Title
The Evolution of the Transcriptional Regulation of Amino Acid Biosynthetic Pathways in Yeasts

Author
Noiman, Liron

Publication Date
2021

Peer reviewed|Thesis/dissertation

The Evolution of the Transcriptional Regulation of Amino Acid Biosynthetic Pathways in Yeasts

by
Liron Noiman

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Alexander Johnson
E2BBE6A63C2745B...
Alexander Johnson
Chair

DocuSigned by:
James S. Fraser
James S. Fraser

DocuSigned by:
Hiten D. Madhani
76DDD3CF67EA4D0...
Hiten D. Madhani

Committee Members

Acknowledgements

The scientific process is filled with conditionals and contingency plans. "*If this experiment shows this result, then we'll do this set of experiments next…*", "*if I can't get this assay to work, I'll try this instead…*", *"if your paper isn't accepted in this journal, we'll send it to this one…"* I have been beyond lucky to find truly unconditional support from mentors, family and friends while navigating this conditional process.

I would be totally remiss if I did not acknowledge my thesis advisor, Sandy, as the first source of unconditional support. Even when my path through grad school got particularly rocky, he displayed an amount of patience and understanding that even took me by surprise. People might think that a mentor who has been in the game as long as Sandy has forgotten what it's like to be a struggling grad student, but I will always remember Sandy telling me, at a point when I'd really dropped the ball, "Well, Liron, I've been doing this for so long, I've seen every kind of mistake…" and then encouraged me to work with what I have and move forward. Truly, if it wasn't for his patience and understanding, this thesis document would not exist. Beyond all that, he has crafted my critical scientific eye, granted me a real appreciation for good writing, and solidified the importance of having an extreme amount of clarity of thought before committing words to a page or to doing a certain experiment. More than any knowledge about any transcription factor, these lessons are the ones that I will cherish the most and I know will be most universally applied.

In addition to Sandy, I have been unconditionally supported by my thesis committee: Jaime Fraser and Hiten Madhani. I met them both within the first days of grad school. Jaime as my macro discussion leader, who (about 20 years removed from

my first first day of school) was the first teacher to ever pronounce my name correctly on the first day of class, and Hiten as my first rotation advisor and ride to Granlibakken during which, I'm pretty sure, on the ride back we'd run out of things to talk about that I ended up doing a pretty bad Ali G impression. Part of me still hopes that that last memory wasn't real but rather a bizarre grad school fever dream, but I'm 99% certain happened. Joking aside, these two have been absolute sources of encouragement and support. Scientifically, their insight often helped me break the "Johnson lab" view of how to do things and offered alternative solutions when experiments weren't going well. Personally, Jaime was the first UCSF professor who seemed to truly empathize with the emotional toll science can take on a person, and coming from a guy who had all the perceived "scientific success" in the world, him confessing his own ups and downs to me made a huge impact. I know I already said this about Sandy, but it applies to Jaime as well: this thesis wouldn't exist without him.

Outside of my committee, I've been lucky enough to benefit from the support, encouragement, and mentorship of Kaveh Ashrafi. Kaveh has the gift of not just listening but truly hearing what someone is saying to him, and making the person on the receiving end of his attention feel not just heard, but understood. In a world filled with people who love to hear themselves talk and to offer their most poignant critiques, Kaveh knows that it's important to give students the time and space to say what we need to say – free from judgement or criticism. A true role model in every sense of the word, Kaveh has deeply impacted my perspective on science and mentorship.

Outside of the faculty, I have relied on the unconditional friendship and support of the Johnson Lab, at large. I was lucky enough to sit back-to-back with Candace Britton for the majority of our PhD experiences, and there is no one else I would want to have my back, literally or figuratively. Whether commiserating or celebrating, laughing or crying, I am consistently in awe of Candace's blend of compassion and practicality. She might not be the loudest voice in the room, but if you're lucky enough to find yourself in a room with her, you should probably shut up and listen to what she has to say, because she usually has the most insight to offer. I was also lucky enough to learn from the best when it comes to all things *Candida albicans* under the guidance of Matt Lohse. Matt, in many ways, is the exact opposite of me as a scientist – organized, goal-oriented, an absolute BEAST at the bench – and this is some of what I find so inspiring about him. In addition to all that, his dark and cynical sense of humor is in direct contrast to his kind and compassionate heart. Chiraj Dalal is one of the kindest and most compassionate people I've ever met, and without his pity on a struggling grad student, I would never would have pulled together a thesis project that worked. Sheena Singh-Babak is practically my "sister from another mister" and we will be friends for life, no question, ride-or-die, the one that will help hide the body. Naomi Ziv has been an incredible source of understanding while also being a teller of hard truths – without her telling me things I didn't always want to hear, but that I knew in my heart were true, I don't think I would have been able to complete this PhD. Kyle Fowler is the best and I don't think you'll find anyone who says otherwise – a model grad student who stays humble with a killer sense of humor – there is no one else with whom I'd rather tell off-color jokes. The rest of the Johnson Lab past and present: Trevor Sorrells, Isabel

Nocedal, Emily Fox, Chris Baker, Lauren Booth, Francesca Del Frate, Haley Gause, Carrie Graham, Nairi Hartooni, Julie Takagi, Lucas Brenes, Christian Perez, Eugenio Mancera, Clarissa Nobile, – you have all inspired me and I am honored to know such an incredible group of intelligent, compassionate, and kind scientists.

If all these people weren't enough, coming to UCSF may have been unconditionally worth it if only because it led me to befriend Elle Roberson and Caitlin Stoddard. The joy our friendship has brought to my life is incalculable and although we've spread out to different corners of the country, I look forward to someday realizing our dream of starting a science commune and co-raising our future children together.

Before UCSF I was already lucky enough to have stumbled my way into a couple labs that provided my first sources of scientific support and mentorship. Christian Mirescu is my scientific foundation, as he was my first in-the-lab mentor when I was a high school and undergrad student in the Gould Lab at Princeton. Nearly 15 years later, Christian is also helping me craft my scientific future. His unconditional belief in me has quite literally restored my own belief in science and for that, I could never thank him enough. While on the subject of the Gould Lab, I should also thank Liz Gould, who took a chance on a high school student with an interest in neuroscience and let me come into the lab and muck around. To date, the only female mentor I've had in science, and you'd think I'd have some big feminist idea to convey here, but I will say that she is the only mentor to ever teach me to snip off the base of a latex glove to use as a hair-tie when you forgot yours at home. Now that's a tip they don't give you at the "Being a Woman in Science" panel.

Dr. Alfredo Quiñones-Hinojosa taught me the importance of doing science for a greater cause and has been an unwavering source of support and kindness for me and my family. The rest of my AQH family: Hugo Guerrero-Cazares, Ahmed Mohyeldin, Sagar Shah and Guillermo Vela, whose friendship, support, and humor I will cherish for the rest of my life.

This brings me to the most unconditional of all the unconditional support I have received: my family. My sister Adi, the first “Dr. Noiman, PhD,” Adi is the only normal one in the family and I’ve never valued her level-headedness more than during these last few years. She is the epitome of living your values without fanfare and without belabored over-analysis. Her ability to see a problem, identify a solution, and execute on that solution is one that I truly envy and maybe in another 30 years, I’ll figure out how she does it. Adi is also the source of my newest favorite distraction and source of joy: my nephew Ben, who is the sweetest, the cutest, and the smartest baby on earth (unbiased opinion!).

Lastly, but without doubt most importantly: my parents. My mom, Edna, and my dad, Shai, who came to this country in pursuit of higher education and have instilled in us the value of education and have always encouraged us to follow our passions, even if the road those passions lead you down is a difficult one. From a young age, my dad has been the #1 “STEMinist” in my life: always encouraging my love of math, science and experimentation. He has also been a role model when it comes to choosing a career you love – the fact that I talk about the regulation of gene expression with the same joy that my dad talks about heat exchangers , extruders, and dust collectors reminds me that I am on the right path. My mom lives and breathes every success and

every set-back I experience, and so, suffice it to say, my PhD journey has been equally hard on her. In addition to Sandy and Jaime, she is the third person without whom this thesis would never have turned into words on a page. Despite being the one humanist in the family, my philosopher mother actually asked the most insightful, the most relevant, and most interesting questions about my work, proving that you don't have to be trained as a scientist to think like one. I'll never be able to sufficiently put into words the magnitude of the love and support my mother has given me so I'll just wrap it up here and say that I love you more than you'll ever know, I appreciate you more than you'll ever know, and this thesis is dedicated to you, mom.

Abstract

The Evolution of the Transcriptional Regulation of Amino Acid Biosynthetic Pathways in Yeasts

Liron Noiman

When evolutionary biology comes to mind, it is often accompanied by contemplating the vast amount of morphological or behavioral differences that exist between species. Whether Darwin's finches or the pattern on an insect wing, we often reference these types of striking apparent visual differences in discussing the power of evolution. However, there are a growing number of examples that suggest that even core, fundamental processes like metabolism and mating are subject to evolutionary changes. In the face of these underlying evolutionary changes, the fundamental logic of these core processes must be conserved. A biosynthetic pathway must be upregulated in the face of nutrient starvation (Ch. 1), and a cellular identity must be conserved in order for mating to occur in yeast (Ch. 2), but how these logical outputs are maintained in spite of evolution acting on the molecules involved in these processes is the outstanding question explored in this work.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 1:
Introduction

Introduction

The contributions of model organisms to our understanding of biology are monumental. Deep mechanistic insights in genetics, cell biology, and physiology have been elucidated by a handful of species like Saccharomyces cerevisiae, Schizosaccharomyces pombe, Caenorhabditis elegans and Drosophila melanogaster. However, a potential pitfall of model organism biology is that our knowledge is biased by taking such a small and sparse sampling of the tree of life. What is found to be true in one species is assumed to be true in related species, often without directly testing those assumptions. Furthermore, the evolutionary trajectory by which differences between species is difficult to assess without direct experimentation.

The ascomycete yeasts or Saccharomycotina are a group of fungi that share a most recent common ancestor encompassing the Saccharomyces cerevisiae clade through the Trichomonascus clades. They usually grow as single cells (although they can exist as hyphal and pseudohyphal morphologies) and reproduce asexually via budding. The most highly characterized member of the saccharomycotina is Saccharomyces cerevisiae. By converting sugar into carbon dioxide or alcohol, S. cerevisiae is most well-known for its culinary contributions to bread-baking and beer-brewing. However, S. cerevisiae has served as a powerful model organism dating back to the early 1900s when cell-free extracts were shown to have enzymatic activity (Bohley & Fröhlich, 2014). Decades later, S. cerevisiae's mating cycle (mating followed by meiosis coupled to spore formation) allowed for genetic cross analysis and isolation of genetic traits (Lindegren & Lindegren, 1943). Finally, S. cerevisiae's readiness to accept bacterial plasmids opened the door for transgenics in a eukaryotic model

organism (Hinnen et al., 1978). This, coupled with the sequencing of its genome, allowed for the generation of whole genome libraries of knockout or tagged S. cerevisiae strains leading to leaps in knowledge in genetics, cell biology and gene regulation.

A more recently developed Saccharomycotina “model” organism is Candida albicans. Separated from S. cerevisiae by hundreds of millions of years, C. albicans’s environmental niche is completely different from that of S. cerevisiae’s. While S. cerevisiae is found predominantly on the bark of deciduous trees or fermenting fruit, C. albicans resides primarily in the human gastrointestinal tract. While a harmless commensal in a majority of healthy individuals, C. albicans can be a harmful pathogen causing mucosal infections such as vaginitis or oropharyngeal thrush, or more severe systemic bloodstream infections in the severely immunocompromised (Kumamoto et al., 2020). In addition to its completely different environmental niche, C. albicans differs from S. cerevisiae in many other aspects of its lifestyle. Unlike S. cerevisiae, C. albicans undergoes a parasexual life cycle, seemingly not undergoing meiosis, and requires undergoing an epigenetic switch between two cell types – the white to opaque switch -- in order to mate (Miller & Johnson, 2002). Despite its apparent in inability to undergo meiosis or retain plasmid DNA, transgenic technology using nutritional markers, drug resistance cassettes, and more recently CRISPR, have rendered C. albicans a valuable additional ascomycete model yeast (Hernday et al., 2010; Nguyen et al., 2017). Thus, directly comparing C. albicans to S. cerevisiae is a powerful approach to understanding evolution.

Beyond *C. albicans* and *S. cerevisiae*, several other ascomycete yeasts have proven to be genetically tractable (Britton et al., 2020; Lombardi et al., 2019; Nocedal et al., 2017) and hundreds more genomes have been sequenced (Shen et al., 2018), allowing for a thorough exploration of the evolutionary trajectories by which different phenotypes arise. Comparative genetic, molecular, and bioinformatic analyses in these species have reinforced the idea that transcription network rewiring is a major molecular mechanism by which phenotypic diversity arises (Li & Johnson, 2010; Nocedal & Johnson, 2015).

Transcription network rewiring refers to changes in the connections between sequence specific DNA binding transcription factors and the target genes they regulate. A connection between a sequence specific DNA binding protein and a target gene is fundamentally determined by the sequence specificity that the transcription factor recognizes and the presence or absence of that site in the regulatory region of that target gene. These connections are subject to rewiring since mutations can occur either in the regulatory regions of the genome, thereby creating or deleting a binding site (Gasch et al., 2004; Kunarso et al., 2010; Wittkopp & Kalay, 2011), or mutations can occur in the transcription factor itself that change the DNA binding specificity of the transcription fact (Jarvela & Hinman, 2015; Lynch & Wagner, 2008). These classes of mutations are characterized as cis- or trans- regulatory evolution, respectively. Detailed molecular studies have demonstrated that the trajectories by which transcription networks evolve ultimately involve both modes of regulatory evolution (Baker et al., 2012; Britton et al., 2020; Lynch et al., 2008; Sorrells et al., 2018).

Ultimately, the rewiring of transcription networks leads to new patterns of gene expression – where one transcription factor governed one set of target genes, now a new transcription factor has “taken over” that set of target genes. Changes in the regulation of gene expression have long been hypothesized to account for diversity between species – dating back to the 70s, Susumo Ohno hypothesized that given how generally intolerant most genes seem to mutation, that regulation of genes must account for the majority of morphological diversity (Ohno, 1972). Supporting this notion, Mary-Claire King and Allan Wilson analyzed the sequences of a core set of human and chimpanzee proteins, and found them to be so similar that they proposed that regulatory mutations must “account for the major biological differences” between the two species (King & Wilson, 1975).

While some of the most striking examples of regulatory evolution and transcription network rewiring result in striking morphological differences in higher metazoans (Gompel et al., 2005; Martin et al., 2016; O’Brown et al., 2015), even seemingly conserved processes like mating (Baker et al., 2012; Booth et al., 2010; Sai et al., 2011; Tsong et al., 2006) and metabolism (Dalal et al., 2016; Gasch et al., 2004; Priyadarshini & Natarajan, 2016; Whiteway et al., 2015) are subject to evolution and result in phenotypic differences between species. In this work, I will focus primarily on the evolution of the transcriptional regulation of amino acid biosynthetic pathways using a comparative approach between *S. cerevisiae* and *C. albicans* (Ch. 1) with a supplemental chapter on the evolution of a mating type regulator (Ch. 2).

References

Baker, C. R., Booth, L. N., Sorrells, T. R., & Johnson, A. D. (2012). Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell*, *151*(1), 80–95. https://doi.org/10.1016/j.cell.2012.08.018

Bohley, P., & Fröhlich, K.-U. (2014). *A PRIZE-WINNING DISCOVERY OF 1896: BUCHNER PROVIDES EVIDENCE OF CELL-FREE FERMENTATION*. /paper/A-PRIZE-WINNING-DISCOVERY-OF-1896-%3A-BUCHNER-OF-Bohley-Fr%C3%B6hlich/08a91e039e7439c779d7b1fa153b79eba12db3b3

Booth, L. N., Tuch, B. B., & Johnson, A. D. (2010). Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature*, *468*(7326), 959–963. https://doi.org/10.1038/nature09560

Britton, C. S., Sorrells, T. R., & Johnson, A. D. (2020). Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit. *Science (New York, N.Y.)*, *367*(6473), 96–100. https://doi.org/10.1126/science.aax5217

Cheatle Jarvela, A. M., & Hinman, V. F. (2015). Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. *EvoDevo*, *6*(1), 3. https://doi.org/10.1186/2041-9139-6-3

Dalal, C. K., Zuleta, I. A., Mitchell, K. F., Andes, D. R., El-Samad, H., & Johnson, A. D. (2016). Transcriptional rewiring over evolutionary timescales changes quantitative and qualitative properties of gene expression. *ELife*, *5*, e18981. https://doi.org/10.7554/eLife.18981

Gasch, A. P., Moses, A. M., Chiang, D. Y., Fraser, H. B., Berardini, M., & Eisen, M. B. (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biology*, *2*(12), e398. https://doi.org/10.1371/journal.pbio.0020398

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., & Carroll, S. B. (2005). Chance caught on the wing: Cis -regulatory evolution and the origin of pigment patterns in Drosophila. *Nature*, *433*(7025), 481–487. https://doi.org/10.1038/nature03235

Hernday, A. D., Noble, S. M., Mitrovich, Q. M., & Johnson, A. D. (2010). Genetics and Molecular Biology in Candida albicans. In *Methods in Enzymology* (Vol. 470, pp. 737–758). Academic Press. https://doi.org/10.1016/S0076-6879(10)70031-8

Hinnen, A., Hicks, J. B., & Fink, G. R. (1978). Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(4), 1929–1933.

King, M.-C., & Wilson, A. C. (1975). Evolution at Two Levels in Humans and Chimpanzees. *Science*, *188*(4184), 107–116.

Kumamoto, C. A., Gresnigt, M. S., & Hube, B. (2020). The gut, the bad and the harmless: Candida albicans as a commensal and opportunistic pathogen in the intestine. *Current Opinion in Microbiology*, *56*, 7–15. https://doi.org/10.1016/j.mib.2020.05.006

Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., & Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics*, *42*(7), 631–634. https://doi.org/10.1038/ng.600

Li, H., & Johnson, A. D. (2010). Evolution of Transcription Networks—Lessons from Yeasts. *Current Biology*, *20*(17), R746–R753. https://doi.org/10.1016/j.cub.2010.06.056

Lindegren, C. C., & Lindegren, G. (1943). Segregation, Mutation, and Copulation in Saccharomyces cerevisiae. *Annals of the Missouri Botanical Garden*, *30*(4), 453. https://doi.org/10.2307/2394308

Lombardi, L., Oliveira-Pacheco, J., & Butler, G. (2019, March 1). *Plasmid-Based CRISPR-Cas9 Gene Editing in Multiple Candida Species*. MSphere; American Society for Microbiology. https://doi.org/10.1128/mSphere.00125-19

Lynch, V. J., Tanzer, A., Wang, Y., Leung, F. C., Gellersen, B., Emera, D., & Wagner, G. P. (2008). Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(39), 14928–14933. https://doi.org/10.1073/pnas.0802355105

Lynch, V. J., & Wagner, G. P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution; International Journal of Organic Evolution*, *62*(9), 2131–2154. https://doi.org/10.1111/j.1558-5646.2008.00440.x

Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C., & Patel, N. H. (2016). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. *Current Biology*, *26*(1), 14–26. https://doi.org/10.1016/j.cub.2015.11.021

Miller, M. G., & Johnson, A. D. (2002). White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell*, *110*(3), 293–302. https://doi.org/10.1016/s0092-8674(02)00837-1

Nguyen, N., Quail, M. M. F., Hernday, A. D., Mitchell, A. P., Kuchler, K., Kadosh, D., & Cormack, B. (2017, April 1). *An Efficient, Rapid, and Recyclable System for CRISPR-Mediated Genome Editing in Candida albicans*. MSphere; American Society for Microbiology. https://doi.org/10.1128/mSphereDirect.00149-17

Nocedal, I., & Johnson, A. D. (2015). How Transcription Networks Evolve and Produce Biological Novelty. *Cold Spring Harbor Symposia on Quantitative Biology*, *80*, 265–274. https://doi.org/10.1101/sqb.2015.80.027557

Nocedal, I., Mancera, E., & Johnson, A. D. (2017). Gene regulatory network plasticity predates a switch in function of a conserved transcription regulator. *ELife*, *6*, e23250. https://doi.org/10.7554/eLife.23250

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *ELife*, *4*, e05290. https://doi.org/10.7554/eLife.05290

Ohno, S. (1972). An argument for the genetic simplicity of man and other mammals. *Journal of Human Evolution*, *1*(6), 651–662. https://doi.org/10.1016/0047-2484(72)90011-5

Priyadarshini, Y., & Natarajan, K. (2016). Reconfiguration of Transcriptional Control of Lysine Biosynthesis in Candida albicans Involves a Central Role for the Gcn4

Transcriptional Activator. *MSphere*, *1*(1). https://doi.org/10.1128/mSphere.00016-15

Sai, S., Holland, L. M., McGee, C. F., Lynch, D. B., & Butler, G. (2011). Evolution of Mating within the Candida parapsilosis Species Group ▿ . *Eukaryotic Cell*, *10*(4), 578–587. https://doi.org/10.1128/EC.00276-10

Shen, X.-X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R.-I., Čadež, N., Libkind, D., … Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, *175*(6), 1533-1545.e20. https://doi.org/10.1016/j.cell.2018.10.023

Sorrells, T. R., Johnson, A. N., Howard, C. J., Britton, C. S., Fowler, K. R., Feigerle, J. T., Weil, P. A., & Johnson, A. D. (2018). Intrinsic cooperativity potentiates parallel cis-regulatory evolution. *ELife*, *7*, e37563. https://doi.org/10.7554/eLife.37563

Tsong, A. E., Tuch, B. B., Li, H., & Johnson, A. D. (2006). Evolution of alternative transcriptional circuits with identical logic. *Nature*, *443*(7110), 415–420. https://doi.org/10.1038/nature05099

Whiteway, M., Tebung, W. A., Choudhury, B. I., & Rodríguez-Ortiz, R. (2015). Metabolic regulation in model ascomycetes – adjusting similar genomes to different lifestyles. *Trends in Genetics*, *31*(8), 445–453. https://doi.org/10.1016/j.tig.2015.05.002

Wittkopp, P. J., & Kalay, G. (2011). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews. Genetics*, *13*(1), 59–69. https://doi.org/10.1038/nrg3095

Chapter 2

Generally conserved, specifically diverged:
The evolution of the transcriptional regulation of amino acid biosynthesis

Introduction

A striking aspect of transcription network rewiring is how truly pervasive it is. Even in processes that appear to be fundamentally conserved, upon deeper digging into the molecular mechanisms underlying these processes, it is common to find that transcription network rewiring has, in fact, occurred. The ability to maintain a fundamental logic to a regulatory system – for example, the need to upregulate a biosynthetic pathway in the face of nutrient starvation – in the face of regulatory evolution suggests that evolution explores multiple “solutions” to the same regulatory problem.

The comparison between S. cerevisiae and C. albicans offer numerous examples of seemingly conserved regulatory logic, whose underpinning molecular mechanisms have, in fact, changed. Even in such a central metabolic process such as glycolysis the two species diverge. While the enzymes involved in the process are largely conserved (Askew et al. 2009), what has diverged is the mechanisms by which these genes are regulated. Crabtree-positive yeasts, that is those that use the fermentation pathway in the presence of oxygen and high amounts of glucose, such as *S. cerevisiae*, repress respiration and up-regulate the glycolytic/fermentation pathway in the presence of excess glucose. In contrast, *C. albicans* will completely oxidize carbohydrates through the respiration pathway in aerobic conditions, only relying on the fermentation pathway in the absence of oxygen. Furthermore, *C. albicans* up-regulates the glycolytic genes in low oxygen conditions while *S. cerevisiae* does not.

These differences in the regulation of glycolytic genes are mediated by different transcription factors. In S. cerevisiae, Gcr1 and Gcr2 together activate glycolytic

enzymes in the presence of excess glucose (Drago Clifton, Weinstock, and Fraenkel 1978; Uemura and Fraenkel 1990). Deleting either regulator decreases the expression levels of the glycolytic genes resulting in growth defects during culture on glucose (D. Clifton and Fraenkel 1981).

However, this Gcr1/2-dependent gene activation has not been documented in widely diverged yeast species and may be unique to S. cerevisiae and close relatives. In particular, C. albicans relies on different regulators for the activation of glycolysis genes: Gal4 and Tye7 (Askew et al. 2009). Deleting both factors resulted in a severe growth defect during culture on several fermentable carbon sources when respiration was inhibited or oxygen was limited. Tye7p provides a strong basal level of glycolytic expression and commits the cell to glycolysis. Gal4p, on the other hand, is a carbon source-dependent regulator that fine-tunes gene expression in response to the presence of fermentable carbon sources (Askew et al. 2009).

In fact, the role of Gal4 across species provides an additional example of evolutionary divergence in the regulation of a central metabolic pathway. In S. cerevisiae, the transcription factor Gal4 is responsible for inducing the expression of the three enzymes responsible for converting galactose to glucose expression in the presence of excess galactose and absence of glucose (Giniger, Varnum, and Ptashne 1985; Traven, Jelicic, and Sopta 2006). While in C. albicans, Gal4 is involved in regulating glycolytic genes, and C. albicans instead relies on different regulators, Rtg1 and Rtg3, to induce the Gal enzymes when galactose is present and glucose is absent (Dalal et al. 2016).

These two examples demonstrate that when it comes to the transcriptional regulation of metabolic pathways what must be conserved is the logic: when a key metabolite is in short supply, the genes involved in synthesizing that metabolite must be upregulated in order to make more of that metabolite. Conversely, when the concentration of a metabolite is high, the enzymes involved in breaking that metabolite down or converting it into another more readily used or stored molecule must be upregulated. While these logical connections are conserved (low concentration of metabolite, increased expression of metabolite's biosynthetic genes), the mechanism by which these biosynthetic genes are upregulated are prone to evolutionary divergence. Specifically, transcription network rewiring – that is, changing the connections between transcription factors and their target genes – can result in different sets of transcription factors regulating a set of target genes in different species (Johnson 2017; Li and Johnson 2010).

The extent to which the transcription factors that regulate amino acid biosynthetic pathways have diverged is less clear. A great deal is known about the regulation of amino acid biosynthetic pathways from experiments in S. cerevisiae. Multiple modes of regulation have been uncovered – transcriptional, post-transcriptional, and allosteric – and all these modes work together to monitor flux and respond to levels of available amino acids. Within transcriptional control of amino acid biosynthetic pathways, two modes of regulation exist – cross-pathway [sometimes referred to as general amino acid control (GAAC)] and pathway-specific control.

GAAC is characterized by enzymes in multiple amino acid biosynthetic pathways induced in response to starvation for any amino acid (A G Hinnebusch 1988). The

majority of experiments have relied on the induction of amino acid starvation using the addition of antimetabolites [e.g., 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of imidazoleglycerol-phosphate dehydratase (HIS3) that catalyzes the sixth step of histidine biosynthesis, and metsulfuron methyl, an inhibitor of acetohydroxyacid synthase (Ilv2) that catalyzes the first step of branched-chain amino acid biosynthesis]. When exposed to 3-AT, cells activate the expression of a large set of genes including representatives in every amino acid biosynthetic pathway, with the exception of cysteine (Alan G Hinnebusch and Natarajan 2002; Natarajan et al. 2001).

The transcriptional activator Gcn4 mediates GAAC. GCN4 expression is induced in starved cells at the translational level by a reinitiation mechanism involving four short upstream open reading frames (Alan G Hinnebusch 2005; Mueller and Hinnebusch 1986). Briefly, upon amino acid starvation, multiple tRNAs become deacylated (Zaborske et al. 2009; 2010). Gcn2 has an auto-inhibited kinase domain that is allosterically activated in starved cells through binding of uncharged tRNAs to an adjacent histidyl-tRNA synthetase-like domain (Dong et al. 2000; Wek, Jackson, and Hinnebusch 1989). The activated Gcn2 kinase phosphorylates the α-subunit of eIF2, resulting in reduced levels of eIF2-GTP-Met-tRNA$_i^{Met}$ ternary complex, and thus decreasing the efficiency of scanning ribosomes to reinitiate translation (Dever et al. 1995). Ultimately, amino acid starvation increases the proportion of ribosomes that reinitiate translation at GCN4, increasing protein levels of GCN4, subsequently activating amino acid biosynthetic pathways, and amino acid output.

While GAAC is effective at up-regulating hundreds of biosynthetic enzymes across multiple biosynthetic pathways, more fine-tuned transcriptional control is

coordinated via pathway-specific transcriptional responses. A well-defined example of pathway-specific transcriptional control comes from what is known about the regulation of the lysine biosynthetic pathway. Lysine is synthesized from α-ketoglutarate via the α-aminoadipate (α-AAS) pathway (Xu et al. 2006). Two fundamental feedback control mechanisms exist in this pathway. Firstly, the first step of the pathway is catalyzed by either Lys20 or Lys21, both of which are feedback-inhibited by the end product, lysine (Andi, West, and Cook 2005). Therefore, when lysine is present in high concentration, flux through the pathway reduces. Conversely, under lysine starvation, flux through the pathway increases. The second mode of feedback control is transcriptional: the expression of the LYS genes is controlled by a pathway-specific transcription factor, Lys14, which, in turn, responds to pathway-intermediate: α-AAS. This pathway intermediate binds and activates the pathway-specific transcription factor Lys14 (Becker et al. 1998; El Alami et al. 2002). As a consequence of a pathway intermediate controlling the capacity of Lys14 to activate gene expression, conditions that increase or decrease the flux through the pathway, positively or negatively, affect LYS gene expression. Increased flux in the pathway results in elevated production of α-AAS, turning on Lys14-dependent expression of all LYS genes. Conversely, due to feedback inhibition of the first step of the pathway (catalyzed by either Lys20 or Lys21), excess lysine reduces the production of α-AAS and causes apparent repression of the LYS genes (Andi, West, and Cook 2005).

In the case of lysine biosynthesis, GAAC and pathway-specific control are integrated as Lys14 is itself a transcriptional target of Gcn4. Thus, upon lysine starvation, the pathway-specific modes of regulation are activated, as is GAAC via the

starvation of a single amino acid. GAAC, in turn, further up-regulates Lys14 and the pathway-specific response.

In S. cerevisiae, other amino acid biosynthetic pathways are co-regulated by pathway-specific and general control. In the case of methionine biosynthesis: a pathway-specific complex is known to regulate the expression of MET genes. The expression of the majority of genes encoding enzymes of methionine genes requires the transcriptional activator Met4 (Lee et al. 2009; Thomas and Surdin-Kerjan 1997). Met4 interacts directly with either of two highly similar zinc-finger proteins, Met31 and Met32, or with the basic-helix-loop-helix protein Cbf1. An additional cofactor, Met28, which also lacks DNA-binding activity, is thought to stabilize DNA-bound Met4 complexes (Blaiseau and Thomas 1998; Kuras, Barbey, and Thomas 1997). Under sulfur-limiting conditions, these interactions enable Met4 to activate transcription through recruitment of the SAGA histone acetyltransferase and Mediator coactivator complexes (Kuras et al. 2002; Leroy, Cormier, and Kuras 2006).

GAAC is thought to have a limited role in MET gene expression under methionine-limiting conditions, however starvation for histidine or tryptophan results in strong Gcn4-dependent induction of several MET genes, including MET1, MET10, MET13, MET16, MET17, MET22, MET28, SUL1, and SUL2 (Natarajan et al. 2001). The fact that Gcn4 induces MET28 suggests that GAAC may indirectly activate MET genes by facilitating the stability of pathway-specific activation complexes.

While many of the details of the regulation of amino acid biosynthetic pathways have been worked out genetically and biochemically in S. cerevisiae, fewer have been examined in C. albicans or other yeast species. However, the evidence accumulated so

far suggests that GAAC mediated by Gcn4 exists across a wide range of fungal species. In C. albicans, specifically, analogous gene expression profiling experiments using 3-AT defined the Gcn4-regulon and found that, similarly to S. cerevisiae, amino acid biosynthetic genes representing all 20 amino acid pathways were enriched upon 3-AT treatment. The majority (~90%) of these 3-AT induced genes were dependent on Gcn4 (Tournu et al. 2005). Of note, C. albicans Gcn4 shares similar translational and Gcn2-mediated regulation with its S. cerevisiae ortholog. In addition, in the filamentous fungi, Neurospora crassa, the Gcn4 ortholog Cpc1 is required for induction of numerous amino acid biosynthetic genes upon amino acid starvation (Tian et al. 2007). While the total conservation of the regulons of the S. cerevisiae, C. albicans and N. crassa Gcn4 orthologs is only about 10% (that is, 32 of the 300-400 genes whose change in a Gcn4-dependent fashion upon 3-AT treatment are the same across all three species), of this deeply conserved Gcn4-regulon two-thirds are amino acid biosynthetic genes (21 of the 32 conserved Gcn4-target genes. The only evidence of GAAC mediated by a transcription factor other than Gcn4 exists lies outside of the Saccharomycotina clade. In the basidiomycete model yeast, S. pombe, Fil1 induces the expression of many amino acid biosynthetic genes in response to 3-AT, and has similar uORF and Gcn2-mediated regulation, however it is in a completely different DNA-binding domain family than Gcn4 – Fil1 is a GATA transcription factor while Gcn4 is a bZip transcription factor (Duncan et al. 2018). Together, this suggests that GAAC mediated by Gcn4 is deeply conserved and is at least ~300 million years old.

While GAAC is conserved, it is unclear to what extent pathway-specific regulation exists in *C. albicans*. Previous work has demonstrated that there has been

transcriptional rewiring at least in the case of lysine biosynthesis. The orthologs of Lys14 – the sequence-specific DNA binding protein responsible for activating the lysine biosynthetic genes in the absence of lysine in *S. cerevisiae* – do not regulate the lysine biosynthetic genes in C. albicans (Pérez et al. 2014). While this work described what the Lys14 orthologs do in C. albicans, it did not answer the question of which transcription factor regulates lysine biosynthesis. Other previous work has suggested the potential for rewiring events in the regulation of methionine biosynthesis, as the deletion of the Met32 ortholog in C. albicans did not produce a methionine auxotrophy (Homann et al. 2009) Again, this study suggests a re-wiring event, but it did not seek to answer the question of which transcription factor regulates the methionine biosynthetic genes in C. albicans. Therefore, the question of which transcription factors regulate the amino acid biosynthetic pathways in C. albicans is still unanswered.

Results

Screen for transcriptional regulators of amino acid biosynthesis in C. albicans

To address this question systematically, we screened for transcriptional regulators of amino acid biosynthesis using a collection of transcription factor knockout mutants. To identify putative pathway-specific transcription factors, we did a preliminary screen comparing growth in synthetic complete media to growth in media lacking amino acids. The logic of this screen is that any transcription factor responsible for the induction of any amino acid biosynthetic pathway would behave as an auxotroph in the absence of amino acids, and would not grow. Due to the strain background of the library, we could not withdraw arginine from the media. We found that 20 strains did not grow in the absence of amino acids, however 2 of these mutant strains also exhibited

slow growth in complete media, so we excluded these 2 strains from further analysis. (Fig 1) It is important to note that this screen was an end-point analysis; strains were grown overnight in media lacking amino acids and optical densities were checked the next day. We were not looking for intermediate growth defects, but rather we wanted to find the transcription factors that were absolutely required for induction of the amino acid biosynthetic pathways.

We took the remaining 18 strains through a secondary screen in which we systematically withdrew a single amino acid one at a time to answer the question of which, if any, of the potential amino acid biosynthetic transcriptional regulators are pathway-specific. If a transcription factor was a pathway-specific regulator, we would expect to see no growth when only the one amino acid is withdrawn. We found that in three conditions (-Lys, -His, -Met), only the Gcn4 deletion strain behaved in a pathway-specific manner (Fig 2). All other mutants in all other conditions demonstrated the ability to grow following over-night culture in media lacking a single amino acid.

The screen was also conducted on plates. The same collection of transcription factor knockout mutants were plated as dilution series on rich plates and plates that were deprived of a single amino acid. Using this approach, the Gcn4 deletion strain was the only mutant that demonstrated growth defects across nearly all of the single drop-out conditions tested, and notably these phenotypes were the most reproducible across both biological and technical replicates. Specifically, and in keeping with the results of the liquid culture assay, the Gcn4 mutant displayed the largest defects in the -His, -Met and -Lys conditions, while less severe growth defects were observable in other conditions (-Ile, -Leu, -Thr, for example). On plates, other mutants did show some

growth defects in certain drop-out conditions, however none were as reproducible nor as severe as the Gcn4 mutant (Fig 3).

The results of both the liquid and plate-based screens together suggest that Gcn4 is a central regulator of amino acid biosynthetic genes in C. albicans, and there are no clear pathway-specific transcriptional regulators. Given that there were some minor defects on plates, there is a chance that other transcriptional regulators do contribute to growth in a pathway-specific manner, however none of those mutants displayed an auxotrophic growth behavior comparable to the Gcn4 mutant.

Species-specific differences in the reliance on Gcn4 for growth

The result of the screens for transcription factors that regulate amino acid biosynthetic pathways in C. albicans converged on Gcn4. Both the liquid overnight and on-plate screens clearly demonstrate that Gcn4 is required for growth in conditions lacking lysine, methionine, or histidine. While the on-plate screen suggested growth rate defects in these and other conditions, we had not measured growth rate directly in either screen. In addition, the requirement of Gcn4 for growth in media lacking lysine in C. albicans is a phenotypic difference than what we see in S. cerevisiae. In S. cerevisiae, a Gcn4 mutant will grow in the absence of lysine, while the Lys14 mutant will not, suggesting that Lys14 plays a larger role in activating the lysine biosynthetic genes. These observations led us to ask if there are other species-specific differences in the reliance on Gcn4 for growth in amino acid dropout conditions.

To answer this question, Gcn4 was deleted in prototrophic strains of C. albicans and S. cerevisiae so we could test all 20 dropout conditions in a plate-based liquid growth assay. The wildtype strains of both species grew without defect across all conditions tested. As expected from the results of the screens, in C. albicans, Gcn4 was required for growth in media lacking histidine, lysine or methionine – the Gcn4 deletion strain did not grow in these conditions (fig 4). Also consistent with the result of the plate-based screen, we observed moderate growth rate defects for the Gcn4 deletion mutant in media lacking leucine, isoleucine or threonine (fig 4). In contrast, S. cerevisiae relied on Gcn4 for growth in media lacking arginine or media lacking tryptophan – consistent with previously reported literature that demonstrates a reliance on Gcn4 for full activation of these biosynthetic pathways.

This direct comparison of C. albicans and S. cerevisiae Gcn4 deletion mutants demonstrates a shift in the reliance on Gcn4 between the two species. It suggests that Gcn4 is the main transcriptional regulator of lysine, methionine and histidine biosynthetic genes in C. albicans. It also suggests that CaGcn4 does not play as central of a role in the activation of arginine and tryptophan biosynthetic genes as it does in S. cerevisiae.

Evolution of pathway-specific v cross-pathway control

Given the species-specific differences in reliance on Gcn4, and the lack of evidence for pathway-specific transcriptional regulation of amino acid biosynthetic pathways in C. albicans, we next asked when did general and pathway-specific

transcriptional regulation evolve. We took a bioinformatic approach to answer this question. We scored the genomes of # yeast species for the presence the motifs of known transcriptional regulators of amino acid biosynthetic genes in the intergenic regions of biosynthetic enzymes. While we generated species-specific motifs when possible – for example, in the case of C. albicans Lys14 orthologs where we had high-confidence data on their target genes (Pérez et al. 2014) – a caveat of this analysis is that it otherwise assumes that orthologous target genes and motifs have conserved their functions across species. We limited this analysis to the lysine, methionine, arginine and branched chain amino acid pathways (leucine, isoleucine, valine) as these were the pathways for which we were able to find information-rich motifs either through motif-generating software like MEME or pre-existing motif libraries.

First, we asked whether the ortholog of known pathway-specific transcriptional regulators in S. cerevisiae exist in C. albicans and closely related species because one explanation for the lack of pathway specific regulation might be the lack of pathway-specific regulators. We already knew that the ratio of orthologs for certain pathway specific regulators would not be 1-to-1 between S. cerevisiae and C. albicans. For example, we already knew that the Lys14 orthologs in C. albicans had undergone two duplication events such that there are now four Lys14 orthologs in C. albicans. In addition, certain S. cerevisiae regulators, specifically Met31/32 and Arg80/81, are the result of the whole genome duplication and would therefore display a 2-to-1 ratio when comparing S. cerevisiae to C. albicans, and this result was clearly demonstrated when looking in a clade-specific manner (Fig 5). However, we found some additional “missing” orthologs when looking at the regulators of methionine biosynthesis. Specifically, Met4

and Met28 – neither of which are the result of the whole genome duplication – do not appear to have a clear ortholog in C. albicans or any of the closely related species in the Candida clade (Fig 5). This could explain the complete reliance on Gcn4 for methionine biosynthesis in C. albicans – some of the essential pathway specific regulators were lost from the C. albicans genome.

We did the same analysis on the biosynthetic enzymes, to ensure that the pathways were likely to be conserved across species. Here we found less variation as compared to the regulators – nearly all the enzymes have identifiable orthologs across species, with isolated losses identified in the case of few species and few enzymes. After establishing the identity of the enzymes and regulators, we scored the intergenic regions of the biosynthetic enzymes for the presence or absence of a given sequence-specific DNA binding motif. If a transcription factor regulates a biosynthetic pathway, we would expect to see an enrichment of that motif across the entire pathway. For example, in the case of S. cerevisiae we would expect to find Lys14 motifs throughout the Lys biosynthetic intergenic regions. This is indeed what we find – the Lys14 motif is highly enriched in the lysine biosynthetic genes in S. cerevisiae and all of S. cerevisiae's closest relatives. In contrast, we see significantly less enrichment for Lys14 motifs in C. albicans and its closest relatives. In addition, we see enrichment for Gcn4 in the Lys biosynthetic genes across both clades, with higher enrichment of Gcn4 in the Candida clade. This finding is consistent with the phenotypic data which demonstrates that Gcn4 is required for growth in conditions lacking lysine in C. albicans, while Lys14 is required for growth in conditions lacking lysine in S. cerevisiae.

To understand whether the regulation of the lysine biosynthetic pathway by Lys14 was an evolutionary loss or gain, we expanded this analysis to all # yeast genomes. We find that the Lys14 motif is enriched in the promoters of the lysine biosynthetic genes only in the species spanning from the Saccharomyces clade to the Hansenula clade. This suggests that pathway-specific regulation by Lys14 was a more recent evolutionary gain. In addition, by scoring the Lys14 motif in the Lys14 promoter, we find that Lys14 auto-regulation may have evolved earlier than the pathway-specific regulation. We find Lys14 enrichment in the Lys14 promoter extending from the Saccharomyces clade through the Wickerhamomyces clade, a larger evolutionary distance than the signature of pathway-specific regulation. By contrast, we find Gcn4 motifs distributed across the entire phylogenetic tree, with perhaps a slight enrichment in the Candida clade. This suggests that Gcn4-mediated control of the lysine biosynthetic pathway is evolutionary ancient, while pathway-specific regulation by Lys14 is a much more evolutionarily novel.

For the other pathways in question, the signature is not nearly as clear as the case of lysine biosynthesis. While the case of methionine-specific regulation may be partially due to the loss of Met4 and Met28, we still find Met31 motifs in many of the genes involved in methionine biosynthesis in the Candida clade. Furthermore, we do not see an enrichment in Gcn4 motifs in the Candida clade as compared to S. cerevisiae.

For the branched chain amino acid pathways, the signature is even less clear. We find Leu3 motifs equally distributed throughout both the Saccharomyces and the Candida leucine, isoleucine and valine pathways. However, the Leu3 ortholog in C.

albicans did not demonstrate an auxotrophic phenotype in the absence of leucine, isoleucine or valine. In addition, there is no enrichment of the Gcn4 motif in these pathways. Admittedly, the Gcn4-dependence of the branched chain amino acid biosynthetic genes is not as strong as the Gcn4-mutant displayed intermediate growth defects in media lacking leucine, isoleucine or valine, so perhaps we would not expect to see an enrichment of Gcn4 in this pathway.

Discussion

In this study, we reveal an apparent lack of C. albicans pathway-specific regulation of amino acid biosynthetic pathways. This is in stark contrast to the pathway-specific regulation of amino acid biosynthetic pathways in S. cerevisiae where at least four pathways (Lys, Met, Arg, and Ser) have been shown to have dedicated pathway-specific transcription factors. Conversely, we demonstrate a conserved and ancient role for Gcn4 in mediating cross-pathway general amino acid control in both C. albicans and S. cerevisiae. This conserved role of Gcn4 in C. albicans has been shown using 3-AT as a mechanism of inducing histidine starvation but has never been put in the context of true amino acid limitation nor in the context of how many transcriptional regulators exist in addition to Gcn4 in C. albicans.

In particular, we have a clear picture of the rewiring of the regulation of the lysine biosynthetic pathway. Previous work showed that the C. albicans Lys14 orthologs do not bind to the promoters of lysine biosynthetic genes (Pérez et al. 2014), while Gcn4 has been shown to bind to the promoters of at least two lysine biosynthetic genes (Lys2, Lys9) (Priyadarshini and Natarajan 2016). In addition, the subsequent neofunctionalization of the Lys14 orthologs via changes in their DNA binding specificity

led them to play a role in gastrointestinal tract colonization and systemic infection (Pérez, Kumamoto, and Johnson 2013). Here we definitively show using comparative genetics that C. albicans Gcn4 is absolutely required for growth in media lacking lysine, while S. cerevisiae Gcn4 is not required under the same conditions. Using bioinformatic analysis across hundreds of published yeast genomes we show a strengthening of Gcn4 binding sites in the promoters of the lysine biosynthetic genes of C. albicans and its close relatives, and a Saccharomyces-clade specific gain of Lys14 binding sites. Thus, the rewiring of the transcriptional regulation of the lysine biosynthetic pathway between S. cerevisiae and C. albicans involved changes in both cis (strengthening of Gcn4 binding sites in C. albicans, gain of Lys14 binding sites in S. cerevisiae) and trans (gene duplication and changes in DNA binding specificity of the Lys14 orthologs in C. albicans.)

Did the role of the Lys14 duplication events in the Candida clade play a role in this apparent shift from Lys14-dependent regulation of the lysine biosynthetic pathway to a Gcn4-dependent mode? Genome-wide studies in both eukaryotes (Lynch and Conery 2000) and prokaryotes (Kondrashov et al. 2002) have shown periods of relaxed selection and accelerated evolution in both paralogs following gene duplication. This is hypothesized to allow for neofunctionalization to occur as one or both paralogs accumulate mutations. It is plausible that in the case of Lys14, which underwent its first duplication in the ancestor of the Candida clade (the CTG clade), that the duplication event led to changes in the DNA binding domain of both paralogs, ultimately leading to their new functions in commensalism and pathogenesis. Our bioinformatic analysis here shows no Lys14 binding sites in the promoters of the Lys biosynthetic genes in species

where more than one copy of Lys14 exists. Conversely, in the genomes where only one copy of Lys14 exists – the clades encompassing S. cerevisiae and K. lactis -- Lys14 binding sites are found in the promoters of the Lys biosynthetic genes. Intriguingly, in the outgroup to S. cerevisiae and C. albicans, (the clade containing the Wickerhameromyces yeast W. anomalus and W. ciferrii), we see a Lys14 binding sites in the Lys14 promoter itself, but not in the promoters of the Lys biosynthetic genes. After running BLAST on each of the C. albicans Lys14 paralogs (and subsequent reciprocal BLAST with any of the plausible hits), it appears that W. anomalus may have at least 2 copies of Lys14. While this may suggest that the duplication event is actually more ancient than we previously reported, it also strengthens this intriguing correlation: wherever we see more than one copy of Lys14, we see a subsequent loss of Lys14 binding sites in the promoters of the lysine biosynthetic pathway. While direct experiments in species like W. anomulus would be necessary to fully conclude anything about the function of the Lys14 paralogs in that species, it is a pattern that suggests that the relaxed selection experienced by the Lys14 paralogs was an important step in the rewiring of transcriptional regulation of lysine biosynthesis.

While transcriptional control is an important regulatory mode for biosynthetic pathways, allosteric control via pathway intermediates is another prominent form of regulation. This is the case for Lys14 which binds and is activated by α-aminoadipate semialdehyde, the product of the Lys2- and Lys5- catalyzed conversion of alpha-aminoadipate into alpha-aminoadipate semialdehyde in the fourth step of the lysine biosynthetic pathway. A functional transcriptional reporter analysis of S. cerevisiae Lys14 identified a 20-amino acid long leucine-rich stretch (positions 298-317 in the S.

cerevisiae protein) which, when deleted, rendered Lys14 into a constitutive activator, suggesting that this reason is required for the α-aminoadipate semialdehyde induced activation. Intriguingly, this leucine-rich region is in a relatively well-conserved region of the protein, however nearly none of the leucine residues found in the S. cerevisiae Lys14 are conserved in the C. albicans Lys14 orthologs (more specifically, there are 6 leucine residues in this 20 amino acid stretch. One leucine is common between S. cerevisiae Lys14 and C. albicans Lys144 and another leucine is common between S. cerevisiae Lys14 and C. albicans Lys142. None of the C. albicans lys14 paralogs seem to have a leucine-rich region in this part of the protein alignment.). This suggests that in addition to critical changes to DNA binding specificity, the Lys14 paralogs in C. albicans may have also lost allosteric control via lysine pathway intermediates – further evidence that pathway-specific control of the lysine biosynthetic pathway is not a feature of amino acid biosynthetic control in C. albicans.

We screened a C. albicans library of 270 transcription factor knockout strains, both on plates and in liquid growth assays, and did not find evidence of pathway-specific transcriptional regulation of amino acid biosynthetic pathways. However, a few important caveats should be noted. Firstly, the screen was designed only to find transcriptional activators of biosynthetic pathways as we looked for strains that were auxotrophic upon withdrawal of amino acids from the growth media. If the predominant mode of transcriptional regulation in C. albicans is via repression, as is the case in the S. cerevisiae serine biosynthetic pathway, we would not have identified those mutants. Furthermore, we applied a stringent threshold as we were looking for strains that truly did not grow in the absence of a given amino acid – as we found for the Gcn4 mutant

across multiple conditions. However, if a transcriptional regulator played a pathway-specific role but displayed a more subtle growth defect – perhaps via redundancy with Gcn4 or another factor – we would not have identified those mutants either. Lastly, while we believe this Candida albicans transcription knockout library to be well-vetted and comprehensive, it is entirely possible that candida-specific families of transcription factors would have been missed in the initial creation of the library. However, this would require that all the pathway-specific regulators that we failed to discover in our screen fall in this candida-specific category of transcriptional regulators. While it is possible, it seems unlikely and would be very unlucky.

If the emerging picture here is that C. albicans lacks pathway-specific transcriptional control of amino acid biosynthetic pathways and relies entirely on cross pathway control mediated by Gcn4, while a species like S. cerevisiae utilizes both pathway-specific and cross-pathway control, it may be worth briefly considering the vast differences in the environmental niche of these two species. C. albicans resides in the human gut while S. cerevisiae is found on the bark of deciduous trees and rotting fruit. An obvious difference between the two species may be nutrient, and in this case specifically amino acid, availability. While it is unclear what the exact concentrations of amino acids in each environment may be, experimental evidence shows that auxotrophic strains of Candida albicans survive, proliferate, and cause disease in mouse models of infection. This suggests, that at least in the mouse gut, amino acids are abundant enough to compensate for the deletion of key biosynthetic enzymes. In addition, C. albicans has an expanded family of secreted aspartic proteases and oligopeptide transporters that have a role in the acquisition of nutrients during

colonization and infection, suggesting that perhaps C. albicans has evolved an enhanced ability to acquire amino acids from the environment. All this, while highly speculative, may indicate that C. albicans is not likely to experience prolonged amino acid starvation induced stress, and that amino acid starvation is not a strong environmental or evolutionary pressure. Thus, it is possible that the evolution of pathway-specific amino acid regulation evolved in the S. cerevisiae lineage as a necessary “fine-tuning” mechanism to survive periods of amino acid starvation, while in C. albicans an “all or nothing” Gcn4-mediated response is sufficient.

Materials & Methods

Strains & Media

The transcription factor knockout collection strains were derived from SN152 (Noble and Johnson, 2005) and constructed by fusion PCR using the His and Leu cassettes as previously described (Hernday et al., 2010; Homann et al., 2009; Noble and Johnson, 2005). OH13 (Homann et al., 2009) was used as a parent strain.

To generate a prototrophic C. albicans Gcn4-deletion strain, the SAT1-flippable cassette was used as described (Sasse & Morschhauser, 2012). Briefly, upstream and downstream homology arms of Gcn4 were amplified and cloned into pMBL162 (Lohse & Johnson, 2010). The plasmid was linearized with SphI and AatII, and transformed in the prototrophic wildtype C. albicans strain SC5314. Nourseothricin resistant colonies were selected on YPD+NAT, and confirmed by PCR flank-check. Colonies were grown overnight in YEP+2% Maltose to induce flipping of the SAT1 cassette. Colonies were plated on YPD, grown for 2 days at 30C before replica plating onto YPD with and

without NAT. Colonies that grew on YPD but not on YPD+NAT were confirmed by PCR flank checks, SAT1-cassette checks, and ORF-checks for loss of Gcn4.

To generate a prototrophic S. cerevisiae Gcn4-deletion strain, a prototrophic diploid derivative of S288C was used as the parental strain, obtained from Naomi Ziv. A KanMX and a G418-resistance marker were amplified with primers with 90bp of homology to Gcn4 and two rounds of a published lithium acetate/single-stranded carrier DNA/polyethylene glycol method (Gietz & Woods, 2002) were used to knockout both copies of Gcn4.

Liquid medium screen for transcriptional regulators of amino acid biosynthetic pathways

The transcription factor knockout collection, and the prototrophic wildtype and Gcn4-deletion strains were pinned onto YPD rich media to ensure that all mutants, even the slow-growing mutants, would grow. Strains were then inoculated into YPD rich media in deep-well 96 well dishes and grown overnight at 30C on an INFORS HT shaker. OD600 was measured on a TECAN plate reader and, after washing 3x in synthetic complete media, the strains were diluted to a starting OD of ~0.1-0.2 in synthetic complete media. Strains were grown for 6-8 hours until an OD of at least ~1.0 had been reached by the majority of strains. Strains were then washed 3x with YNB+20% Glucose + Arginine, and diluted to a starting OD600 of .01 for overnight growth in YNB+20%Glucose+Arg at 30C. Overnight cultures were diluted 1:10 in order to be in the linear range of the TECAN plate reader, and OD600 was measured for the entire library.

This protocol was repeated with hits from the screen described above, however instead of doing the overnight growth in YNB+20% Glucose+Arg, the strains were grown in “single amino acid selection” media (YNB+20%Glucose+Arg+all but one amino acid, exact recipe in the “Strains & Media” section).

Plate screen for transcriptional regulators of amino acid biosynthetic pathways

The same transcription factor knockout library was grown overnight in synthetic complete media at 30C. Strains were diluted ~100-1000 fold in the morning and grown for ~4-6 hours in synthetic complete media. Strains were spun and washed 3x in YNB+20% glucose without amino acids to wash out any trace amino acids from the growth media. Strains were then serially diluted 10-fold in YNB+20% Glucose five times, and then plated as a dilution series on synthetic complete media, and synthetic complete media lacking all but one amino acid. Strain growth was monitored visually.

Growth Curves

Cells were grown overnight in synthetic complete media at 30C. The next morning, cells were diluted to a starting OD600 = .1 and grown for 6 hours until reaching mid-log (OD600 = 0.6-0.9). Cells were then washed 3x in YNB+Glucose, and then diluted to a starting OD600 = 0.02 in a 96 well plate in either complete media or media lacking a single amino acid. Cells were grown on a TECAN plate reader, with shaking, at 30C for 36hrs. Strains and conditions were done in triplicate. The growth rate was calculated by empirically determining the conversion between optical density and cell density by dilution series, plotting the growth curves on a semi-log plot and fitting the

linear part of the curve. Doubling times were calculated using the equation: $P = P_0 e^{\lambda t}$, where t = doubling time and λ = growth rate.

Bioinformatics

Promoters were extracted from genome sequences for each species published in (Shen et al. 2018), and taken to be 1000 base pairs (or the maximum number of bases if there were less than 1000 bases) upstream of the start codon for each gene annotated in that dataset.

The Met31 and Leu3 motifs are taken from the JASPAR database (Fornes et al. 2019) with identifiers MA0333.1 and MA0324.1 respectively. Both motifs were originally described in (Badis et al. 2008). The Lys14 motif was created using MEME to search for enriched binding.

The raw score for a given motif with respect to a given promoter is calculated using the pssm.calculate method for the Biopython motif object with the background base frequency distribution set empirically based on all promoters for the species, and with a pseudocount of 0.10% of the total counts used to define the motif. After calculating the raw score for all positions in the promoter, the maximum score was determined. The distribution of maximum raw scores for all promoters longer than 200bp for a given species was then calculated. This distribution was used to produce a normalized maximum score between 0 and 1 for a given motif with respect to a promoter using the empirical cumulative distribution function (python's statsmodels.distributions.empirical_distribution.ECDF function) on the maximum raw

score for that promoter. Where there were paralogs, the highest score for the paralogs is reported.

Figures

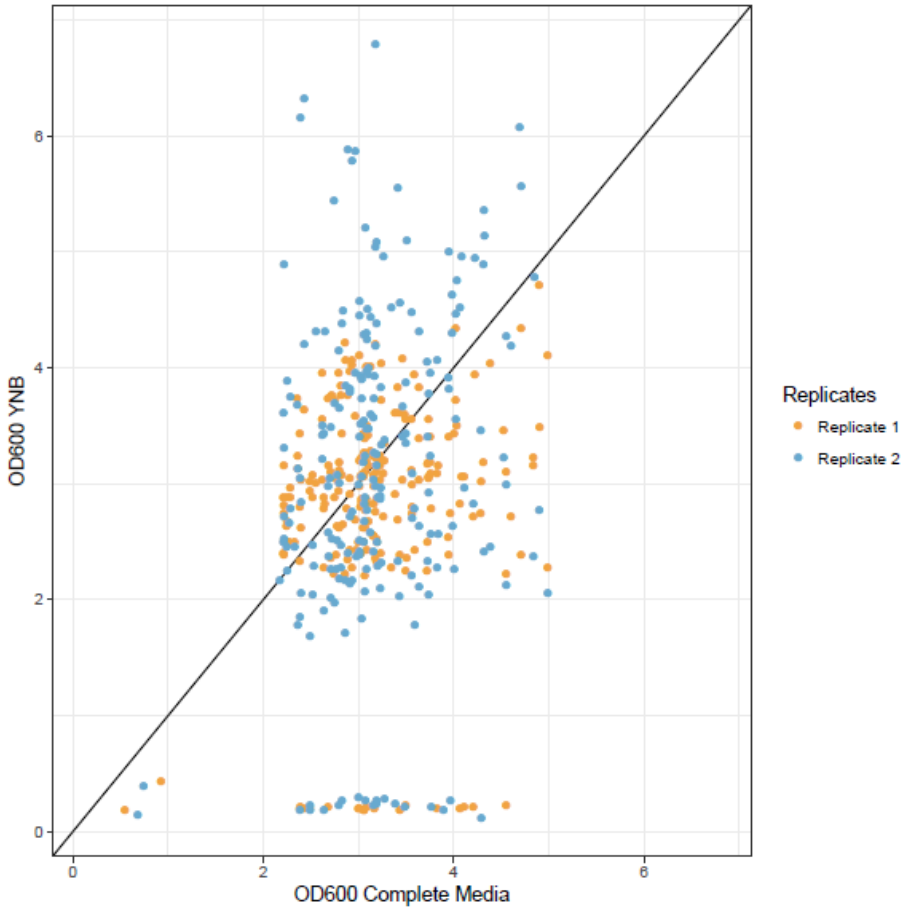


Figure 2. 1: Primary screen of C. albicans transcription factor knockout library to detect potential transcriptional regulators of amino acid biosynthetic pathways. Transcription factor knockout mutants were grown in complete media and in yeast nitrogen base lacking all amino acids except for arginine overnight at 37C. Each point represents one mutant strain, technical replicates are plotted in orange and blue. Mutant strains that grew in complete media but did not grow in yeast nitrogen base were carried into the secondary screen in which each amino acid was withdrawn individually.

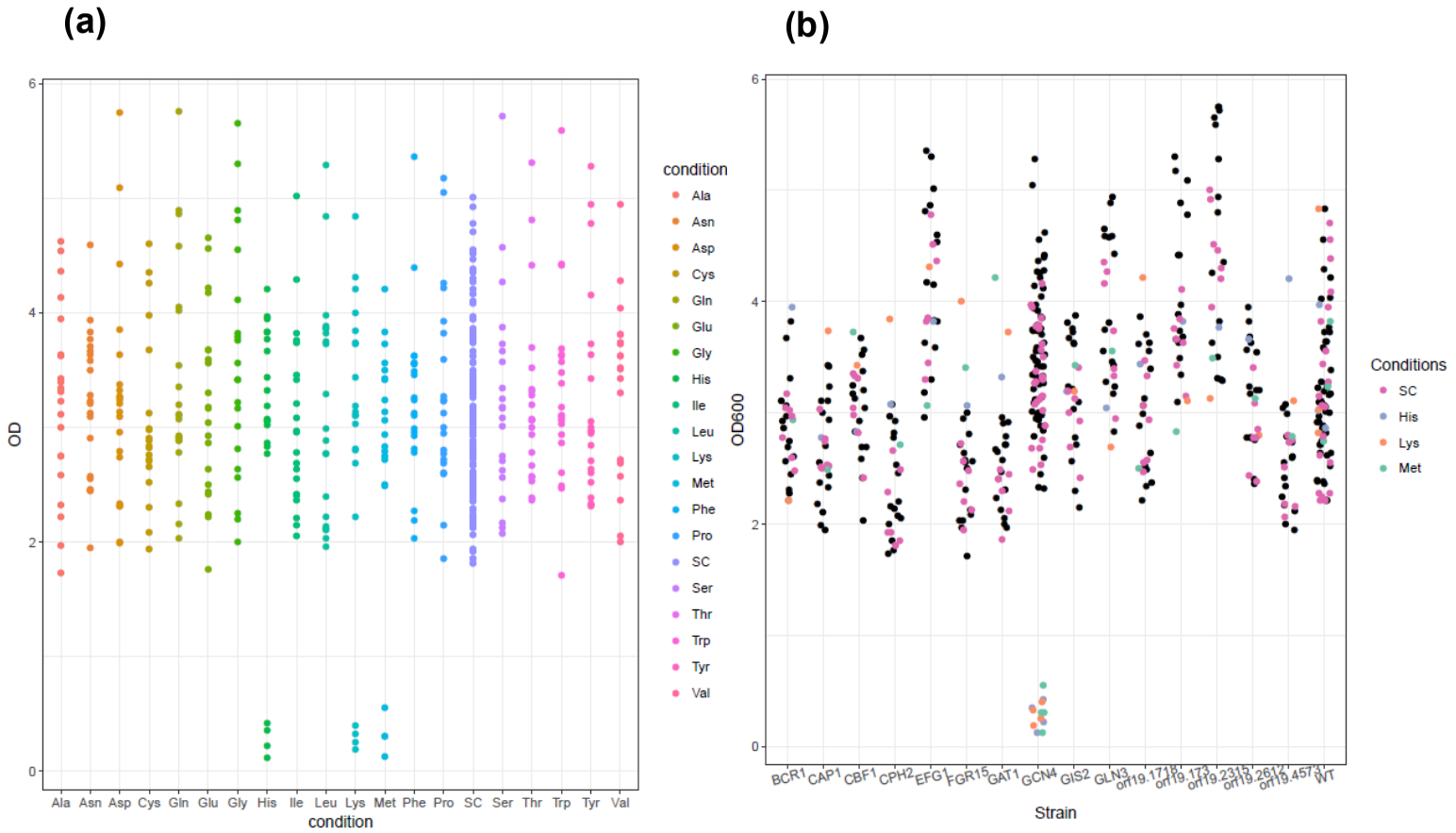


Figure 2. 2: Gcn4 is the only C. albicans transcription factor identified as a regulator of amino acid biosynthetic pathways. Mutant strains identified in the primary screen (Fig 1) were grown overnight in complete media and in media lacking just one amino acid. Media lacking arginine was omitted due to the Arg- background of the library. Gcn. The OD600 after overnight growth at 37C is plotted by (a) condition and (b) strain background.

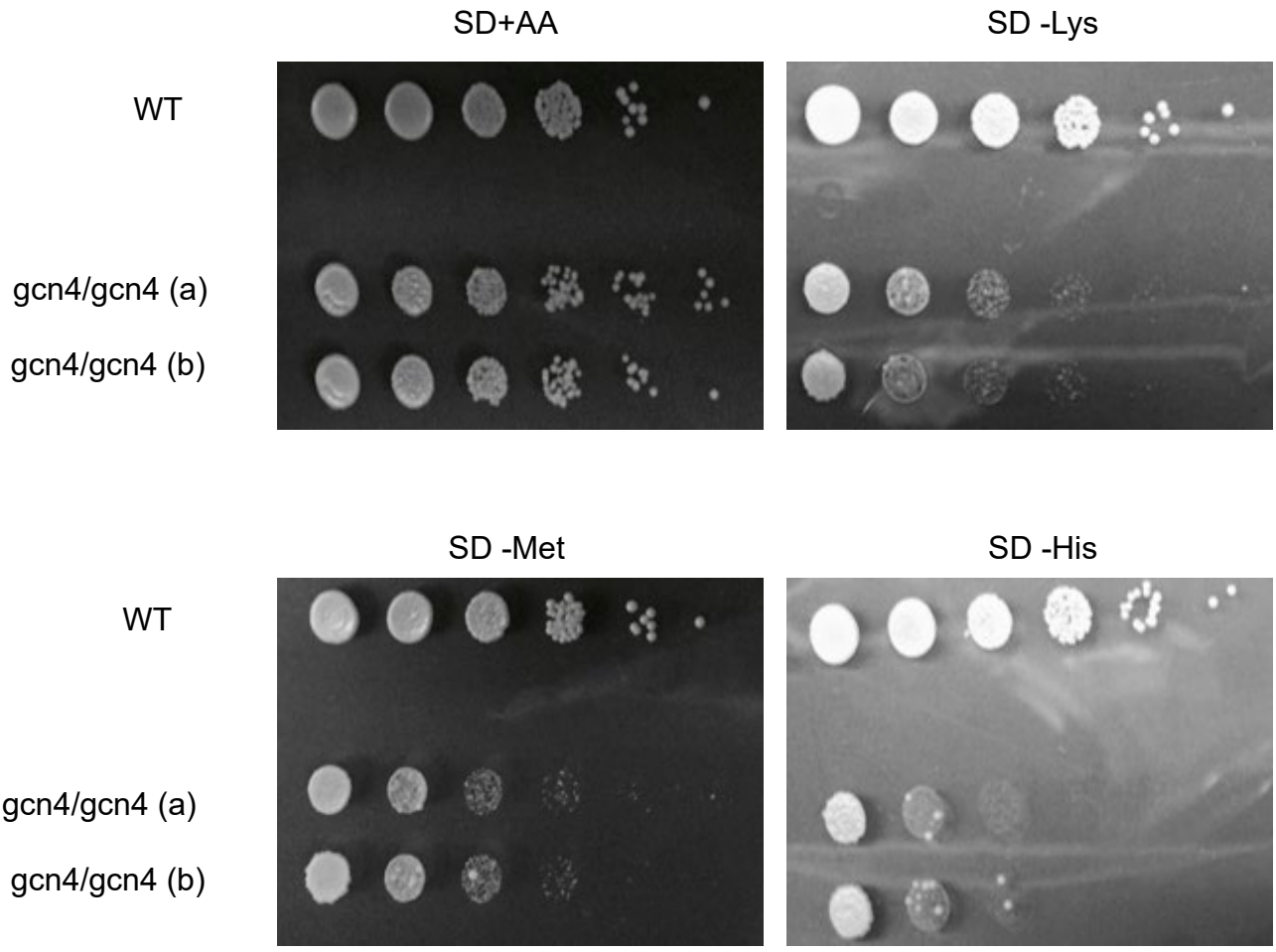
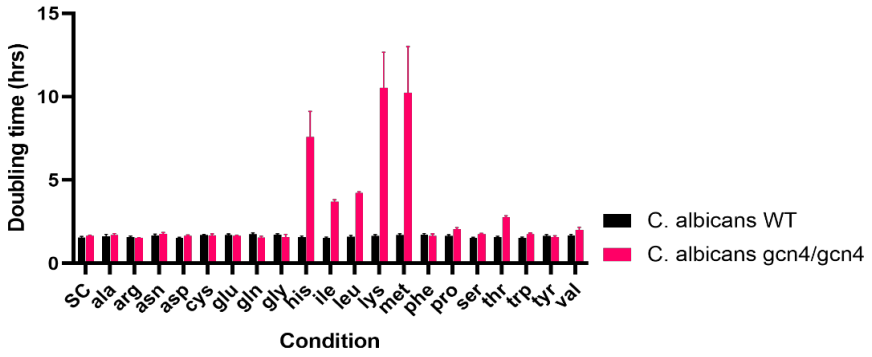


Figure 2. 3: Gcn4 is required for growth in media lacking Lysine, Methionine or Histidine in C. albicans. Strains were plated as a dilution series on complete media or media lacking one amino acid. These results support the results of the overnight liquid screens in Fig 1 & 2. No other mutant strain displayed such a severe and reproducible growth defect as Gcn4.

C.



S. cerevisiae

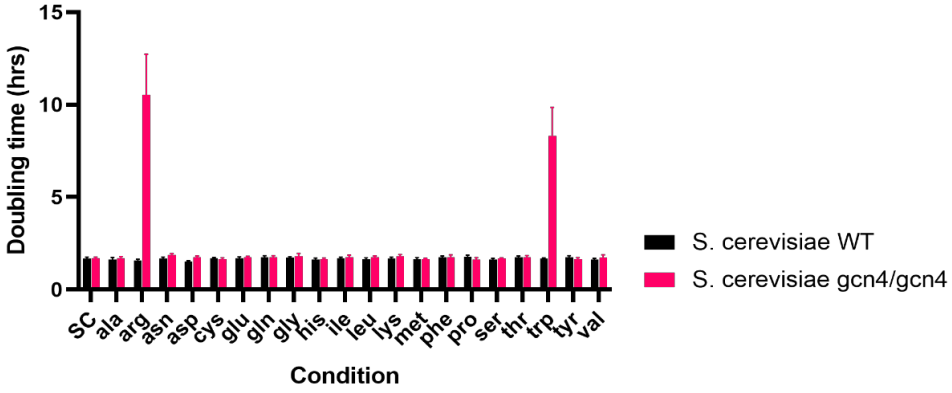


Figure 2. 4: **Species-specific requirements on Gcn4 for growth in media lacking single amino acids.** Prototrophic wildtype and gcn4-deletion strains were compared for both (a) C. albicans and (b) S. cerevisiae. Strains were grown for 24 hrs on a TECAN plate reader with constant shaking and temperature control (37C for C. albicans and 30C for S. cerevisiae). Doubling times were calculated by converting optical density to cell densities, plotting the data on a semi-log plot and calculating the slope of the exponential part of the curve. Doubling times are plotted for wildtype strains (black) and gcn4-deletion strains (pink). Error bars show the standard error of the mean for three replicates.

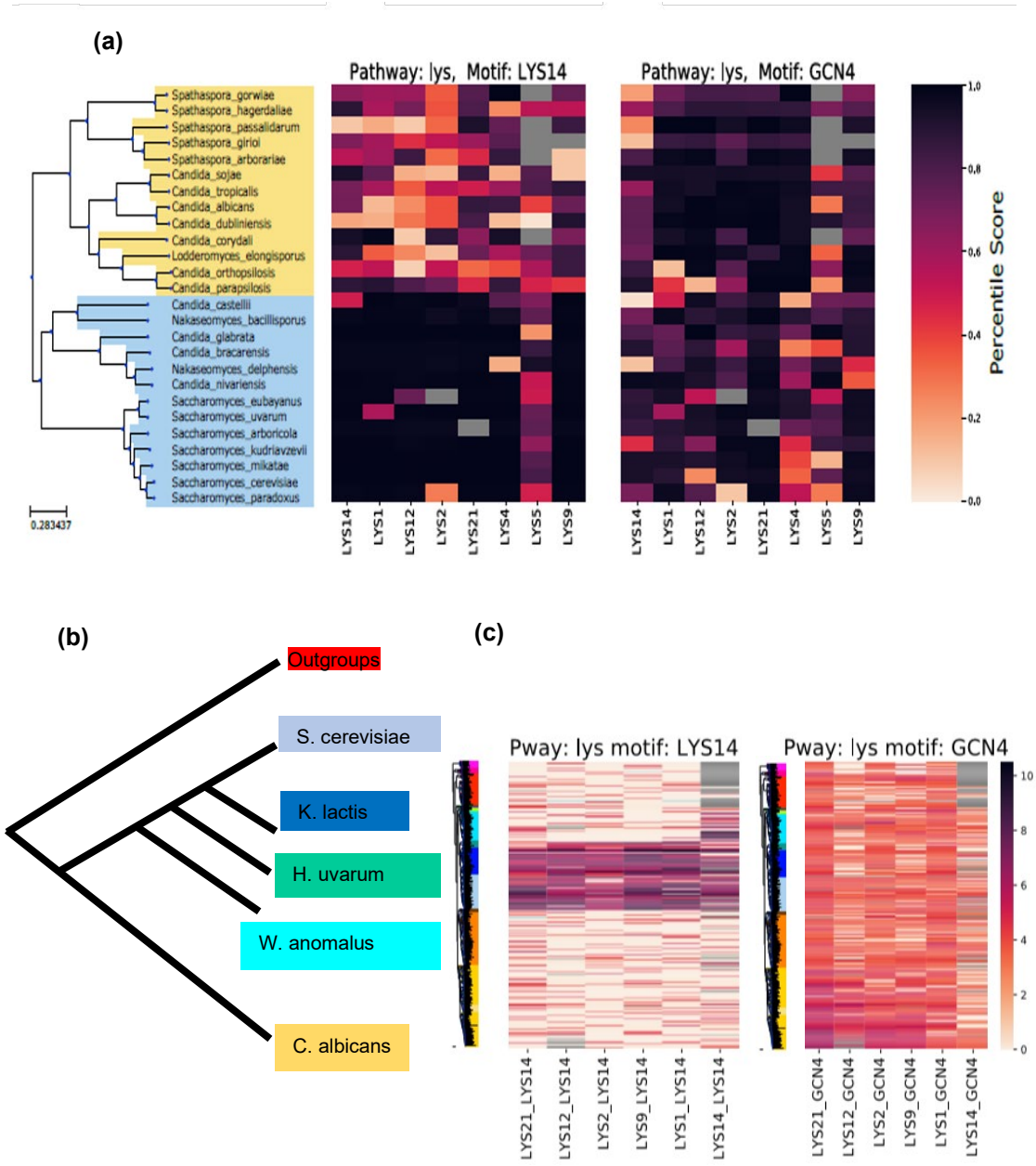


Figure 2. 5: Cis-regulatory evolution of the lysine biosynthetic pathway in ascomycetes a) Promoters of *S.cerevisiae* lysine pathway genes have strong scores for the LYS14 motif, but the promoters for *C. albicans* do not. Percentile of the max motif interaction score for the LYS14 motif and the Gcn4 motif for the promoter of each lysine pathway gene in each species based on the distribution of LYS14 binding scores for all promoters in each species. Grey shading indicates that there was no ortholog for the gene in the indicated species. b) Schematic of phylogenetic tree representative of all species in (Shen et al., 2018). c) Normalized motif scores for the lysine biosynthetic pathway for all species in (Shen et al., 2018)

References

Andi, Babak, Ann H. West, and Paul F. Cook. 2005. “Regulatory Mechanism of Histidine-Tagged Homocitrate Synthase from Saccharomyces Cerevisiae I. KINETIC STUDIES.” *Journal of Biological Chemistry* 280 (36): 31624–32. https://doi.org/10.1074/jbc.M502846200.

Askew, Christopher, Adnane Sellam, Elias Epp, Hervé Hogues, Alaka Mullick, André Nantel, and Malcolm Whiteway. 2009. “Transcriptional Regulation of Carbohydrate Metabolism in the Human Pathogen Candida Albicans.” PLoS Pathogens. Public Library of Science (PLoS). October 1, 2009. https://doi.org/10.1371/journal.ppat.1000612.

Badis, Gwenael, Esther T. Chan, Harm van Bakel, Lourdes Pena-Castillo, Desiree Tillo, Kyle Tsui, Clayton D. Carlson, et al. 2008. “A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters.” *Molecular Cell* 32 (6): 878–87. https://doi.org/10.1016/j.molcel.2008.11.020.

Becker, B., A. Feller, M. el Alami, E. Dubois, and A. Piérard. 1998. “A Nonameric Core Sequence Is Required Upstream of the LYS Genes of Saccharomyces Cerevisiae for Lys14p-Mediated Activation and Apparent Repression by Lysine.” *Molecular Microbiology* 29 (1): 151–63. https://doi.org/10.1046/j.1365-2958.1998.00916.x.

Blaiseau, Pierre-Louis, and Dominique Thomas. 1998. “Multiple Transcriptional Activation Complexes Tether the Yeast Activator Met4 to DNA.” *The EMBO Journal* 17 (21): 6327–36. https://doi.org/10.1093/emboj/17.21.6327.

Clifton, D., and D. G. Fraenkel. 1981. “The Gcr (Glycolysis Regulation) Mutation of Saccharomyces Cerevisiae.” *Journal of Biological Chemistry* 256 (24): 13074–78.

Clifton, Drago, Shelley B. Weinstock, and Dan G. Fraenkel. 1978. “Glycolysis Mutants in Saccharomyces Cerevisiae.” *Genetics* 88 (1): 1–11.

Dalal, Chiraj K, Ignacio A Zuleta, Kaitlin F Mitchell, David R Andes, Hana El-Samad, and Alexander D Johnson. 2016. “Transcriptional Rewiring over Evolutionary Timescales Changes Quantitative and Qualitative Properties of Gene Expression.” *ELife* 5 (September): e18981. https://doi.org/10.7554/eLife.18981.

Dever, T. E., W. Yang, S. Aström, A. S. Byström, and A. G. Hinnebusch. 1995. “Modulation of TRNA(IMet), EIF-2, and EIF-2B Expression Shows That GCN4 Translation Is Inversely Coupled to the Level of EIF-2.GTP.Met-TRNA(IMet) Ternary Complexes.” *Molecular and Cellular Biology* 15 (11): 6351–63. https://doi.org/10.1128/mcb.15.11.6351.

Dong, J., H. Qiu, M. Garcia-Barrio, J. Anderson, and A. G. Hinnebusch. 2000. “Uncharged TRNA Activates GCN2 by Displacing the Protein Kinase Moiety from a Bipartite TRNA-Binding Domain.” *Molecular Cell* 6 (2): 269–79. https://doi.org/10.1016/s1097-2765(00)00028-9.

Duncan, Caia D. S., María Rodríguez-López, Phil Ruis, Jürg Bähler, and Juan Mata. 2018. “General Amino Acid Control in Fission Yeast Is Regulated by a Nonconserved Transcription Factor, with Functions Analogous to Gcn4/Atf4.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (8): E1829–38. https://doi.org/10.1073/pnas.1713991115.

El Alami, Mohamed, André Feller, André Piérard, and Evelyne Dubois. 2002. “The Proper Folding of a Long C-Terminal Segment of the Yeast Lys14p Regulator Is Required for Activation of LYS Genes in Response to the Metabolic Effector.” *Molecular Microbiology* 43 (6): 1629–39. https://doi.org/10.1046/j.1365-2958.2002.02854.x.

Fornes, Oriol, Jaime A. Castro-Mondragon, Aziz Khan, Robin van der Lee, Xi Zhang, Phillip A. Richmond, Bhavi P. Modi, et al. 2019. “JASPAR 2020: Update of the Open-Access Database of Transcription Factor Binding Profiles.” *Nucleic Acids Research*, November. https://doi.org/10.1093/nar/gkz1001.

Giniger, E., S. M. Varnum, and M. Ptashne. 1985. “Specific DNA Binding of GAL4, a Positive Regulatory Protein of Yeast.” *Cell* 40 (4): 767–74. https://doi.org/10.1016/0092-8674(85)90336-8.

Hinnebusch, A G. 1988. “Mechanisms of Gene Regulation in the General Control of Amino Acid Biosynthesis in Saccharomyces Cerevisiae.” *Microbiological Reviews* 52 (2): 248–73.

Hinnebusch, Alan G. 2005. “Translational Regulation of GCN4 and the General Amino Acid Control of Yeast.” *Annual Review of Microbiology* 59 (January): 407–50. https://doi.org/10.1146/annurev.micro.59.031805.133833.

Hinnebusch, Alan G, and Krishnamurthy Natarajan. 2002. “Gcn4p, a Master Regulator of Gene Expression, Is Controlled at Multiple Levels by Diverse Signals of Starvation and Stress.” *Eukaryotic Cell* 1 (1): 22–32.

Homann, Oliver R., Jeanselle Dea, Suzanne M. Noble, and Alexander D. Johnson. 2009. “A Phenotypic Profile of the Candida Albicans Regulatory Network.” *PLoS Genetics* 5 (12): e1000783. https://doi.org/10.1371/journal.pgen.1000783.

Johnson, Alexander D. 2017. “The Rewiring of Transcription Circuits in Evolution.” *Current Opinion in Genetics & Development*, Evolutionary genetics, 47 (December): 121–27. https://doi.org/10.1016/j.gde.2017.09.004.

Kondrashov, Fyodor A., Igor B. Rogozin, Yuri I. Wolf, and Eugene V. Koonin. 2002. “Selection in the Evolution of Gene Duplications.” *Genome Biology* 3 (2): RESEARCH0008. https://doi.org/10.1186/gb-2002-3-2-research0008.

Kuras, Laurent, Régine Barbey, and Dominique Thomas. 1997. “Assembly of a BZIP–BHLH Transcription Activation Complex: Formation of the Yeast Cbf1–Met4–Met28 Complex Is Regulated through Met28 Stimulation of Cbf1 DNA Binding.” *The EMBO Journal* 16 (9): 2441–51. https://doi.org/10.1093/emboj/16.9.2441.

Kuras, Laurent, Astrid Rouillon, Traci Lee, Regine Barbey, Mike Tyers, and Dominique Thomas. 2002. “Dual Regulation of the Met4 Transcription Factor by Ubiquitin-Dependent Degradation and Inhibition of Promoter Recruitment.” *Molecular Cell* 10 (1): 69–80. https://doi.org/10.1016/s1097-2765(02)00561-0.

Lee, Traci A., Paul Jorgensen, Andrew L. Bognar, Caroline Peyraud, Dominique Thomas, and Mike Tyers. 2009. “Dissection of Combinatorial Control by the Met4 Transcriptional Complex.” *Molecular Biology of the Cell* 21 (3): 456–69. https://doi.org/10.1091/mbc.e09-05-0420.

Leroy, Christophe, Laëtitia Cormier, and Laurent Kuras. 2006. “Independent Recruitment of Mediator and SAGA by the Activator Met4.” *Molecular and*

Cellular Biology 26 (8): 3149–63. https://doi.org/10.1128/MCB.26.8.3149-3163.2006.

Li, Hao, and Alexander D. Johnson. 2010. “Evolution of Transcription Networks — Lessons from Yeasts.” *Current Biology* 20 (17): R746–53. https://doi.org/10.1016/j.cub.2010.06.056.

Lynch, M., and J. S. Conery. 2000. “The Evolutionary Fate and Consequences of Duplicate Genes.” *Science (New York, N.Y.)* 290 (5494): 1151–55. https://doi.org/10.1126/science.290.5494.1151.

Mueller, P. P., and A. G. Hinnebusch. 1986. “Multiple Upstream AUG Codons Mediate Translational Control of GCN4.” *Cell* 45 (2): 201–7. https://doi.org/10.1016/0092-8674(86)90384-3.

Natarajan, K, M R Meyer, B M Jackson, D Slade, C Roberts, A G Hinnebusch, and M J Marton. 2001. “Transcriptional Profiling Shows That Gcn4p Is a Master Regulator of Gene Expression during Amino Acid Starvation in Yeast.” *Molecular and Cellular Biology* 21 (13): 4347–68. https://doi.org/10.1128/MCB.21.13.4347-4368.2001.

Pérez, J. Christian, Polly M. Fordyce, Matthew B. Lohse, Victor Hanson-Smith, Joseph L. DeRisi, and Alexander D. Johnson. 2014. “How Duplicated Transcription Regulators Can Diversify to Govern the Expression of Nonoverlapping Sets of Genes.” *Genes & Development* 28 (12): 1272–77. https://doi.org/10.1101/gad.242271.114.

Pérez, J. Christian, Carol A. Kumamoto, and Alexander D. Johnson. 2013. “Candida Albicans Commensalism and Pathogenicity Are Intertwined Traits Directed by a

Tightly Knit Transcriptional Regulatory Circuit." *PLoS Biology* 11 (3): e1001510. https://doi.org/10.1371/journal.pbio.1001510.

Priyadarshini, Yumnam, and Krishnamurthy Natarajan. 2016. "Reconfiguration of Transcriptional Control of Lysine Biosynthesis in Candida Albicans Involves a Central Role for the Gcn4 Transcriptional Activator." *MSphere* 1 (1). https://doi.org/10.1128/mSphere.00016-15.

Shen, Xing-Xing, Dana A. Opulente, Jacek Kominek, Xiaofan Zhou, Jacob L. Steenwyk, Kelly V. Buh, Max A. B. Haase, et al. 2018. "Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum." *Cell* 175 (6): 1533-1545.e20. https://doi.org/10.1016/j.cell.2018.10.023.

Thomas, D., and Y. Surdin-Kerjan. 1997. "Metabolism of Sulfur Amino Acids in Saccharomyces Cerevisiae." *Microbiology and Molecular Biology Reviews* 61 (4): 503–32.

Tian, Chaoguang, Takao Kasuga, Matthew S Sachs, and N Louise Glass. 2007. "Transcriptional Profiling of Cross Pathway Control in Neurospora Crassa and Comparative Analysis of the Gcn4 and CPC1 Regulons." *Eukaryotic Cell* 6 (6): 1018–29. https://doi.org/10.1128/EC.00078-07.

Tournu, Hélène, Gyanendra Tripathi, Gwyneth Bertram, Susan Macaskill, Abigail Mavor, Louise Walker, Frank C Odds, Neil A R Gow, and Alistair J P Brown. 2005. "Global Role of the Protein Kinase Gcn2 in the Human Pathogen Candida Albicans." *Eukaryotic Cell* 4 (10): 1687–96. https://doi.org/10.1128/EC.4.10.1687-1696.2005.

Traven, Ana, Branka Jelicic, and Mary Sopta. 2006. “Yeast Gal4: A Transcriptional Paradigm Revisited.” *EMBO Reports* 7 (5): 496–99. https://doi.org/10.1038/sj.embor.7400679.

Uemura, HIROSHI, and D. G. Fraenkel. 1990. “Gcr2, a New Mutation Affecting Glycolytic Gene Expression in Saccharomyces Cerevisiae.” *Molecular and Cellular Biology* 10 (12): 6389–96.

Wek, R. C., B. M. Jackson, and A. G. Hinnebusch. 1989. “Juxtaposition of Domains Homologous to Protein Kinases and Histidyl-TRNA Synthetases in GCN2 Protein Suggests a Mechanism for Coupling GCN4 Expression to Amino Acid Availability.” *Proceedings of the National Academy of Sciences* 86 (12): 4579–83. https://doi.org/10.1073/pnas.86.12.4579.

Xu, Hengyu, Babak Andi, Jinghua Qian, Ann H. West, and Paul F. Cook. 2006. “The Alpha-Aminoadipate Pathway for Lysine Biosynthesis in Fungi.” *Cell Biochemistry and Biophysics* 46 (1): 43–64. https://doi.org/10.1385/CBB:46:1:43.

Zaborske, John M., Jana Narasimhan, Li Jiang, Sheree A. Wek, Kimberly A. Dittmar, Florien Freimoser, Tao Pan, and Ronald C. Wek. 2009. “Genome-Wide Analysis of TRNA Charging and Activation of the EIF2 Kinase Gcn2p.” *Journal of Biological Chemistry* 284 (37): 25254–67. https://doi.org/10.1074/jbc.M109.000877.

Zaborske, John M, Xiaochen Wu, Ronald C Wek, and Tao Pan. 2010. “Selective Control of Amino Acid Metabolism by the GCN2 EIF2 Kinase Pathway in Saccharomyces Cerevisiae.” *BMC Biochemistry* 11 (January): 29. https://doi.org/10.1186/1471-2091-11-29.

Chapter 3

Conserved function, diverged specificity: How does a DNA binding protein diversify binding specificity?

Introduction

The evolution of transcription networks – that is changes to the connections between transcription factors and their given target genes – is a well-established source of evolutionary novelty (Carroll, 2000; Johnson, 2017; Li & Johnson, 2010; Peter & Davidson, 2011; Prud'homme et al., 2007). Changes to these connections can occur via changes to regulatory sequences of DNA via the loss or gain of transcription factor binding sites (cis-regulatory evolution), or it can occur via changes to the transcription factor itself via mutations to its DNA binding domain or cofactor interacting region (trans regulatory evolution). While arguments have been made that trans-regulatory evolution are more infrequent compared to cis-regulatory evolution – usually based on the notion that mutations to DNA binding proteins could have deleterious effects on numerous downstream target genes -- numerous examples of trans-regulatory evolution contributing to novel phenotypes have been documented in the literature (Britton et al., 2020; Jarvela & Hinman, 2015; Lynch & Wagner, 2008). These examples demonstrate that protein evolution plays a key role in the rewiring of transcription networks and the subsequent phenotypic changes that arise due to these rewiring events.

While it is apparent that protein evolution contributes to the evolution of transcription networks, the questions surrounding transcription factor evolution are still valid: how does a DNA binding protein change its DNA binding specificity without either losing its ancestral function or causing off-target deleterious effects? While gene duplication followed by neofunctionalization is one well-accepted answer to this question (that is, a gene duplicates and one copy maintains the ancestral function while the second

copy is ‘free to explore’ new functions), the answer is not quite as clear when dealing with a gene that has not appeared to have undergone duplication.

One close examination of this scenario used the plant transcription factor LEAFY as an example (Sayou et al., 2014). Sayou et al. identified three different DNA binding specificities in LEAFY proteins across different plant species: a type I specificity in the majority of land plants, type II specificity in mosses, and type III in algae and hornworts. They inferred the amino acid sequence of the last common ancestor of these three groups and found that this ancestral LEAFY demonstrated promiscuous binding; it was able to bind all three of the modern-day DNA binding motifs. This suggests that one mechanism by which a sequence-specific DNA binding protein can alter its specificity without losing its ancestral function is via evolutionary intermediates that can bind to the ancestral target genes while simultaneously gaining new target genes. Presumably these promiscuous intermediates eventually resolve by “handing off” one set of these target genes to a different regulator or by simply expanding the size of the transcription network.

While fully disproving the occurrence of a gene duplication event is challenging, and was a subject of some debate in the case of LEAFY (Brockington et al., 2015; Brunkard et al., 2015), independent of whether a gene duplication did or did not occur, the target genes of LEAFY in the modern species are not fully defined and so it is also unclear the extent to which these changes in DNA binding specificity altered downstream function. On the other hand, in the case of MATα1, a regulator of yeast mating type in ascomycete yeasts, we know that the function and downstream target genes have been conserved (Tsong et al., 2006), despite an apparent change in DNA binding specificity (Baker et al., 2011).

Ascomyete mating type is controlled by mating-type specific transcription factors. These set of mating-type specific transcription factors are differential expressed across two cell types: ‘a’ cells and ‘α’ cells. While the molecular mechanism and the transcription factors that defines one cell type (the ‘a’ cell type) has been shown to evolve (Baker et al., 2012; Britton et al., 2020), the mechanism by which alpha cells are defined is conserved. For an α-cell to be an α-cell, MATα1, together with its cofactor MCM1, activate the α-specific genes. These core alpha-specific genes are the same over the hundreds of millions of years that separate S. cerevisiae and C. albicans: the a-factor receptor (encoded by Ste3), the alpha-factor pheromone (encoded by MFα1 and MFα2), and the alpha-specific agglutinin (encoded by AGα1) (Tsong et al., 2006).

Despite this conservation of function, Baker et al. demonstrate biochemically that the S. cerevisiae MATα1 and the C. albicans MATα1 protein have divergent DNA binding specificity – the S. cerevisiae protein cannot bind to or activate transcription off the C. albicans cis-regulatory site and the same is true in the converse. This, coupled with the apparent lack of gene duplication events between the two species – in no extant species is there more than one copy of MATα1 – presents an interesting case study for understanding how a transcription factor retains its function while diversifying its binding specificity. Did MATα1 evolve via promiscuous intermediates as has been argued in the case of LEAFY? Or, instead, do we see a distinct specificity for ancestral MATα1 proteins that evolved in a stepwise pattern?

With this in mind, we set out to attempt to answer this question. Ultimately, the project was left behind in the face of technical challenges, however it is useful to revisit the planned approach and to consider alternative approaches in the future. The original

plan was to utilize ancestral protein reconstruction (Thornton, 2004) to infer the sequences of ancestral MATα1 protein. At minimum, the ancestral MATalpha1 proteins of interest are: the last common ancestor of S. cerevisiae and C. albicans, the last common ancestor of the C. albicans CTG clade, and the last common ancestor of the S. cerevisiae clade (Fig 1). Other ancestral nodes could also be interrogated, especially now that we have many more sequenced genomes than when this project originally began, but these three are essential to understanding the trajectory of DNA binding diversification. Ancestral proteins would be recombinantly expressed and purified and, along with extant MATα1 proteins from S. cerevisiae and C. albicans, would be tested for DNA binding affinity and specificity. We originally proposed to use mechanically induced trapping of molecular interactions (MITOMI) to measure the affinities and specificities of every possible permutation of the MATα1 binding site (Le et al., 2018; Rockel et al., 2012) in conjunction with x-ray crystallography of both extant and ancestral MATα1 proteins on their preferred cis-regulatory sites. Conceptually, our approach was nearly identical to the LEAFY study mentioned previously. These methodologies combined would give a comprehensive view of the trajectory by which a DNA binding protein could diversify its binding specificity without compromising its regulatory function.

Results

Alpha1 ancestral protein reconstructions

Since a central component of this project was to use ancestral protein reconstruction and resurrection to map the trajectory by which MATalpha1 has diversified its binding specificity, we wanted to verify that ancestors could be reconstructed with

confidence. To this end, MATalpha1 orthologs were identified from 44 ascomycota spanning the evolutionary distance from S. cerevisiae to C. albicans and using Y. lipolytica, N. crassa, H. capsulatum, and A. terreus as outgroup species. These orthologs were run through the PhyloBot software and ancestors were reconstructed. Predictably, the statistical support for the DNA-binding domain reconstructions were much higher than the full-length proteins, and this statistical support gave us license to proceed as we believed that the ancestors at the nodes of interest – that is going as far back as the last common MATalpha1 ancestor of S. cerevisiae and C. albicans – could be reconstructed with enough confidence (Fig 1).

Alpha1 Expression and Purification

Since relatively little biochemistry has been done with S. cerevisiae alpha1 (and none with C. albicans), it was unclear which portion of the protein would express, purify, and bind DNA effectively. With this in mind, we decided to make multiple expression constructs. We input the MATalpha1 primary protein sequences in the Protein Homology/analogy Recognition Engine (PHYRE) server to predict regions of secondary structure. We then designed primers to span every possible combination of alpha-coils that also include the DNA binding domain (Fig 2a). In total, 27 constructs spanning from full-length protein to DNA binding domain alone were PCR amplified (Fig 2b) and cloned into pET-28 – a bacterial expression vector with a T7lac promoter and a cleavable N-terminal His tag. These constructs were transformed into BL-21 competent E. coli and tested for recombinant protein expression.

As a first pass, E. coli were grown to an OD600 of .4 at 37C before addition of IPTG (a lactose analog that binds the lac repressor thereby allowing for expression of the

construct). Cells were exposed to .4mM IPTG for 4 hours at 37C, then harvested and cell lysates were run on polyacrylamide gels. Nearly all of the constructs for both S. cerevisiae and C. albicans alpha1 protein demonstrated strong expression (Fig 2c), and so a handful of constructs were tested for purification. Following nickel column purification, it became apparent that all of the protein constructs were insoluble and were lost in the insoluble fraction prior to purification.

In an attempt to overcome the insolubility problem, we tried a variety of parameters that might lower the expression and subsequent insolubility of the recombinant proteins: lower induction temperature, lower concentration of inducer, and shorter length of induction period (separately and in combinations). Unfortunately, both the expression and insolubility of these constructs were robust to all conditions tested (Fig 3). Therefore, we decided to take a denaturation and re-folding approach using urea as a denaturing agent.

Following exposure of cell lysates to 8M urea and purification on an Ni-NTA column, the samples were dialysed into a refolding buffer containing 500 mM arginine. The samples were run on an HPLC size exclusion column to verify that the refolding did not result in re-aggregation. Upon obtaining purified S. cerevisiae and C. albicans alpha1 protein, the next step was to functionally characterize DNA binding using gel shift assays.

Gel Shifts

We took the constructs that were successfully purified and used them in a DNA binding gel shift assay. Here, again, we ran into significant technical difficulties. Upon using annealed, synthetic DNA oligos and MATalpha1 protein constructs alone, we saw no evidence of DNA binding, despite varying the binding buffers ionic strength, pH, and

composition multiple times. We hypothesized that one of two factors may be contributing to the lack of DNA binding: 1. The presence of unannealed single stranded DNA interfering with DNA binding and/or 2. The requirement of Mcm1, MATalpha1's known cooperative co-factor. We tested the first idea by cloning the DNA binding oligonucleotides into a vector, purifying the vector, excising the double stranded DNA binding site, and gel-purifying the excised binding site. These excised binding sites were radiolabeled and tested in the gel shift, however, this did not appear to resolve the issue. We tested the second idea (the requirement of Mcm1), by using purified S. cerevisiae Mcm1 at a set concentration along with the alpha1 dilution series. However, none of these combinations improved DNA binding.

From these attempts we concluded that while the re-folding technique resulted in a non-aggregated protein, that perhaps the re-folded alpha1 constructs were not folded back into an active form and they lost their DNA binding and/or Mcm1-interacting abilities. While other biochemical and non-biochemical approaches may be valuable for future exploration (see future directions section), at this point, the project was put aside in pursuit of other projects.

Discussion & future directions

The technical problems encountered here, while challenging, are by no means insurmountable. The following section outlines potential approaches one could take in attempt to answer the question of how MATalpha1 diversified its binding specificity.

Firstly, to pursue the same biochemical approach, it is necessary to overcome the protein insolubility displayed by the His-tag, bacterially expressed recombinant proteins. The first avenue worth exploring is the usage of a solubility tag like maltose binding protein (MBP), for example. MBP is believed to interact with hydrophobic amino acid residues present in unfolded proteins to prevent aggregation or proteolysis. While MBP is one of the most frequently used solubility tags, others exist and could also be tried. Changing the tag would require cloning new expression constructs and also optimizing purification conditions. For MBP, the pMAL series of expression vectors, such as pMAL-c6T, are readily available through NEB (or can very likely be gifted from another lab) and are designed to produce cleavable MBP-His fusion tags in the cytoplasm. The fusion tag allows for a two-step purification protocol using amylose resin elution, TEV protease cleavage, and nickel column purification which could yield a highly pure end-product. The pMAL series of plasmids are still IPTG-inducible E. coli expression constructs, so the induction protocol and conditions could remain the same as have already been to yield expression of recombinant protein.

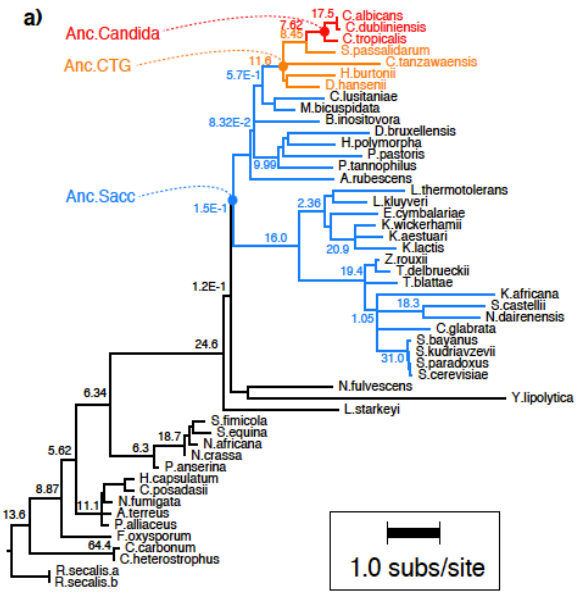
The second parameter to vary with an eye towards keeping the same biochemical approach is to utilize a eukaryotic expression system such as S. cerevisiae or Pichia pastoris. Expressing a yeast protein in a yeast cell may increase proper protein folding due to the presence of chaperones and/or necessary post-translational modifications thereby resulting in more soluble and more biochemically active recombinant protein. Obviously, moving to a eukaryotic system would require new expression construct design and new induction protocols, although a His-tag and subsequent nickel column purification could still be used. Yeast expression systems often use galactose-inducible

promoters, although other options like phosphate- or copper-inducible promoters can also be employed. The pYES series of expression vectors are commonly used for recombinant protein expression in S. cerevisiae and come with a variety of His or His-V5 antibody fusion tags and are a galactose-inducible system.

The biochemical approach was favored at the outset because we believed that it would be informative to do structural analysis of extant and ancestral MATalpha1 proteins bound to their preferred cis-regulatory sequences. However, in addition or instead of a biochemical approach, an in vitro cell-based fluorescent reporter assay could be employed to address by which trajectory MATalpha1 diversified its binding specificity. A pool of reporter constructs could be designed such that every possible permutation of the MATalpha1 cis-regulatory site (with an adjacent Mcm1-site) was inserted upstream of a GFP reporter. This reporter strategy has been employed numerous times by members of the Johnson lab. These reporters would be cloned into a-cells which have been transformed with a MATalpha1 version of interest. For example, if S. cerevisiae MATalpha1 was cloned into S. cerevisiae a-cells which also had the S. cerevisiae MATalpha1 wildtype reporter, we would expect this strain to express high levels of GFP. We would expect a similar result for C. albicans MATalpha1 matched with its wildtype cis-regulatory reporter. We would quantitatively assess levels of GFP expression across all the reporter constructs in the modern species to measure how specific or promiscuous the modern day MATalpha1 proteins are (promiscuous DNA binding protein would result of GFP expression across many iterations of the cis-regulatory site, while a specific binder will only activate GFP off a narrow number of similar DNA-sequences). At the very least, we expect that the S. cerevisiae MATalpha1 would not activate transcription in the

presence of the C. albicans MATalpha1 binding site and vice versa (Baker 2011). In the case of the ancestral MATalpha1 proteins, we would be interested to see which permutations of the MATalpha1 binding site lead to transcriptional activation and whether the trajectory from the last common MATalpha1 ancestor of S. cerevisiae and C. albicans to the modern day proteins involved many promiscuous intermediates or a specific step-wise changes in DNA binding specificity.

Figures:



Full Length

Ancestor	mean P	% sites, P > 0.8	% sites, P > 0.9
Anc.Sacc	0.476	24.9%	20.5%
Anc.CTG	0.603	37.6%	34.3%
Anc.Candida	0.805	68.4%	60.6%

Alpha box

Ancestor	mean P	% sites, P > 0.8	% sites, P > 0.9
Anc.Sacc	0.729	48.2%	48.2%
Anc.CTG	0.858	72.7%	69.1%
Anc.Candida	0.950	93.2%	88.1%

Figure 3. 1: Ancestral protein reconstructions. MATalpha1 amino acid sequences were found using a combination of BLAST and synteny gene order browsers (YGOB, CGOB). A phylogenetic tree was constructed based on these MATalpha1 sequences, and the sequences were run through the PHYLOBOT pipeline to generate ancestral MATalpha1 sequences.

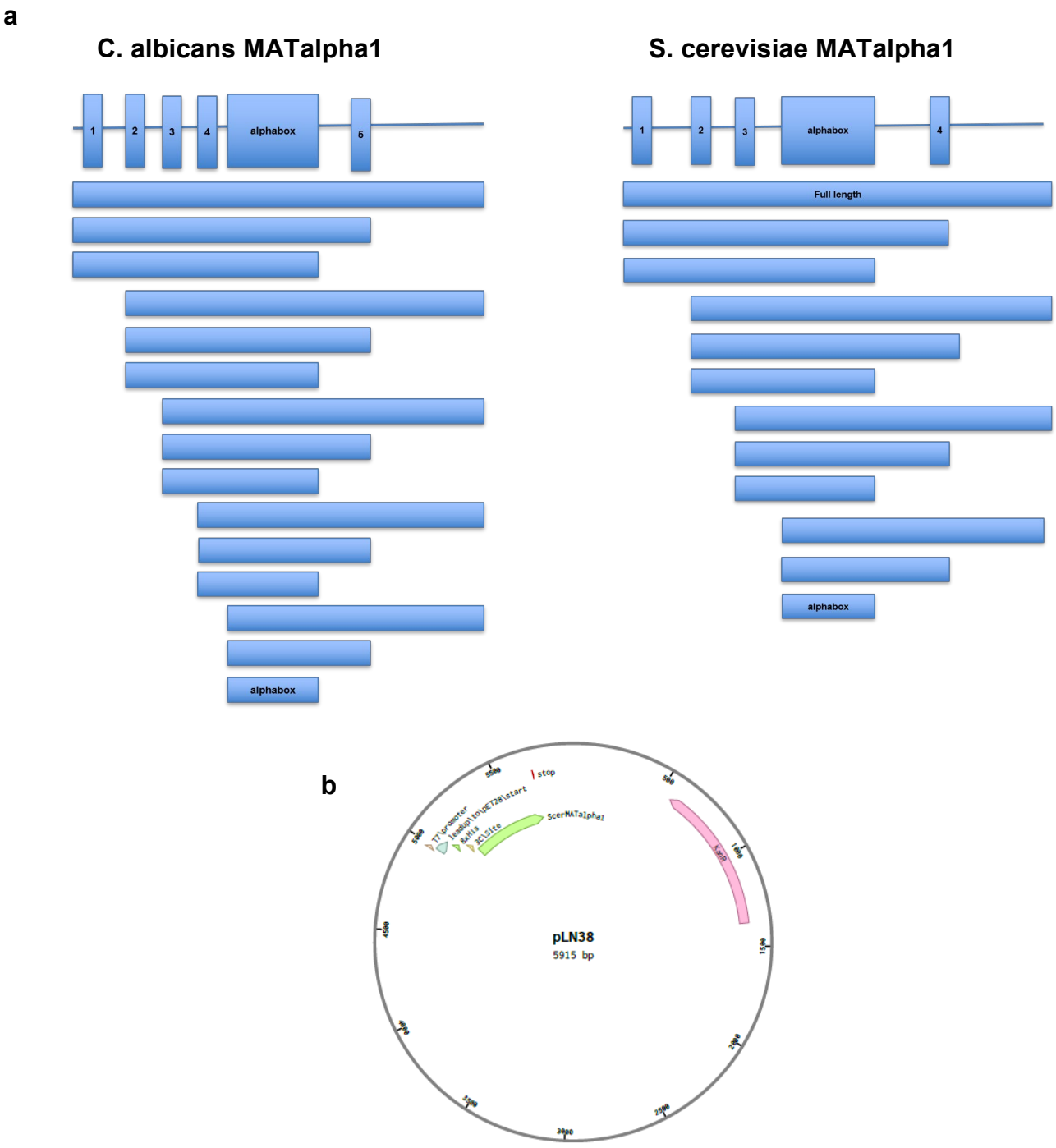


Figure 3. 2. MATalpha1 expression constructs schematics & plasmid map.

a) A variety of MATalpha1 constructs from both C. albicans and S. cerevisiae were PCR amplified using predicted secondary structure as the boundaries of the constructs. b) a plasmid map for the full-length S. cerevisiae MATalpha1 expression construct. The constructs depicted in 2a were cloned into the pLIC-H3 expression plasmid backbone. In this example, the green bar is the full-length S. cerevisiae MATalpha1.

References

Baker, C. R., Booth, L. N., Sorrells, T. R., & Johnson, A. D. (2012). Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell*, *151*(1), 80–95. https://doi.org/10.1016/j.cell.2012.08.018

Baker, C. R., Tuch, B. B., & Johnson, A. D. (2011). Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(18), 7493–7498. https://doi.org/10.1073/pnas.1019177108

Britton, C. S., Sorrells, T. R., & Johnson, A. D. (2020). Protein-coding changes preceded cis-regulatory gains in a newly evolved transcription circuit. *Science (New York, N.Y.)*, *367*(6473), 96–100. https://doi.org/10.1126/science.aax5217

Brockington, S. F., Moyroud, E., Sayou, C., Monniaux, M., Nanao, M. H., Thévenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y., Wong, G. K.-S., Weigel, D., Dumas, R., & Parcy, F. (2015). Response to Comment on “A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity.” *Science*, *347*(6222), 621–621. https://doi.org/10.1126/science.1256011

Brunkard, J. O., Runkel, A. M., & Zambryski, P. C. (2015). Comment on “A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity.” *Science*, *347*(6222), 621–621. https://doi.org/10.1126/science.1255437

Carroll, S. B. (2000). Endless forms: The evolution of gene regulation and morphological diversity. *Cell*, *101*(6), 577–580. https://doi.org/10.1016/s0092-8674(00)80868-5

Cheatle Jarvela, A. M., & Hinman, V. F. (2015). Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. *EvoDevo*, *6*(1), 3. https://doi.org/10.1186/2041-9139-6-3

Johnson, A. D. (2017). The rewiring of transcription circuits in evolution. *Current Opinion in Genetics & Development*, *47*, 121–127. https://doi.org/10.1016/j.gde.2017.09.004

Le, D. D., Shimko, T. C., Aditham, A. K., Keys, A. M., Longwell, S. A., Orenstein, Y., & Fordyce, P. M. (2018). Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(16), E3702–E3711. https://doi.org/10.1073/pnas.1715888115

Li, H., & Johnson, A. D. (2010). Evolution of Transcription Networks—Lessons from Yeasts. *Current Biology*, *20*(17), R746–R753. https://doi.org/10.1016/j.cub.2010.06.056

Lynch, V. J., & Wagner, G. P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution; International Journal of Organic Evolution*, *62*(9), 2131–2154. https://doi.org/10.1111/j.1558-5646.2008.00440.x

Peter, I. S., & Davidson, E. H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell*, *144*(6), 970–985. https://doi.org/10.1016/j.cell.2011.02.017

Prud'homme, B., Gompel, N., & Carroll, S. B. (2007). Emerging principles of regulatory evolution. *Proceedings of the National Academy of Sciences*, *104*(suppl 1), 8605–8612. https://doi.org/10.1073/pnas.0700488104

Rockel, S., Geertz, M., & Maerkl, S. J. (2012). MITOMI: A microfluidic platform for in vitro characterization of transcription factor-DNA interaction. *Methods in Molecular Biology (Clifton, N.J.)*, *786*, 97–114. https://doi.org/10.1007/978-1-61779-292-2_6

Sayou, C., Monniaux, M., Nanao, M. H., Moyroud, E., Brockington, S. F., Thévenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y., Wong, G. K.-S., Weigel, D., Parcy, F., & Dumas, R. (2014). A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science (New York, N.Y.)*, *343*(6171), 645–648. https://doi.org/10.1126/science.1248229

Thornton, J. W. (2004). Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nature Reviews. Genetics*, *5*(5), 366–375. https://doi.org/10.1038/nrg1324

Tsong, A. E., Tuch, B. B., Li, H., & Johnson, A. D. (2006). Evolution of alternative transcriptional circuits with identical logic. *Nature*, *443*(7110), 415–420. https://doi.org/10.1038/nature05099

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
55607DFEC1A947E... Author Signature

3/22/2021
Date