequency switching in Candida albicans. Clin Microbiol Rev *5*, 183-203.

Soll, D. R., Lockhart, S. R., and Zhao, R. (2003). Relationship between Switching and Mating in Candida albicans. Eukaryot Cell *2*, 390-397.

Souciet, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de Montigny, J., Dujon, B., *et al.* (2000). Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett *487*, 3-12.

Sprague, G. F., Jr., and Thorner, J. (1992). Pheromone response and signal transduction during the mating process of *Saccharomyces cerevisiae*. In *The Molecular and Cellular Biology of the Yeast*, E. W. Jones, J. R. Pringle, and J. R. Broach, eds. (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press), pp. 657-744.

Srikantha, T., Lachke, S. A., and Soll, D. R. (2003). Three Mating Type-Like Loci in Candida glabrata. Eukaryot Cell *2*, 328-340.

Turgeon, B. G., Bohlmann, H., Ciuffetti, L. M., Christiansen, S. K., Yang, G., Schafer, W., and Yoder, O. C. (1993). Cloning and analysis of the mating type genes from

Cochliobolus heterostrophus. Mol Gen Genet *238*, 270-284.

Valencia, M., Bentele, M., Vaze, M. B., Herrmann, G., Kraus, E., Lee, S. E., Schar, P., and Haber, J. E. (2001). NEJ1 controls non-homologous end joining in Saccharomyces cerevisiae. Nature *414*, 666-669.

Vershon, A. K., and Johnson, A. D. (1993). A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. Cell *72*, 105-112.

Whitfield, L. S., Lovell-Badge, R., and Goodfellow, P. N. (1993). Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature *364*, 713-715.

Wilson, R. B., Davis, D., Enloe, B. M., and Mitchell, A. P. (2000). A recyclable Candida albicans URA3 cassette for PCR product-directed gene disruptions. Yeast *16*, 65-70.

Witthuhn, R. C., Harrington, T. C., Wingfield, B. D., Steimel, J. P., and Wingfield, M. J. (2000). Deletion of the MAT-2 mating-type gene during uni-directional mating-type switching in Ceratocystis. Curr Genet *38*, 48-52.

Wong, S., Butler, G., and Wolfe, K. H. (2002). Gene order evolution and paleopolyploidy in hemiascomycete yeasts. Proc Natl Acad Sci U S A *99*, 9272-9277.

Wong, S., Fares, M. A., Zimmermann, W., Butler, G., and Wolfe, K. H. (2003). Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata. Genome Biol *4*, R10.

Yun, S. H., Arie, T., Kaneko, I., Yoder, O. C., and Turgeon, B. G. (2000). Molecular organization of mating type loci in heterothallic, homothallic, and asexual Gibberella/Fusarium species. Fungal Genet Biol *31*, 7-20.

Appendices to Chapter 2

Appendix A. White and Opaque genes identified in microarrays.

White and Opaque-specific genes were determined by comparison of white- and opaque-phase strains bearing all combinations of the *MTL* locus. 35 individual experiments (70 microarrays) were averaged to give the fold-upregulated figures.

Complete list of Opaque specific genes

orf19	orf6	Fold upregulated, Opaques	Name
orf19.4520	orf6.4991	2.7	YDR248C
orf19.854	orf6.1903	2.7	UGA1
orf19.5033	orf6.7976	2.8	APG12
orf19.6705	orf6.6814	2.8	YBL060W
orf19.6788	orf6.8843	2.8	
orf19.11826	orf6.3993	2.9	YOL070C
orf19.1313	orf6.4875	2.9	CDR3a
orf19.6830	orf6.8801	3.0	PXEL
orf19.5268	orf6.7674	3.0	NUT2
orf19.2317	orf6.5533	3.0	
orf19.4150	orf6.5229	3.1	YBR014C
orf19.5029	orf6.7980	3.1	YDR061W
orf19.3434	orf6.6471	3.1	YGR067C-1
orf19.13909	orf6.7544	3.2	
orf19.2837	orf6.4498	3.2	ALG5
orf19.937	orf6.8186	3.2	
orf19.5613	orf6.727	3.3	RNH1
orf19.6443	orf6.7811	3.3	ECI1-2
orf19.7303	orf6.8384	3.5	
orf19.3708	orf6.5306	3.5	SAP2-2
orf19.5455	orf6.8691	3.6	NRF1
orf19.2133	orf6.5367	3.6	LIP4
orf19.13150	orf6.3792	3.6	DIT2-2
orf19.7495	orf6.8747	3.6	
orf19.4591	orf6.7890	3.7	CAT2
orf19.12440	orf6.3143	3.7	HYR1-1
orf19.13685	orf6.5501	3.7	
orf19.2459	orf6.3561	3.8	
orf19.3444	orf6.6461	3.8	AZR1
orf19.2030	orf6.5969	3.9	
orf19.4584	orf6.7883	3.9	PHO11-2
orf19.10922	orf6.6107	3.9	CAR1-1
orf19.9167	orf6.1648	3.9	
orf19.7043	orf6.7969	3.9	YLR050C
orf19.3727	orf6.897	4.0	PHO11-1
orf19.5121	orf6.6329	4.0	OPT2-4
orf19.4157	orf6.2621	4.1	SPS19-2
orf19.6254	orf6.2269	4.1	ANT1
orf19.7071	orf6.8482	4.1	
orf19.94	orf6.3706	4.1	
orf19.6330	orf6.5523	4.1	

orf19.6008	orf6.9029	4.1	YLL032C
orf19.8015	orf6.6279	4.1	GCV2-1
orf19.7077	orf6.8476	4.1	FRE7b
orf19.9674	orf6.5361	4.1	
orf19.6083	orf6.6258	4.2	
orf19.5993	orf6.9044	4.2	KEL3
orf19.2638	orf6.3825	4.2	
orf19.12259	orf6.1471	4.3	
orf19.5190	orf6.3501	4.3	CA35A5.03
orf19.5992	orf6.9045	4.4	NAF1
orf19.3282	orf6.8282	4.4	
orf19.7017	orf6.7943	4.4	YOX1
orf19.4486	orf6.2885	4.4	
orf19.6028	orf6.3156	4.4	CLN2
orf19.3754	orf6.4212	4.4	PHO11-3
orf19.6600	orf6.6212	4.4	YLR193C
orf19.4121	orf6.3192	4.5	TES1
orf19.5561	orf6.3250	4.5	STE23
orf19.6996	orf6.7742	4.5	MNT2-2
orf19.1600	orf6.1649	4.5	
?	orf6.34	4.5	
orf19.4833	orf6.7378	4.6	MLS1
orf19.10847	orf6.4730	4.6	
orf19.6003	orf6.9034	4.6	
orf19.1651	orf6.827	4.8	
orf19.1562	orf6.6419	4.8	
orf19.2247	orf6.5899	4.8	
orf19.3127	orf6.5864	4.9	CZF1
orf19.3897	orf6.2342	4.9	
orf19.8334	orf6.2636	4.9	
orf19.2515	orf6.3920	4.9	
orf19.7513	orf6.8765	5.0	ALK2-A
orf19.3400	orf6.4331	5.0	
orf19.1066	orf6.4923	5.0	YPL067C
orf19.9381	orf6.1326	5.1	UME6
?	orf6.7397	5.1	
orf19.8614	orf6.5799	5.2	GCA1a
orf19.8920	orf6.1342	5.2	YDL124W-1
orf19.2936	orf6.4000	5.2	
orf19.2886	orf6.1819	5.3	CEK1 / ERK1
orf19.11211	orf6.1832	5.3	PHO12
orf19.4459	orf6.2833	5.4	YNL234W-1
orf19.6706	orf6.6815	5.4	GYP7
orf19.4381	orf6.6326	5.5	VTC2
orf19.5557	orf6.4127	5.5	
orf19.4476	orf6.4311	5.6	
orf19.9328	orf6.925	5.6	PHO23
orf19.341	orf6.7225	5.6	TPO1
orf19.3824	orf6.7519	5.7	
orf19.9673	orf6.864	5.7	
orf19.9755	orf6.3574	5.8	
orf19.10239	orf6.1970	5.8	
orf19.12670	orf6.3419	5.8	
orf19.1701	orf6.3042	5.8	RKI1
orf19.7521	orf6.8773	5.8	
orf19.2838	orf6.4499	5.9	TM1624
orf19.5141	orf6.5957	5.9	

orf19.11376	orf6.2344	5.9	CHT2
?	orf6.6673	5.9	HGT1-3
orf19.2633	orf6.1438	6.0	QUTD-2
orf19.2157	orf6.684	6.1	NAGPD
orf19.2475	orf6.4960	6.1	
orf19.5585	orf6.4427	6.2	
orf19.5453	orf6.8689	6.2	
orf19.3589	orf6.2047	6.2	
orf19.9167	orf6.1648	6.2	
orf19.2461	orf6.4974	6.3	PIRIN-4
orf19.3612	orf6.3282	6.3	PST1
orf19.12077	orf6.7906	6.3	SPAC23H3.04
orf19.4551	orf6.6697	6.3	YAT1
orf19.4651	orf6.305	6.4	
?	orf6.4255	6.4	
orf19.3363	orf6.1784	6.5	VTC4
orf19.1612	orf6.1337	6.5	
orf19.6983	orf6.7729	6.5	REG1
orf19.8419	orf6.1353	6.5	STE4
orf19.7094	orf6.8459	6.5	SNF3-2
orf19.7966	orf6.1276	6.6	
orf19.4899	orf6.2026	6.6	ROT2
orf19.11377	orf6.2343	6.7	
orf19.1767	orf6.2870	6.8	UBP8
orf19.9988	orf6.294	6.8	
?	orf6.2695	6.8	
orf19.5509	orf6.4357	6.9	
orf19.11755	orf6.3107	6.9	MNN1
orf19.10303	orf6.3604	6.9	PRY1
orf19.8864	orf6.2900	6.9	
orf19.11750	orf6.3112	7.0	PUT1
orf19.6518	orf6.7642	7.0	YMR110C-1
orf19.1539	orf6.2306	7.1	
orf19.6073	orf6.7617	7.2	HMX1
orf19.3029	orf6.7766	7.2	MRP5
orf19.3746	orf6.5476	7.3	OPT2-1
orf19.5169	orf6.6967	7.3	AMD2
orf19.1005	orf6.581	7.6	
orf19.5305	orf6.8294	7.8	CA38F10.05-2
orf19.5716	orf6.3803	8.1	SAP6-1
orf19.5952	orf6.9085	8.2	
orf19.2431	orf6.1394	8.3	
orf19.3982	orf6.5658	8.6	YOL157C
orf19.9179	orf6.1937	8.6	
orf19.6311	orf6.5504	8.6	
orf19.9498	orf6.5596	8.7	SDS22-2
orf19.12211	orf6.4019	8.7	
orf19.1766	orf6.2871	8.7	SCJ4.49
orf19.2701	orf6.5489	8.7	
orf19.159	orf6.144	8.8	
orf19.10860	orf6.2296	8.8	YOR246C-1
orf19.8043	orf6.2094	8.9	
orf19.5597	orf6.4439	9.1	POL5a
orf19.3749	orf6.4217	9.1	OPT2-6
orf19.12437	orf6.1556	9.1	CRZ1-4
orf19.2590	orf6.4108	9.4	BIO4
orf19.467	orf6.3214	9.5	

orf19.4082	orf6.6854	9.7	DDR48
orf19.4873	orf6.3832	9.7	
orf19.12603	orf6.5960	9.7	
orf19.7053	orf6.8500	9.7	GAC1-1
orf19.9703	orf6.2470	9.7	NAG1
orf19.6741	orf6.8890	9.8	YLR414C
orf19.4189	orf6.3646	9.8	PHO5YAR
orf19.531	orf6.6743	9.8	
orf19.1705	orf6.3122	9.9	POT1
orf19.10042	orf6.3911	9.9	
orf19.5902	orf6.9135	10.1	RAS1-2
orf19.7434	orf6.8594	10.1	GLG2-1
orf19.4971	orf6.1555	10.3	
orf19.1585	orf6.5335	10.6	ZRT2
orf19.3668	orf6.6479	10.7	HGT1-1
orf19.4211	orf6.5010	10.9	FET3-2a
orf19.6998	orf6.7744	11.0	GTT1-1
orf19.6066	orf6.7610	11.2	YMR110C-2
orf19.3794	orf6.3067	11.3	ZAP1
orf19.6731	orf6.6880	11.3	
orf19.2038	orf6.5977	11.7	
orf19.5307	orf6.991	12.0	JEN1-2
orf19.3232	orf6.8232	12.1	
orf19.4647	orf6.3344	12.7	HAP3
orf19.11233	orf6.5478	12.8	OPT2-2
orf19.6420	orf6.7834	12.9	
orf19.2810	orf6.4583	12.9	
orf19.6484	orf6.3710	13.0	
orf19.716	orf6.2637	13.1	CLOST1
orf19.2941	orf6.4005	13.4	SCW4
orf19.2254	orf6.5892	13.6	
orf19.2062	orf6.7493	13.7	EC-SOD
orf19.1830	orf6.1220	13.8	
orf19.8165	orf6.6744	13.9	
orf19.6334	orf6.5527	14.4	DAN4
orf19.13011	orf6.3246	15.1	D3HI
orf19.7596	orf6.8944	15.2	PMU1-3
orf19.460	orf6.2854	15.7	FUS3 / CEK2
orf19.3931	orf6.1915	16.2	SFC1
orf19.12767	orf6.8296	17.2	JEN1-3
orf19.2253	orf6.5893	19.0	
orf19.629	orf6.425	20.8	
orf19.1171	orf6.3438	21.4	
orf19.11981	orf6.4197	21.5	SADH
orf19.5144	orf6.311	21.9	
orf19.6844	orf6.6993	22.2	ICL1b
orf19.535	orf6.6747	22.3	
orf19.6001	orf6.9036	22.7	SAP3
orf19.12002	orf6.3044	24.3	HGT1-2
orf19.6078	orf6.241	24.5	POLRT-1
orf19.1172	orf6.3437	25.0	PHO84
orf19.534	orf6.6746	25.2	
orf19.9115	orf6.821	25.2	
orf19.6805	orf6.8826	27.1	YDR089W
orf19.1655	orf6.3140	29.0	POX1
orf19.3950	orf6.2417	29.4	MSM1
orf19.4599	orf6.7898	30.4	PHO89

orf19.6224	orf6.1289	30.7	RSB1
orf19.9486	orf6.1281	31.2	
?	orf6.450	31.8	
orf19.8167	orf6.449	32.7	
orf19.4934	orf6.6056	36.5	OPS4-1
orf19.742	orf6.4131	40.3	
orf19.2292	orf6.1364	41.5	OPT2-5
orf19.3718	orf6.2195	41.7	OPT2-7
orf19.356	orf6.7240	42.4	
orf19.1370	orf6.1148	42.8	
orf19.1239	orf6.5078	43.6	OPS4-2
orf19.3377	orf6.1151	45.2	
orf19.4884	orf6.3050	45.4	YEL007W
orf19.9997	orf6.3562	46.8	
orf19.5714	orf6.4644	47.6	SAP2-1
orf19.539	orf6.6751	47.8	LAP3
orf19.3902	orf6.5146	53.0	
orf19.3926	orf6.4050	56.6	RNY1
orf19.6274	orf6.6033	68.6	
orf19.2833	orf6.4494	72.9	
orf19.893	orf6.7284	75.1	
orf19.8473	orf6.1902	95.3	
orf19.11555	orf6.3206	98.5	
orf19.11553	orf6.3209	193.2	HYR1-12

Complete list of White specific genes

orf19	orf6	Fold upregulated, Whites	Name
orf19.97	orf6.2500	2.6	
orf19.2651	orf6.3352	2.7	
orf19.12987	orf6.2205	2.8	YPL184C
orf19.7522	orf6.8774	2.9	ARCT
orf19.6906	orf6.7170	3.0	ASC1
orf19.7150	orf6.8050	3.0	NRG1
orf19.3618	orf6.3288	3.1	FLO1
orf19.6577	orf6.7565	3.1	TPO1-3
orf19.6874	orf6.7138	3.1	FLBD
orf19.3362	orf6.1783	3.2	NDI1
orf19.4907	orf6.3754	3.2	YCR061W-2
orf19.6656	orf6.6765	3.2	SPBC23G7.13
orf19.7021	orf6.7947	3.3	GPH1b
?	orf6.1865	3.3	PMT3
orf19.4358	orf6.3780	3.3	YDL157c
orf19.85	orf6.1558	3.3	
orf19.1880	orf6.7040	3.3	HEM15
orf19.851	orf6.1543	3.3	
orf19.5686	orf6.4564	3.4	
orf19.7655	orf6.9003	3.4	RPO21c
orf19.4114	orf6.6822	3.4	FAA2-1
orf19.3043	orf6.7780	3.5	TGL2
orf19.10132	orf6.4097	3.5	HEM1
orf19.2602	orf6.3597	3.5	OPT1-1a
orf19.5800	orf6.6634	3.5	
orf19.3802	orf6.1474	3.5	PMT6
orf19.3641	orf6.2072	3.6	
orf19.9534	orf6.2482	3.7	SPBC1271.09
orf19.13636	orf6.2272	3.7	GLT1a
orf19.1664	orf6.5739	3.7	BGL2-3
orf19.2165	orf6.1313	3.7	
orf19.2119	orf6.4742	3.7	NDT80-1
orf19.5636	orf6.4505	3.7	RBT5
orf19.6489	orf6.1074	3.8	YDL222C-1
?	orf6.332	3.8	FCY22
orf19.903	orf6.7274	3.9	GPM1-1
orf19.4942	orf6.6048	3.9	
orf19.473	orf6.2283	3.9	TPO4
orf19.7585	orf6.8933	3.9	INO1b
orf19.1736	orf6.348	4.0	
orf19.7852	orf6.475	4.0	
orf19.2179	orf6.6084	4.0	ARN1
orf19.2947	orf6.6669	4.1	SNZ1
orf19.386	orf6.6278	4.1	MHT1
orf19.5805	orf6.6639	4.1	DLD1
orf19.5118	orf6.6332	4.1	SDS24a
orf19.320	orf6.5638	4.2	SPCC736.13-2
orf19.8937	orf6.1000	4.2	FCY21a
orf19.271	orf6.3178	4.3	ADHX-1
orf19.3406	orf6.4325	4.3	YHL008Ca
orf19.10651	orf6.5876	4.3	BCLHH
orf19.4771	orf6.3982	4.3	
orf19.13055	orf6.1045	4.3	

orf19.796	orf6.2996	4.3	HYM1-2
orf19.2371	orf6.4769	4.4	GAG-1
orf19.1799	orf6.2231	4.4	GAP1-6
orf19.6655	orf6.6764	4.5	
orf19.4304	orf6.4609	4.5	GAP1-2
orf19.3442	orf6.6463	4.5	EBP1-3
orf19.395	orf6.6269	4.6	ENO1
orf19.3448	orf6.6457	4.6	
orf19.6637	orf6.6534	4.7	PH1107
orf19.8745	orf6.3128	4.7	GAD1
orf19.7851	orf6.474	4.7	HSP150
orf19.8242	orf6.4822	4.7	
orf19.3419	orf6.6106	4.7	MAE1
orf19.6814	orf6.8817	4.8	TDH3
orf19.321	orf6.5637	4.8	MUP3a
orf19.3104	orf6.6941	4.8	
orf19.2337	orf6.5571	4.8	ALP1
orf19.7098	orf6.8455	4.8	YKL070W
orf19.5032	orf6.7977	4.8	SIM1
orf19.2846	orf6.4940	4.9	
orf19.5673	orf6.6913	4.9	OPT2-8
orf19.3119	orf6.6926	4.9	
orf19.7600	orf6.8948	5.0	SFA1
orf19.3575	orf6.5754	5.0	CDC19
orf19.219	orf6.5154	5.0	HSP150
orf19.2063	orf6.7492	5.0	
orf19.112	orf6.2564	5.1	CAN1-1
orf19.5770	orf6.6188	5.1	YGL114W
orf19.11350	orf6.550	5.1	
orf19.5193	orf6.3409	5.1	YIR035C
orf19.12233	orf6.927	5.1	IPT1
orf19.9304	orf6.1348	5.1	
orf19.3461	orf6.995	5.4	
orf19.2726	orf6.2518	5.6	YKL051W
orf19.3656	orf6.3305	5.6	COX15
orf19.2584	orf6.4685	5.7	OPT1-2
orf19.10952	orf6.2792	5.7	
orf19.4824	orf6.7387	5.8	
orf19.3651	orf6.3310	5.8	PGK1
orf19.6576	orf6.7564	6.0	
orf19.7251	orf6.8151	6.0	WSC4
orf19.111	orf6.2565	6.0	CAN1-1
orf19.4774	orf6.3985	6.0	AOX1-2
orf19.278	orf6.4807	6.1	MTRNC
orf19.6659	orf6.6768	6.1	GAP1-4
orf19.10965	orf6.3764	6.1	
orf19.946	orf6.8195	6.2	MET14
orf19.6201	orf6.5423	6.2	RBT4
orf19.5025	orf6.7984	6.2	MET3a
orf19.13583	orf6.537	6.2	RBT4
orf19.4309	orf6.4614	6.3	GRE2-1
orf19.2898	orf6.6124	6.5	YNL275W
orf19.7214	orf6.8114	6.5	YBR056W
orf19.9017	orf6.795	6.6	PLB2-2
orf19.889	orf6.7288	6.7	THI20
orf19.5334	orf6.8323	6.8	CTH1
orf19.8243	orf6.4821	6.8	EFG1b

orf19.1395	orf6.5348	6.8	YER053C
orf19.4618	orf6.7917	6.8	FBA1
orf19.10399	orf6.4390	6.8	MNN4
orf19.3652	orf6.3309	6.9	YCR013C
?	orf6.5700	7.0	
orf19.1906	orf6.1265	7.2	
orf19.2765	orf6.6587	7.4	
orf19.7252	orf6.8152	7.4	
orf19.6937	orf6.7323	7.5	PTR2-2
orf19.3182	orf6.4479	7.7	GIS2
orf19.7085	orf6.8468	7.8	
orf19.734	orf6.6009	7.9	GLK1-1
orf19.542	orf6.6754	8.0	HXK2
orf19.610	orf6.2978	8.0	EFG1a
orf19.1408	orf6.1017	8.2	GLK1-2
orf19.24	orf6.5918	8.3	RSB1
orf19.4940	orf6.6050	8.3	HIP1
?	?	8.4	CRD2
orf19.7514	orf6.8766	8.4	PCK1
orf19.10506	orf6.1981	8.7	YNL274C-1
orf19.2023	orf6.2376	8.8	HXT6
orf19.4889	orf6.3333	9.0	HOL1-1
orf19.10357	orf6.4500	9.2	FLIP-1
orf19.13582	orf6.538	9.4	
orf19.3038	orf6.7775	9.4	TPS2
orf19.369	orf6.1059	9.5	PLB2-1b
orf19.2332	orf6.5576	9.5	
orf19.1443	orf6.796	9.6	PLB3-3
orf19.8536	orf6.1243	9.8	
orf19.6116	orf6.5097	9.9	GLK1-3
orf19.3040	orf6.7777	10.0	EHT1
?	?	10.0	FCY1
orf19.1258	orf6.5725	10.3	
orf19.4980	orf6.2627	10.4	SSA4-1
orf19.9571	orf6.187	10.6	HXT3
orf19.5911	orf6.9126	10.7	CMK2-2
orf19.8744	orf6.2780	11.2	SPAC637.03
orf19.6770	orf6.8861	11.2	
orf19.8780	orf6.4626	11.2	
orf19.4679	orf6.5457	11.3	AGP2
orf19.6540	orf6.5274	11.5	
orf19.5674	orf6.6914	11.7	CSA1-4
orf19.8434	orf6.6705	11.7	
orf19.7374	orf6.8534	11.8	YAF1-1
orf19.2020	orf6.2379	12.2	HXT7-1
orf19.12101	orf6.2054	12.7	ERG25-1
orf19.54	orf6.4552	13.0	
orf19.5737	orf6.4916	13.4	ALS5-2
orf19.308	orf6.4446	14.0	SNG1-1
orf19.6816	orf6.8815	14.2	YJR096W-1
?	orf6.790	14.5	FET3-1a
orf19.2849	orf6.4943	14.8	AQY1
orf19.4044	orf6.4410	15.3	MUM2
orf19.1979	orf6.2481	15.6	GIT1-3
orf19.4943	orf6.6047	16.5	PSA2
orf19.3967	orf6.3504	17.7	PFK1
orf19.7310	orf6.8391	18.5	MSC1

orf19.13162	orf6.1116	18.8	
orf19.33	orf6.5927	19.2	
orf19.5818	orf6.4041	20.2	SUR2
orf19.8487	orf6.2588	20.8	ADAEC
orf19.12241	orf6.1906	20.9	DAK2
orf19.711	orf6.522	21.2	
orf19.8096	orf6.2179	21.3	HYR1-3
orf19.3669	orf6.6480	21.4	SHA3
orf19.2608	orf6.4277	22.4	ADH2
orf19.1862	orf6.1262	23.1	YHR087W
orf19.5113	orf6.6337	23.8	ADH1-1
orf19.3160	orf6.1668	26.8	HSP12-2
orf19.5437	orf6.8673	26.9	HOR2
orf19.2048	orf6.5987	28.1	HXK1
orf19.4555	orf6.3074	28.8	
orf19.12199	orf6.3649	29.8	
orf19.6882	orf6.7146	30.7	OSM1-2
orf19.1264	orf6.5731	34.5	CFL1-6
?	orf6.7251	36.7	RME1
orf19.822	orf6.6713	38.3	
orf19.11284	orf6.1866	39.2	MNN2-2
orf19.3707	orf6.2156	42.6	YHB1
orf19.508	orf6.7063	44.1	QDR1
orf19.4556	orf6.3075	48.0	ALS2a
orf19.5079	orf6.6297	51.2	CDR4
orf19.2122	orf6.1311	54.7	
orf19.1691	orf6.4148	56.0	YNR018W-1
orf19.1868	orf6.7052	66.7	
orf19.5288	orf6.3827	76.7	BDH1

Appendix B. Strains used in Chapter 2

a1	α1	α2	ura- ADE+ CHY	ura- ADE+ RRY	ura- ADE+ CHY	URA+ ade- CHY	URA+ ade- RRY	URA+ ade- CHY	URA+ ADE+ ATY	ura- ADE+ ATY
+	+	+	CAI4	RRY1	-	y476	RRY37	-	CAF2-1	CAI4
-	+	+	y257	RRY10	y257RS	y477	RRY38	y477	218	227,228
+	-	+	y341	RRY11	-	y479	RRY39	-	226	234,235
-	-	+	y343	RRY12	-	y482	RRY41	-	245,246	257,258
+	+	-	y414	RRY13	-	y480	RRY43	-	225	239,240
-	+	-	y415	RRY14	-	y483	RRY45	-	250,251	269,270
+	-	-	y444	RRY15	y439RS	-	RRY47	MMY278	248,249	263,264
-	-	-	y575	RRY16	-	y579	RRY49	-	319-323	-

a1	α1	α2	a2	White ura-ADE+	Opaque (ATY)
+	+	+	+	RRY1	-
-	+	+	+	ATY464,465	480
+	-	+	+	RRY11	-
-	-	+	+	ATY466,467	481
+	+	-	+	RRY13	482
-	+	-	+	ATY468,469	483
+	-	-	+	RRY15	484
-	-	-	+	ATY470,471	485
+	+	+	-	ATY458	-
-	+	+	-	ATY472,473	486
+	-	+	-	ATY459	-
-	-	+	-	ATY491,493	
+	+	-	-	ATY460,461	487
-	+	-	-	ATY475,476	488
+	-	-	-	ATY462	489
-	-	-	-	ATY478,479	490

This table lists all strains used in Chapter 2 by phenotype and by strain name. Strains are available in the CHY (Christina Hull), RRY (Ryan Raisner), or ATY (Annie Tsong) collections, unless indicated otherwise (in red). Strains y257RS and y439RS were constructed by Matt Miller. CHY strains were constructed by Ryan Raisner and Christina Hull. RRY strains were constructed by Ryan Raisner. ATY strains were constructed by myself. For details on strain construction, see Hull, 1999, Hull et al, 2000, and materials and methods from this chapter.

Appendix C. Identification of the a2 gene: a cautionary tale

There once was a mating-type locus -
Making sense of three genes nearly broke us -
The alleles were contrary,
ORF finding was hairy,
A fourth gene was no hocus-pocus.

When the *MTL* was first identified by Christina Hull, she annotated three genes, encoding a1, α1, and α2. Using deletion constructs designed by Christina, Ryan constructed a series of gene deletion strains (RRY1-15) that were deleted for all combinations of *MTLa1*, *MTLα1*, and *MTLα2*. When I used these strains in quantitative mating assays, I found that a1 was required for **a**-type mating.

To confirm these results, I constructed a series of gene deletion strains using independently designed deletion constructs. I also deleted the genes in a different order than they had been deleted by Ryan. In these new strains, **a**-type mating did not require a1 (Table 1).

On close inspection, we found that Ryan's construct deleted 141 more basepairs upstream of *MTLa1* than my construct, which was a clean deletion of *MTLa1*.

Soon after the discovery of this disparity, we received an e-mail from Ken Wolfe from University of Dublin that he had identified a potential ORF in the *MTL***a**-locus that had very weak homology to HMG-domain transcriptional regulators from other species, including a2 from *K. lactis*. This ORF existed in two short exons (Figure 1), explaining why it had been overlooked by Christina and by other annotation efforts. We found that Ryan's construct deleted the last 8 amino acids of this new ORF as well as its entire 3'-UTR, while my construct left the new ORF and its 3'-UTR intact (Figure 1).

I created a new knock-out construct that deleted this new ORF, and found that the new ORF, *MTLa2*, was responsible for positive regulation of **a**-type mating in *C. albicans*.

a1 α1 α2	RRY	ATY
+ + +		
− + +	α++++	α++++, a+/-
+ − +		
− − +		a++++
+ + −	a+ and α+	a+ and α+
− + −	α++++	a+ and α+
+ − −	a++++	a++++
− − −		a++++

Table 1. Mating behavior of RRY and ATY strains. MTL configuration is shown in column 1. Column 2 and 3 show the phenotypes of strains constructed by Ryan Raisner (column 2) and myself (column 3) in non-quantitative mating assays with **a**- and α-testers. "+" indicates that a strain is able to mate as the indicated strain. Strains constructed by Ryan Raisner (RRY) show a different phenotype than strains constructed by myself (ATY). Namely, RRY strains deleted for *MTLa1* are unable to mate as **a** strains, but ATY strains deleted for *MTLa1* are able to mate efficiently as **a** strains ("a++++").

Figure 1.

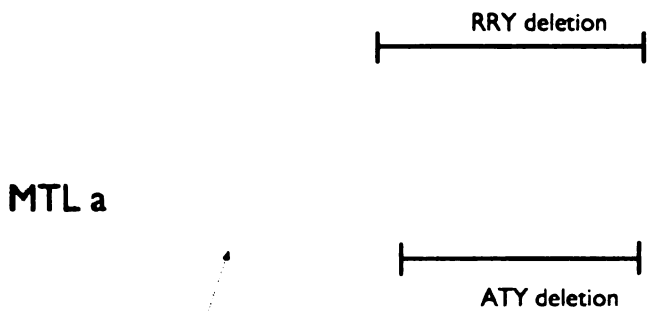


Figure 1. Accidental functional deletion of a2. Figure 1 shows a 3-frame ORF map of a portion of *MTLa*. *MTLa1* exons are shown in light green, and *MTLa2* exons are shown in yellow (indicated by the arrow). The RRY deletion of *MTLa1* deletes all of *MTLa1* as well as the last 8 amino acids of *MTLa2*. The ATY deletion is a precise deletion of *MTLa1*, and leaves *MTLa2* and its 3'UTR intact. This difference accounts for the phenotypic difference between ATY and RRY strains. Subsequent deletion of *MTLa2* in an ATY background shows that a2 is required for **a**-type mating.

Chapter 3

Successive modification of *cis*- and *trans*-elements in combinatorial transcriptional circuit evolution

The work in this chapter was carried out by myself and Brian Tuch. I determined that asgs were positively regulated in the common ancestor of *C. albicans* and *S. cerevisiae*. I also used experimental methods to determine the asgs in *C. albicans*, and validated the *C. albicans* asg operator *in vivo*. Brian determined the PSSMs of the *S. cerevisiae*, *K. lactis*, and *S. cerevisiae* asg operators, and analyzed the evolution of the *cis-* and *trans-* elements across 16 yeast species. Brian also aligned the sequences of Mcm1 and α2, and mapped *K. lactis* α2 onto the *S. cerevisiae* α2 structure with input from Matt Jacobson.

ABSTRACT

Evolution of gene regulation is thought to underlie the extraordinary variety of animal forms. However, technical limitations in animal research models often restrict detailed analysis of gene regulatory evolution to either *cis*- or *trans*-elements. Here, we study evolution of the combinatorial transcriptional circuit that governs mating in the yeast lineage. We show that a group of orthologs, the "**a**-specific genes" (**asgs**), was regulated by an activator in a fungal ancestor, but is now regulated by a repressor in modern *Saccharomyces cerevisiae*. Using a combination of experimental and informatic approaches, we then identify relevant changes in all participating *cis*- and *trans*-regulatory elements. These changes comprise the transition from positive to negative regulation of **asgs** in its entirety; moreover, they suggest specific mechanisms by which fitness barriers during the transition were traversed. By examining **asg** regulation in multiple extant yeast species, we also begin to deduce a temporal sequence to the changes that we have observed in *cis*- and *trans*-regulatory elements.

INTRODUCTION

Darwinian evolution posits that natural selection, acting on heritable, random, "successive, slight variations" in organisms over billions of years, can result in novel features of increasing complexity or specialization (Darwin, 1859). While physiological and metabolic innovation have been linked to new protein function (Chen et al., 1997; Copley, 2000, and references therein), recent work in evolution and development suggests that complexity and variation in animal form are often attributable to changes in transcriptional regulation throughout development (Carroll, 2001; Davidson, 2001; Gerhart and Kirschner, 1997).

Change in transcriptional regulation is appealing as a major evolutionary mechanism on several levels. First, it partially accounts for recent comparative genome analyses, which indicate that increases in gene number do not scale with increases in organismal complexity (Levine and Tjian, 2003). Second, it explains the "Hox paradox," the observation that orthologous transcription factors often direct development of analogous structures in highly diverged species, even when those structures have few shared molecular components (Davidson, 2001). Third, transcriptional regulation itself is particularly amenable to evolution on multiple levels. In addition to being more flexible in spatial and directional requirements than coding sequence, eukaryotic promoters tend to be modular in form; that is, discrete promoter regions are self-sufficient in directing a defined portion of an expression program. This modularity permits individual sub-programs to evolve independently of one another, decreasing the likelihood that a mutation will have a pleiotropic effect on fitness (Carroll, 2005). Finally, it is thought that the combinatorial nature of transcriptional regulation allows a high degree of

specificity to be achieved via integration of multiple interactions, each of which individually has low affinity or specificity (Ptashne and Gann, 1998). This model predicts that simple changes in protein-DNA or protein-protein affinities can have strong but specific consequences in functional output (Buchler et al., 2003), and thus be rapidly subject to selection.

In the past several years, many remarkable examples of transcriptional circuit evolution have been uncovered, including changes in both *cis*- and *trans*- elements (for examples, see Gasch et al., 2004; Gompel et al., 2005; Ihmels et al., 2005; Landry et al., 2005; Ludwig et al., 1998; Ronshaugen et al., 2002; Tanay et al., 2005; Wittkopp et al., 2004). Due to the complexity and technical limitations in model organisms, however, detailed analysis is often limited in scope to a subset of the *cis*- or *trans*-elements involved in a specific change. In many cases, *trans*- or *cis*- elements are unidentified, or are difficult to evaluate at a comparable level of detail.

Because gene regulation depends on the interplay between *cis*-elements and cognate *trans*-factors, we wished to study an example of regulatory circuit evolution in which all involved *cis*- and *trans*-elements could be identified and analyzed in detail. Here, we present an analysis of evolution in a combinatorial transcriptional circuit which regulates mating in multiple yeast species of the ascomycete lineage. Availability of genome sequences, informatic tools, and genetic and reverse genetic methods allows comprehensive mapping of transcription factors to the gene products they regulate; furthermore, genome sequences in closely related yeasts can be used to identify functional *cis*-elements (Cliften et al., 2003; Kellis et al., 2003). The transcriptional circuit controlling mating-type in *S. cerevisiae* is particularly well-studied; all known

transcription factors involved in the circuit have been identified, and the targets of their regulation have not only been identified, but their biological roles characterized as well (Herskowitz, 1992).

We first identify a group of orthologs (**a**-specific genes, or **a**sgs) required for mating that is regulated via activation in most characterized descendants of a fungal ancestor, but is regulated via repression in the lineage that carries *S. cerevisiae* and its close relatives. By examining the regulation of **a**sgs in modern yeasts that diverged at successively later times from a common ancestor, we then deduce specific, sequential changes in both *cis*- and *trans*-regulatory elements that constitute the transition from positive to negative regulation.

RESULTS

a-type mating is negatively regulated in modern *S. cerevisiae*, but was positively regulated in its ancestor

Mating in ascomycetes is regulated by a single locus in the genome called the "Mating-type locus," or MAT locus. In the yeasts we consider here, the MAT locus exists as one of two alleles, "**a**" or "α." Cells which express only the **a** allele ("**a**-type") can mate with those that express only the α allele ("α-type"), resulting in formation of a non-mating cell which expresses both alleles ("**a**/α-type," figure 1a).

The MAT loci encode transcription factors that control mating-type by turning on or off specific batteries of mating-type specific genes, including those involved with pheromone production and pheromone sensing. The **a**-specific genes (**a**sgs) are required for **a**-type mating, the α-specific genes (αsgs) are required for α-type mating, and the

haploid specific genes (hsgs) are required for both **a** and α cells to mate. In the bakers' yeast *S. cerevisiae*, mating type is controlled by three transcription factors called a1 (encoded at MAT**a**), α1, and α2 (encoded at MATα; Figure 1a). a1 and α2 are homeodomain proteins, whereas α1 is an α-domain protein. The means by which these three factors define three mating-types have been meticulously studied over the past several decades (Figure 1b; for review, see Herskowitz, 1992; Johnson, 1995).

In *S. cerevisiae* **a** cells, **a**sgs are contitutively activated by the ubiquitous MADS box transcription factor Mcm1. Mcm1 is expressed in all three cell types, and to restrict expression of **a**sgs to **a** cells, **a**sgs are repressed when Mcm1 binds in conjunction with α2, a situation that arises only in α and a/α cells. In α cells, the α-domain protein α1 activates αsgs, again in conjunction with Mcm1. In a/α cells, a1 from the **a** locus cooperates with α2 from the α locus to form a heterodimer which represses hsgs. The co-expression of a1 and α2 only in a/α cells ensures that a and α cells can mate, but that a/α cells cannot (Figure 1b).

Orthologs of the transcription factors a1, α1, and α2 are widespread in the hemiascomycete lineage (Butler et al., 2004; Coppin et al., 1997; Fabre et al., 2005). However, an additional gene encoding for an HMG-domain transcription factor has been identified in the *MATa* loci of several related species. This gene is present in yeasts as closely related to *S. cerevisiae* as *Zygosaccharomyces rouxii,* and as divergent as *Schizosaccharomyces pombe*, yet is missing in *S. cerevisiae* and its closest relatives (Butler et al., 2004). In several yeasts, the HMG-domain transcription factor has been experimentally determined to be an activator of **a**-type mating (A.E.T., unpublished data, Debuchy et al., 1993; Kelly et al., 1988; Kurischko et al., 1999; Philley and Staben, 1994;

Tsong et al., 2003; Turgeon et al., 1993).

Given the role of an HMG-domain transcription factor as an activator of **a**-type mating over a wide phylogenetic range of yeasts, the parsimonious interpretation is that HMG-domain transcription factor was recently lost by the *S. cerevisiae* branch, and that the repressive role of α2 in **a**-type mating is a recent innovation within the *S. cerevisiae* branch (Figure 1c). This innovation preserves the logic of the overall regulatory circuit.

Identification of ancestral a-specific genes

To understand the historical events leading to the negative regulation of **a**sgs in *S. cerevisiae*, we sought to approximate the ancestral *cis*-element responsible for positive regulation of **a**sgs. We reasoned that extant yeasts which retain the ancestral regulatory logic of positive regulation of **a**-type mating by an HMG-domain protein may also have retained *cis*-elements close to the ancestral form.

Candida albicans is a fungal pathogen of humans that last shared a common ancestor with S. cerevisiae some 200 to 800 million years ago (Calderone, 2002; Hedges, 2002). Many molecular biological tools are available for studying *C. albicans*; additionally, the genome has been fully sequenced and annotated, allowing extensive genome-wide analysis. The mating-type regulatory circuit in *S. cerevisiae* and *C. albicans* shares several features. In both organisms, αsgs are activated by α1, and in both organisms, mating is repressed by the a1/α2 heterodimer. Furthermore, many of the genes regulated by these transcription factors are orthologous. Unlike *S. cerevisiae*, however, **a**sgs in *C. albicans* are positively regulated by the HMG-domain transcription factor a2. Thus, *C. albicans* fulfills the criterion of an extant yeast which has retained the

ancestral form of asg regulation.

To find the *cis*-regulatory site associated with asgs in *C. albicans*, we first experimentally identified the asgs. Previously, it was shown that detectable asg expression in *C. albicans* **a** cells requires exposure to α-factor, the pheromone from the opposite mating-type. Although the genes induced by α-factor have been identified (Bennett et al., 2003), we were initially unable to parse the genes specifically dependent on a2, the **a**-specific genes, from the general pheromone response genes.

In principle, determining the genes induced in α cells on exposure to **a**-factor would reveal the subset of genes that specifically require activation by a2 (Figure 2a); however, **a**-factor has not yet been identified in *C. albicans*. In *S. cerevisiae*, the pheromone-response MAP kinase cascade is identical between **a** and α cells, except for the pheromone receptor itself (Bender and Sprague, 1986). By analogy, we predicted that an engineered *C. albicans* α strain expressing α-factor receptor should respond to α-factor as wildtype α strains respond to **a**-factor. We therefore constructed an α strain which expresses the α-factor receptor *STE2* under the promoter of the α-specific gene *STE3*. This strain was made in an α-factor gene deletant background to prevent activation of the receptor by endogenously encoded α-factor (Figure 2b).

The constructed α cells responded to α-factor by forming mating projections, indistinguishable from those formed in mixtures with **a** cells secreting **a**-factor (data not shown), showing that the strategy successfully "fooled" α cells into responding to α-factor. When we compared the transcriptional profile of the induced α strain to that of an induced **a** strain, we identified a group of genes induced only in **a** strains as well as a group of genes induced only in α strains (Figure 2c). The vast majority of pheromone-

induced genes (60 of 65) were induced in both mating-types, and were previously described (Bennett et al., 2003). The genes induced only in **a** strains includes *HST6, BAR1, RAM2, ASG7, AXL1*, and *DAL5* (orf19 numbers are indicated in figure 2c). We also subsequently identified *STE2* as **a**-specific in expression (see below), bringing the total number of **a**sgs to seven. Four of these genes (*ASG7, BAR1, STE2,* and *STE6*) are also **a**-specific in *S. cerevisiae*, suggesting that they were **a**-specific in their common ancestor. We note that *S. cerevisiae RAM2* and *AXL1* encode functions needed specifically by **a** cells, but unlike their orthologs in *C. albicans*, expression of these genes are not cell-type regulated.

Identification of a MADS box binding site in *C. albicans* asg promoters

For ancestral **a**sgs to undergo the transition from positive to negative regulation, *cis*-elements involved in activation by a2 were likely lost, while *cis*-elements involved in repression by α2 must have been gained. To identify regulatory elements involved in activation by a2, we submitted 1000 bp upstream of each *C. albicans* **a**sg ortholog to MEME (Bailey and Elkan, 1994). A single element was present in the promoter of each **a**-specific gene except *DAL5*, and in every case was within ~500 bp of the start codon.

The regulatory element present at *C. albicans* **a**sg promoters is highly specified at 26 bp, and includes several remarkable features (Figure 3a). First, it shows the hallmarks of an Mcm1 binding site, which is characterized by an AT-rich core of six basepairs flanked by a CC on one side and a GG on the opposite side. Mcm1 is part of a highly conserved family of transcription factors called the MADS box family, and the precise residues involved in DNA binding have been identified by both structural and mutational

analyses (Acton et al., 2000; Tan and Richmond, 1998). *C. albicans* Mcm1 and *S. cerevisiae* Mcm1 are 84% identical in their DNA binding domains; furthermore, every key residue shown to contact DNA is conserved, suggesting that the observed site in *C. albicans* is indeed an Mcm1 binding site.

As mentioned above, Mcm1 is central to asg expression in *S. cerevisiae*. In addition, the MADS box protein *map1* is involved in mating-type gene expression in *S. pombe* (Yabana and Yamamoto, 1996); based on these data, it seems likely that a MADS box protein was involved in a-type mating control in the common ancestor of *S. cerevisiae, C. albicans*, and *S. pombe*.

Identification of the a2 binding site in *C. albicans* asg promoters

Second, the putative Mcm1 site in *C. albicans* asg promoters is tightly associated with a highly conserved 6 bp motif of the consensus sequence CATTGT (Figure 3a). The spacing between the 6 bp motif and the putative Mcm1 site is invariantly 5 bp. Notably, the motif is similar to the demonstrated binding site for the HMG domain proteins ste11 and MatMc, both of which are involved in activation of asg orthologs in *S. pombe*, as well as to the binding site for the HMG domain protein MT a-1, which is required for activation of asg orthologs in *Neurospora crassa* (Philley and Staben, 1994). ste11 and MatMc both bind a core sequence of CTTTGTT, while MT a-1 binds a core sequence of CTTTG (Kjaerulff et al., 1997); the motif observed in *C. albicans* asg promoters is a single basepair A:T transversion from both of these core sequences (Figure 3b).

This highly conserved motif also shows remarkable similarities to the consensus α2 monomer site of *S. cerevisiae* (CATGT), differing only in a single basepair indel (Figure

3b, Smith and Johnson, 1992). This observation is consistent with previous speculations that the transition from positive to negative regulation of asgs might occur via co-option of the a2 binding site by the α2 protein (Scannell and Wolfe, 2004). The similarity of the sites is striking, given that a2 and α2 belong to different protein families.

Biological validation of the asg operator in *C. albicans*

To test whether the conserved element upstream of *C. albicans* asgs represents a biologically functional activator, we fused either a wildtype or a mutant fragment of the *STE2* promoter to a codon-optimized GFP reporter (Figure 3c-d, Care et al., 1999). The mutant promoter differed from the wildtype in the putative a2-binding element by two basepairs: the element was mutated from CATTGT to CATAAT, a change predicted to destroy the a2 binding site. Both constructs were integrated at the RP10 locus of *C. albicans*.

The wildtype promoter was able to activate GFP expression on exposure to α-factor (Figure 3c), while the mutant promoter showed no induction in GFP expression under the same conditions (Figure 3d), demonstrating that this *cis*-element is required for a2-dependent activation of asgs.

Analysis of *cis*- asg regulation across species

Having identified the operator associated with positive regulation of asgs, we next wished to know when the transition from positive to negative regulation occurred in evolutionary time. We first inferred a robust phylogeny of 16 yeast species (Figure 4b) using methods similar to Rokas et al. (Materials and Methods, Rokas et al., 2003). We

then identified orthologs of *C. albicans* and *S. cerevisiae* a-specific genes across these yeasts (Materials and Methods). Using position specific scoring matrices (PSSMs) constructed from the *S. cerevisiae* or *C. albicans* asg operators (Figure 4a, Appendix A) we scanned the asg ortholog promoters (defined as 2000 bp upstream of the putative translational start site) in the 16 yeasts, and recorded the maximum log-odds score (Figure 4c-d).

S. cerevisiae-like operators were clearly found in asg ortholog promoters in organisms as far diverged as *S. castellii.* Past *S. castellii*, the presence of an *S. cerevisiae*-like operator in asg promoters was diminished, though present in some *C. glabrata, K. lactis, E. gossypii,* and *K. waltii* promoters (Figure 4c).

The *C. albicans* PSSM yielded a nearly converse pattern (Figure 4d). Organisms that branch with *C. albicans* have a2-Mcm1 sites upstream of their asg orthologs; in contrast, this matrix recovered no significant matches in the species close to *S. cerevisiae*, correlating with the loss of a2 in these species (Butler et al., 2004). As with the *S. cerevisiae* promoter, a *C. albicans*-like promoter was present in a small number of asg orthologs in the *K. lactis* branch.

Identification of the asg operator in the *K. lactis* branch

Since neither the *C. albicans* nor the *S. cerevisiae* matrices elicited strong matches in the *K. lactis*, *E. gossypii* and *K. waltii* asg promoters, we used MEME to determine if this evolutionary branch has a unique asg operator. We submitted promoters from the intersection of the *C. albicans* and *S. cerevisiae* asg sets (*ASG7, BAR1, STE2,* and *STE6*), as they were most likely to be a-specific in an ancestor yeast, and therefore a-specific in

K. lactis, *E. gossypii*, and *K. waltii*. Due to difficulties in motif finding inherent in small input sets, we submitted the four asg promoters from all three yeasts to MEME at once (12 total promoters).

We were surprised to find a promoter element which has elements of both the *C. albicans* a2-Mcm1 site, as well as the *S. cerevisiae* α2-Mcm1 site (Figure 4e). As in the *S. cerevisiae* and *C. albicans* sites, the *K. lactis* site contains an Mcm1 binding site. One flank of the Mcm1 site matches the a2 site of *C. albicans* in sequence and spacing. We have confirmed that a-type mating is dependent on a2 in *K. lactis* (unpublished results, A.E.T.). Unlike the *C. albicans* element, however, the *K. lactis* site is also specified on the opposite side of the Mcm1 element; this side matches the core α2 sequence, which is conserved in all α2 sites in *S. cerevisiae*, including both α2-Mcm1 and a1-α2 contexts. Its spacing from the Mcm1 site is identical to the spacing in the *S. cerevisiae* operator.

To summarize, the *S. cerevisiae* lineage asgs have an Mcm1 site with flanking α2 sites at defined distances. The *C. albicans* lineage asgs have an Mcm1 site a defined distance from a single a2 site. The *K. lactis* lineage asgs appear to have a hybrid site, an Mcm1 site with an α2 site to one side and an a2 site to the other.

Analysis of *trans*- elements in asg regulation across species

With asg associated *cis*-elements from multiple species in hand, we next investigated correlated changes in the *trans*-factors α2 and Mcm1. In modern *S. cerevisiae*, α2 and Mcm1 interact to repress asgs. In contrast, α2 does not repress asgs in *C. albicans* (Tsong et al., 2003) or any other known yeasts that diverged prior to the *C. albicans*/*S. cerevisiae* split, suggesting that the α2-Mcm1 interaction is a recent

innovation. Structural studies of α2-Mcm1, in conjunction with extensive mutational studies of α2-Mcm1 *in vivo,* have identified the critical residues required for α2-Mcm1 interaction and for DNA binding (Acton et al., 2000; Mead et al., 2002; Mead et al., 1996; Tan and Richmond, 1998). To determine when the α2-Mcm1 interaction arose, we aligned orthologs of α2 and Mcm1 across the 16 yeast species, then searched for conservation of the interaction interface.

The DNA-binding domain of Mcm1, which is contacted by α2, is highly conserved across all species analyzed (Figure 5a). Many proteins besides α2 contact this portion of Mcm1, so this degree of conservation is not surprising. In contrast, the portion of α2 that contacts Mcm1 varies considerably across the 16 yeasts (Figure 5b). A critical 9-residue "linker" region required for α2-Mcm1 interaction is highly conserved from *S. cerevisiae* to *C. glabrata*, and 5 of the 9 residues in this linker region are conserved in *K. lactis*. This linker region shows no conservation in the yeasts that branch with *C. albicans*, consistent with biological observations that α2 has no role in asg regulation in *C. albicans*.

Despite considerable conservation of α2-Mcm1 interaction residues in *K. lactis* α2 and Mcm1, homology modeling onto the *S. cerevisiae* crystal structure reveals that *K. lactis* lacks a highly favorable pi-stacking interaction that is present in *S. cerevisiae* (Figure 5c, Materials and Methods). Disrupting this interaction in *S. cerevisiae* causes a 20-fold decrease in the cooperativity of α2-Mcm1 DNA binding activity (Mead et al., 1996), suggesting that *in vivo*, the *K. lactis* α2-Mcm1 interaction may be significantly weaker than the *S. cerevisiae* α2-Mcm1 interaction.

To summarize, parallel evidence for the transition between positive control of the

asgs (the situation in modern *C. albicans* is inferred for in ancestral yeasts) and negative control (the situation in the modern *S. cerevisiae* lineage) is seen in the *cis*-acting sequences as well as in the α2-Mcm1 interaction surface.

DISCUSSION

In this work, we identify a group of genes that was positively regulated in an ancestral yeast, but is negatively regulated in modern *S. cerevisiae*. Orthologs of these genes (the asgs) are required for mating in fungal lineages that span at least 1.3 billion years of evolution (Hedges, 2002; Lengeler et al., 2000). We identify changes in both the *cis*- and *trans*- elements regulating this circuit that constitute the transition from positive to negative regulation. Our analysis has major mechanistic implications for how fitness barriers during the regulatory transition were overcome, which we discuss here. We also find that the *cis*- and *trans*-regulatory elements in *K. lactis* have properties of both the ancestral and the modern state. Combining this observation with phylogneetic information suggests a temporal sequence in which the observed changes occurred.

Mcm1 has facilitated the activator-to-repressor transition

The presence of Mcm1 at ancestral asg promoters significantly decreases fitness barriers in the transition from positive to negative regulation – i.e., the swapping of one regulator for another. Historically, we know that several changes occurred in asg regulation during this transition: 1) a2 was lost in the *S. cerevisiae* lineage, 2) asg expression in modern *S. cerevisiae* depends only upon Mcm1, but required both a2 and Mcm1 in an ancestor, and 3) α2 represses the asgs *via* an interaction with Mcm1, a mode

of regulation absent in *C. albicans*. Transitions 2) and 3) can both be facilitated by the presence of Mcm1.

For instance, in *S. cerevisiae*, αsg expression requires both α1 and Mcm1, while asg expression requires only Mcm1. This disparity is due to a higher A/T content flanking Mcm1 sites at asgs, allowing Mcm1 to bind and activate in the absence of a co-factor (Acton et al., 2000). For asg expression to change from requiring both a2 and Mcm1 to requiring only Mcm1, an increase in A/T content flanking the asg promoter Mcm1 site could be sufficient to confer binding of Mcm1 without a2. This is a much smaller evolutionary change than evolving a new 10bp Mcm1 site and associated A/T rich flanks *de novo*. Consistent with this, the A/T content flanking Mcm1 sites in both *K. lactis* and *C. albicans* asg promoters is lower than that in *S. cerevisiae* (Figure 4).

The presence of Mcm1 at asgs facilitate the recruitment of α2 as a repressor by widening the interaction surfaces available to α2. A sub-optimal α2 binding site in front of asgs could be compensated for by an interaction with Mcm1, or vice versa, allowing the optimization of the α2 binding site to occur independently from the evolution of an α2-Mcm1 interaction.

In general, the presence of Mcm1 in a combinatorial interaction with a co-activator or co-repressor could act to relax the operator sequence constraints for functional activation or repression. Relaxing the constraints would result in a higher level of operator sequence turnover, which in turn would affect the recruitment of new factors to asg promoters, or drive compensatory co-evolution of *trans*-factors. Consistent with this, we have identified a potential change in α1 binding specificity between *C. albicans* and *S. cerevisiae*. This change is also tightly associated with the presence of

Mcm1.

a2 and α2 occupy a similar binding site

We have identified and verified the a2 binding site present in *C. albicans* asg promoters. This site is remarkably similar to the *S. cerevisiae* α2 half-site, differing only by a single base deletion; further, the spacing between the a2 and Mcm1 elements is identical to the spacing between the *S. cerevisiae* α2 and Mcm1 elements. We know from studies of the *C. albicans* a1/α2 site that α2 binding specificity is identical in *C. albicans* and *S. cerevisiae* (Hull and Johnson, 1999), suggesting that α2 could co-opt the ancestral a2 site with a small change to the *cis* element. The *S. cerevisiae* asg operator consists of two α2 monomer sites flanking the Mcm1 dimer site (figure 4a). Thus, a co-option mechanism reduces the barrier to *de novo* evolution of α2 binding sites by half.

Changes in *trans*- regulation of asgs

The α2-Mcm1 interaction is strongly conserved in yeasts as far diverged as *C. glabrata*, and a substantial fraction of the critical interaction residues in both α2 and Mcm1 are conserved in the *K. lactis* branch. Taken with homology modeling of *K. lactis* sequence onto the *S. cerevisiae* structure, our analysis suggests that a weak interaction between Mcm1-α2 arose prior to the divergence of *K. lactis* and *S. cerevisiae*, and was subsequently strengthened in the *S. cerevisiae* lineage.

Ordering the pathway

We have identified a *cis*- element upstream of *K. lactis* asgs that encodes a α2 core

binding site. Strikingly, this site appears coincidently with conservation of α2-Mcm1 interaction residues, suggesting the co-evolution of *cis*- and *trans*-factors in this instance of new regulation (Figures 4-5). Because *K. lactis* has also retained positive regulation of the asgs by a2, we can tentatively define the succession of events leading to repression of asgs in modern *S. cerevisiae:* a2-Mcm1 activated asgs in an ancestor (Figure 6a,d). Subsequently, the α2-Mcm1 interaction evolved, coincident with evolution of an α2-Mcm1 binding site in the asg operator (Figure 6b,d). Finally, a2 was lost and both the α2-Mcm1 interaction and the α2-Mcm1 operator specificity were strengthened, resulting in the regulation observed in modern *S. cerevisiae* (Figure 6c,d).

We have not ruled out a model in which the *K. lactis* site represents an ancestral mode of regulation, and that the exclusively positive (*C. albicans*) and exclusively negative (*S. cerevisiae*) regulation are derived from a dual regulatory strategy (Figure 6e). However, exclusively positive regulation has been biologically demonstrated in yeasts which represent several independent, ancient-branching lineages, including the *C. albicans* branch, the *Yarrowia lipolytica* branch, the *Neurospora crassa* branch, and the *S. pombe* branch (Coppin et al., 1997; Kelly et al., 1988; Kurischko et al., 1999; Tsong et al., 2003). This would imply that in such a scenario, evolution of the exclusively positive regulation occurred independently at least four times. Furthermore, many of the yeasts that employ exclusively positive regulation of asgs also encode α2 at their MAT loci, which functions as part of the a1/α2 heterodimer but has no role in asg repression. In contrast, none of the yeasts that employ exclusively negative regulation of asgs encode a2 at their MAT loci, suggesting that the conversion to an exclusively negative regulatory scheme was concomittant with, and possibly selected for due to the loss of a2. This

ambiguity may be resolved as more ascomycete genomes are sequenced.

Unanswered questions

Evolutionary change reflects an interplay between genetic variation and environmental selection. Despite the changes we have identified in transcriptional circuitry, however, the logical output of all examined circuits is identical: asgs are activated in **a** cells, but repressed in α cells. Therefore a great challenge will be to identify the selective pressures that drove evolution of new modes mating-type regulation. Interestingly, it has been shown that positive and negative regulation result in different systemic properties in terms of kinetic response to environmental inputs and evolutionary stability, even when the two modes encode identical logical output (Savageau, 1983; Wall et al., 2004). We also note that sex-determination across all phyla is less evolutionarily conserved than other developmental programs (de Bono and Hodgkin, 1996; Haag et al., 2002; Kulathinal et al., 2003; O'Neil and Belote, 1992; Whitfield et al., 1993), potentially because changes in sex-determination pathways closely track with reproductive isolation and speciation (Singh and Kulathinal, 2000). Thus, the selection underlying evolution of the mating-type regulatory circuit in yeasts may include sexual as well as environmental factors.

MATERIALS AND METHODS

Strain construction

To express *STE2* (α-factor receptor) in *C. albicans* α cells, we integrated *STE2*

under the promoter of the α-specific gene *STE3* (Tsong et al., 2003), yielding strains ATY496 and 497. The *STE2* ORF was amplified using primers 5'-aacattgagctccgatcaacaaccagtattcc-3' and 5'-tgtgaccgcggttcaatgcctgtaccgtggc-3'. The product was cut with *SacI* and *SacII* and ligated into pDDB57 (Wilson et al., 2000) cut with *SacI* and *SacII*, yielding pAT103. A fragment of pAT103 containing *STE2* and *URA3* was amplified using primers 5'-TTAATACCTTAGCATACATAGAGAACTTTATTTTGGCTTCTTAATAATATTTAAAGCAAAGTTTTCCCAGTCACGACGTT-3' and 5'-TAGACTTGTTTTTTTTTTATTTATTATATTTTTCATCAACAAGTAACAGGATACGATGGATGTGGAATTGTGAGCGGATA-3'. This PCR product was purified and transformed into MMY563, an α strain deleted for *MFα* (Bennett et al., 2003). Integration of *STE2* under the *STE3* promoter was confirmed by PCR. The 5' junction was tested using 5'- TTGAGGTATTTGCTGGTGCC-3' and 5'-gttgttgatcgGAGCTCccc-3'. The 3' junction was confirmed with primers 5' TCGTGGCCATACTAACGCCC-3' and 5'-ggcttattatgacacctgg-3'.

To validate the *C. albicans* asg element, we used wildtype or mutant *STE2* promoters to drive expression of GFP. A 2023-bp region of the *STE2* promoter containing the putative a2-Mcm1 site was amplified using the primers 5'-ATTACGCGCGGATCCAAGCTTcagattagaagtcaaaagcgctttccactacc-3' and ATO379 5'-TTTACGCAAGGATCCtgaagaaagataaatgaaagaggaatactgg-3'. This product was digested with *BamHI* and cloned into pRSC4b digested with *BamHI* (Care et al., 1999). The pRSC4b plasmid encodes a version of GFP codon-optimized for *C. albicans* (Cormack et al., 1997), the *URA3* marker, and the *RP10* gene with an engineered *StuI* site. The

resulting plasmid, pAT199, was cut with *StuI* and transformed into the **a**-strain RRY15, yielding the strain ATY631. Integration at the *RP10* locus was confirmed by PCR. The 5' integration event was tested with primers 5'-GATTTTGTTACAGCGTAACCAGTGCG-3' and 5'-GATTTATGAAAGTTTTCAGCTCTAGTCACG-3'. The 3' integration event was tested with primers 5'- GGAATGTGGTAGTCGATATTCAGGG-3' and 5'-ccaactccaggtgtcataataagccaatc-3'.

To create the mutation in the putative a2 site, we used the Quikchange Multi-Site Directed Mutagenesis kit (Stratagene) with primer 5'-ttttttgcagcaatCATTaaCAAATCCAAAAACAGTAATTttcc-3' to mutagenize pAT199. The mutagenized plasmid, pAT205, was integrated at the *RP10* locus of RRY15 as described above, yielding the strain ATY634. Strains were switched the opaque phase as previously described (Miller and Johnson, 2002), then induced with 10 μg/ml α-factor (stock concentration of 10 mg/ml, dissolved in 10% DMSO).

Preparation of cultures and cDNA for microarray experiments

α-factor inductions and microarray protocols were described previously in detail (Bennett et al., 2003), with the following modification: here, strains were grown to OD1.0 in YEPD+55μg/ml adenine, then induced with 10 μg/ml α-factor, using either a stock concentration of 10 mg/ml dissolved in 10% DMSO or 100 μg/ml dissolved in water in order to separate effects of DMSO and α-factor.

Phylogeny reconstruction

A robust phylogeny of the 16 yeast species was inferred using methods similar to Rokas et al. (Rokas et al., 2003). This method requires identification of orthologous genes. At the time of this report, 4 of the 16 yeast genomes were not yet annotated; ORFs for these four genomes were annotated by translating genomic nucleotide sequence in all six reading frames and retaining any contiguous amino acid sequence longer than 100 residues (i.e. no stop codon). To build the phylogeny, groups of orthologous genes containing one and only one representative from each of the 16 yeasts were chosen at a stringent branch length cutoff (0.4; see *Orthologous ORF mapping* below). This yielded 139 groups of orthologous genes that, showing no evidence for deletion or duplication events, were more likely than other groups of orthologous genes to preserve the underlying speciation signal. The 16 sequences within each group were then multiply aligned with ClustalW (Higgins and Sharp, 1988). The resulting 139 alignments were concatenated and columns containing gaps were dropped, producing a single alignment with 37,963 columns. Finally, a maximum likelihood species tree was estimated employing the TREE-PUZZLE algorithm with default parameters (Figure 4a, Schmidt et al., 2002). Demonstrating the robustness of this inference, a tree with identical topology was generated when the neighbor-joining method of ClustalW (Higgins and Sharp, 1988) was applied to the same dataset (not shown).

Orthologous ORF mapping

Reciprocal best BLAST hit (RBBH) is often used in comparative genomics studies to pair orthologous genes between species (for instance, in Tanay et al., 2005). However, in its unmodified form this method can not take advantage of the added information

offered by multiple sequence alignment. Furthermore, RBBH does not treat the one-to-many and many-to-many evolutionary relationships that can arise due to gene duplication and which have been shown to be particularly prominent within the yeast lineage under study (Dietrich et al., 2004; Kellis et al., 2004). Because our primary interest is to examine the evolution of the regulation of genes that can be traced to a single gene in an ancestral organism, we devised the following method for mapping orthologs. We ran PSI-BLAST for each *S. cerevisiae* ORF "query" sequence against a database of all ORF sequences from each of the 16 fungal species, employing an E-value cutoff of 10^{-5} and the Smith-Waterman alignment option (Altschul et al., 1997). The sequences returned by PSI-BLAST were then multiply aligned with ClustalW (using the fast alignment option) and a neighbor joining tree (NJ) was inferred, again using ClustalW (Higgins and Sharp, 1988). Finally, the resulting NJ tree was traversed to extract a set of orthologous genes in the following manner: Start at the leaf node for the query sequence and ascend the tree, incrementing a level counter for each node ascended. At each internal node descend. If a leaf node is reached, the gene is from a species not yet seen at a lower level, and the branch length traversed is less than a cutoff (1.0), then add that gene to the set of orthologous genes. This procedure was repeated for each *S. cerevisiae* sequence, resulting in a 16 species many-to-many ortholog map.

Ortholog mapping in this automated fashion was employed for the purposes of reconstructing the fungal phylogeny (see *Phylogeny reconstruction*). Due to the small sample size of asgs (7 asgs in *S. cerevisiae* and 6 asgs in *C. albicans)*, we were able to hand-curate the asgs incorporating additional biological information due to its small sample size). Manual mapping maximized the accuracy and coverage of our ortholog

mapping. For example, *MFA1* and *MFA2*, two asgs from *S. cerevisiae*, are less than 40 amino acids long and were therefore not annotated as ORFs in several of the fungal sequencing projects. Using TBLASTN we identified putative *MFA1/MFA2* orthologs and added them to our sequence database and ortholog map. The *C. albicans* ortholog to *S. cerevisiae BAR1*, which encodes an a-specific aspartyl protease, could not be identified on the basis of protein sequence data alone. However, orf19.9629 belonging to the family of aspartyl proteases clustered with the *C. albicans* asgs in the microarray experiments (Figure 2c) and had an a2-Mcm1 motif in its promoter region (Figure 3a). Additionally, we have experimental evidence that this ORF is a functional homolog (R.J. Bennett, personal communication). On this basis, we could confidently assign this *C. albicans* ORF as the ortholog to *S. cerevisiae* Bar1 and in turn assign ORFs from other *Candida* species to the Bar1 ortholog group as well.

Alignment of α2 and Mcm1 orthologs

Putative orthologs of *S. cerevisiae* Mcm1 from the 15 other fungal genomes were taken directly from our 16 species ortholog map. Because MATα2 contains an intron that is sometimes not properly annotated, putative orthologs had to be selected more carefully. For each genome we undertook an iterative process of TBLASTN query, nucleotide sequence extraction, splice site prediction and translation that yielded MATα2 orthologs from each species except *E. gossypii*, *C. lusitaniae*, *D. hansenii*, *C. guillermondii* and *S. pombe*. It is likely that the MATα2 ortholog was not found in these species either because the MATα mating-type locus was not sequenced (e.g., *E. gossypii*) or because the gene is simply not present in that genome (e.g., *C. lusitaniae* and *D. hansenii*).

Although the *K. delphensis* genome has not been sequenced, the MATα2 gene has been sequenced (Butler et al., 2004) and was therefore added to our set of MATα2 orthologs. *K. delphensis* branches with *C. glabrata.* The translated Mcm1 and MATα2 sequences were then aligned with T-COFFEE (Notredame et al., 2000) using default parameters.

Homology mapping of *K. lactis* sequences onto *S. cerevisiae* a2-Mcm1 structure

We modeled the *K. lactis* α2-Mcm1 complex with the Protein Local Optimization Program (PLOP, Matt Jacobson) using the crystal structure of the *S. cerevisiae* α2-Mcm1 complex (PDB ID: 1MNM; Tan S 1998) as a template. First, pairwise sequence alignments of α2 and Mcm1 (between *K. lactis* and *S. cerevisiae*) were extracted from their respective multiple alignments (see *Alignment of α2 and Mcm1 orthologs*). The *K. lactis* α2 and Mcm1 chains were then modeled independently with PLOP utilizing both the pairwise alignments and the *S. cerevisiae* α2-Mcm1 complex. After modeling, the chains were recombined and amino acid side chains were optimized.

FIGURE 1

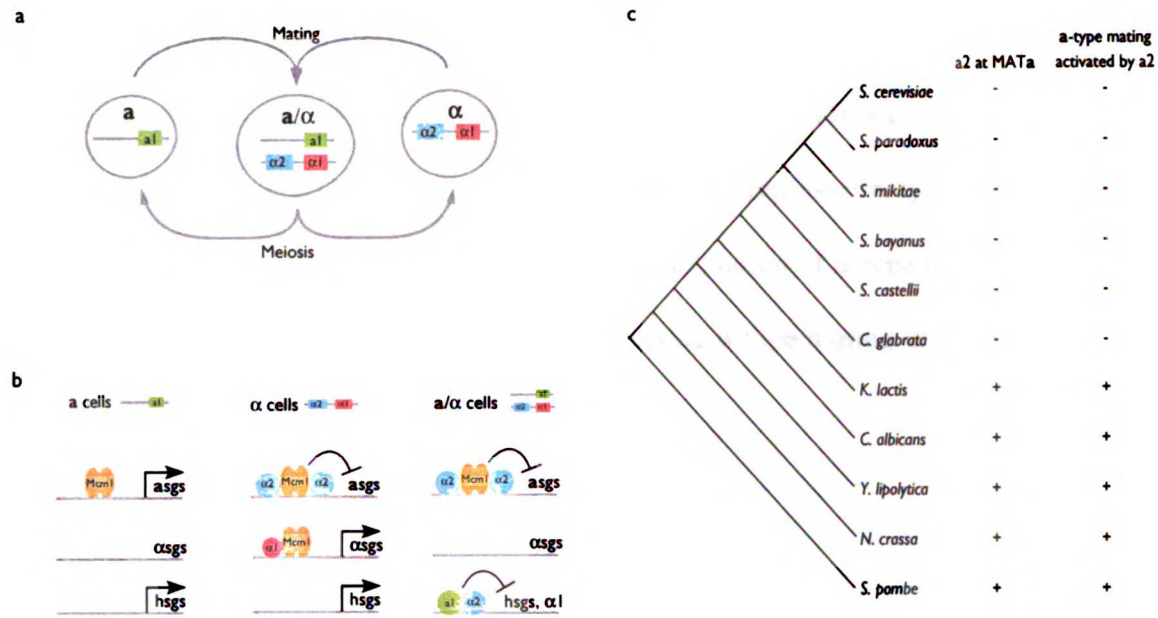


Figure 1. a-type mating is negatively regulated in modern *S. cerevisiae*, but was positively regulated in its ancestor.

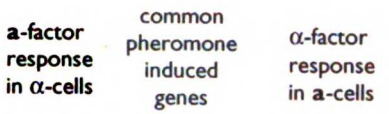
a. In *S. cerevisiae*, the MAT locus exists as one of two alleles, "**a**" or "α." Cells expressing only the **a**-locus (**a**-cells) can mate with cells expressing only the α-locus (α-cells), yielding a non-mating conjugant that expresses both loci (**a**/α cells).

b. The MAT loci encodes three transcription factors that control mating-type by turning on or off the **a**sgs, the αsgs, and the hsgs. In **a**-cells, the ubiquitous regulator Mcm1 activates **a**sgs, and the hsgs are constitutively active. In α cells, α2 and Mcm1 work together to repress **a**sgs, and α1 from MATα activates αsgs. In **a**/α cells, a1 from the **a**-locus and α2 from the α-locus cooperate to repress the hsgs and α1. α2 and Mcm1 repress the **a**sgs, and the αsgs are off, as α1 is repressed.

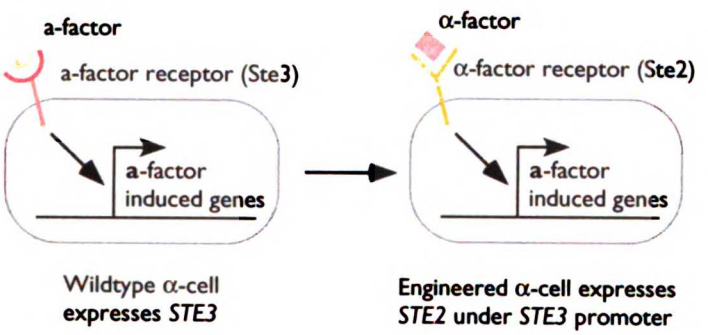
c. An HMG-domain transcription factor, a2, is present in the MAT**a** loci over, but is missing in *S. cerevisiae* and its closest relatives. a2 has been experimentally determined to be an activator of **a**-type mating in many yeasts, including *K. lactis*, *C. albicans*, *N. crassa*, *Y. lipolytica*, and *S. pombe* (for references, see text), but it was recently lost by the *S. cerevisiae* branch (Butler et al., 2004). Instead, α2 represses **a**-type mating in *S. cerevisiae*. The parsimonious interpretation is that activation of **a**-type mating represents the ancestral state, and that the role of α2 in repressing **a**-type mating is a recent innovation.

FIGURE 2

a


a-factor response in α-cells
common pheromone induced genes
α-factor response in a-cells

b


a-factor
a-factor receptor (Ste3)
a-factor induced genes
α-factor
α-factor receptor (Ste2)
a-factor induced genes
Wildtype α-cell expresses STE3
Engineered α-cell expresses STE2 under STE3 promoter

c

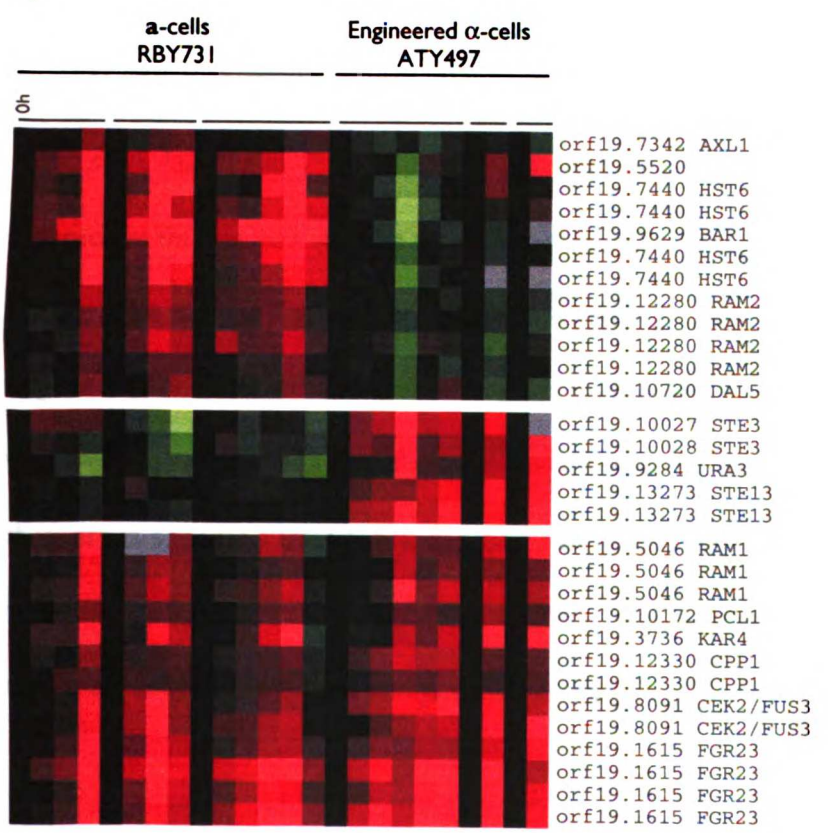

a-cells RBY731
Engineered α-cells ATY497
0h
orf19.7342 AXL1
orf19.5520
orf19.7440 HST6
orf19.7440 HST6
orf19.9629 BAR1
orf19.7440 HST6
orf19.7440 HST6
orf19.12280 RAM2
orf19.12280 RAM2
orf19.12280 RAM2
orf19.12280 RAM2
orf19.10720 DAL5
orf19.10027 STE3
orf19.10028 STE3
orf19.9284 URA3
orf19.13273 STE13
orf19.13273 STE13
orf19.5046 RAM1
orf19.5046 RAM1
orf19.5046 RAM1
orf19.10172 PCL1
orf19.3736 KAR4
orf19.12330 CPP1
orf19.12330 CPP1
orf19.8091 CEK2/FUS3
orf19.8091 CEK2/FUS3
orf19.1615 FGR23
orf19.1615 FGR23
orf19.1615 FGR23
orf19.1615 FGR23

Figure 2. Identification of the a-specific genes in *C. albicans*

a. Detection of **a**-specific genes in *C. albicans* requires induction of **a**-cells by α-factor. α-factor induces both the **a**-specific genes as well as general pheromone response genes; parsing the **a**-specific genes requires determining the α cell response to **a**-factor.

b. Because **a**-factor has not yet been identified in *C. albicans*, we have engineered an α strain which expresses α-factor receptor *STE2* under the promoter of the α-specific gene *STE3* (ATY497). When exposed to α-factor, this genotypically α strain upregulates it's **a**-factor-induced genes.

c. We compare the transcriptional profiles of wildtype **a** cells (RBY731) and engineered α cells (ATY497) on α-factor induction. Six α-factor induction time-courses, indicated by the solid bars, were performed. In the first two time-courses, wildtype **a** cells was sampled at 0h, 1h, 2h, and 4h (Data from these two time-courses previously published in Bennett et al., 2003). In the third time-course, wildtype **a** cells were sampled at 0h, 10m, 30m, 1h, 2h, and 4h. In the fourth time-course, engineered α-cells were sampled at 0h, 10m, 30m, 1h, 2, and 4h. In the fifth and sixth time-courses, engineered α-cells were sampled at 0h and 4h. Results were clustered as previously described (Bennett et al., 2003). The top panel shows genes that were upregulated in wildtype **a** cells, but not engineered α-cells, indicating that they are **a**-specific. The middle panel shows genes that were upregulated in engineered α-cells, but not in wildtype **a** cells, indicating that they are α-specific. *URA3* is under the *STE3* promoter in the engineered α-cells, as it was used as a marker for integration of *STE2*. The bottom panel shows a subset of genes that were induced in both **a** and α cells, representing the general pheromone response.

FIGURE 3

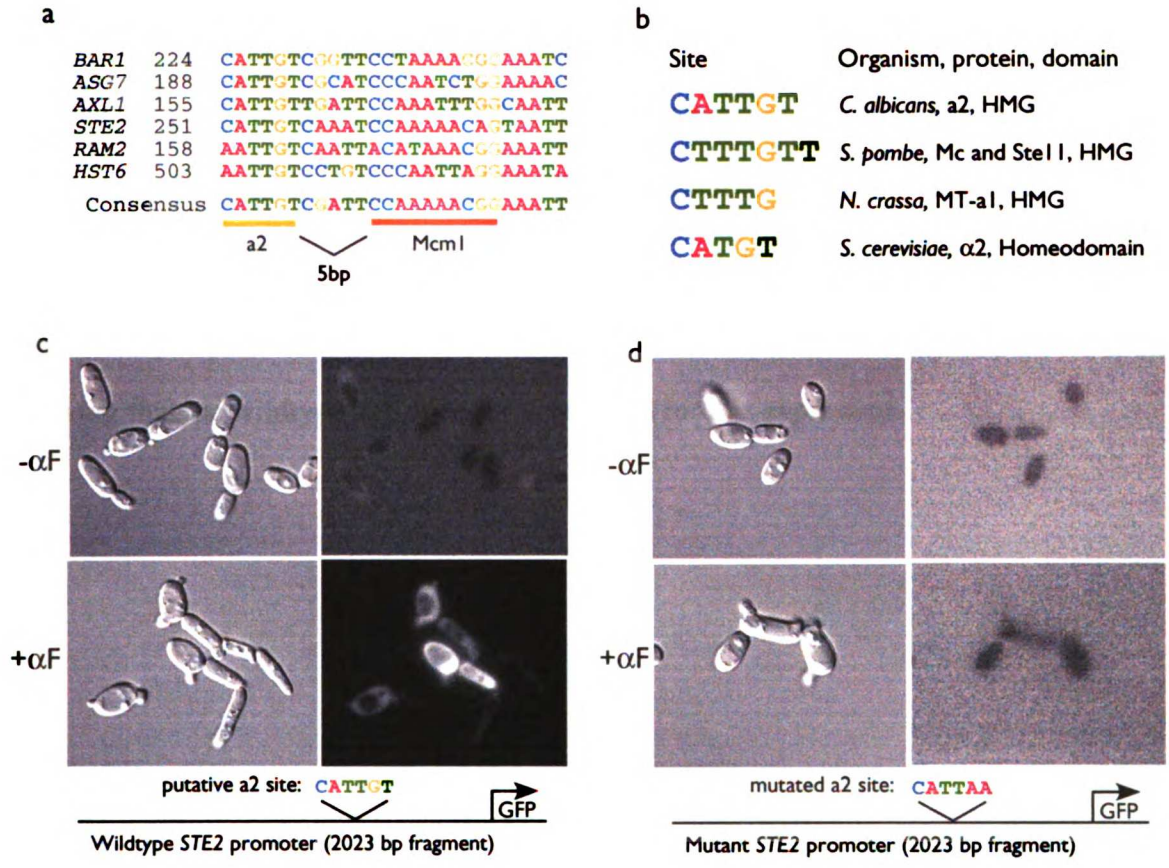


Figure 3. Identification and validation of the *C. albicans* asg operator

a. We submitted 1000 bp upstream of each *C. albicans* asg to MEME. A single motif was found in the promoters of 6 asgs. The element was in the same orientation in all promoters, and the distance from the translation start site is indicated next to the gene name. The element contains a highly conserved 6-bp site, indicated by the yellow line, and a putative Mcm1 binding site, indicated by the orange line, separated by 5 bp.

b. The 6-bp motif is similar to the demonstrated binding sites of MTa-1 from *Neurospora crassa* as well as MatMc and ste11 from *S. pombe*, all of which activate a-type mating in

their respective organisms (for references, see text). It is also similar to the binding site for *S. cerevisiae* α2.

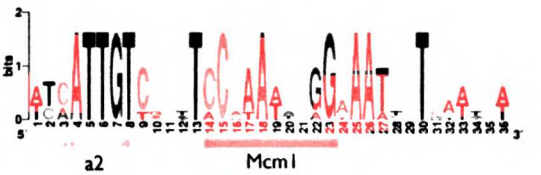
c,d. We fused either a wildtype (c) or mutant (d) *STE2* promoters to a GFP reporter. The mutant promoter differed from the wildtype promoter in two basepairs in the putative a2-binding element. The top panels in c and d represent uninduced cells. On addition of α-factor, both strains form mating projections (bottom left panels). However, only the wildtype *STE2* promoter is able to activate expression of GFP (bottom right panels), showing that the wildtype asg operator is required for asg expression.

FIGURE 4

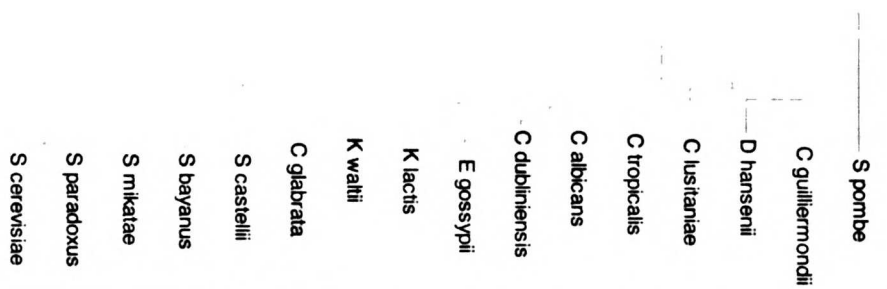
a.

S. cerevisiae α2-Mcm1 motif

C. albicans a2-Mcm1 motif



b.



c.

YDR461W	MFA1	7.99	8.05	8.31	8.53	8.64	0.95	-0.56	2.97	3.11							
YNL145W	MFA2	8.78	7.08	8.22	7.01	8.63											
YGL032C	AGA2	7.79	8.81	8.43	8.38	8.35	1.80	-1.27	1.75	3.40							
YJL170C	ASG7	7.08	2.11	7.08	7.00	3.13		-0.55	-0.65	2.30	-1.38	-1.38	-0.84	1.51	-0.68		
YIL015W	BAR1	8.54	7.45	7.43	7.75	8.64	3.45	0.39	1.63	0.82	-0.35	-1.52	-0.71		-0.02		
YFL026W	STE2	7.86	6.45	7.86	5.83	6.19	-0.31	0.09	1.47	5.61	-1.76	-0.32	-0.11	-1.60	-1.80	-1.12	2.07
YKL209C	STE6	7.19	7.29	8.27	6.49	2.17	1.96	-1.28	0.32	0.21	0.11	0.41		-2.39	-1.76	1.88	0.64
YKL019W	RAM2	-0.65	-0.84	-0.96	-1.27	-1.59	-1.71	-1.72	-1.11	-1.15	0.98	-0.20	1.83	0.51	0.86	-1.26	-1.26
YPR122W	AXL1	0.38	0.99	0.82	-0.28	3.09	-1.38	-1.87	-1.16	-1.44	-1.35	-0.45	-0.07	-0.94	-1.65	0.65	0.34

d.

YDR461W	MFA1	-0.67	-1.15	-0.43	0.66	-0.52	-0.17	1.47	0.72	1.53							
YNL145W	MFA2	-0.10	-0.40	-0.22	-0.52	0.58											
YGL032C	AGA2	-0.36	-0.66	-0.49	0.04	-1.38	-0.54	-0.08	0.80	0.45							
YJL170C	ASG7	-0.32	-0.75	-0.33	-0.10	0.33		-0.37	-0.32	2.10	3.24	3.32	3.13	0.97	-0.27		
YIL015W	BAR1	-0.92	-0.79	-1.18	-0.38	0.42	0.07	-0.61	0.81	0.58	5.99	6.44	0.14		0.29		
YFL026W	STE2	-0.61	-1.13	-0.51	0.81	-0.52	-1.01	1.41	0.48	4.55	3.71	4.19	3.50	1.58	2.07	1.86	0.72
YKL209C	STE6	0.59	0.66	0.29	-0.75	0.75	1.52	0.78	0.19	1.12	4.06	4.82		-0.10	1.50	4.37	-0.42
YKL019W	RAM2	-0.43	0.49	0.07	-0.30	-0.52	-1.49	-0.53	-0.76	-0.46	4.13	4.56	1.37	-0.62	-0.57	0.32	-0.97
YPR122W	AXL1	-0.24	-0.43	-0.84	1.22	-0.25	-1.09	0.43	1.00	-0.26	4.61	4.89	0.09	-0.02	0.26	2.73	-0.64

e. *K. lactis* motif

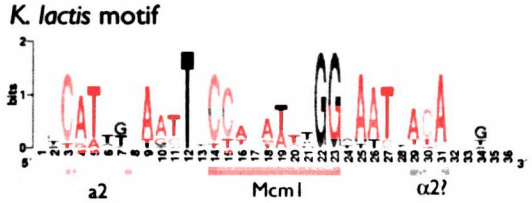
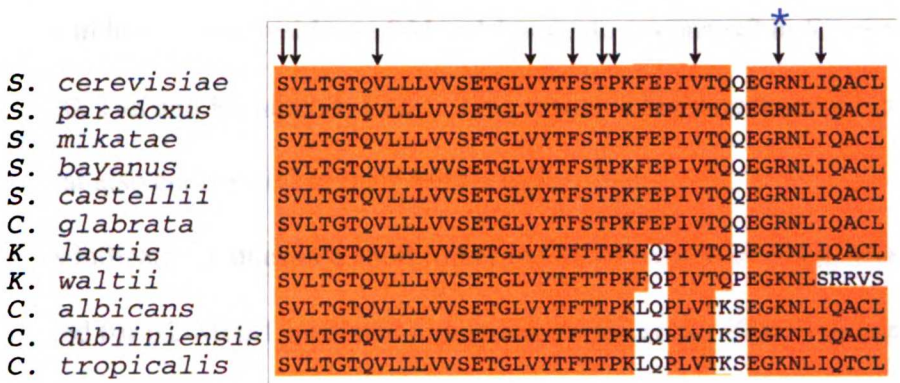


Figure 4. Analysis of *cis*- asg regulation across species

a. *S. cerevisiae* α2-Mcm1 and *C. albicans* a2-Mcm1 position specific scoring matrices (PSSMs) were determined using the seven *S. cerevisiae* asgs or six *C. albicans* asgs. The Mcm1 site is represented by an orange bar. The α2 site is represented by blue, and the a2 site by yellow. b. Methods similar to those of Rokas et al. (Rokas et al., 2003) were used to infer a phylogeny of 16 sequenced yeasts. Promoters of asg orthologs from 16 yeasts were then scanned with either the *S. cerevisiae* PSSM (c.) or *C. albicans* PSSM (d.). Maximum log-odds scores are shown. Red indicates a strong match for the given PSSM. e. Promoters from the *K. waltii, K. lactis,* and *E. gossypii* orthologs of *ASG7, BAR1, STE2,* and *STE6* were pooled and submitted to MEME (Bailey and Elkan, 1994). The recovered site has elements of both the *S. cerevisiae* and *C. albicans* sites. An a2-like site (yellow) and an α2-like site (blue) flank an Mcm1 site.

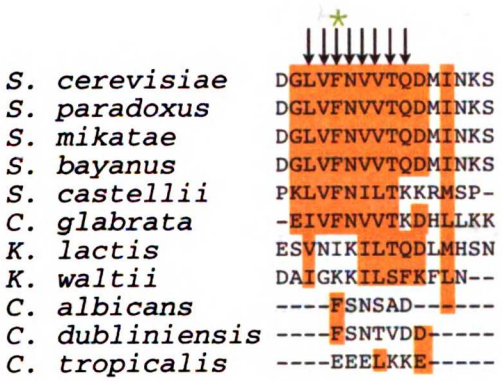
FIGURE 5

a.



Mcm1 alignment, *S. cerevisiae* residues 51-94

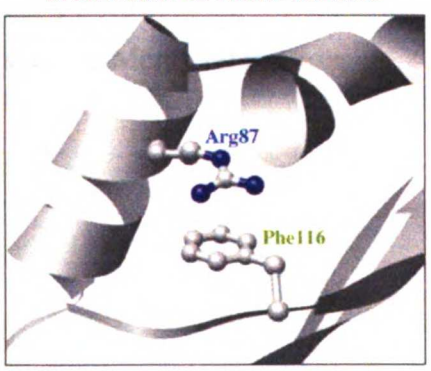
b.



α2 alignment
S. cerevisiae
residues 112-127

c.

S. cerevisiae α2-Mcm1 interface



K. lactis α2-Mcm1 interface (model)

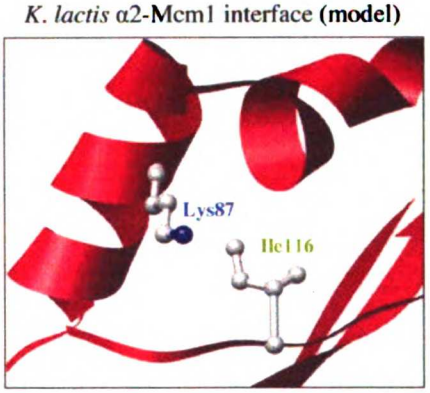


Figure 5.

a. Arrows indicate those residues of Mcm1 known to contact α2 in *S. cerevisiae* (Acton et al., 2000, Tan and Richmond, 1998). This region is highly conserved in the yeast species that also encodes α2 at their *MATα* loci.

b. Arrows indicate residues in a short linker region of α2 that contacts Mcm1, and that are required for α2-Mcm1 mediated repression (Mead et al., 1996, Tan and Richmond, 1998). This region is well-conserved to *C. glabrata*. *K. lactis* and *K. waltii* α2 also show limited conservation in this region. The homeodomain of α2 is alignable in all species (not shown).

c. The left panel shows the structure of the interface between the *S. cerevisiae* α2 linker region and Mcm1 (Tan and Richmond, 1998). Arg87 from Mcm1 (indicated by the blue asterisk in the alignment) and Phe116 (indicated by the green asterisk in α2) form a highly favorable pi-stacking interaction. In the right panel, *K. lactis* sequences are modeled onto the *S. cerevisiae* structure. The Arg87-Phe116 interaction is not present at this interface, suggesting that the *K. lactis* α2-Mcm1 interaction is weaker than the *S. cerevisiae* α2-Mcm1 interaction.

Figure 6

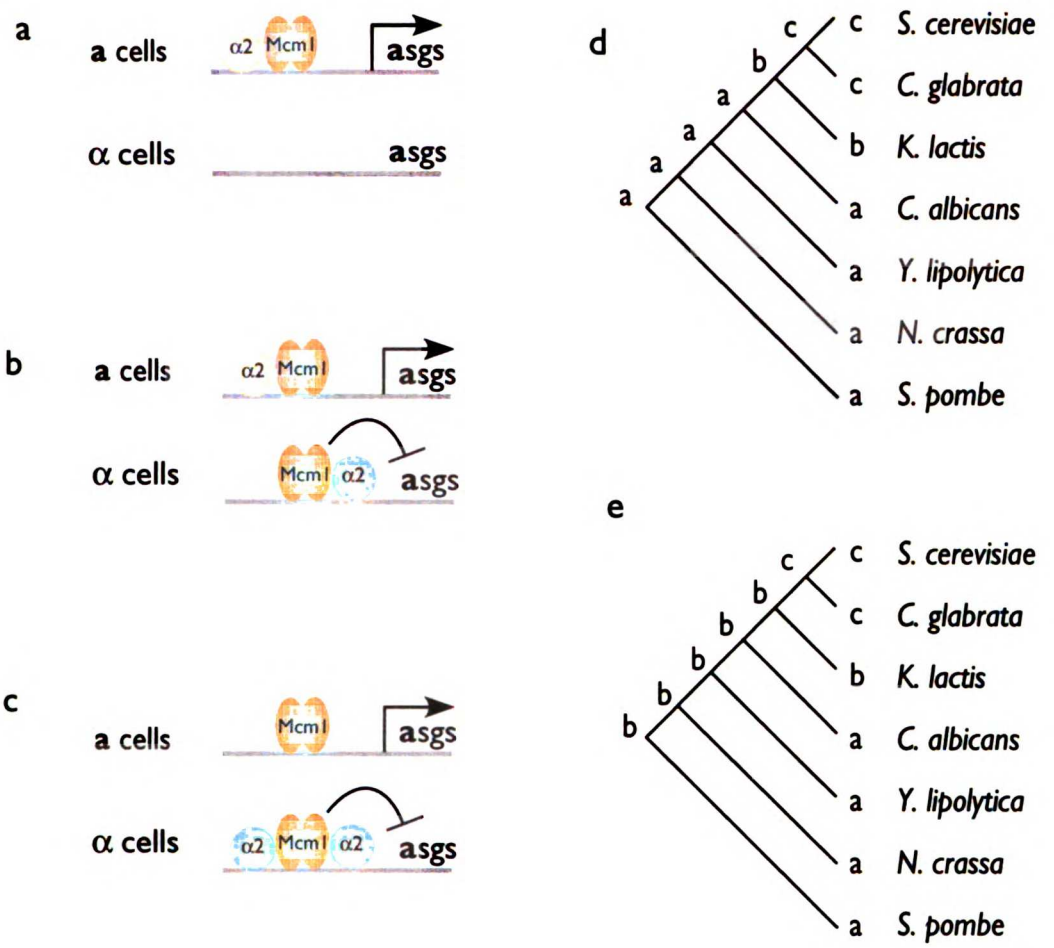


Figure 6. Ordering the changes in *cis*- and *trans*- regulatory elements

a. In an ancestral yeast, a2 and Mcm1 cooperated to activate **asgs** in **a**-cells. This scheme of **asg** regulation persists in modern *C. albicans*.

b. Based on evidence from *cis*- and *trans*- elements in *K. lactis, K. waltii*, and *E. gossypii*, the **asgs** appear to have undergone a state in which they were both positively regulated by a2-Mcm1 in **a** cells and negatively regulated by α2-Mcm1 in α cells.

c. In modern *S. cerevisiae*, **asgs** are activated by Mcm1 in **a**-cells, and repressed by α2-Mcm1 in α cells.

d,e. The asg regulatory schemes illustrated in parts a, b, and c of this figure are mapped onto extant species arranged by their phylogenetic relationships. *C. albicans, Y. lipolytica, N. crassa,* and *S. pombe* are known to activate asgs using an a2 homolog, and their asg regulation most closely resembles scheme "a." *K. lactis* appears to fit into scheme "b," and *S. cerevisiae* and *C. glabrata* fit scheme "c."

We also map the asg regulatory schemes onto ancestral nodes. There are two possibilities: **d.** We map scheme "a" as an ancestral state and scheme "b" as a transitional state, first appearing in the common ancestor of *K. lactis* and *S. cerevisiae.* Scheme "c" is the most derived state, first appearing in the ancestor of *C. glabrata* and *S. cerevisiae.* **e.** We have not ruled out a model in which scheme "b" represents the ancestral state. In this case, both "c" and "a" are derived forms. Scheme "c" arose once in the ancestor of *C. glabrata* and *S. cerevisiae.* In this scenario, scheme "a" must have arisen independently at least four times in the lineages that carry *S. pombe, N. crassa, Y. lipolytica,* and *C. albicans.*

ACKNOWLEDGMENTS

We are very grateful to Mike Lorentz and Gerald Fink for the collaboration that produced the DNA microarrays used in this paper, to Stefan Ästrøm for providing unpublished *K. lactis* strains and much advice on their handling, and to Kenneth Wolfe generously bringing the *C. albicans* a2 ORF to our attention. We are also grateful to the Stanford Genome Technology Center (http://www-sequence.stanford.edu/group/candida), the Broad Institute (http://www.broad.mit.edu/annotation/fungi/fgi/), the Sanger Center (http://www.sanger.ac.uk/Projects/Fungi/) and Genolevures (http://cbi.labri.fr/Genolevures/index.php) for providing sequence data. Sequencing of the *C. albicans* genome was supported by NIDR and the Burroughs Wellcome Fund. We thank Diane Inglis, Andrew Uhl, and Burk Braun for their extensive efforts in genome annotation, Bethann Hromatka for overseeing microarray printing, and Richard Bennett, Rebecca Zordan, Andrew Uhl, Matt Lohse, and members of the Johnson lab for many fruitful discussions. This work was supported by grants from the Burroughs Wellcome Fund (993218) and the NIH (R01 AI49187) to A.D.J.. A.E.T. is a Howard Hughes Medical Institute Predoctoral Fellow.

REFERENCES

Acton, T. B., Mead, J., Steiner, A. M., and Vershon, A. K. (2000). Scanning mutagenesis of Mcm1: residues required for DNA binding, DNA bending, and transcriptional activation by a MADS-box protein. Mol Cell Biol *20*, 1-11.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol *2*, 28-36.

Bender, A., and Sprague, G. F., Jr. (1986). Yeast peptide pheromones, a-factor and alpha-factor, activate a common response mechanism in their target cells. Cell *47*, 929-937.

Bennett, R. J., Uhl, M. A., Miller, M. G., and Johnson, A. D. (2003). Identification and Characterization of a Candida albicans Mating Pheromone. Mol Cell Biol *23*, 8189-8201.

Buchler, N. E., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. Proc Natl Acad Sci U S A *100*, 5136-5141.

Butler, G., Kenny, C., Fagan, A., Kurischko, C., Gaillardin, C., and Wolfe, K. H. (2004). Evolution of the MAT locus and its Ho endonuclease in yeast species. Proc Natl Acad Sci U S A *101*, 1632-1637.

Calderone, R. A. (2002). Candida and Candidiasis (Washington, D.C.: ASM Press).

Care, R. S., Trevethick, J., Binley, K. M., and Sudbery, P. E. (1999). The MET3 promoter: a new tool for Candida albicans molecular genetics. Mol Microbiol *34*, 792-798.

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

Carroll, S. B., Grenier J. K., Weatherbee, S. D. (2001). From DNA to Diversity (Malden, Massachusetts: Blackwell Science, Inc.).

Chen, L., DeVries, A. L., and Cheng, C. H. (1997). Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc Natl Acad Sci U S A *94*, 3811-3816.

Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A., and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science *301*, 71-76.

Copley, S. D. (2000). Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. Trends Biochem Sci *25*, 261-265.

Coppin, E., Debuchy, R., Arnaise, S., and Picard, M. (1997). Mating types and sexual development in filamentous ascomycetes. Microbiol Mol Biol Rev *61*, 411-428.

Cormack, B. P., Bertram, G., Egerton, M., Gow, N. A., Falkow, S., and Brown, A. J. (1997). Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in Candida albicans. Microbiology *143 (Pt 2)*, 303-311.

Darwin, C. R. (1859). The Origin of Species (New York, New York: Gramercy).

Davidson, E. H. (2001). Genomic Regulatory Systems (San Diego, CA: Academic Press).

de Bono, M., and Hodgkin, J. (1996). Evolution of sex determination in caenorhabditis: unusually high divergence of tra-1 and its functional consequences. Genetics *144*, 587-595.

Debuchy, R., Arnaise, S., and Lecellier, G. (1993). The mat- allele of Podospora anserina contains three regulatory genes required for the development of fertilized female organs. Mol Gen Genet *241*, 667-673.

Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., *et al.* (2004). The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. Science *304*, 304-307.

Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B., and Fairhead, C. (2005). Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. Mol Biol Evol *22*, 856-873.

Gasch, A. P., Moses, A. M., Chiang, D. Y., Fraser, H. B., Berardini, M., and Eisen, M. B. (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol *2*, e398.

Gerhart, J., and Kirschner, M. (1997). Cells, Embryos, and Evolution (Malden, Massachusetts: Blackwell Science, Inc.).

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., and Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature *433*, 481-487.

Haag, E. S., Wang, S., and Kimble, J. (2002). Rapid coevolution of the nematode sex-determining genes fem-3 and tra-2. Curr Biol *12*, 2035-2041.

Hedges, S. B. (2002). The origin and evolution of model organisms. Nat Rev Genet *3*, 838-849.

Herskowitz, I., Rine, J., and Strathern, J. (1992). Mating-type determination and mating-type interconversion in *Saccharomyces cerevisiae.*, In The Molceular Cellular Biology fo the Yeast *Saccharomyces*, J. R. P. E.W. Jones, J.R. Broach, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 583-657.

Higgins, D. G., and Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene *73*, 237-244.

Hull, C. M., and Johnson, A. D. (1999). Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans. Science *285*, 1271-1275.

Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yanai, I., McClellan, M., Berman, J., and Barkai, N. (2005). Rewiring of the yeast transcriptional network through the evolution of motif usage. Science *309*, 938-940.

Johnson, A. D. (1995). Molecular mechanisms of cell-type determination in budding yeast. Curr Opin Genet Dev *5*, 552-558.

Kellis, M., Birren, B. W., and Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature *428*, 617-624.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241-254.

Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988). Four mating-type genes control sexual differentiation in the fission yeast. Embo J *7*, 1537-1547.

Kjaerulff, S., Dooijes, D., Clevers, H., and Nielsen, O. (1997). Cell differentiation by interaction of two HMG-box proteins: Mat1-Mc activates M cell-specific genes in S.pombe by recruiting the ubiquitous transcription factor Ste11 to weak binding sites. Embo J *16*, 4021-4033.

Kulathinal, R. J., Skwarek, L., Morton, R. A., and Singh, R. S. (2003). Rapid evolution of the sex-determining gene, transformer: structural diversity and rate heterogeneity among sibling species of Drosophila. Mol Biol Evol *20*, 441-452.

Kurischko, C., Schilhabel, M. B., Kunze, I., and Franzl, E. (1999). The MATA locus of the dimorphic yeast Yarrowia lipolytica consists of two divergently oriented genes. Mol Gen Genet *262*, 180-188.

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., and Hartl, D. L. (2005). Compensatory cis-trans Evolution and the Dysregulation of Gene Expression in Interspecific Hybrids of Drosophila. Genetics.

Lengeler, K. B., Davidson, R. C., D'Souza, C., Harashima, T., Shen, W. C., Wang, P., Pan, X., Waugh, M., and Heitman, J. (2000). Signal transduction cascades regulating fungal development and virulence. Microbiol Mol Biol Rev *64*, 746-785.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.

Ludwig, M. Z., Patel, N. H., and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development *125*, 949-958.

Mead, J., Bruning, A. R., Gill, M. K., Steiner, A. M., Acton, T. B., and Vershon, A. K. (2002). Interactions of the Mcm1 MADS box protein with cofactors that regulate mating in yeast. Mol Cell Biol *22*, 4607-4621.

Mead, J., Zhong, H., Acton, T. B., and Vershon, A. K. (1996). The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. Mol Cell Biol *16*, 2135-2143.

Miller, M. G., and Johnson, A. D. (2002). White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. Cell *110*, 293-302.

Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217.

O'Neil, M. T., and Belote, J. M. (1992). Interspecific comparison of the transformer gene of Drosophila reveals an unusually high degree of evolutionary divergence. Genetics *131*, 113-128.

Philley, M. L., and Staben, C. (1994). Functional analyses of the Neurospora crassa MT a-1 mating type polypeptide. Genetics *137*, 715-722.

Ptashne, M., and Gann, A. (1998). Imposing specificity by localization: mechanism and evolvability. Curr Biol *8*, R897.

Rokas, A., Williams, B. L., King, N., and Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature *425*, 798-804.

Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. Nature *415*, 914-917.

Savageau, M. A. (1983). Regulation of differentiated cell-specific functions. Proc Natl Acad Sci U S A *80*, 1411-1415.

Scannell, D. R., and Wolfe, K. (2004). Rewiring the transcriptional regulatory circuits of cells. Genome Biol *5*, 206.

Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics *18*, 502-504.

Singh, R. S., and Kulathinal, R. J. (2000). Sex gene pool evolution and speciation: a new paradigm. Genes Genet Syst *75*, 119-130.

Smith, D. L., and Johnson, A. D. (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.

Tan, S., and Richmond, T. J. (1998). Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature *391*, 660-666.

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

Tsong, A. E., Miller, M. G., Raisner, R. M., and Johnson, A. D. (2003). Evolution of a combinatorial transcriptional circuit: a case study in yeasts. Cell *115*, 389-399.

Turgeon, B. G., Bohlmann, H., Ciuffetti, L. M., Christiansen, S. K., Yang, G., Schafer, W., and Yoder, O. C. (1993). Cloning and analysis of the mating type genes from Cochliobolus heterostrophus. Mol Gen Genet *238*, 270-284.

Wall, M. E., Hlavacek, W. S., and Savageau, M. A. (2004). Design of gene circuits: lessons from bacteria. Nat Rev Genet *5*, 34-42.

Whitfield, L. S., Lovell-Badge, R., and Goodfellow, P. N. (1993). Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature *364*, 713-715.

Wilson, R. B., Davis, D., Enloe, B. M., and Mitchell, A. P. (2000). A recyclable Candida albicans URA3 cassette for PCR product-directed gene disruptions. Yeast *16*, 65-70.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. Nature *430*, 85-88.

Yabana, N., and Yamamoto, M. (1996). Schizosaccharomyces pombe map1+ encodes a MADS-box-family protein required for cell-type-specific gene expression. Mol Cell Biol *16*, 3420-3428.

Appendices to Chapter 3

Appendix A: Position Specific Scoring Matrices

C. albicans a2-Mcm1 PSSM

```
ALPHABET= ACGT
probability matrix: alength= 4
0.400     0.100     0.100     0.400
0.100     0.300     0.100     0.500
0.300     0.500     0.100     0.100
0.700     0.100     0.100     0.100
0.100     0.100     0.100     0.700
0.100     0.100     0.100     0.700
0.100     0.100     0.700     0.100
0.100     0.100     0.100     0.700
0.100     0.600     0.100     0.200
0.300     0.200     0.400     0.100
0.400     0.200     0.200     0.200
0.300     0.100     0.200     0.400
0.100     0.100     0.100     0.700
0.200     0.600     0.100     0.100
0.100     0.700     0.100     0.100
0.400     0.300     0.100     0.200
0.600     0.100     0.100     0.200
0.700     0.100     0.100     0.100
0.400     0.100     0.100     0.400
0.400     0.200     0.100     0.300
0.200     0.300     0.200     0.300
0.200     0.100     0.600     0.100
0.100     0.100     0.700     0.100
0.500     0.200     0.100     0.200
0.700     0.100     0.100     0.100
0.700     0.100     0.100     0.100
0.200     0.100     0.100     0.600
0.200     0.300     0.100     0.400
0.400     0.200     0.200     0.200
0.100     0.100     0.100     0.700
0.100     0.400     0.300     0.200
0.500     0.200     0.100     0.200
0.400     0.100     0.100     0.400
0.100     0.200     0.300     0.400
0.300     0.300     0.200     0.200
0.400     0.100     0.100     0.400
```

S. cerevisiae α2-Mcm1 PSSMs

Two matrices were used, one which scores a 5bp space between the Mcm1 and α2 site, and one which scores a 6bp space. The spacing is variable in *S. cerevisiae* asg operators.

```
6 bp spacing:

ALPHABET= ACGT
probability matrix: alength= 4
0.250 0.250 0.167 0.333
0.333 0.083 0.333 0.250
0.167 0.167 0.500 0.167
0.167 0.583 0.083 0.167
0.583 0.083 0.250 0.083
0.083 0.167 0.083 0.667
0.083 0.083 0.750 0.083
0.083 0.167 0.083 0.667
0.750 0.083 0.083 0.083
0.500 0.250 0.083 0.167
0.083 0.083 0.083 0.750
0.083 0.083 0.083 0.750
0.750 0.083 0.083 0.083
0.167 0.667 0.083 0.083
0.083 0.750 0.083 0.083
0.083 0.250 0.417 0.250
0.417 0.167 0.167 0.250
0.667 0.083 0.083 0.167
0.333 0.083 0.083 0.500
0.417 0.250 0.083 0.250
0.417 0.167 0.333 0.083
0.083 0.083 0.750 0.083
0.083 0.083 0.750 0.083
0.750 0.083 0.083 0.083
0.750 0.083 0.083 0.083
0.750 0.083 0.083 0.083
0.083 0.167 0.083 0.667
0.083 0.083 0.083 0.750
0.200 0.100 0.100 0.600
0.667 0.083 0.083 0.167
0.083 0.750 0.083 0.083
0.750 0.083 0.083 0.083
0.167 0.167 0.083 0.583
0.083 0.083 0.667 0.167
0.167 0.167 0.333 0.333
0.167 0.250 0.083 0.500
```

5 bp spacing

ALPHABET= ACGT
probability matrix: alength= 4
0.250 0.250 0.167 0.333
0.333 0.083 0.333 0.250
0.167 0.167 0.500 0.167
0.167 0.583 0.083 0.167
0.583 0.083 0.250 0.083
0.083 0.167 0.083 0.667
0.083 0.083 0.750 0.083
0.083 0.167 0.083 0.667
0.750 0.083 0.083 0.083
0.500 0.250 0.083 0.167
0.083 0.083 0.083 0.750
0.083 0.083 0.083 0.750
0.750 0.083 0.083 0.083
0.167 0.667 0.083 0.083
0.083 0.750 0.083 0.083
0.083 0.250 0.417 0.250
0.417 0.167 0.167 0.250
0.667 0.083 0.083 0.167
0.333 0.083 0.083 0.500
0.417 0.250 0.083 0.250
0.417 0.167 0.333 0.083
0.083 0.083 0.750 0.083
0.083 0.083 0.750 0.083
0.750 0.083 0.083 0.083
0.750 0.083 0.083 0.083
0.750 0.083 0.083 0.083
0.083 0.167 0.083 0.667
0.083 0.083 0.083 0.750
0.667 0.083 0.083 0.167
0.083 0.750 0.083 0.083
0.750 0.083 0.083 0.083
0.167 0.167 0.083 0.583
0.083 0.083 0.667 0.167
0.167 0.167 0.333 0.333
0.167 0.250 0.083 0.500
0.417 0.083 0.250 0.250

Appendix B: *K. lactis* mating-type regulation

Background

In Chapter 3, we observe that during the transition from positive to negative regulation, asgs appear to have undergone a transitional state in which they were both positively regulated by a2 and negatively regulated by α2. This observation is based on evidence in *K. lactis*, *K. waltii*, and *E. gossypii* that the asg operator has an α2 core site as well as an a2 binding site, as well as informatic evidence in *K. lactis* and *K. waltii* for a weak α2-Mcm1 interaction.

A key prediction of this observation is that deleting a2 in an **a**-cell should reduce **a**-type mating, and that deleting α2 in an α-cell should derepress **a**-type mating (i.e., allow weak **a**- and α-type mating).

Quantitative mating of *K. lactis* Δa2 and Δα2 strains

To test this hypothesis, I generated a2 deletants in **a**-strains and α2 deletants in α-strains (wildtype strains courtesy of Stefan Ästrøm), and tested the ability of these strains to mate against wildtype **a** strains and wildtype α strains. On quantitative mating, I found that deletion of a2 reduces the ability of **a**-strains to mate with α-strains by approximately 100-fold (Table 1). However, when I examined the MAT configurations of the mating products by PCR, I found that all of the products encoded wildtype a2 and had lost the Δa2 allele at their MAT loci, indicating that the silent copy of a2 had restored the MAT locus to its wildtype configuration through gene conversion. This suggests that a2 is strictly required for **a**-type mating in *K. lactis*.

Corroborating this observation, when Δα2 α-strains were mated with wildtpe α-

strains, no mating products were recovered (data not shown), suggesting that Δα2 does not derepress a-type mating in an α strain (Table 1):

Parent 1	Parent 2	Mating frequency
ATY607-a-WT	ATY606-α-WT	1.6E-03
ATY607-a-WT	ATY606-α-WT	1.4E-03
ATY619-a-Δa2	ATY606-α-WT	1.5E-05
ATY620-a-Δa2	ATY606-α-WT	4.9E-06
ATY621-a-Δa2	ATY606-α-WT	7.3E-06
ATY622-a-Δa2	ATY606-α-WT	1.4E-05
ATY607-a-WT	ATY625-α-Δα2	3.7E-04
ATY607-a-WT	ATY626-α-Δα2	3.8E-04
ATY607-a-WT	ATY627-α-Δα2	6.7E-04
ATY607-a-WT	ATY628-α-Δα2	1.4E-03
ATY644-α-WT	ATY625-α-Δα2	<1.0E-07
ATY644-α-WT	ATY626-α-Δα2	<1.0E-07
ATY644-α-WT	ATY627-α-Δα2	<1.0E-07
ATY644-α-WT	ATY628-α-Δα2	<1.0E-07
ATY644-α-WT	ATY646-α-WT	<1.0E-07
ATY644-α-WT	ATY647-α-WT	<1.0E-07
ATY644-α-WT	ATY648-α-WT	<1.0E-07
ATY644-α-WT	ATY649-α-WT	<1.0E-07

Table 1. Quantitative mating of a- and α *K. lactis* strains

The first and second columns show the input strains in quantitative mating assays. The format is <strain name>-<MAT locus>-<genotype>. The third column shows the mating frequency. Wildtype strains in equal quantities mate at a frequency of 1/1000. The mating frequency of Δa2 strains is reduced ~100x, while Δα2 strains mate at wildtype frequencies as α-cells, and do not mate as **a**-cells.

Discussion

My quantitative mating results in *K. lactis* show suggest that a2 is required for **a**-type mating, and that α2 has no role in repressing **a**-type mating. However, our results do not rule out the possibility that a2 activates a subset of **a**sgs, and that α2 represses a subset of **a**sgs, the union of which is required for **a**-type mating. For instance, if a2 activated α-factor receptor and α2 repressed **a**-factor, a2 would be required for mating while a deletion of α2 would not derepress **a**-type mating, even if it derepressed an individual **a**sg. Such a result would reconcile the quantitative mating results with the informatic results. This hypothesis is testable by quantitative RT-PCR, which I have attempted without success.

It is also possible that the **a**sgs are solely activated by a2 in *K. lactis*, and the informatics merely show a vestigial dual-regulatory state in the *K. lactis* lineage.

Materials and Methods

Strains ATY606 and ATY607 were gifts from Stefan Ästrøm, Wennergren Institute, Stockholm University. ATY606 is mat α, *nej1::LEU2, ade1, trp1, leu2, metA1, uraA1.* ATY607 is mat **a**, *nej1::LEU2, lysA1, trp1, leu2, metA1, uraA1.* To knock out a2, ATO368 (5'aatttcacctaaactggtttcaaacgacagaattaaattcactgaattttgatgtataccAGCAGACAAGCCCGTCAGGG3') and ATO369 (5'ccagtgtggttaaagatagtcgattccctctgtgcagtccaatctcaagatccatcaatcCTCCTTACGCATCTGTGCGG3') were used to amplify a fragment from pRS306, the URA3 marked Sikorski vector (Sikorski and Hieter, Genetics 122: 19-27). This fragment was transformed into ATY607, yielding ATY619-622. The 5' and 3' junctions of integration were confirmed

by PCR, as well as the locus of integration. To knock out α2, ATO376 (5'acatttgagctgaagaattataccaaattcttaaaataaattccagtgaaggacaaccccAGCAGACAAGCCCG TCAGGG3') and ATO377 (5'tgaattgattaaatacttgcgagagtaccttcattatttatggttaacaatttgagtgtgCTCCTTACGCATCTGTG CGG3') were used to amplify a fragment from pRS306. This fragment was transformed into ATY606, yielding ATY625-628. Again, 5', 3', and locus integration was confirmed by PCR.

In order to test the ability of the Δα2 strains AT625-629 with wildtype MATα strains, it was necessary to construct differently marked MATα strains. ATO437 (5' tattccagaagatgattcatttctttatcgtagtagattgacagaagcgtaaaggtatatAGCAGACAAGCCCGTCA GGG3') and ATO438 (5' ggaagacggagctgtgcgtcaagcggttaagggccgtgctcgtaacttaaatggaacaaaCTCCTTACGCATCTG TGCGG3') were designed to replace the *HIS3* gene of *K. lactis* with either URA3 or TRP1. Using ATO437 and 438, URA3 from pRS306, or TRP1 from pRS304 was amplified (Sikorski). Fragments were then transformed into ATY606, yielding ATY644 (ATY606 but *his3::TRP1*), and ATY646 (ATY606 but *his3::URA3*). Fragments were also transformed into ATY607, yielding ATY638-639 (ATY607 but *his3::TRP1*) and ATY640-644 (ATY607 but *his3::URA3*). Integration at the *HIS3* locus was confirmed by PCR of the 5' and 3' junctions, as well as by replica plating on –his, -trp, or –ura SD media.

Transformation

K. lactis transformation was optimized based on an *S. cerevisiae* protocol. Strains were inoculated in YEPD, 30° overnight. 1 ml was then inoculated 10mls of YEPD, 30°, and grown for 5.5 hrs. Cells were washed in LiAce/TE and resuspended in 100 μl of LiAce/TE (100 mM Lithium Acetate, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA). 2 μg of PCR product and 10 μl of 10 mg/ml salmon sperm DNA were added to 50 μl of cells, with 800 μl of PEG solution (40% PEG, 100mM Lithium Acetate). Cells were incubated for 60 minutes at 30° on a roller drum. 40 μl of DMSO was added, and the cells were heat shocked at 42° for 15 minutes. Cells were then washed and resuspended in 200 μl ddH20. The entire transformation was plated on the appropriate drop-out plates.

Quantitative mating conditions

Quantitative mating was previously described (Miller and Johnson, 2002). For *K. lactis*, I used an equivalent of 0.3 OD of cells of each mating-type, instead of 1.5 ODs as used for *C. albicans*. The filter was then placed on YEPD at 30° for 24 hours.

Appendix C. Identification of the α1 site in *C. albicans*: an additional role for Mcm1 in facilitating a major transcriptional regulatory change

To identify the αsg operator in *C. albicans*, we submitted the promoters of *STE3* and *MFα* orthologs from *C. albicans*, *C. dubliniensis*, and *C. tropicalis* to MEME (Bailey and Elkan, 1994). A highly conserved element was found upstream of each ortholog (Figure 1). The element shows hallmarks of the 2-fold symmetric Mcm1 dimer site (purple and orange), as well as a highly conserved element (AAAGGAA) that is likely the α1 site, but has no resemblance to the *S. cerevisiae* α1 site (TCAATG). Although I have not yet biologically validated this site, it appears that α1 binding sites are entirely unconserved between *C. albicans* and *S. cerevisiae*, despite α1's biologically identical role. In both organisms, α1 activates α-type mating in conjunction with Mcm1; moreover, the targets of gene activation are orthologous.

The presence of Mcm1 at the α-specific gene promoters may have buffered otherwise negative fitness consequences of changes in α1 specificity. While we cannot speculate as to the mechanism for this change, it is remarkable that Mcm1 is central to this conversion, as it is also central to the conversion of asg regulation (Figure 2).

Mcm1 is a member of the MADS box family, a family of proteins which is as prominent in plant development as homeodomain proteins are in animal development. In *S. cerevisiae*, Mcm1 is a ubiquitous, essential transcription factor that is required for cellular processes as diverse as cell-cycle regulation, arginine synthesis, and mating-type regulation. It has been proposed that combinatorial control is advantageous due to the large number of regulatory outcomes that can be achieved with a small number of

transcription factors; it seems likely that the prevalence of combinatorial control also reflects its role in facilitating evolution of transcriptional regulation.

FIGURE 1

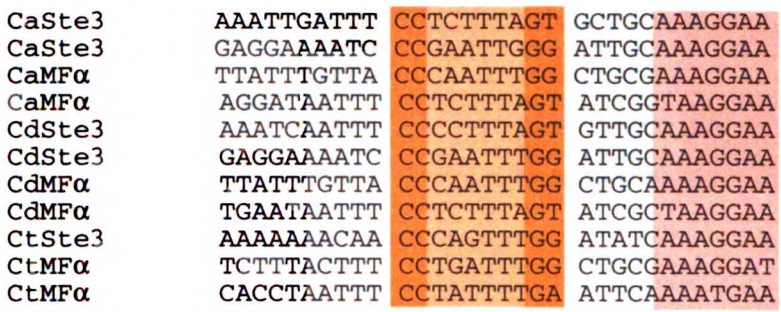


Figure 1. Identification of a potential αsg operator in *Candida* species. MEME identified an element that was present twice the promoters of *STE3* and *MFα* in orthologs from *C. albicans*, *C. dubliniensis*, and *C. tropicalis*. The elements show hallmarks of the 2-fold symmetric Mcm1 dimer site (orange), and indicate a highly conserved element (AAAGGAA) that may be the α1 site.

FIGURE 2

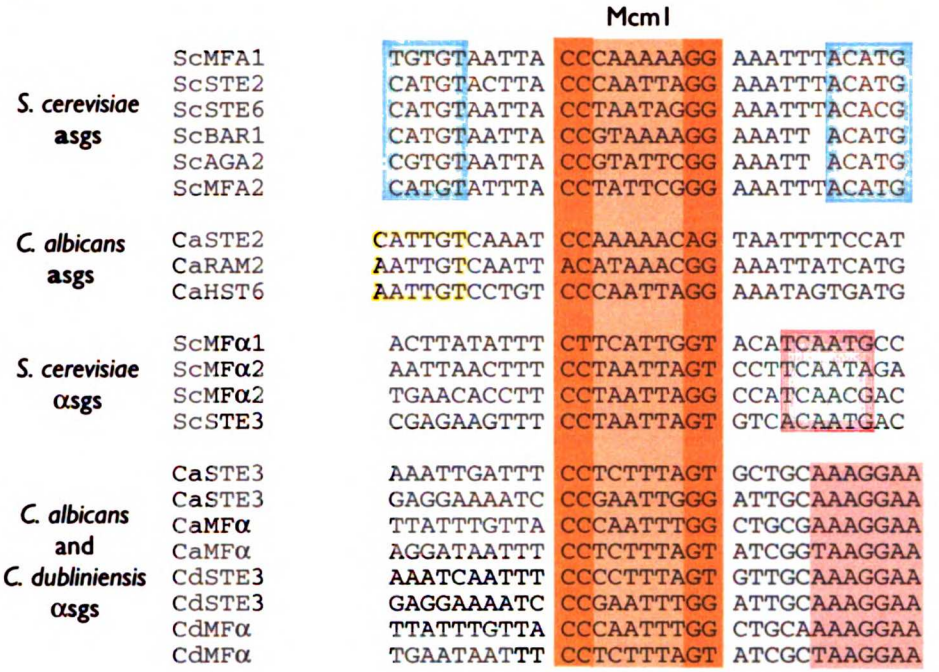


Figure 2. Mcm1 is central to two major changes in regulatory circuitry. Mcm1 likely facilitated the conversion between negative regulation facilitated by α2-Mcm1 in *S. cerevisiae* (top panel: α2 site is highlighted in blue, and Mcm1 site is highlighted in orange) and positive regulation facilitated in an ancestor (represented in extant *Candida* species, 2nd panel: a2 site is highlighted in yellow). Mcm1 is also central to a potential change in α1 binding specificity, as the putative α1 site in *Candida* species (bottom panel, pink) bears no resemblance to that in *S. cerevisiae* (3rd panel, pink).

Chapter 4

Discussion and Future Directions

Summary of results

Developing new regulation of existing genes is thought to underlie the remarkable variety of animals forms. Despite recent identification of spectacular examples illustrating evolution of transcriptional regulation in animals, specific mechanisms underlying evolution of transcriptional regulation have remained elusive. This is due in part to technical limitations in animal models of study, and in part to the extraordinary complexity of animal development. Evolutionary changes described are often limited in scope to a subset of the *cis*- or *trans*-elements involved; in many cases, corresponding *trans*- or *cis*-elements are simply unknown, or cannot be evaluated in detail.

In this thesis, I have examined evolution of transcriptional regulation by dissecting a simple combinatorial regulatory circuit that governs mating-type in the ascomycete branch of the fungal lineage. I first determine the means by which four conserved transcription factors regulate mating-type in the pathogenic yeast *C. albicans*. I then closely compare the *C. albicans* circuit to that of *S. cerevisiae*, identifying multiple classes of changes that have arisen since their divergence from a common ancestor. Each of these changes represents an example of evolution of transcriptional circuitry.

Among the many changes identified, I focus on a group of orthologs, the **a**-specific genes, that is positively regulated in *C. albicans*, but is negatively regulated in *S. cerevisiae*. I demonstrate that positive regulation represents the ancestral form, and that the *S. cerevisiae* mode of negative regulation is a recently derived innovation. Using a combination of experimental and informatic approaches, I identify the changes in all participating *cis*- and *trans*-regulatory elements that effect the change in **a**sg regulation.

I then examine the participating *cis*- and *trans*-elements in a series of 16 yeasts.

The resulting data allows a tentative assignment of temporal sequence to the changes that I have observed in the individual elements that define the circuit.

Determination of the *C. albicans* mating-type regulatory circuit

At the outset of this project, mating in *C. albicans* had just been discovered. We knew that mating-type was specified by the *MTL* locus, as it is in *S. cerevisiae* (Hull and Johnson, 1999; Hull et al., 2000; Magee and Magee, 2000). At the time, we believed that mating-type was specified by three transcription factors, a1, α1, and α2. These three factors have orthologs in the *S. cerevisiae MAT* locus, and their sequence, overall configuration in the MAT loci, and intron positions are conserved between *S. cerevisiae* and *C. albicans.*

Additionally, Matt Miller had demonstrated that in order for *C. albicans* to mate, it must first undergo a transition to a mating-competent form, called the "opaque phase." Further, he showed that the transition to the opaque phase is repressed by the combination of a1-α2, much as repression of mating in *S. cerevisiae* is repressed by a1-α2 (Miller and Johnson, 2002). Although the opaque phase had previously been described in detail, prior to Matt's discovery its relationship to mating was not even guessed (Soll, 1988; Soll, 1992).

Determining the *C. albicans* mating-type regulatory circuit involved several steps. First, differences in two constructs intended to delete a1 from the *MTL* locus functionally revealed a fourth activity at the *MTL*, which was later identified as an HMG-domain transcription factor with no ortholog in *S. cerevisiae.* Upon discovery of this ORF, I created precise deletions of each of the four transcription factors at the *MTL*, singly and

in all possible permuted combinations. By studying the mating behaviors of each of these strains and their transcriptional profiles by microarray, I was able to determine the overall role of each transcription factor in specifying mating-type, as well as their specific regulatory targets (Tsong et al., 2003).

This analysis revealed that the four *MTL* transcription factors regulate almost 500 genes in total. In contrast, the *S. cerevisiae MAT* locus uses three transcription factors to regulate approximately 40 genes (Herskowitz et al., 1992).

Several classes of changes in mating-type regulation have occurred: changes in the regulation of individual target genes, the requirement for α-factor induction in asg expression in *C. albicans* but not *S. cerevisiae*, incorporation of an additional level mating control in the form of the white-opaque switch, and replacement of positive regulation of the asgs by negative regulation in *S. cerevisiae*.

Discussion of changes in mating-type circuitry

The transition from positive to negative regulation of asgs has been discussed in detail in Chapter 3. Here, I review some of the additional changes I have identified in mating-type circuitry.

Changes in the regulation of individual target genes

The a1-α2 regulated battery in *S. cerevisiae* consists of more than 20 genes, but in *C. albicans*, a1-α2 directly regulates only 8 genes (indirectly, it regulates almost 450 via the white-opaque switch). Several genes in the *C. albicans* and *S. cerevisiae* batteries are overlapping (*GPA1, STE4, FUS3, FAR1, STE18*). Notably, all of these overlapping

genes are involved in mating. *GPA1*, *STE4*, and *STE18* form the heterotrimeric G-protein coupled receptor involved in transducing pheromone signal. *FUS3* encodes a MAP kinase in the pheromone signal transduction pathway, and *FAR1* mediates cell-cycle arrest in response to pheromone in *S. cerevisiae*, and is also involved in the morphological response to pheromone. In addition, many genes are specific to either *C. albicans* or *S. cerevisiae*, which I discuss below.

STE2*, *FGR23*, and *TOS9* are a1-α2 regulated in *C. albicans*, but not *S. cerevisiae

STE2 encodes α-factor receptor. In *C. albicans*, **a**-specific genes are only differentially expressed between **a** and α cells upon α-factor induction. Because α-factor induction requires α-factor receptor, a low level of Ste2 is required in **a** cells; this is accomplished by directly repressing *STE2* with a1-α2. As a result, *STE2* is expressed at low levels in both **a** and α white cells. In some *C. albicans* strains, white **a** cells have been shown to respond in a limited fashion to α-factor induction; however, full α-factor induction requires a switch to the opaque state.

Whether *STE2* also has a role in white α-cells is unknown; however, genes with orthologs required for **a**-factor processing (*STE23*) and α-factor processing (*STE13*) have been identified in a screen for genes affecting filamentous growth (Uhl et al., 2003), suggesting that *STE2* could be involved in cell-cell communication during filamentous growth.

Another gene identified in the screen for filamentous growth, *FGR23*, is highly repressed by a1-α2 in *C. albicans*. *FGR23* has no known homology, and a transposon insertion near *FGR23* in **a**/α cells causes decreased filamentous growth specifically on

Spider medium (thought to mimic starvation conditions). Interestingly, *FGR23* is also among the most highly induced genes on pheromone induction in both **a** and α cells, suggesting a prominent role for this gene in mating.

The gene *TOS9* has recently been shown to play a major role in regulating the white-opaque transition, a transition that is absent in *S. cerevisiae* (R. Zordan, personal communication). The white-opaque transition is thought to be a recent innovation in the *C. albicans* lineage, as only those species closest to *C. albicans* have been observed to undergo this phenotypic switch.

Genes regulated by a1-α2 in *S. cerevisiae*, but not *C. albicans*

Genes regulated by a1-α2 in *S. cerevisiae* but not *C. albicans* include *HO, NEJ1, RME1,* and *AXL1*. Many of the genes reflect aspects of *S. cerevisiae* biology, such as differences between haploids and diploids. These differences are irrelevant in *C. albicans,* which is thought to undergo a diploid-tetraploid life cycle (Bennett and Johnson, 2005). For instance, *HO* is required for mating-type switching in haploid *S. cerevisiae*. However, *HO* has no homolog in *C. albicans*; furthermore, *C. albicans* does not possess silent mating cassettes, and recombination between the *MTL* alleles is unlikely to be site-specific (Legrand et al., 2004).

In another example, *NEJ1* is required for non-homologous end-joining (NHEJ) in *S. cerevisiae*. *NEJ1* is repressed in diploid **a**/α cells by a1-α2, and expressed in **a**- and α-cells (Kegel et al., 2001). Because *S. cerevisiae* spends a substantial portion of its life cycle in the haploid phase, NHEJ is often the primary source of double-strand break repair as homologous chromosomes are only available for DNA repair in G2. In contrast,

in the diploid phase, a homologous chromosome is always available as a template for homologous recombination. Thus, a high level of *NEJ1* allows haploids to use NHEJ, while repression of NEJ1 by a1-α2 causes diploids default to homologous recombination, a DNA repair mechanism with higher fidelity. Such regulation of NHEJ is extraneous in *C. albicans*, as haploid *C. albicans* are not known to exist; in fact, the arrangement of recessive lethal alleles in *C. albicans* may make a haploid state nearly impossible (R. Bennett, personal communication).

Likewise, *RME1* is a repressor of meiosis that is repressed in a-α cells (Covitz et al., 1991). Repression of the *RME1* in a/α cells allows meiosis to proceed under specific environmental conditions. Meiosis in *C. albicans* has not been observed yet; furthermore, orthologs of meiotic regulators from *S. cerevisiae* such as *NDT80* and *UME6* appear to have different or additional roles in *C. albicans*, suggesting that if meiosis exists, its regulation is considerably different (A.E.T. and R. Bennett, unpublished data).

A double change in regulation of *AXL1* may reflect a "trickle-up" change

A particularly intriguing case of a change in transcriptional regulation is that of *AXL1*. *AXL1* is an **a**sg in *C. albicans*, but is an a1-α2 repressed gene (hsg) in *S. cerevisiae*. *AXL1* is required for processing of **a**-factor in *S. cerevisiae*, and this activity is likely conserved in *C. albicans* given its regulation. However, in *S. cerevisiae*, *AXL1* has the additional role of specifying the haploid-specific axial budding pattern, a biological trait thought to take advantage of the rapid mating-type switching in *S. cerevisiae* (Fujita et al., 1994). In contrast, *C. albicans* white **a**, α, **a**-α cells and opaque

a and α cells all display an identical bipolar budding pattern (A.E.T., unpublished data), perhaps reflecting the lack of mating-type switching.

It will be interesting to see when *AXL1*-directed axial budding evolved, and compare the timing of this event to when *AXL1* became a1-α2 repressed. Recently, it has proposed that changes in gene regulation may be a mechanism by which non-transcriptional changes are made permanent (Alonso and Wilkins, 2005). In other words, a change in a protein's regulation at a non-transcriptional level may "trickle-up" to the transcriptional level, as a change occuring at a non-transcriptional level could be evolutionarily reinforced or expedited if it were linked to a transcriptional change. *AXL1* may represent an example of such "trickle-up" regulatory changes.

Incorporation of an Additional Level of Mating Control

The incorporation of the white-opaque switch into the *C. albicans* mating circuitry is undoubtedly the most dramatic of the differences between *S. cerevisiae* and *C. albicans* mating. The white-opaque switch, apparent even to the naked eye, is accompanied by massive changes in cell shape, and wall and membrane architecture and components (Soll, 1992; Soll et al., 2003). This transition was recently shown to be under mating-type control, being repressed by a1-α2 (Miller and Johnson, 2002).

Although a1-α2 likely repress only 8 genes directly in *C. albicans*, Matt and I have demonstrated that through the white-opaque switch, a1-α2 indirectly regulates nearly 450 genes. A handful of genes involved in mating are among those indirectly repressed by a1-α2; however, most opaque-specific genes concern many aspects of cell biology such as metabolic specialization, cell surface properties, and virulence (Lan et al., 2002). Why

such a vast phenotypic transition is under a1-α2 regulation is unclear at present. It is likely that white-opaque switching is a recent innovation, brought about by the pressure of life in a mammalian host. However, white-opaque switching is under the control of an ancient regulatory circuit, and this coupling ensures that mating is closely integrated with host environments.

How could a1-α2 take control of 450 genes?

The example of *NEJ1* in *S. cerevisiae* may be illustrative for how the white-opaque switch could come under mating-type regulation. In *S. cerevisiae,* ploidy is highly related to mating-type in *S. cerevisiae*; **a** and α cells are generally haploid, while **a**/α cells are diploid. But this is not strictly the case, as some wild strains undergo a diploid-tetraploid life-cycle (for example, the strain DH1-1A). Thus, mating-type is a proxy that generally coincides with ploidy, rather than an exact predictor of ploidy. Clearly, however, the coincidence of ploidy and mating-type is a sufficiently precise predictor that a1-α2 regulation was selected for in *NEJ1*, rather than under another regulatory system – for example, one that could exactly measure homolog number (such as mechanisms that count the number of X chromosomes in *Drosophila* and *C. elegans* sex-determination), or one linked to sensing synapsis of homologous chromosomes (for instance, in meiotic apparatus).

Similarly, it is possible that conditions appropriate for mating in a mammalian host or in the environment almost always coincide with those requiring characteristics encoded by the opaque phase, and that during evolution of the circuit, regulation of mating served as a proxy to predict specific environmental conditions. For example,

many elements of the phosphate scavenging pathway are upregulated in opaque cells (Chapter 2, Appendix A). Corroborating this, a secreted acid phosphatase assay shows that opaque cells resemble Δ*pho84 S. cerevisiae* strains in their constitutive expression of secreted acid phosphatases under all conditions (*PHO84* encodes a transporter of phosphate, and in its absence, cells always behave as though starved for phosphate); in contrast, white cells appropriately repress phosphatase expression in high phosphate media, and activate phosphatase expression in low phosphate conditions (Figure 1, in collaboration with Dennis Wykoff). It is possible that host conditions favoring mating almost always coincide with low phosphate conditions, and that a1-α2 regulation of the phosphate pathway is a simple way to integrate this information.

Such a scenario could help to explain why so many disparate aspects of cell biology are incorporated into the switch. It will be interesting to see if the genes expressed in the opaque or white phase fall into coregulated modules (for instance, the phosphatase module) under other conditions; if so, the white-opaque transition could prove to be an ideal system in which to study the role of transcriptional modularity in evolution.

Why does mating require the white-opaque switch?

The use of mating as a proxy for predicting other environmental predictions could also explain why mating now strictly requires the white-opaque transition. Once a non-mating module was integrated into the a1-α2 regulatory circuit, a mating-specific gene could fall under this pathway. For instance, the transcriptional regulator Pho23 (a deacetylase) is among the several phosphatase-related genes regulated by a1-α2. A gene required for mating, such as *STE4,* could come under Pho23 regulation and lose its

original a1-α2 site. This would preserve the circuit logic, since Pho23 itself is under a1-α2 regulation. However, such a change would "cement" the acid phosphatase module into a1-α2 regulation, because losing regulation of the phosphatase module would simultaneously cause loss of *STE4* regulation.

Careful dissection of the white and opaque transcriptional networks will certainly shed light on the extent of its modularity and interconnectedness. While modularity in transcriptional networks has been studied in theoretical terms (Kashtan and Alon, 2005), and in general terms (Ihmels et al., 2004), the roles of network structure in evolution remain largely uncharacterized. The white-opaque switch has the potential to serve as a powerful model for studing the role of modularity in evolution of developmental programs. Techniques are already available - a combination of determining the targets of the multiple transcription factors controlled by the white-opaque switch, combined with studies of transcriptional modules in *C. albicans* (Ihmels et al., 2005), and potentially comparative studies with *C. dubliniensis*, *C. tropicalis,* or other closely related species, would bring studies of modularity to a concrete and biologically interpretable level.

The white-opaque transition and incorporation of alternate mechanisms of mating regulation

The requirement of an epigenetic switch prior to mating in *C. albicans* raises many intriguing questions regarding regulatory tiers. For example, as in *S. cerevisiae,* α1 in *C. albicans* is required for activation of α-specific genes; however, in *C. albicans* this activation only occurs in the opaque phase, even though α1 is expressed at detectable in the white phase (it is upregulated only 2x in the opaque phase; A.E.T., unpublished

results). An additional factor present only in the opaque phase appears to be required for αsg activation; thus, mating in *C. albicans* requires traversing a tier of regulation absent in *S. cerevisiae.*

Initially, we thought that the additional factor might be Ste12, since in *S. cerevisiae*, a phosphorylated form of Ste12 is involved in upregulation of αsgs and asgs during pheromone response. However, *STE12* is not upregulated on the white-to-opaque transition. It will be very interesting to see if Ste12 is phosphorylated on the white-to-opaque transition. If so, the opaque transition would have some characteristics analogous to a pheromone-induced *S. cerevisiae* cell. The large-scale changes in opaque cell wall and membrane structure may reflect a state that is more analogous to a pheromone induced *S. cerevisiae* cell than a resting cell.

On a similar note, the *FAR1* gene is a1-α2 repressed in *S. cerevisiae.* On pheromone induction, *FAR1* is not only further upregulated, but the protein is also exported from the nucleus and localized to the cell membrane (Blondel et al., 1999). In *C. albicans*, *FAR1* is also a1-α2 repressed; however, it is neither upregulated upon the white-opaque transition nor upon pheromone induction. The absence of *FAR1* upregulation on either the white-opaque transition or on pheromone induction implies that functionally analogous regulation does not occur at a transcriptional level. As with Ste12, it will be informative to see if Far1 localization changes upon the white-to-opaque transition, a change associated with pheromone induction in *S. cerevisiae*.

Many other examples of regulatory tier evolution are illustrated in Table 1, and some changes are quite intriguing (e.g., *GPA1* is upregulated in a and α white cells, but its trimeric partner *STE4* is upregulated in a and α opaque cells).

Discussion of asg evolution

I have demonstrated that asgs were activated by an HMG domain protein in the common ancestor of *C. albicans* and *S. cerevisiae*, and that asgs are repressed by the homeodomain protein α2 in *S. cerevisiae*. At the end of Chapter 2, I presented a pathway by which the positive-to-negative regulatory transition may have occurred (Chapter 2, Figure 6). I speculated that the loss of a2 occurred prior to the recruitment of Mcm1 into the asg circuit, and that α2 was later recruited from the α locus to repress asgs. I reasoned that a loss of a2 early in the transition would create an enormous selective pressure to rapidly generate an alternative mechanism of asg regulation.

However, my subsequent experiments and informatics analyses demonstrate that Mcm1 was in fact already involved in asg regulation in the common ancestor of *C. albicans* and *S. cerevisiae*; furthermore, informatic data from *K. lactis*, *K. waltii*, and *E. gossypii* – all of which have retained a2 – strongly suggest that α2 was involved in asg repression even before loss of a2. Identification of the a2 binding site also suggested that α2 could co-opt the a2 site after a single indel – a much smaller change than evolving two α2 sites *de novo*.

These results are meaningful on several levels. First, they demonstrate that it is possible to infer detailed molecular mechanisms of evolution from extant yeasts, even in the absence of molecular fossils. Second, such a level of certainty and detail regarding a molecular evolutionary change is unusual, as many examples of evolution are studied in far more complex systems. Due to the simplicity of mating-type regulation, and to the vast literature regarding *S. cerevisiae* mating-type regulation in particular, the level of

detail we have achieved is such that in principle, it would be possible to engineer the *C. albicans* mating-type circuitry into *S. cerevisiae* (or vice versa) with a minimum number of changes to the native system.

Third, I have begun to deduce a specific order in which the changes occurred, which in itself is a novel achievement in molecular evolution. In essence, I have had the advantage of using outgroups for progressively larger phylogenetic distances – e.g., *C. albicans* serves as an outgroup for a comparison between *K. lactis* and *S. cerevisiae*, while *S. pombe*, *Y. lipolytica,* and *N. crassa* all serve as outgroups for a comparison between *C. albicans* and *S. cerevisiae* (Figure 2).

Fourth, the order of events determines the fitness barriers faced during the transition; for example, my previous model in which a2 was lost at before α2-Mcm1 repression developed predicted a transition state in which the cells were sterile – an enormous fitness obstacle, followed by another transition state in which the cells were considerably impaired in mating. The current model, on the other hand, suggests that the transition occurred in the absence of such extreme barriers; given the identical logical outputs of all observed mating-type circuits, any natural selection must have been far more subtle. At the end of Chapter 3, I suggest demand theory (Savageau, 1983) and sexual selection (Singh and Kulathinal, 2000) as two possible causes of natural selection. The current model could also be taken as consonant with a neutral theory of evolution, which proposes that most changes occur through random genetic drift and purifying selection.

Evolvability

The identification of Mcm1 at the center of two changes in evolution – the transition from positive to negative regulation of asgs, and the probable change in α1 binding specificity – raises the ever-present question of evolvability. Mcm1 is a member of the MADS box family, a family of proteins that is widely used in plant morphogenesis and patterning. It is analogous in function and predominance in plant development to the role of homeobox proteins in animals. Combinatorial circuitry has been proposed to be a prominent form of regulation for many reasons (discussed in Chapter 1). It is possible that the presence of Mcm1 at asgs widens the number of changes that can be tolerated while preserving function, increasing the possibility of transcription factor binding site turnover. Brian Tuch is currently developing models to test this hypothesis.

In another potential example of evolvability, the short linker region of α2 that contacts Mcm1 has been shown to be disordered (Vershon and Johnson, 1993); however, it adopts a beta fold upon binding DNA with Mcm1 (Tan and Richmond, 1998). Remarkably, the region directly next to the disordered linker can adopt either a beta sheet or an alpha-helix in complex with Mcm1, putting it in a very small class of proteins with so-called chameleon sequences (Tan and Richmond, 1998). On a whole, disordered protein sequences have a higher rate of evolution than ordered sequences, which could potentially contribute to the use of this region in the Mcm1 interaction. Whether certain proteins or certain protein characteristics – such as the presence of disordered or chameleon domains – are particularly adaptable in nature remains an open question. However, the idea that some domains are inherently adaptable has considerable precendent in non-transcriptional systems (Dueber et al., 2004).

Summary

By studying molecular evolution in a highly tractable model system with a large body of supporting literature, I have been able to identify a substantial number of specific evolutionary changes in a single transcriptional circuit. Further, I have dissected one particular change in detail. The findings of this thesis provide a point of departure for many future studies in molecular evolution, some of which have the potential to uncover general principles of evolution. The model system I have described can be used to study evolvability (the roles of Mcm1 and α2 facilitating evolutionary transitions), changes in modularity and interconnectedness of transcriptional network structure when evolving major transcriptional programs (the white-opaque transition), and the role of transcriptional evolution in reinforcing evolution at alternate levels (co-evolution of Axl1 function and regulation).

Table 1.

	Repressed by a1/α2		Up in a,α opaques		Induced by α factor		Induced by a-factor	
Gene:	**C.a. whites**	**S.c.**	**C.a. opaques**		**C.a. opaques**	**S.c.**	**C.a. opaques**	**S.c.**
FUS3	Yes, 2x	Yes	Yes, 15x		Yes, 15x	Yes	Yes, 15x	Yes
STE2	Yes, 2x	No	No		Yes, 13x	Yes	No	No
FGR23	Yes, 16x	n/a	No		Yes, 37x	n/a	Yes, 40x	n/a
TOS9	Yes, 2x	No	Yes, 45x		No	No	No	No
GPA1	Yes, 8x	Yes	No		No	Yes	No	Yes
FAR1	Yes, 4x	Yes	No		No	Yes	No	Yes
STE4	No	Yes	Yes, 6x		No	No	No	No
STE18	unknown, has site	Yes	unknown		No	No	No	No
STE2	Yes, 2x	Yes	No		Yes, 13x	Yes	No	No
HST6	No	Yes	No		Yes, 33x	Yes	No	No
RAM2	No	No	No		Yes, 5x	No	No	No
BAR1	No	Yes	No		Yes, 60x	Yes	No	No
ASG7	No	No	No		Yes, 30x	Yes	No	No
AXL1	No	Yes	No		Yes, 2x	Yes	No	No
STE3	No	Yes	Yes, 300x		No	No	Yes, 4x	?
MFα	No	Yes	Yes, 1000x		No	No	No	?
STE13	No	No	No		No	No	Yes, 3x	?

Table 1. Differences in regulatory tiers in *C. albicans* and *S. cerevisiae*

Regulation of *C. albicans* hsgs, asgs, and αsgs are compared to regulation of orthologous genes in *S. cerevisiae.*

FIGURE 1. Secreted acid phosphatase activity

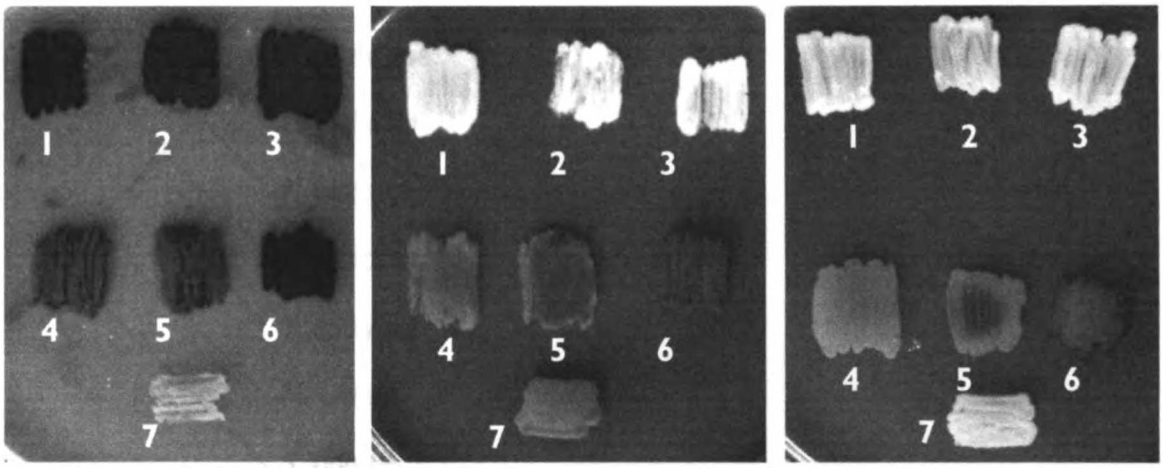


Figure 1. Secreted acid phosphatase activity of white and opaque *C. albicans* strains. 1 - **a**/α white, 2 - **a** white, 3 - α white, 4 - **a** opaque, 5 - α opaque, 6 - *S. cerevisiae* Δ*pho84*, 7 - *S. cerevisiae* Δ*pho4*. Higher acid phosphatase activity results in darker patches. On no phosphate media, all strains (except *S. cerevisiae* Δ*pho4*, which does not transcribe acid phosphatases) have strong activity. On medium and high phosphate, White strains (1,2,3) repress acid phosphatase activity. However, opaque strains (4,5) continue to secrete acid phosphatases. Their consitutive activity matches that of *S. cerevisiae* Δ*pho84* strains, which are unable to sense phosphate in the media and therefore behave as though phosphate-starved, even in high-phosphate conditions.

FIGURE 2.

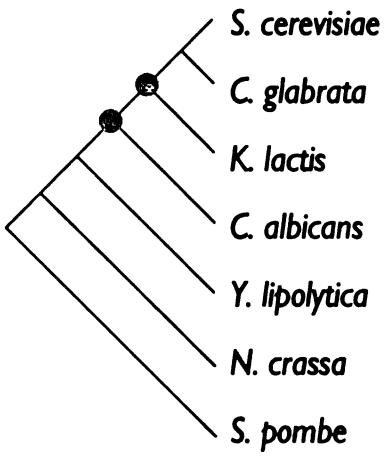


Figure 2. Use of multiple outgroups in evolutionary comparisons

The sequencing of multiple organisms in the same evolutionary lineage, combined with the inference of their phylogenetic relationships, allows use of outgroups for progressively larger distances. An outgroup is used to determine which characteristic between two related organisms is closer to an ancestral form. For instance, *S. cerevisiae* uses negative regulation of asgs, and *K. lactis* uses positive regulation. Because the outgroup *C. albicans* uses positive regulation via an orthologous regulator, positive regulation likely represents the ancestral state (*S.c.* and *K.l.* ancestral node in red), and negative regulation is an innovation in *S. cerevisiae*. In contrast, *K. lactis* encodes a fifth transcriptional regulator, α3, at its mating-type locus, while *S. cerevisiae* does not (Astrom et al., 2000). Because *C. albicans* also lacks α3, it is likely that *S. cerevisiae* represents the ancestral state with regard to α3, and that α3 is an innovation in *K. lactis.* Likewise, *Y. lipolytica, N. crassa,* and *S. pombe* may be used to determine ancestral states when inferring the common ancestor of *S. cerevisiae* and *C. albicans* (*S.c.* and *C.a.* ancestral node, in red).

REFERENCES

Alonso, C. R., and Wilkins, A. S. (2005). Opinion: the molecular elements that underlie developmental evolution. Nat Rev Genet *6*, 709-715.

Astrom, S. U., Kegel, A., Sjostrand, J. O., and Rine, J. (2000). Kluyveromyces lactis Sir2p regulates cation sensitivity and maintains a specialized chromatin structure at the cryptic alpha-locus. Genetics *156*, 81-91.

Bennett, R. J., and Johnson, A. D. (2005). Mating in Candida albicans and the search for a sexual cycle. Annu Rev Microbiol *59*, 233-255.

Blondel, M., Alepuz, P. M., Huang, L. S., Shaham, S., Ammerer, G., and Peter, M. (1999). Nuclear export of Far1p in response to pheromones requires the export receptor Msn5p/Ste21p. Genes Dev *13*, 2284-2300.

Covitz, P. A., Herskowitz, I., and Mitchell, A. P. (1991). The yeast RME1 gene encodes a putative zinc finger protein that is directly repressed by a1-alpha 2. Genes Dev *5*, 1982-1989.

Dueber, J. E., Yeh, B. J., Bhattacharyya, R. P., and Lim, W. A. (2004). Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry. Curr Opin Struct Biol *14*, 690-699.

Fujita, A., Oka, C., Arikawa, Y., Katagai, T., Tonouchi, A., Kuhara, S., and Misumi, Y. (1994). A yeast gene necessary for bud-site selection encodes a protein similar to insulin-degrading enzymes. Nature *372*, 567-570.

Herskowitz, I., Rine, J., and Strathern, J. (1992). Mating-type determination and Mating-type interconversion in Saccharomyces cerevisiae, Vol II (Cold Spring Harbor: Cold Spring Harbor Laboratory Press).

Hull, C. M., and Johnson, A. D. (1999). Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans. Science *285*, 1271-1275.

Hull, C. M., Raisner, R. M., and Johnson, A. D. (2000). Evidence for mating of the "asexual" yeast Candida albicans in a mammalian host. Science *289*, 307-310.

Ihmels, J., Bergmann, S., and Barkai, N. (2004). Defining transcription modules using large-scale gene expression data. Bioinformatics *20*, 1993-2003.

Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yanai, I., McClellan, M., Berman, J., and Barkai, N. (2005). Rewiring of the yeast transcriptional network through the evolution of motif usage. Science *309*, 938-940.

Kashtan, N., and Alon, U. (2005). Spontaneous evolution of modularity and network motifs. Proc Natl Acad Sci U S A *102*, 13773-13778.

Kegel, A., Sjostrand, J. O., and Astrom, S. U. (2001). Nej1p, a cell type-specific regulator of nonhomologous end joining in yeast. Curr Biol *11*, 1611-1617.

Lan, C. Y., Newport, G., Murillo, L. A., Jones, T., Scherer, S., Davis, R. W., and Agabian, N. (2002). Metabolic specialization associated with phenotypic switching in Candidaalbicans. Proc Natl Acad Sci U S A *99*, 14907-14912.

Legrand, M., Lephart, P., Forche, A., Mueller, F. M., Walsh, T., Magee, P. T., and Magee, B. B. (2004). Homozygosity at the MTL locus in clinical strains of Candida albicans: karyotypic rearrangements and tetraploid formation. Mol Microbiol *52*, 1451-1462.

Magee, B. B., and Magee, P. T. (2000). Induction of mating in Candida albicans by construction of MTLa and MTLalpha strains. Science *289*, 310-313.

Miller, M. G., and Johnson, A. D. (2002). White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. Cell *110*, 293-302.

Savageau, M. A. (1983). Regulation of differentiated cell-specific functions. Proc Natl Acad Sci U S A *80*, 1411-1415.

Singh, R. S., and Kulathinal, R. J. (2000). Sex gene pool evolution and speciation: a new paradigm. Genes Genet Syst *75*, 119-130.

Soll, D. R. (1988). High-frequency switching in Candida albicans and its relations to vaginal candidiasis. Am J Obstet Gynecol *158*, 997-1001.

Soll, D. R. (1992). High-frequency switching in Candida albicans. Clin Microbiol Rev *5*, 183-203.

Soll, D. R., Lockhart, S. R., and Zhao, R. (2003). Relationship between Switching and Mating in Candida albicans. Eukaryot Cell *2*, 390-397.

Tan, S., and Richmond, T. J. (1998). Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex. Nature *391*, 660-666.

Tsong, A. E., Miller, M. G., Raisner, R. M., and Johnson, A. D. (2003). Evolution of a combinatorial transcriptional circuit: a case study in yeasts. Cell *115*, 389-399.

Uhl, M. A., Biery, M., Craig, N., and Johnson, A. D. (2003). Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen C.albicans. Embo J *22*, 2668-2678.