Genetic Basis of Natural Variation in Cellular Decision Making in Wild Yeasts

Genetic basis of natural variation in cellular decision making in wild yeasts

A dissertation presented

By

Kayla Lee

To

The department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

In the subject of

Molecular and Cellular biology

Harvard University

Cambridge, Massachusetts

March 2018

Dissertation Advisor: Dr. Michael Springer Kayla Lee

Genetic basis of natural variation in cellular decision making in wild yeasts

Abstract

In nature, microbes often need to decide which of many potential nutrients to consume. This decision-making process is complex, involving both intracellular constraints and the organism's perception of the environment. What factors does a cell take into account as it makes this decision? And how is it molecularly encoded? In this thesis, we quantified differences in the concentration of induction galactose-responsive (GAL) genes across wild *S. cerevisiae* strains. To begin to mimic the complexity of natural environments, we grew cells in mixtures of two sugars, glucose and galactose. Galactose induces the well-studied pathway, while glucose represses the pathway. Using bulk segregant analysis, we mapped genomic loci underlying phenotypic variation in the GAL pathway, identified causal genes, and did a series of allele-replacements to understand the specific effect of these genes. By analyzing crosses of phenotypically different strains, we identified a locus containing the galactose sensor, *GAL3*, as a gene that can explain upwards of 90% of the variation in the decision to induce GAL genes in a single cross. This gene can also modulate the time required for cells to switch from utilizing glucose to galactose. In a related study, we showed that this gene and other genes not specific to the GAL pathway tune two independent features of the GAL response, which can switch a strain from being bimodal to unimodal. Our results suggest that signaling pathways can be highly variable across strains and thereby might allow for rapid adaption in fluctuating environments. The independent tuning of specific and non-specific pathway genes can help cells adapt to complex environments at various timescales.

Table of Contents

Acknowledgments

It took a village.

Graduate school has taught me many lessons. A few of those lessons have been the peculiarities of yeast and their utilization of a six-carbon sugar galactose, but most of them have been lessons about me. Graduate school has been a journey where I have learned much about myself and even more about the importance of the people who love and care about me the most. I can't submit this thesis without thanking the people who have been with me since the beginning and the friends that I have met throughout.

Thank you all for being my village.

First and foremost, I have to thank my parents, Morris and Leslie Lee. My parents have always served as an inspiration for me. They started my love for creativity, self-expression, and the sciences.

Next, I have to thank my younger siblings, Trey Lee and MaKenna Lee. Despite the increase in distance, my time in graduate school has brought as all closer. I am so grateful to have such a supportive, encouraging, and inspiring family.

I am positive that I would not have made it through graduate school without the love and kindness of my boyfriend, Christopher Wade. Throughout my ups, downs, and extreme lows he has been incredibly supportive of me and my dreams. As I continued to learn more about myself, as I learned to manage my anxiety attacks, my failed therapy sessions, and what it means to be a black woman in the sciences in America—you have been there unconditionally.

And as far as lifelong friends, there are my line sisters. The girLS who have been a text message, FaceTime, or plane ride away. Thank you for being a listening ear, a roof over my head, and the best cheering squad I could ever ask for. A specific thank you to Colby Mason, Alia Barnes, LaTonia Kelley, Moriah Lark, Jasmine Stephens, Janelle Trueheart, Kierra Greene, Tyler Hood, Antoinette Trott, Dannielle Brown, Natalie Harris, Akilah Lamar, Sherene Lattimore, and Erin Phillips.

I am grateful for my various scientific adventures that have brought me to close friends, specifically Amber Anders. I'll always remember how intertwined our stories have been—we still have so much living to do.

To my professors, scientific network, my community, and my support system: thank you Dr. Edison Fowlks for introducing me to HHMI and the world of biomedical research. Thank you Dr. Indu Sharma and Dr. Raymond Samuel for continuously supporting me throughout graduate school. Your kind words, mentorship, and never-ending support has been critical for my success. Thank you Eboney Hearn, Francie LaTour, Bruce Birren, and Gisselle Velez-Ruiz for showing me the wildly creative science done at the Broad Institute of MIT and Harvard, which ultimately led to my application to graduate school at Harvard. Thank you to HHMI, who has been supporting my scientific studies since 2010 and the HHMI Gilliam Fellows, who are truly some of the most incredible scientists you will ever encounter.

To my BodyScapes gym family—thank you for helping me finally be willing to call Boston something like home. Marina Kovalenko, thank you for being both a student and a teacher for me. You've taught me to be fearless, to follow my dreams, and always reminded me to smile a little more. You helped get me through my last years of graduate school, and I am so grateful that our paths crossed.

Thank you to my advisor, Mike Springer, and the entire Springer Lab. Throughout my time in graduate school, I have watched the lab turn over to new students who all bring exciting and unique perspectives to our research. I am grateful that my time in the Springer Lab introduced me to Yarden Katz, I will always admire your science, your beliefs, and how you intertwine the two. A special thank you to Jue Wang, one of my main collaborators in graduate school. Thank you to the rest of the Springer Lab (old and new), Christine DeGennaro, Yoni Savir, Sarah Boswell, Bo Hua, Chiara Ricci-Tam, Julius Palme, Jiayin Hong, and Nicoletta Commins.

I know we are all more than familiar with the African proverb, it takes a village to raise a child. Well, it took way more than a village to get me through graduate school. I will be forever grateful to my village and the child I am still becoming.

Chapter 1: Introduction

Natural Variation and Yeast as a Model System

The nutrient composition of natural environments can fluctuate, and organisms must induce metabolic pathways that allow them to utilize the available nutrients. Cells need to sense and respond to their environment, and these responses lead to downstream fitness outputs. How cells choose to respond to their environment is critical to their survival. We often think of these decisions in the context of development or cell fate response; but with single-celled microbes, a decision to or not to induce a specific pathway can be the difference between life or death—their survival is dependent on making the right choices in various environments. This makes the budding yeast a good model to begin to understand the factors involved in cellular decision making.

Humans have taken advantage of the fermentative properties of yeast for over thousands of years in brewing and baking [1], while the budding yeast has been a model organism for laboratory experiments for over a century [2]. Early in the molecular biology of yeast, several strains and their derivatives were adopted as laboratory favorites, among these s288c, W303, and CEN.PK [3]. These laboratory isolates of yeast formed the foundation for our understanding of enzyme activity, linkage mapping, and genetic recombination [4]. Yeast makes an ideal model system because it is eukaryotic, unlike some of its microbial counterparts; it has a generation time of about ninety minutes in ideal conditions, and it is easy to manipulate genetically [4]. In addition, in 1996 s288C became the first eukaryote to be sequenced, which opened the door a deeper, genomic understanding of more complex organisms [5]. Recently, scientists learned that there is a rich ecological history that exists within the *S. cerevisiae* species and once scientists started paying attention to other strains, it became realized that s288C was an outlier in both phenotypic and genetic variation [6].

A history of natural isolates of yeast

Traditionally, yeasts have been found in ecological niches related to alcoholic fermentation or human consumption, like natural fermentation, wine, or baking—suggesting that most lineages of yeast are associated with human production [1]. The fermentative property to convert sugars into alcohol has involved independently in yeast several times, possibly as a strategy to outcompete other microorganisms [7]. Strains of *S. cerevisiae* have also been isolated from clinical patients, although less commonly than their distant relative *C. albicans.* These clinically relevant strains are often able to grow at higher temperatures, likely due to their adaptation to human hosts [8]. Rather than being isolated from a single ancestor these strains represent various, independent opportunities that strains have colonized human hosts [9]. In addition to human association, strains of yeast have also been isolated from wild environments, like oak trees, soil, and rotting fruit [9-11]. These environments and ecological niches that strains have been isolated from have introduced a new understanding and appreciation for the budding yeast and the population genomics of these various lineages.

By studying the genetic variation of natural isolates of yeast, scientists have started to understand the evolutionary history and population structure of yeast strains [6,9-11]. On average, the frequency of polymorphisms between each strain is about 2.8 SNPs per kilobase (kb) [9]. Scientists have used a comprehensive view of genomic data to elucidate the complex population structure of *S. cerevisiae* [11]. Strains that share the same phylogenetic relationship across their entire genome are considered clean or non-mosaic lineages [11]. Some of these strains fall within their geographic boundaries (North American, Malaysian, West African), while other strains are related to their source like the Wine/European or the sake lineage [11]. The independent lineages linked to fermentation (Wine/European and Sake) suggest independent human domestication events [9,12]. The genome of mosaic strains represent a combination of these lineages and are likely a result of human association or recent genetic recombination events [11]. Sequence segments from

many of the mosaic strains are not represented within the identified clean lineages, suggesting that there is some unsampled sequence space of *S. cerevisiae* [11]. Supporting this, a group of researchers conducted a field study of Chinese isolates of *S. cerevisiae*, and their collection of yeast represented a population structure that doubled the existing genetic variation of known isolates [10].

These studies have created the foundation for our understanding of population genomics across various yeast species and strains. We often hope that these population studies can inform us of the evolutionary history of the species. Liti et al. [11] has found that the existing patterns of variation support a mutation-selection drift model [13,14], rather than adaptive selection. There is a high quantitative overlap between phenotypic and genomic clustering based on identified polymorphisms [6,11]. In many cases, the phenotypic clustering matched phylogenetic properties better than the ecological source [6].

Notably, the population structure of *S. cerevisiae* is much more complex than its most recent relative, *S. paradoxus*. [10]. While common polymorphisms in *S. cerevisiae* are found across lineages, most SNPs in *S. paradoxus* are private within each population. In addition, linkage disequilibrium values suggest that recombination events happen more in *S. cerevisiae*, which is likely a result of *S. cerevisiae* strains having more opportunities to mate or recombine [11]. These increased recombination events are also matched with more phenotypic variation within the species, which could be a result of the increased number of ecological niches that *S. cerevisiae* occupy. Despite the longstanding history between humans and yeast, strains of *S. cerevisiae* that are most like their most recent relative, *S. paradoxus*, have been isolated from natural environments and not sources related to fermentation [9].

A major goal of my thesis has been to take advantage of natural variation that exists in natural isolates of the budding yeast *S. cerevisiae*. Here, I briefly describe how recent decades' research has expanded to a wide range of natural isolates that have been used to brew beer, make your favorite wine, or have been found in clinical patients. Various labs have taken advantage of this phenotypic

and genetic natural variation to determine the genetic basis of complex traits using linkage analysis, microarrays, and other genetic tools.

Natural variation as a tool to understand gene networks

Many labs have taken advantage of the diversity that exists within isolates of *S. cerevisiae* to uncover genetic and molecular basis of natural variation. Perturbation and deletion screens are often used to help identify the function of the deleted or mutated gene. While we have learned a lot from these types of screens, by taking advantage of natural variation we can start to understand the axes of variation nature tunes. Commonly studied lab strains and wild isolates are often used in genetic linkage studies to identify polymorphisms between the pair.

Brem et al. (2002) combined genetic analysis and genome wide expression patterns to begin to understand the genetic basis of expression levels [15]. They found that patterns of differential expression were linked to one or more genes, which had complex patterns of inheritance. In general, detected loci were either cis-acting regulators of single genes or trans-acting regulators of many genes. They concluded that even in simple microorganisms, inheritance can be polygenic and highly complex, which study highlighted the potential for yeast as a system and linkage studies for quantitative traits.

Gerke et al. (2009) used sporulation efficiency as a quantitative trait to understand how multiple and interacting loci translates into phenotypic variation [16]. They found single nucleotide changes in several transcription factors and showed that selection has helped shaped this variation. This supports previous work, which suggests that trans-acting regulators of many genes can produce quantitative variation [16-19].

These studies and many others [6,9,20-27] have formed the foundation for using natural isolates of yeast and functional genomics as a tool to understand the genetic and molecular basis of complex traits. Other major lessons from these studies are the existence of transgressive QTLs, linked QTNs, and an understanding of the genetic architecture of complex traits in all organisms.

Similar to other studies, this thesis takes advantage of natural variation to show the degree of variation we see in nature, the molecular and physiological basis of the variation, and begins to explore the evolutionary timeframe that this variation has occurred. To give context to this natural variation, we focus on cellular decision making in the well characterized galactose (GAL) utilization pathway.

The Galactose Signaling Pathway

The Galactose Signaling Pathway as a model for quantitative traits

The GAL signaling pathway is a well understood genetic regulatory circuit and has contributed to our understanding of cell circuitry and signaling. The GAL regulatory network is necessary for converting galactose into glucose-6-phosphate via the Leloir Pathway. Components necessary for galactose metabolism were identified by genetic screens in the 1960s. Over time, the GAL pathway has served as a model to understand gene expression, bimodality, and bistability of signaling networks.

The components of the GAL pathway

The GAL pathway is composed of a combination of enzymatic and regulatory genes. *GAL1*, *GAL5*, *GAL10*, and *GAL7* are enzymatic genes of the pathway clustered on chromosome II [28]. GAL1 is the first metabolic gene in the Leloir pathway and is often used as a readout for GAL pathway induction [29-32]. *GAL10* and *GAL7* are adjacent genes and have a bidirectional promoter. *GAL2*, *GAL3*, *GAL4*, and *GAL80* make up the regulatory machinery of the pathway [28]. In the presence of galactose, GAL2p regulates the uptake of galactose from the external environment. Both galactose and glucose can enter through the GAL2p transporter, although galactose has a higher affinity. *GAL4* is the transcriptional activator of the pathway. In the presence of galactose, ATP and galactose bind with GAL3p, which then leads to a conformational change that then binds with GAL80p and relieves

GAL80p of its transcriptional repression of GAL4p [28,33-37]. *GAL80*, maintains the negative feedback of the pathway. In the absence of galactose, GAL80p is bound to the transcriptional activator, GAL4p, which prevents the induction of the pathway. The regulatory components of the GAL pathway make simple feedback loops that to create dynamic, ultra-sensitivity within the pathway [38,39].

Modeling quantitative traits in yeast

By using a combination of natural variation and the GAL signaling pathway as a model, we can change the abundance of various carbon sources and use properties of the pathway like induction, growth, and fitness as quantitative traits. With genetic linkage analysis and allele replacements, we can then determine the underlying genetics of these traits. Taking advantage of the genetic tractability of yeast, natural isolates, and GAL pathway, form the basis of this thesis. I've spent most of my introduction presenting the foundation for my thesis—natural variation and quantitative traits. My thesis will focus on understanding the genetic basis of two features of cellular decision making in mixed nutrient environments: 1) **the induction of signaling pathways**, and 2) **the population dynamics of cells.**

Chapter 2 of my thesis, *Polymorphisms in the yeast galactose sensor underlie a natural continuum of nutrient- decision phenotypes*, was inspired by a previous observation in the lab. By looking at hundreds of mixtures of glucose and galactose, Escalante-Chong et al. (2015) showed that induction of the GAL pathway is dependent on the ratio of glucose and galactose [31]. It was previously thought that in the presence of both glucose and galactose, GAL genes were only activated when the glucose concentration dropped below some threshold. Their results suggested that competitive binding upstream of the canonical signaling pathway is the source of the ratio response. Their work presents the possibility that there are likely other biological examples of ratio-sensing,

which was previously overlooked because of limited sampling conditions. To further confirm this phenotype, they expanded their search to other strains of *S. cereviasiae.* In addition to strains maintaining their ratio response, the 'decision front' or concentration regime where 20% of cells were induced varied between the different strains. This intriguing result became the foundation for chapter two.

Chapter 3 of my thesis, *Nature uses specific and non-specific GAL genes to quantitatively tune bimodal signaling response of the glucose-galactose circuit*, was inspired by my earlier observations in the Springer Lab. The idea of 'bimodality' or simply the presence of two, distinct modes of a population of cells has been well studied in biology—specifically in the GAL pathway [32,38-41]. My early observations showed that some natural isolates of yeast looked unimodal in intermediate concentrations of glucose and galactose, while others looked bimodal. In fact, in some strains this bimodality operated in different phenotypic spaces. This phenotype, which has been well studied and generally associated with positive feedback loops [32,38], seemed to be a bit more complex. We approached this study with genetic mapping, using a similar method to the previous chapter—and our results highlighted that nature tunes various genetic knobs to produce unique phenotypes.

References

1. Vaughan-Martini A, Microbiology AMJOI, 1995. Facts, myths and legends on the prime industrial microorganism. Springer.

2. Liti G. The fascinating and secret wild life of the budding yeast S. cerevisiae. eLife. 2015;4: 1705–9. doi:10.7554/eLife.05835

3. Mortimer RK, Genetics JJ, 1986. Genealogy of principal strains of the yeast genetic stock center. Genetics Soc America.

4. Duina AA, Miller ME, Keeney JB. Budding Yeast for Budding Geneticists: A Primer on the Saccharomyces cerevisiaeModel System. Genetics. 2014;197: 33–48. doi:10.1534/genetics.114.163188

5. Goffeau A, Barrell BG, Bussey H, Davis RW, 1996. Life with 6000 genes. Science.

6. Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, et al. Trait Variation in Yeast Is Defined by Population History. Kruglyak L, editor. PLoS Genet. 2011;7: e1002111–15. doi:10.1371/journal.pgen.1002111

7. dowska ERE, Hellborg L, Ishchuk OP, Orhan F, Galafassi S, Merico A, et al. Parallel evolution of the make consume strategy in Saccharomyces and Dekkera yeasts. Nature Communications. Nature Publishing Group; 2011;2: 302–7. doi:10.1038/ncomms1305

8. Klingberg TD, Lesnik U, Arneborg N, Raspor P, Jespersen L. Comparison of Saccharomyces cerevisiae strains of clinical and nonclinical origin by molecular typing and determination of putative virulence traits. FEMS Yeast Research. 2008;8: 631–640. doi:10.1111/j.1567-1364.2008.00365.x

9. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. Nature. 2009;458: 342–345. doi:10.1038/nature07670

10. Wang Q-M, Liu W-Q, Liti G, Wang S-A, Bai F-Y. Surprisingly diverged populations of Saccharomyces cerevisiaein natural environments remote from human activity. Molecular Ecology. 2012;21: 5404–5417. doi:10.1111/j.1365-294X.2012.05732.x

11. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population genomics of domestic and wild yeasts. Nature. 2009;458: 337–341. doi:10.1038/nature07743

12. Fay JC, Benavides JA. Evidence for Domesticated and Wild Populations of Saccharomyces cerevisiae. PLoS Genet. 2005;1: e5–6. doi:10.1371/journal.pgen.0010005

13. Elyashiv E, Bullaughey K, Sattath S, Rinott Y, Przeworski M, Sella G. Shifts in the intensity of purifying selection: An analysis of genome-wide polymorphism data from two closely related yeast species. Genome Research. 2010;20: 1558–1573. doi:10.1101/gr.108993.110

14. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang S-P, et al. A Catalog of Neutral and Deleterious Polymorphism in Yeast. Pritchard JK, editor. PLoS Genet. 2008;4: e1000183–15. doi:10.1371/journal.pgen.1000183

15. Brem RB. Genetic Dissection of Transcriptional Regulation in Budding Yeast. Science. 2002;296: 752–755. doi:10.1126/science.1069516

16. Gerke J. Genetic interactions between transcription factors cause natural variation in yeast. Science. 2009;323: 495–498. doi:10.1126/science.1166426

17. Gerke J, Lorenz K, Ramnarine S, Cohen B. Gene–Environment Interactions at Nucleotide Resolution. Mackay TFC, editor. PLoS Genet. 2010;6: e1001144–11. doi:10.1371/journal.pgen.1001144

18. Brem RB, Storey JD, Whittle J, Kruglyak L. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature. 2005;436: 701–703. doi:10.1038/nature03865

19. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, et al. Dissecting the architecture of a quantitative trait locus in yeast. Nature. 2002;416: 326–330. doi:10.1038/416326a

20. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. Nature Publishing Group; 2010;464: 1039–1042. doi:10.1038/nature08923

21. Ehrenreich IM, Bloom J, Torabi N, Wang X, Jia Y, Kruglyak L. Genetic Architecture of Highly Complex Chemical Resistance Traits across Four Yeast Strains. Gasch AP, editor. PLoS Genet. 2012;8: e1002570–9. doi:10.1371/journal.pgen.1002570

22. Taylor MB, Ehrenreich IM. Genetic Interactions Involving Five or More Genes Contribute to a Complex Trait in Yeast. Fay JC, editor. PLoS Genet. 2014;10: e1004324–8. doi:10.1371/journal.pgen.1004324

23. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. Nature Publishing Group; 2013;: 1–6. doi:10.1038/nature11867

24. Smith EN, Kruglyak L. Gene–Environment Interaction in Yeast Gene Expression. Mackay T, editor. PLoS Biol. 2008;6: e83–15. doi:10.1371/journal.pbio.0060083

25. Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet. 2006;7: 862–872. doi:10.1038/nrg1964

26. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nature Publishing Group. Nature Publishing Group; 2015;16: 197–212. doi:10.1038/nrg3891

27. Bergstrom A, Simpson JT, Salinas F, Barre B, Parts L, Zia A, et al. A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. Molecular Biology and Evolution. 2014;31: 872–888. doi:10.1093/molbev/msu037

28. Sellick CA, Campbell RN, Reece RJ. Chapter 3 Galactose Metabolism in Yeast—Structure and Regulation of the Leloir Pathway Enzymes and the Genes Encoding Them. Elsevier; 2008. pp. 111–150. doi:10.1016/S1937-6448(08)01003-4

29. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population Diversification in a Yeast Metabolic Program Promotes Anticipation of Environmental Shifts. Siegal ML, editor. PLoS Biol. 2015;13: e1002042–24. doi:10.1371/journal.pbio.1002042

30. Wang J, Atolia E, Hua B, Savir Y, Escalante-Chong R, Springer M. Natural Variation in Preparation for Nutrient Depletion Reveals a Cost–Benefit Tradeoff. Siegal ML, editor. PLoS Biol. 2015;13: e1002041–31. doi:10.1371/journal.pbio.1002041

31. Escalante-Chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ, et al. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. Proceedings of the National Academy of Sciences. 2015;112: 1636–1641. doi:10.1073/pnas.1418058112

32. Acar M. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435: 228–232. doi:10.1038/nature03524

33. Lavy T, Yanagida H, Tawfik DS. Gal3 Binds Gal80 Tighter than Gal1 Indicating Adaptive Protein Changes Following Duplication. Molecular Biology and Evolution. 2016;33: 472–477. doi:10.1093/molbev/msv240

34. Diep CQ, Tao X, Pilauri V, Losiewicz M, Blank TE, Hopper JE. Genetic Evidence for Sites of Interaction Between the Gal3 and Gal80 Proteins of the Saccharomyces cerevisiae GAL Gene Switch. Genetics. 2008;178: 725–736. doi:10.1534/genetics.107.074799

35. Diep CQ. Intragenic Suppression of Gal3C Interaction With Gal80 in the Saccharomyces cerevisiae GAL Gene Switch. Genetics. 2005;172: 77–87. doi:10.1534/genetics.105.050807

36. Melcher K, Xu HE. Gal80-Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression. The EMBO Journal. 2001.

37. Pilauri V. Gal80 Dimerization and the Yeast GAL Gene Switch. Genetics. 2005;169: 1903–1914. doi:10.1534/genetics.104.036723

38. Venturelli OS, El-Samad H, Murray RM. Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. Proceedings of the National Academy of Sciences. 2012;109: E3324–E3333. doi:10.1073/pnas.1211902109

39. o MSAN, Leidy C, Pedraza JM. Tuning the range and stability of multiple phenotypic states with coupled positive–negative feedback loops. Nature Communications. Nature Publishing Group; 2013;4: 1–8. doi:10.1038/ncomms3605

40. Stockwell SR, Rifkin SA. A living vector field reveals constraints on galactose network induction in yeast. 2016 Sep pp. 1–16. doi:10.1101/073817

41. Nature MF, 2000. Feedback control of intercellular signalling in development. naturecom.

Chapter 2: Polymorphisms in the yeast galactose sensor underlie a natural continuum of nutrient-decision phenotypes

Kayla B. Lee¶, Jue Wang¶, Julius Palme, Renan Escalante-Chong, Bo Hua, and Michael Springer

¶These authors contributed equally to this work.

From the paper: http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006766

Abstract

In nature, microbes often need to "decide" which of several available nutrients to utilize, a choice that depends on a cell's inherent preference and external nutrient levels. While natural environments can have mixtures of different nutrients, phenotypic variation in microbes' decisions of which nutrient to utilize is poorly studied. Here, we quantified differences in the concentration of glucose and galactose required to induce galactose-responsive (GAL) genes across 36 wild *S. cerevisiae* strains. Using bulk segregant analysis, we found that a locus containing the galactose sensor *GAL3* was associated with differences in GAL signaling in eight different crosses. Using allele replacements, we confirmed that *GAL3* is the major driver of GAL induction variation, and that *GAL3* allelic variation alone can explain as much as 90% of the variation in GAL induction in a cross. The *GAL3* variants we found modulate the diauxic lag, a selectable trait. These results suggest that ecological constraints on the galactose pathway may have led to variation in a single protein, allowing cells to quantitatively tune their response to nutrient changes in the environment.

Introduction

The nutrient composition of natural environments can fluctuate and organisms must induce metabolic pathways that allow them to utilize the available nutrients [1-3]. Recent studies have found that closely related microbes vary in both the types of nutrients they can utilize and the efficiency at which they do so [4,5]. However, most studies have focused on differences in growth in single-nutrient environments (e.g., growth in "pure" glycerol). Natural environments, on the other hand, often contain multiple nutrients that cells need to choose between, and suboptimal nutrient decisions can have severe fitness consequences [6-9]. Hence, it is likely that cells have been selected not only to utilize nutrients efficiently, but to decide which subsets of nutrients to utilize.

Signaling pathways sense which nutrients are present and control the decision of which transcriptional network to activate. The majority of plasticity in gene expression patterns has been linked to changes in transcriptional regulatory networks. While transcription factor binding sites are typically conserved, the location of the binding sites in the genome can rapidly evolve [10]. Chromatin immunoprecipitation followed by sequencing in yeast [11-13], mice and human [14], and flies [15] have shown a surprisingly small conservation in the genes and sites that were bound by transcriptional regulators between species. Even when the regulated genes are conserved, the transcription factors that regulate them can change [16-19]. The development of genomic tools has greatly aided the interspecific comparison of regulatory binding sites.

There are relatively few cases where adaptive changes in signaling networks have been linked to molecular and genetic variation. By contrast, changes in transcription regulatory networks have been easier to identify due to the development of high-throughput genomic and computational approaches. Additionally, studies are often biased towards finding changes in transcriptional regulatory networks based on the phenotypes assayed, i.e. fitness in 'extreme' environments. Still, there are multiple situations where upstream signaling changes must have occurred. For instance, in

the galactose-utilization pathway (GAL) in *C. albicans*, Rtg1p and Rtg3p activate GAL genes while Gal4p is involved in glucose regulation; in *S. cerevisiae*, Gal4p activates GAL genes, while Rtg1p and Rtg3p are involved in glucose regulation [20]. This implies that the upstream signaling networks that sense and transduce galactose and glucose signals have also changed. Furthermore, duplication and divergence can shape signaling networks. For example in the GAL pathway in yeast, duplication and divergence allowed the sensing and catabolic activity of a single ancestral protein to be separated into two paralogous proteins [21]; this divergence likely had profound consequences for how yeast were able to 'perceive' galactose. Hence, it is likely that cellular decision-making can also evolve, but the degree of variation, its molecular and physiological basis, and the evolutionary timeframe at which it occurs has yet to be resolved.

To begin to address these questions we characterized differences in natural isolates of the budding yeast, *S. cerevisiae* in the *decision* to induce the GAL pathway in mixtures of glucose and galactose. In the presence of high concentrations of glucose, the preferred carbon source, yeast cells repress the GAL pathway [22,23]. In the presence of galactose alone, cells activate GAL genes. In mixtures of both glucose and galactose, cells must "decide" whether to induce GAL-associated genes. In such mixed environments, cells show a complex response [24] where the induction of the pathway is dependent on the ratio of glucose and galactose [25]. These observations, combined with the deep molecular understanding in the literature [26,27], make the GAL pathway an excellent model for studying natural variation in cellular decision-making.

Here, we use single-cell measurements to quantify differences in GAL decision-making across closely related natural isolates of *S. cerevisiae*, followed by bulk segregant analysis and allele replacements to find the genetic determinants of this variation. We found that the glucose concentration needed to induce GAL genes varies ~100-fold across yeast strains. Even though this phenotypic variation is continuous, a large proportion of it can be explained by differences in a single gene, the galactose sensor *GAL3*. Changing the *GAL3* allele produces a measurable difference in the

diauxic lag length, a trait that was previously shown to be selectable [8]. These results highlight the fact that cellular decision-making has the potential to be rapidly shaped by selective pressures in the environment.

Results

The decision to induce GAL pathway varies across *S. cerevisiae* natural isolates

To enable measurement of the GAL signaling response, we generated a fusion of the *GAL1* promoter from *S. cerevisiae* and yellow fluorescent protein (*GAL1pr-YFP*) (Fig. 1A). *GAL1* is the first metabolic gene in the galactose utilization pathway [28], and this promoter has been used by numerous studies as a faithful readout of pathway activity [9,25,29,30]. The reporter construct was integrated into the neutral HO locus [31] in 42 different *S. cerevisiae* strain backgrounds (S2.1 Fig.) [32,33]. These 42 strains span a range of phylogenetic and ecological diversity [32,33]. Six of these strains either did not grow in galactose, likely due to inactivation of the pathway [4], and thus were not characterized further. We focused on determining the GAL response phenotype of the remaining 36 strains (S1 Table).

To survey the natural variation in the inducibility of GAL genes in mixtures of glucose and galactose, we measured the GAL reporter response in a titration of glucose concentrations from 2% to 0.004% w/v on a background of constant 0.25% galactose (Fig. 1B, Materials and methods). Cells were first pre-grown for 14-16 hours in 2% raffinose (which does not induce or repress the GAL pathway), and then transferred to glucose + galactose and grown for 8 hours at low densities. We previously showed that this protocol is sufficient for cells to reach steady-state without depleting the carbon sources [25]. Finally, single-cell YFP fluorescence was measured by flow cytometry. To account for well-to-well variability or variability in our glucose titration, each of the 36 query strains were co-cultured with a reference strain, YJM978, containing *TDH3pr-mCherry* (this constitutive

fluorophore allowed us to distinguish the query and reference strains) and *GAL1pr-YFP* (Materials and Methods).

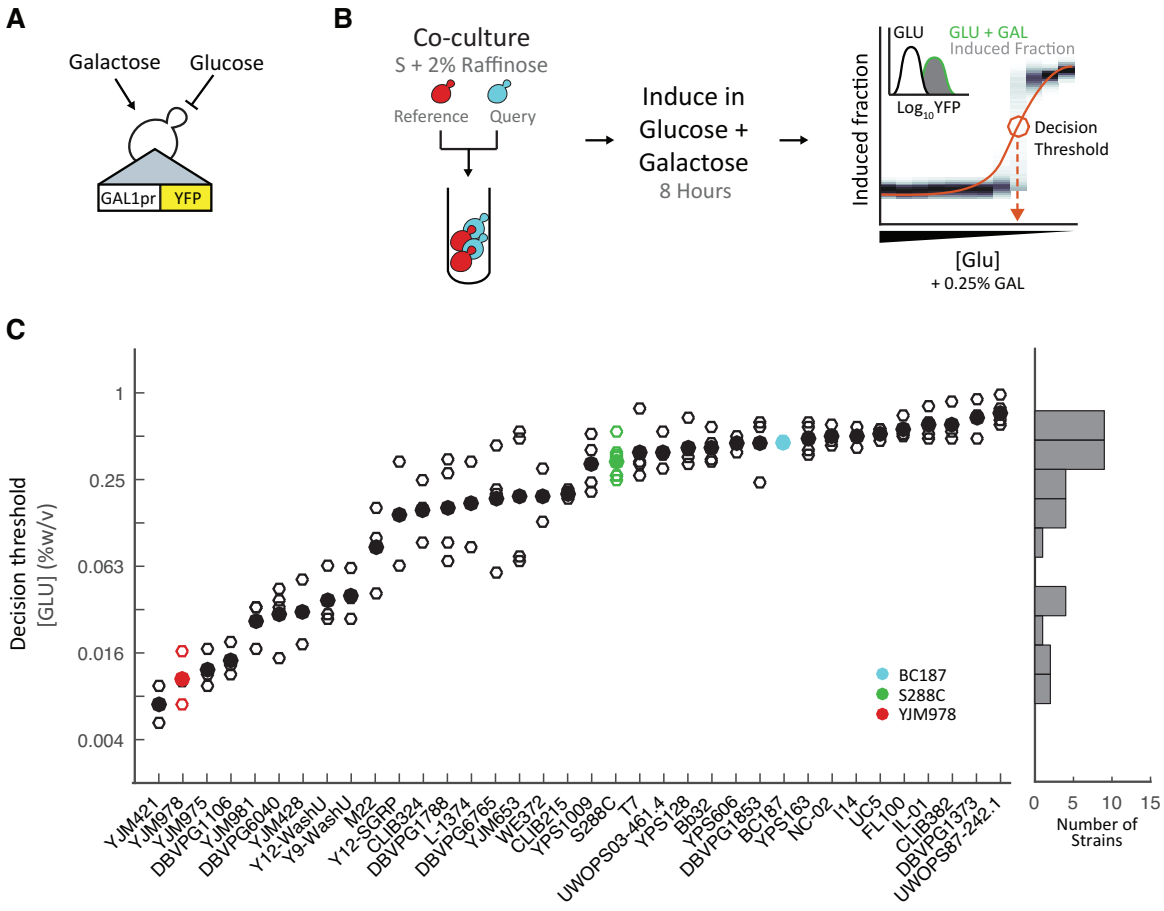


Figure 2.1. Natural isolates of *S. cerevisiae* vary in the decision to induce the GAL pathway.

(A) Schematic of the reporter construct (*GAL1pr-YFP*). (B) Schematic of co-culture pre-growth, glucose and galactose induction, and flow cytometry measurement for a glucose titration. The decision threshold, the concentration of glucose where 50% of the cells are induced, is indicated by circle and dashed line. (C) Decision threshold for 36 lab and natural isolates of *S. cerevisiae*. Histogram shows the distribution of the mean decision threshold for all strains assayed.

Qualitatively, there were large strain-to-strain differences in the concentration of glucose at which cells induced the GAL pathway (Fig. 2.1, S2.4 Fig.). We also observed bimodal expression in some strains and conditions, a likely consequence of cellular heterogeneity and ultra-sensitivity in the GAL circuit [30,34,35]. This complicates quantitative analysis, because a metric such as the mean expression (which is implicit, for example, in a bulk assay) would convolute both the number of cells that are inducing and the expression level of the cells that have 'decided to' induce, two factors that may vary independently in bimodal responses. Hence, to compare the GAL pathway response between natural isolates, we defined a metric, the "decision threshold", as the concentration at which 50% of cells have greater-than-basal expression of the GAL reporter (Materials and Methods). This metric is similar to those used in previous work [29,36], and focuses on *when* a cell decides to induce a pathway while differentiating it from *how strongly* a cell responds once induced. The decision threshold is highly reproducible across replicate measurements for all of our natural isolates (S2.3 Fig.).

Quantitatively, the decision threshold varies over a range of 108 ± 0.7-fold glucose concentrations across our strains (Fig. 2.1, S2.4 Fig.). The Hawaiian cactus strain UWOPS87-242.1, was most inducible, with a decision threshold of 0.74±0.2% glucose (mean ± standard error mean), while the clinical isolate, YJM421, was least inducible, with a decision threshold of 0.01±0.01% glucose (mean ± S.E.M.). Half of the strains have decision thresholds within a 8.1-fold range centered at 0.25% glucose (Fig. 2.1C). This glucose concentration corresponds to a galactose+glucose ratio of ~1:1. The distribution of decision thresholds appears continuous; there are significantly more than two distinct decision thresholds given the reproducibility of our measurements (Materials and Methods).

Strain differences in decision threshold could be due to differences in sugar signaling, utilization, or both. If sugar utilization is a factor, we expect the decision threshold to be correlated to growth rates in glucose or galactose. We measured the growth rates of the 36 natural isolates

during mid-exponential growth in either 0.5% glucose or 0.5% galactose (S2.5 Fig., [6]). Despite substantial variation in single-sugar growth rates across our strains (0.23-fold in glucose and 0.16-fold in galactose), neither growth in pure galactose or glucose are correlated with the decision threshold (glucose $r^2 = 0.2$, galactose $r^2 = 0.001$). This implies that while both sugar utilization and signaling can vary between strains, evolution has the potential to select these two traits independently.

Previous studies have determined the correlation between genotypic diversity and either phenotypic diversity or ecological niche. For example, analysis of 600 traits in yeast by Warringer et al. identified a correlation between phylogeny and phenotype [4]. These studies can be used to assess whether traits are more likely to be neutral or undergoing selective constraint. To determine if the decision threshold is correlated with phylogeny, we began by comparing the 13 closely related strains of the wine/European clean lineage. Despite the close phylogenetic relationship of these strains, this lineage represents the most phenotypic diversity. The two most phenotypically distinct strains in this lineage, YJM978 and DBVPG1373, have a 48±0.3-fold difference (mean ± S.E.M.). More broadly, we compared the pairwise genetic distances (determined by RAD-SEQ [37]) to pairwise phenotypic distance (Materials and Methods). We did not find a significant correlation (r^2=-0.08) between genetic distance and decision threshold. This level of correlation with genetic distance is comparable to that of many other traits [4] (S2.6 Fig., p-value=0.17 by ANOVA). Finally, we tested for and found a signification association between ecological niche and decision threshold (S2.7 Fig., p-value=2.95e-5 by ANOVA).

Bulk segregant analysis identifies one major-effect locus underlying natural variation in the GAL decision threshold

To investigate the genetic basis of the observed variation in GAL decision threshold, we performed bulk-segregant analysis using a variant of the X-QTL method (Fig. 2.2A) [38-40]. We crossed eight strains that span the phenotypic and phylogenetic diversity of *S. cerevisiae* in a round-

robin design (Fig. 2.2). This design is known to efficiently sample parental genetic variation and allow downstream linkage analyses to detect loci with a range of effect sizes [38]. Pools of segregants from each cross were grown in a glucose + galactose condition that maximally differentiates the parental phenotypes. The 5% least and 5% most induced cells ("OFF" and "ON" segregant pools) were isolated by fluorescence-activated cell sorting (FACS) and sequenced in bulk to determine the parental allele frequencies in each pool. We used the MULTIPOOL software [41] to determine statistical significance for allele frequency differences between OFF and ON pools across the genome (Materials and Methods), and called significant loci as regions where the peak log-odds-ratio was greater than 10 (Fig.2. 2B). This cutoff had a low false-discovery rate in a previous study, and correlated well with allele frequency difference, a proxy for locus effect size, in our data [38] (S2.8 Fig.).

Over all 8 crosses, we found 16 loci where segregant pools differ in allele frequency at LOD > 10 (Fig. 2.2B). One locus centered at 460 kb on chromosome IV (henceforth, "chrIV:460") was the only locus to exceed the LOD cutoff in all 8 crosses, as well as the most significant locus in each cross (Fig. 2.2B). The 2-LOD support interval for this locus in the YJM978 x BC187 cross, defined as the genomic region where LOD decreases by 2 from its peak, is 10 kb wide and contains six genes (Fig. 2.2C). This includes *GAL3*, whose product directly binds galactose and positively regulates the GAL pathway [42]. The support interval for chr:460 looked similar in other crosses (S2 Table). One other locus, at chrXIV:462, reached LOD > 10 in two crosses; the remaining significant loci were confined to a single cross. We did detect additional loci in multiple crosses using a less stringent cutoff of LOD > 5; however, chrIV:460 remained the only locus significant in all crosses (S2.8 Fig., S2.2 Table, Materials and Methods).

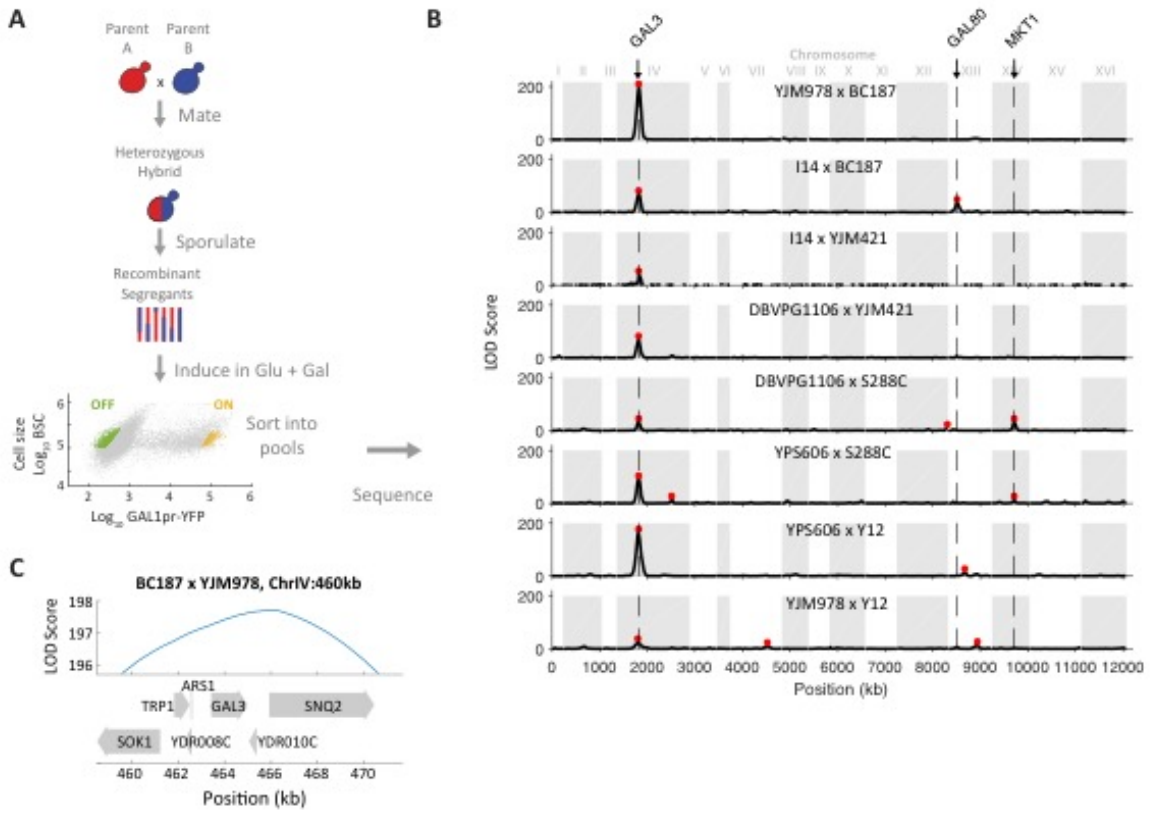


Figure 2.2. Bulk segregant analysis identifies one major-effect locus underlying natural variation in decision threshold.

(A) Schematic of bulk segregant analysis. Meiotic segregants from heterozygous hybrids were sorted by FACS into 'ON' and 'OFF' pools based on *GAL1pr-YFP* expression and then sequenced. (B) LOD score of allele frequency difference between 'ON' and 'OFF' segregant pools versus genomic position (red asterisks: LOD > 10). A region of chromosome IV containing *GAL3* was associated with the difference between the 'ON' and 'OFF' phenotype in all 8 crosses. Potential candidate genes for other loci include *GAL80*, *MKT1*, and others listed in Table S2. (C) Genes found within the 2-LOD support interval around the peak LOD score from ChrIV:460kb plotted with the LOD score for the BC187xYJM978 cross.

In principle, a round-robin cross design is expected to detect each locus in more than one cross. The fact that we identified several alleles in only one cross is potentially explained by a lack of statistical power, epistasis, or gene-by-environment effects [38]. Indeed, a potential caveat to the apparent importance of *GAL3* is that in a pooled segregant analysis, large effect QTLs might mask the presence of smaller-effect genes [43]. Furthermore, low sequencing depth of some of our segregant pools may have limited our power to detect small-effect alleles (Materials and Methods). However, even the lowest sequencing depth we obtained (25x) is still sufficiently powered to map alleles with effects as low as 5% of phenotypic variance [44]. More importantly, we performed a complementary analysis of segregants to directly measure the contribution of *GAL3* to the phenotypic variance in a cross (see below). Finally, alleles that were identified in only one cross may arise from the different conditions used for sorting each cross (Materials and Methods), i.e. gene-by-environment effects. Given its importance, we chose to focus on the chrIV:460 locus for further characterization.

GAL3 is the causative allele and major driver of variation in the GAL signaling response

To determine if *GAL3* was the causative allele on chrIV:460 with a predictable and quantitative impact on the decision threshold, we replaced the endogenous *GAL3* allele of strains YJM978, BC187, and S288C with alleles from eleven natural isolates spanning the observed range of phenotypic variation (Fig. 2.1). Allele replacements were constructed by deleting the 3283bp *GAL3* locus, which includes 890 bp upstream, 911 bp downstream, and the 1563 bp *GAL3* ORF in haploid parental strains and then replacing the deleted locus with the homologous ~3283bp *GAL3* locus from other strains using the CRISPR-Cas9 system [45] (Materials and Methods). Replacement of *GAL3* alleles in the YJM978 background recapitulated the ~95-fold range of decision threshold of the natural isolates that served as *GAL3* allele donors. Additionally, the decision thresholds of allele-replacement and *GAL3* donor strains were well-correlated in this background (r^2 of 0.58). Similarly, *GAL3* alleles in the S288C background had a ~55-fold range and r^2 of 0.60; *GAL3* alleles in the BC187

background had a ~138-fold range and r^2 of 0.63. In total, this confirms the significant impact that the *GAL3* locus has on variation in the decision threshold (Fig. 2.1, Fig. 2.3A-C, S2.9 Fig.).

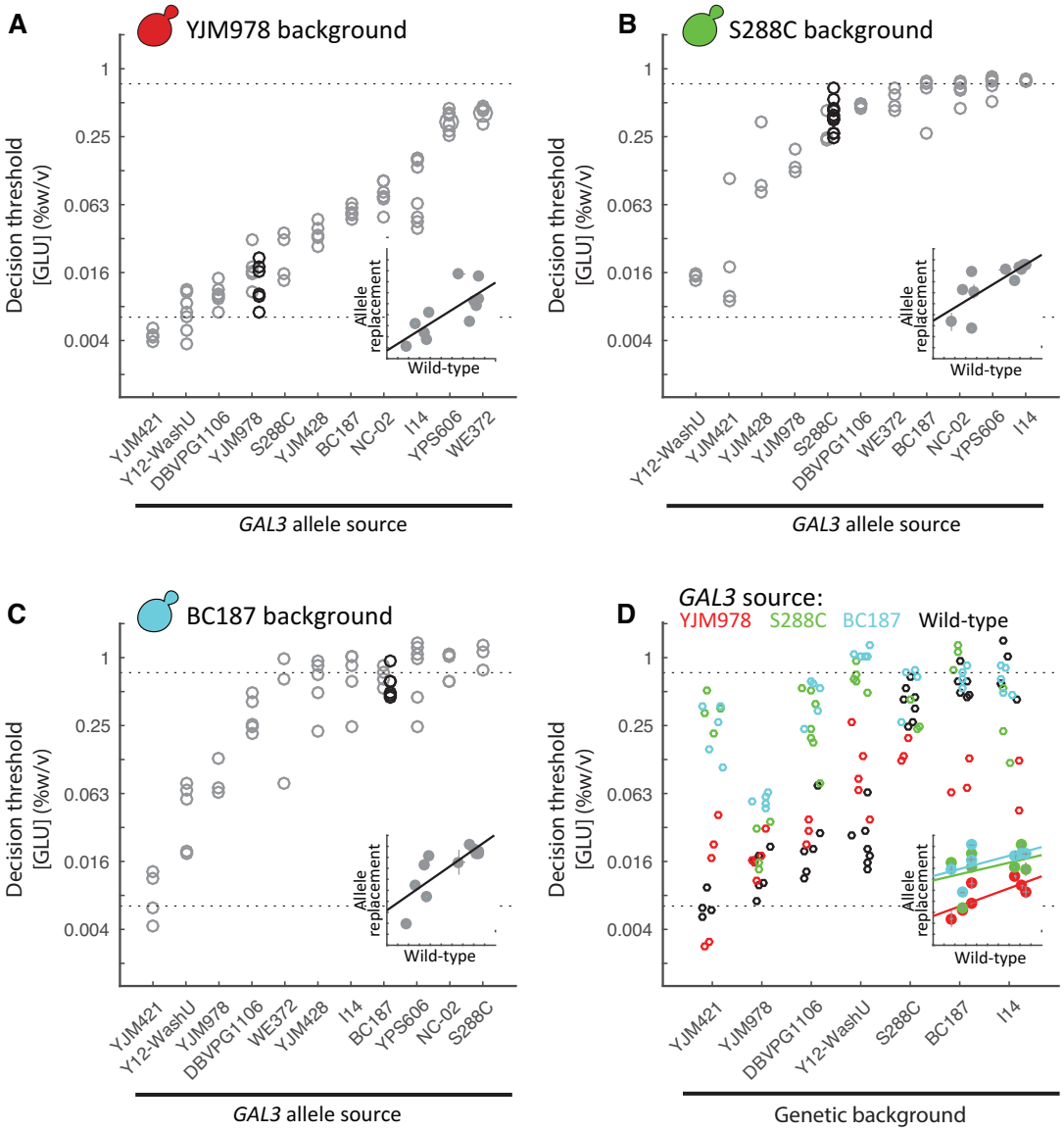


Figure 2.3. *GAL3* allele largely sets the decision threshold.

Decision threshold of eleven different *GAL3* homologous replacements in three genetic backgrounds: (A) YJM978, (B) S288C, and (C) BC187. Decision threshold of wild-type strain is indicated by the black circle. Inset: Correlation plot of natural isolate versus allele replacement decision threshold; error bar represents S.E.M. (D) Decision threshold of three allelic variants of *GAL3* (Mean and S.E.M. of at least two replicates) inserted into various genetic backgrounds: $GAL3^{YJM978}$ (red), $GAL3^{S288C}$ (green),

Figure 2.3 (continued)

and *GAL3*^{BC187} (blue), haploid wild-type strain (black). Strains are ordered based on wild-type decision threshold. Inset: Correlation plot of natural isolate (decision threshold of background strain) versus the decision threshold of the allele replacement, error bar represents S.E.M.

While different *GAL3* alleles were able to confer a range of phenotypes in a particular strain background, the three strain backgrounds also displayed different decision thresholds for a given *GAL3* allele. This suggests that genes other than *GAL3* also affect the decision threshold, even for the BC187xYJM978 cross. To assess the magnitude of this background effect, we measured the decision threshold in seven different strain backgrounds where the *GAL3* locus has been replaced with an allele from YJM978, S288C, or BC187 (Fig. 2.3D, S2.10 Fig.). Across the seven backgrounds, *GAL3*^{YJM978} allele-replacement strains varied in decision threshold over a ~14-fold range, *GAL3*^{S288C} strains over ~20-fold, and *GAL3*^{BC187} strains over ~49-fold (Fig. 2.3), and the correlation (r^2) in decision threshold between these allele-replacement strains and their strain background donors was 0.60, 0.28, and 0.12, respectively. These results confirm that strain background strongly influences decision threshold. However, it is also clear that *GAL3* allele still has a stronger effect, because both the phenotypic range and correlations to donor strain were lower for strain background than for *GAL3* allele. This can also be seen by the fact that the *GAL3*^{BC187} and *GAL3*^{S288C} strains have decision thresholds that are similar to each other but systematically higher than *GAL3*^{YJM978}, regardless of strain background.

The *GAL3* allele accounts for 70-90% of the phenotypic variance in a cross between strains with extreme opposite decision thresholds

The allele replacements show that *GAL3* is a major driver of natural variation in the decision threshold, but also suggests that other genes play a significant role. To quantify the relative contribution of *GAL3* allele versus other genes to variation in decision threshold, we analyzed the variance in decision threshold across meiotic segregants from YJM978 x BC187 hybrids with

different combinations of *GAL3* alleles. This method is relatively insensitive to the metric chosen and potential non-linear relationships between genotype and phenotype. We chose this cross because the *GAL3* locus was the only significant locus from BSA, and thus our calculation should yield a rough upper bound on the *GAL3* contribution in other strains. We constructed three hybrid strains: a 'wild-type' hybrid (YJM978 x BC187), a hybrid with *GAL3* only from YJM978 (YJM978 x BC187 *gal3Δ::GAL3*^{YJM978}) and a hybrid with *GAL3* only from BC187 (YJM978 *gal3Δ::GAL3*^{BC187} x BC187). The decision threshold of at least 58 meiotic segregants was measured for each hybrid in duplicate (Fig. 2.4, S2.11 Fig.). Consistent with *GAL3* having a large effect, we found that converting a single allele in each hybrid greatly reduced the phenotypic variation of the segregant populations.

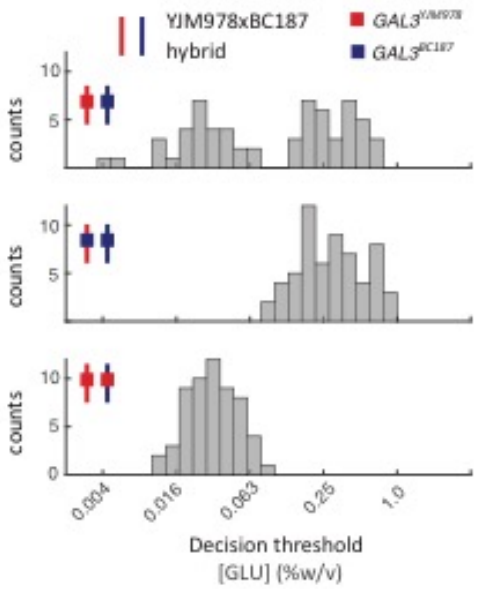


Figure 2.4. The *GAL3* allele accounts for 70-90% of the decision threshold.

Decision threshold of segregants produced from hybrid (top), hybrid with *GAL3*^{YJM978} allele homologously replaced with *GAL3*^{BC187} (middle), and hybrid with *GAL3*^{BC187} allele homologously replaced with *GAL3*^{YJM978} (bottom). Hybrids are indicated by small schematic, the line represents the

Figure 2.4 (continued)

genetic background and the filled in box represents the origin of the *GAL3* allele (YJM978: red, BC187: blue)

To quantify the effect of GAL3, we used a variance-partitioning model with additive effects. We assumed that the total variance of each segregant population (V_P) can be separated into several contributions: $V_P = V_G + V_E + V_{EG} + V_D + V_I$. We assumed no interactions between gene and environment (V_{EG}=0) and no epistatic interactions (V_I=0). Additionally, there is no dominance as we used haploid strains (V_D=0) and the environmental variability is equal to the measurement noise because the strains are isogenic and are grown in identical environments (V_E=ε^2). Since we know that *GAL3* is a major driver of the decision threshold phenotype, we partitioned V_G into two components: the variance due to the background (V_{BG}) and the variance due to *GAL3* (V_{GAL3}). Hence the total variance could be simplified to $V_P = \varepsilon^2 + V_{GAL3} + V_{BG}$ (Fig. 2.4, S2.11 Fig.). By definition, in the allele swap segregants (hybrids 2 and 3) $V_{GAL3} = 0$.

Based on this variance-partitioning model (Materials and Methods), we can estimate the contribution of the *GAL3* allele by dividing V_{GAL3} with the sum of V_{GAL3} and V_{BG} or the total genetic variance. We can estimate the V_{BG} by comparing segregants from hybrid 2 and 3 or from the 'wild-type' hybrid, which will give us an upper and lower bound of *GAL3* allelic contribution. Using the segregant population from hybrid 2 and 3, the *GAL3* allele contributes 86% of the genetic variance between YJM978 and BC187. Two segregants from hybrid 1 ('wild-type' hybrid) have a decision threshold lower than what we would have expected from segregants of hybrid 2 (S2.11 Fig.). These two strains increase the background variance, which ultimately reduces the effect of *GAL3*. Using the segregant population from hybrid 1, we estimate that *GAL3* explains 67% of the variance between YJM978 and BC187. These two 'outliers' could potentially result from a rare combination of alleles between the strains, implying that we under sampled the distribution from hybrid 2. These

calculations suggest that *GAL3* could contribute anywhere from 70-90% to the variance of the decision threshold phenotype.

Polymorphisms within *GAL3*

To further explore how polymorphisms in *GAL3* might contribute to the decision threshold phenotype, we analyzed the sequences of 55 natural isolates of *S. cerevisiae* [32,46,47]. We identified 8 synonymous and 19 nonsynonymous polymorphisms in the coding region of *GAL3*, which represent 26 unique haplotypes (Table S3). The natural isolates that we assayed (Fig. 2.1) included 21 of these unique haplotypes, where we excluded the haplotypes from the 6 strains that cannot utilize galactose. To determine whether the *GAL3* haplotype is predictive of the decision threshold, we tested for and found a significant association between decision threshold and *GAL3* haplotype (S2.12 Fig., p-value=0.04 by ANOVA). However, strains that share *GAL3* haplotypes also tend to share population history (i.e. genomic background) and ecological niche. In particular, YPS163, YPS606, YSP218, and T7 were all isolated from North American oak trees and make up the North American lineage; S288C and FL100 are both mosaic lab strains; YJM978, YJM981, and YJM975, are clinical isolates in the Wine/European lineage. Due to the correlation between phylogeny and *GAL3* haplotype, follow-up investigations using a larger and more diverse set of strains are needed to determine the extent to which decision threshold can be determined solely from the *GAL3* haplotype.

The *GAL3* polymorphisms we observed can in principle affect the expression level, regulation, or function of the protein. Using mutfunc, a database that predicts the consequences of mutations in a protein, we found that 13 of the 19 nonsynonymous SNPs are predicted to affect protein function (Table S4). This includes non-conservative amino-acid substitutions in the Gal3p dimerization interface and the Gal3-Gal80p interface [48]. Gal3p and Gal80p are both homodimers and the Gal3p-Gal80p interaction, which is crucial to the mechanism of GAL pathway activation, is thought to depend on this homodimerization [49,50]. We are less able to predict the impact of promoter variation, but we found 13 SNPs in the promoter (500 bp upstream of the start codon), none of which

were in known transcription factor binding sites (Table S3). Furthermore, we did not find a significant association between the *GAL3* promoter haplotype and decision threshold (S2.12 Fig., p-value=0.98 by ANOVA). Follow-up investigations to characterize the effects of each SNP in *GAL3* will provide mechanistic insight into how the GAL response can be tuned quantitatively by polymorphisms in a single gene.

To determine if *GAL3* or any other genes in the canonical pathway are subject to adaptive evolution, we performed a McDonald-Kreitman analysis [51] using DnaSP [52]. The McDonald-Kreitman test compares intraspecies variation with the divergence between two species. If the ratio of nonsynonymous to synonymous variation between species is equal, there is neutral selection, while any act of natural selection will result in a shift of these two ratios. This test suggests that *GAL3*, *GAL80*, and *GAL5* are under strong purifying selection (Table S5). Our analysis is consistent with two studies that analyzed polymorphism and divergence data between *S. cerevisiae* and *S. paradoxus*, which suggested that there is strong evidence for purifying selection across the yeast genome [53,54].

GAL3 tunes the glucose-galactose diauxic lag

We next asked whether variation in *GAL3* produces selectable variation in phenotype. Diauxic growth is a classical phenotype observed when cells are grown in two sugars [3]. Cells undergo two phases of growth separated by a period with little growth, known as the "diauxic lag", during which cells induce the genes required to metabolize the second sugar. Previously, our lab has shown that diauxic lag length varies across natural yeast isolates vary, and that *GAL1* transcriptional reporter level before the lag is negatively correlated with lag length [6]. Here, we further show that decision threshold is correlated to *GAL* reporter expression (S2.13 Fig.), and likely as a result, also negatively correlated with diauxic lag length (Fig. 2.5A). This suggests that changing *GAL3* alleles will also change the diauxic lag.

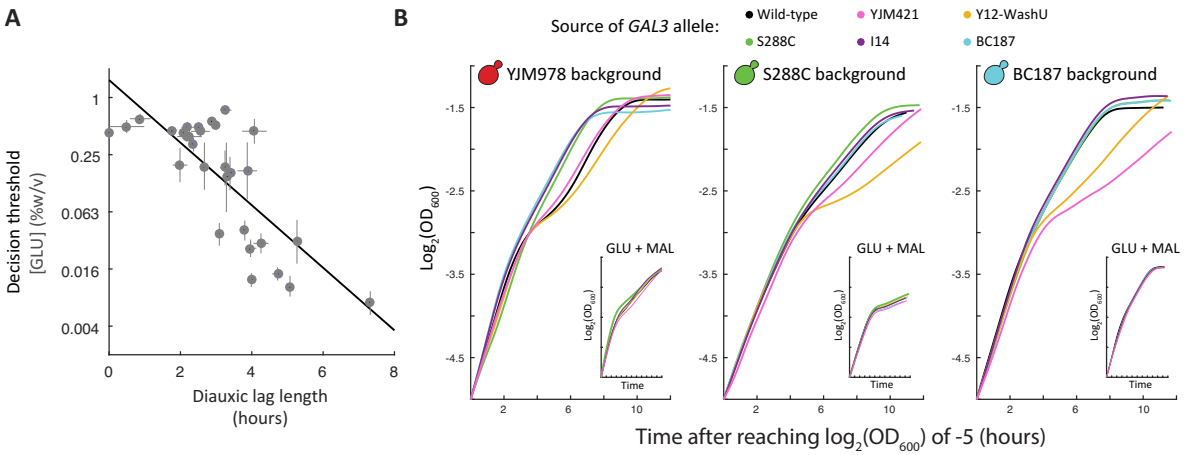


Figure 2.5. Changing *GAL3* alleles specifically affects the glucose-galactose diauxic lag.

(A) Growth curves (OD_{600} versus time) of allele replacement strains in three genetic backgrounds: YJM978 (red), S288C (green), BC187 (blue) of cells grown in a mixture 0.25% glucose and 0.25% galactose. Cultures grew for 6 to 8 hours before entering the diauxic lag. A single replicate is shown (additional replicates are shown in Figure S14). Inset: Growth curves (OD_{600} versus time) of the same strains grown in a mixture of 0.25% glucose and 0.25% maltose.

(B**)** The decision threshold (as measured in Fig. 1) is inversely correlated with the diauxic lag length (as measured in [6]).

To determine if *GAL3* allele affects diauxic lag across our natural isolates, we performed diauxic shift experiments on allele replacement strains representing six *GAL3* alleles (I14, YJM421, Y12-WashU, BC187, and S288c) in three strain backgrounds (YJM978, S288C and BC187) (Fig. 5B). As expected, simply changing the *GAL3* allele in either the YJM978, BC187, or S288C background was sufficient to change the diauxic lag (Fig. 2.5B, S2.14 Fig.). Additionally, *GAL3* alleles from short-lag strains S288C, BC187, and I14 (which also have higher decision thresholds) tended to reduce diauxic lag length when introduced into long-lag strain backgrounds YJM978, DBVPG1106, and YJM421, and vice versa. Previously, strains evolved to have an altered diauxic lag in glucose+maltose also had altered lag in glucose+galactose [8]. To determine if the *GAL3* alleles we identified had a specific effect

on GAL regulation, we also measured diauxic lag in glucose+maltose. This showed that *GAL3* allele only affects diauxic lag in glucose+galactose and not in glucose+maltose (Fig. 2.5B, inset).

Discussion

Natural genetic variation in the GAL pathway

Genetically and phenotypically diverse natural isolates of yeast have become a powerful system to determine the genetic basis of complex traits. Analyzing natural variation in the well-characterized GAL pathway has the potential to allow us to connect molecular variation, phenotypic variation, and selection. Two recent studies also explored variation in the GAL response across budding yeasts. Peng et al. used combinatorial promoter swaps of GAL regulatory components (*GAL2*, *GAL3*, *GAL4*, *GAL80*) between *S. cerevisiae* and *S. paradoxus* to show that *GAL80* promoter variation was responsible for differences in the GAL response [36]. Roop et al. used a combination of promoter and ORF swaps to show that variation in multiple GAL pathway genes underlies regulatory differences between *S. cerevisiae* and *S. bayanus* [55]. In our study, we used bulk-segregant linkage mapping across diverse *S. cerevisiae* strains to find that most of the variation in GAL regulation is caused by polymorphisms in a single gene, *GAL3*.

Why did each of the three studies identify different genetic loci? A potential explanation is the difference in genetic distance between the strains/species analyzed. We analyzed variation within *S. cerevisiae*, Peng et al. analyzed variation between the closely related species *S. cerevisiae* and *S. paradoxus* [36], and Roop et al. analyzed variation between the more distantly related species *S. cerevisiae* and *S. bayanus* [55]. One hypothesis, based on studies of evolution of development, holds that phenotypic changes on short timescales (i.e. between closely related organisms) are more likely to be caused by nonsynonymous coding-sequence mutations [56,57]. These are favored because of their large phenotypic effects, but they come at a cost of increased pleiotropy. On a longer timescale, *cis*-regulatory mutations are enriched, presumably because they are less pleiotropic and allow finely

tuned regulation of fitness-enhancing activities [56,57]. Results from Peng et al., Roop et al., and our study are largely inconclusive or weakly inconsistent with pleiotropy being the driving force between the sources of variation. Variation in the GAL response between *S. cerevisiae* and *S. paradoxus* is driven by promoter variation in *GAL80* [58], while variation between *S. cerevisiae* and *S. bayanus* was driven by a combination of promoter and ORF changes [55]. Furthermore, many causative genes were identified between *S. cerevisiae* and *S. bayanus*, while a single gene drove most of the variation within *S. cerevisiae* and between *S. cerevisiae* and *S. paradoxus*. Based on genome-wide expression profiles, there is no evidence that a *GAL80* or *GAL3* variants should be more pleiotropic than simultaneously varying all pathway components [59]. Broader investigations of multiple *Saccharomyces* species will help clarify the relationship between evolutionary distance and the repertoire of mutations.

There are possibilities other than pleiotropy that could cause the difference in genes identified. The different phenotypes assayed in each study could be controlled by different components in the GAL pathway. However, we believe our assays measure highly correlated underlying traits. Peng et al. supplemented their media with mannose to avoid the confounding effects of carbon limitation at low galactose concentrations [36]. The effect of mannose on galactose utilization has not extensively been studied in *S. cerevisiae*, but in other systems mannose can be utilized as a preferred carbon source [60]. Therefore, we expect that the decision threshold in mannose and glucose are likely correlated. Roop et al. compared batch growth in a mixture of glucose and galactose a condition that leads to a diauxic lag in *S. cerevisiae* but not in *S. bayanus.* While natural variation in diauxic growth could have involved many pathways, we showed previously that glucose-galactose diauxic lag is driven by the timing of GAL pathway induction [6]. Here we extended this by showing that diauxic lag is correlated with the decision threshold (Fig. 2.5A) and primarily modulated by variation in *GAL3* (Fig. 2.5B). Overall, our work here and recent findings in the literature suggest that all these traits are highly interconnected.

Is the observed variation in the GAL pathway the result of neutral drift or selection? There are three lines of evidence that suggest the GAL pathway is under selection. First, previous analysis has used the QTL cis-regulatory sign test [61] to argue that the GAL pathway has undergone selection between *S. cerevisiae* and *S. bayanus* [55]. Second, we show via the McDonald-Kreitman test that several of the genes in the GAL pathway within *S. cerevisiae* are significantly enriched for nonsynonymous polymorphisms, and therefore likely under adaptive constraint (Table S2.5). Third, there are at least five genes that affect variability in the GAL response [29,35,36,55]. Given this knowledge we can ask whether the GAL pathway is under selection in a manner similar to the cis-regulatory sign test. Instead of looking for concordant expression changes, we look for enrichment of independent functional mutations in an unexpectedly small subset (i.e. one gene) of multiple possible target genes. Specifically, what is the chance of eight independent alleles of *GAL3* being the main driver of variation in all eight of our crosses given a mutational target size of 5 genes (p-value<1e-6, permutation test). While there are caveats with this method, e.g. what is the true number of potential QTN for each gene, the potential mutational target size is probably much larger than five genes. A recent study of deletion mutants found that upwards of 40% of genes in the yeast genome have the potential to influence the GAL response (Hua et al., under review). Together, we believe these lines of evidence support the hypothesis that the GAL pathway is under selection.

The interplay between selection, pleiotropy, and natural variation is further highlighted by experimental evolution studies in yeast [8] and *E. coli* [62,63]. In mixtures of carbon sources, microbes first consume a preferred nutrient, followed by a " diauxic lag" where cells must induce the genes necessary to metabolize the second, less preferred nutrient [3]. New et al. found that yeast strains evolved in rapid shifts between glucose and maltose also had a shorter diauxic lag in a mixture of the two sugars [8]. Similarly, *E. coli* passaged in glucose-acetate mixtures evolved into both short-lag and long-lag subpopulations [62,63]. These results show that diauxic lag length is a readily evolvable trait. However, in both previous cases, the evolved phenotypes were due to mutations in

global metabolic regulators. For example, New et al. obtained evolved isolates with weakened catabolite repression, via mutations in the glucose-sensing genes *HXK2* and *STD1*, while maltose regulatory genes were unchanged [8]. These mutations are pleiotropic, and thus the evolved strains had shorter diauxic lags in both galactose and maltose. By contrast, we did not find a strong role for general catabolite repression underlying natural variation in GAL regulation, even though the potential mutational target size is large. Instead, our *GAL3* allele replacements specifically tune the glucose-galactose diauxic lag and do not affect the glucose-maltose lag (Fig. 2.5B, inset). This raises the possibility that in natural environments, where evolution has had longer to act, mutations that perturb global metabolic regulation (as in *STD1* or *HXK2*) may be more detrimental than mutations that tune a particular sugar preference (as in *GAL3*). Hence, as predicted, the frequency of pleiotropic mutations may be an important difference between evolution at short versus long timescales [56,57].

Genetic complexity of quantitative traits

A number of labs have analyzed phenotypic variation in response to a range of environmental conditions [4,64,65] and delved into the genetic basis of variation in specific traits such as heat tolerance [66], gene/protein expression [67,68], sporulation efficiency [69-71], colony morphology [72-75], sulfur uptake [76], and carbon regulation [36,55]. The vast majority of these studies used growth or expression level [67] as readout. Collectively these studies have yielded insight into the nature of quantitative traits [77]. But, these readouts can potentially miss the complexities of response to fluctuating environments. For example, cells grown in mixtures of glucose and galactose must choose when to induce GAL genes, a property that varies between natural isolates and is distinct from the growth rate on pure glucose and pure galactose (S2.5 Fig.). Is the decision to induce the GAL pathway similar to other phenotypic traits?

In principle, multiple different pathways and genes can shape natural variation of a trait. A round-robin BSA analysis of MAPK-pathway-mediated stress tolerance in yeast showed that genes both inside and outside the assayed pathway can have large effects on intraspecies phenotypic

variation [38]. Previous X-QTL analyses of various traits in yeast have identified a handful of QTLs per trait [38,39,78]. The largest throughput single study found a median of 12 loci for 46 phenotypic traits [79]. While neither previous analysis of the GAL pathway performed a BSA, the number of causative alleles identified through swaps and the total amount of variation explained by these alleles suggests a similar number of genes affect the GAL pathway as other traits [36,55]. Similar to other traits [80], despite the strong correlation between the decision threshold of *GAL3* allele-replacements and their corresponding *GAL3* donor strains, the genetic background still plays a strong role in phenotypic variation (Fig. 2.3). QTLs outside the GAL pathway might explain why swapping the main regulators of the GAL pathway between *S. cerevisiae* and *S. bayanus* was only able to partially interconvert the phenotypes [55]. Taken together, these results suggest the GAL pathway is similar to other quantitative traits. But, taken alone, our BSA of the GAL pathway suggests that the GAL phenotype is a simpler genetic trait than many of the previously analyzed traits. While other studies have found a small number of QTLs drive the majority of variation in a cross, e.g. sporulation efficiency, when these traits are analyzed in a different cross unique QTLs are often found [70]. Our variation stands out in that there appears to be an allelic series of a single gene, *GAL3*, driving the variation. Similar to the genes whose alleles drive variation in sporulation efficiency, *GAL3* is positioned in a 'signal transduction bottleneck' [69]. Unlike sporulation where multiple genes critical for decision making were identified [69], we found variation is driven by a subset, i.e. one, of potential decision making proteins. Taken collectively with the previous analysis of variation in the GAL pathway, an intriguing possibility is this difference might not arise from 'genetic simplicity' of the GAL response. Instead, *GAL3* might control the variability in a subset of the phenotype of the GAL response, i.e. the decision threshold, while other members of the GAL pathway might control different aspects of a broader phenotypic response, e.g. diauxic lag. Our future work will directly test this hypothesis.

Genetic interactions, often referred to as epistasis, play a role in many QTL-mapping studies [81,82]. Our study highlights two types of epistasis. First, similar to many other systems [80], the quantitative effect of each *GAL3* allele is influenced by the genetic background. While, the directional effect of the *GAL3* alleles from S288C, BC187, and YJM978 are largely preserved, in the YJM978 and S288C background, the effects of the allele replacements are diminished, and the background dominates the resulting phenotype, which is highly compressed (Fig. 2.3D). This suppression of variation in certain strain backgrounds has been seen in other systems, such as colony morphology [83,84], and can result from the interaction of two or many genes. Second, there appears to be a 'maximum' achievable decision threshold of around 1% glucose in 0.25% galactose. Hence, when *GAL3* alleles with decision thresholds above 0.25% glucose in 0.25% galactose are placed into a background with a high decision threshold (e.g. S288C or BC187), the effect of the *GAL3* allele appears to saturate. This behavior is phenomenologically reminiscent of epistatic interactions in peaked fitness landscapes where beneficial mutations have diminishing effects [85,86] or the apparent saturating interaction between gene expression and fitness [87]. Further elucidation of these examples of epistasis in the GAL pathway will likely provide new insights into basic principles of quantitative genetics.

In conclusion, while other genes contribute, the repeated and sizeable role of *GAL3* in this study stands out compared to other QTL analyses in yeast. Using BSA, we identified the galactose sensor *GAL3* as a major driver of this phenotypic variation, accounting for 70-90% of the variation in a single cross. Hence, polymorphisms in a single gene in the canonical GAL pathway are sufficient to create a continuum of natural variation. An intriguing possibility is that in *S. cerevisiae*, variation in *GAL3* may allow strains to vary the diauxic lag in a non-pleiotropic manner. Environments that fluctuate in a 'predicable' manner might be expected to select for a pathway architecture that allow strains to evolve on this fluctuating axis [56]. Further analysis of the GAL pathway should help to elucidate the interplay of molecular variation, phenotypic variation, and selection.

Materials and methods

Strains and media

Strains were obtained as described in [6]. Strains used in this study can be found in Table S1. All strains from the collection and those assayed in Fig. 2.1 were homozygous diploids and prototrophic. An initial set of 42 strains were assayed in a gradient of glucose (2% to 0.004% by two-fold dilution) in a background of 0.25% galactose. Strains W303 and YIIC17-E5 were excluded from downstream analysis due to poor growth in our media conditions. Strain 378604X was also excluded due to a high basal expression phenotype that was an outlier in our collection. All experiments were performed in synthetic minimal medium, which contains 1.7g/L Yeast Nitrogen Base (YNB) (BD Difco) and 5g/L ammonium sulfate (EMD), plus D-glucose (EMD), D-galactose (Sigma), or raffinose (Sigma). Cultures were grown in a humidified incubator (Infors Multitron) at 30°C with rotary shaking at 230rpm (tubes and flasks) or 999rpm (600uL cultures in 1mL 96-well plates).

Flow cytometry assay

GAL induction experiments were performed in a 2-fold dilution series of glucose concentration, from 1% to 0.004% w/v, with constant 0.25% galactose. 2% glucose and 2% galactose conditions were also included with each glucose titration experiment. To assess and control for well-to-well variation, experiments were performed as a co-culture of a "query" strain to be phenotyped and a "reference" strain that was always SLYB93 (natural isolate YJM978 with constitutive mCherry segmentation marker).

To start an experiment, cells were struck onto YPD agar from -80C glycerol stocks, grown to colonies, and then inoculated from colony into YPD liquid and cultured for 16-24 hours. Query and reference strains were then co-innoculated at a 9:1 ratio by volume in a dilution series (1:200 to 1:6400) in S + 2% raffinose medium. The raffinose outgrowths were incubated for 14-16 hours, and then their optical density (OD_{600}) was measured on a plate reader (PerkinElmer Envision). One

outgrowth culture with OD_{600} closest to 0.1 was selected for each strain, and then washed once in S (0.17% Yeast Nitrogen Base + 0.5% Ammonium Sulfate). Washed cells were diluted 1:200 into glucose + galactose gradients in 96-well plates (500uL cultures in each well) and incubated for 8 hours. Then, cells were processed by washing twice in Tris-EDTA pH 8.0 (TE) and resuspended in TE + 0.1% sodium azide before transferring to a shallow microtiter plate (CELLTREAT) for measurement.

Calculating the decision threshold (F_{50}) metric

Flow cytometry was performed using a Stratedigm S1000EX with A700 automated plate handling system. Data analysis was performed using custom MATLAB scripts, including Flow-Cytometry-Toolkit (https://github.com/springerlab/Flow-Cytometry-Toolkit, https://github.com/springerlab/Induction-Gradient-Toolkit). All experiments were co-cultured with a reference strain and were manually segmented using a fluorescent channel (mCherry or BFP) and side scatter channel (SSC). *GAL1*pr-YFP expression for each segmented population was collected and the induced fraction for each concentration of sugars was computed as shown previously in Escalante et al. [25]. The decision threshold for each glucose titration was calculated from the induced fraction of the ten sugar concentrations. The decision threshold was reported as the glucose concentration were 50% of the cells were induced.

Filtering reference and query data

To account for well-to-well variability or variability in our glucose titration, each of the query strains were co-cultured with a reference strain, YJM978, containing TDH3pr-mCherry. This constitutive fluorophore was used to segment the query and reference strains. Three filters were used to discard bad samples. 1) The 5% truncated standard deviation was calculated. Samples where the reference strains response was double this truncated deviation from the mean reference response were discarded without analyzing the co-cultured query strain (39 of 480 total

experiments) (S2.2 Fig.). 2) Query strains where the data was of poor quality such that we could not make an accurate calculation of F50, typically for low counts or cultures that did not induce (8 of 441). 3) Query strain values that were over 1.5 standard deviations from the mean of the other replicates, (21 of 433) (S2.3 Fig.). This 1.5 standard deviation cut-off was determined based on calculating the difference of each sample from the mean and fitting this to a normal distribution assuming outliers (S2.3 Fig.). All strains were measured at least twice; replicates were performed on different days.

Estimation of the number of unique GAL phenotypes

To estimate a lower bound for the number of distinct GAL phenotypes, we compared our measurement noise from replicate measurements to the range of variation between strains. By simply dividing the range by the measurement noise or by asking on a pair-wise manner which phenotypes are statistically distinguishable, the number of phenotype is at least five.

Crossing and generating segregants

To prepare parent strains for crossing and sporulation, diploid natural isolates bearing the *hoΔ::GAL1pr-YFP-hphNT1* reporter cassette were sporulated and random spores were isolated. Mating type was determined by a test cross. We then introduced a constitutive fluorescent marker in tandem with the GAL reporter, to obtain MAT**a**; *hoΔ::GAL1pr-YFP-mTagBFP2-kanMX4* or MATα; *hoΔ::GAL1pr-YFP-mCherry-natMX4* parent strains. To the MAT**a** parent we also introduced a pRS413-derived plasmid bearing *STE2pr-AUR1-C* and *hphNT1*. This plasmid is maintained by hygromycin selection but also allows selection for MAT**a** cells by Aureobasidin A [88]. This plasmid design is inspired by a similar mating-type selection plasmid used in a recent study [38].

To generate segregant pools, we prepared a diploid hybrid and sporulated it as follows. We crossed a parent with *BFP-kanMX* with the mating type selection plasmid to a parent with *mCherry-natMX4* and isolated a $G418^{R}Nat^{R}Hyg^{R}$ diploid hybrid with the plasmid. We sporulated the hybrid by

culturing it to saturation in YPD, diluting 1:10 in YP+2% potassium acetate and incubating at 30C for 8 hours. Cell were then washed and resuspended in 2% potassium acetate and incubated at 30C until >20% of cells were tetrads, or about 72 hours. We incubated ~$5x10^6$ tetrads in 100uL water with 50U of zymolyase 100T (Zymo Research) for 5 hours at 30C, and then resuspended tetrads in 1mL of 1.5% NP-40 and sonicated for 10 seconds at power setting 3 on a probe sonicator (Fisher Scientific Model 550).

To reduce the size of recombination blocks and improve the resolution of linkage mapping [89], we then performed the following "intercross" protocol 4 times: 1) Spores were isolated using the Sony SH800 Cell Sorter selecting for $4x10^6$ BFP+ or mCherry+ (but not +/+ or -/-). 2) The sorted cells were grown into 100uL YPD + 40ug/mL tetracycline. 3) Cells were incubated for 16 hours at 30C without shaking. 4) 5mL of YPD + 200ug/mL G418 + 100ug/mL ClonNat + 200ug/mL Hygromycin B was added and cells were incubated for 48 hours at 30C with shaking. 5) Cultures were sporulated and spores were isolated by zymolyase treatment and sonication as described above. Steps 1-5 were repeated 4 times, resulting in a sonicated suspension of spores that had undergone 5 generations of meiosis since the parents. These spores were resuspended in YPD + 0.5ug/mL AbA and incubated at 30C for 16 hours to select for MAT**a** haploids. This haploid culture was split to create a frozen glycerol stock, and was used as the inoculum for phenotypic isolation by FACS (as described above).

Sorting-based bulk-segregant analysis

To sort segregant pools for bulk genotyping, the intercrossed MAT**a**-selected segregants were inoculated from a saturated YPD culture into S + 2% raffinose + AbA at dilutions of 1:200, 1:400, 1:800, and 1:1600, and incubated at 30C for 16-24 hours. The outgrowth culture with OD_{600} closest to 0.1 was selected for each strain, washed once in S, and diluted 1:200 into S + 0.25% glucose + 0.25% galactose + AbA. The glucose-galactose culture was incubated at 30C for 8 hours, and then a Sony SH800 sorter was used to isolate pools of 30,000 cells with the 5% lowest ("OFF") and highest

("ON") YFP expression, among cells whose Back Scatter (BSC) signal was between 10^5 and $3x10^5$. This BSC gate was used to minimize the effects of cell size on expression level as cell with similar BSC have similar cell size. The sorted cells were resuspended in YPD + AbA and incubated at 30C until saturation, about 16-24 hours. An aliquot of this culture was saved for -80C glycerol stocks, and another was used to prepare sequencing libraries.

To sequence the segregant pools, genomic DNA was extracted from 0.5mL of saturated YPD culture of each segregant pool using the PureLink Pro 96 kit (Thermo Fisher K182104A). From these genomic preps, sequencing libraries were made using Nextera reagents (Illumina FC-121-1030) following a low-volume protocol [90]. The input DNA concentration was adjusted so that resulting libraries had mean fragment sizes of 200-300bp as measured on a BioAnalyzer. Libraries were multiplexed and sequenced in an Illumina NextSeq flow cell.

Genome sequences of round-robin parents

Non-S288C parental genomes for the bulk segregant analysis were obtained from the literature: I14 from [38]; BC187, YJM978, DBVPG1106, and Y12 from [91]; YPS606 from [92]. We sequenced our parent strains at ~1x depth and verified their SNP patterns against these datasets. We initially obtained an unpublished sequence for YJM421 from the NCBI Sequencing Read Archive (accessions SRR097627, SRR096491), but this did not match our strain (it appeared similar to YJM326 instead). A RAD-seq SNP profile of YJM421 [37] partially matched our YJM421, but the RAD-seq data displayed heterozygosity. Because we crossed our YJM421 strain to both I14 and DBVPG1106, for which we have high-quality genomes, we could do the linkage mapping given only one parental genome. However, we confirmed that the YJM421 parent used for both crosses were the same strain, by looking at SNPs in the segregant pools of the two crosses that did not match the other parent. Our current hypothesis is that the YJM421 isolate we obtained from the Fay lab (and which was genotyped by RAD-seq in Cromie et al. [37]) was a heterozygous diploid, a haploid spore of which we used as the parent in our round robin cross.

Linkage mapping and loci detection

To perform linkage analysis, we aligned raw reads for parent strains (from the literature) and segregant pools (from our experiments) to the sacCer3 (S288C) reference genome using `BWA-MEM` on the Harvard Medical School Orchestra cluster (http://rc.hms.harvard.edu, see Orchestra High Performance Compute Cluster note below). We identified SNPs between cross parents and determined allele counts at each SNP in segregant pools using `samtools mpileup` and `bcftools call -c`. Using custom MATLAB code, we removed SNPs where read depth was less than 2 or higher than 1000 to avoid alignment artifacts. After filtering, average sequencing depth per pool ranged from 25x to 71x, with a median of 48x.

To calculate LOD scores for allele frequency differences between OFF and ON pools, we input filtered allele counts to the `mp_inference.py` script (MULTIPOOL Version 0.10.2; [41]) with the options `-m contrast -r 100 -c 2200 -n 1000`, following previous practice [38]. A value of n=1000 likely underestimates our segregant pool size and will lead to conservative LOD estimates. An exception to this is the I14xYJM421 cross, which displayed unusually low spore viability (~20%), possibly due to a Dobzhansky-Muller incompatibility [93]. Thus we used n=200 for this cross.

We defined significant loci as LOD peaks where LOD > 10 (Fig. 2.2B). Previous bulk segregant analyses using MULTIPOOL used a less stringent cutoff of LOD > 5 [38,39]. This corresponded to a false discovery rate of 5% in one study [39], but led to a much higher number of unreplicated locus calls in another study [38]. Given that our segregant pools underwent multiple rounds of meiosis (and potentially diversity-reducing selection), we chose to use the more conservative LOD > 10. The choice of LOD does not affect our main conclusions about *GAL3*; even the lowest LOD for the chrIV:460 locus (in YJM978 x Y12) is 24 and thus highly significant (S2 Table). Besides this locus, other moderately significant loci may still be biologically relevant, and so we provide a list of LOD peaks and their corresponding support intervals at LOD > 5 (S2 Table). We clustered these peaks as a single locus if they occur within 20kb of each other from different crosses (S2.8 Fig., S2.2 Table).

CRISPR/Cas9 allele replacement

Allele replacement strains were constructed by knocking out *GAL3* (-890bp from start to +911bp from the stop) with KANMX4 followed by CRISPR/Cas9-mediated markerless integration of the heterologous allele. Initially, strains were prepared by introducing Cas9 on a CEN/ARS plasmid (SLVF11); this plasmid is derived from a previous one [94], but the auxotrophic *URA3* marker was replaced with *AUR1-C* to allow Aureobasidin A selection on prototrophic natural isolates. Then, a donor DNA, a guide RNA insert, and a guide RNA backbone were simultaneously transformed into the strain [45]. The donor DNA contained the new allele (PCR amplified from the desired natural isolate genome), its flanking sequences, and an additional 40bp of homology to target it to the correct genomic locus. The guide RNA insert was a linear DNA containing a SNR52 promoter driving a guide RNA gene containing a 20bp CRISPR/Cas recognition sequence linked to a crRNA scaffold sequence, plus 40bp of flanking homology on both ends to a guide RNA backbone. The guide RNA backbone was a 2u plasmid containing natMX4 (pRS420). This was linearized by NotI + XhoI digestion before transformation. Allele re-integration transformations were plated on cloNAT to select for in vivo assembly of the guide RNA into a maintainable plasmid, and Aureobasidin A to select for presence of Cas9. Successful re-integration was verified by colony PCR and Sanger sequencing was performed on a subset of strains and on all donor DNAs to verify the sequence of allelic variants.

Determining *GAL3* allelic effect by analyzing segregant variance

To estimate the effect of *GAL3* allele on decision threshold, we performed a variance partitioning analysis on decision thresholds of segregants from each of 3 hybrids (Fig. 2.4). Two heterozygous hybrids with homozygous *GAL3* alleles were constructed by mating CRISPR/Cas9 generated allele replacement strains (YJM978 $GAL3^{BC187}$ or BC187 $GAL3^{YJM978}$) to either BC187 or YJM978 wildtype haploids. A "wildtype" hybrid heterozygous at all loci (BC187 x YM978) was also analyzed. These 3 hybrids were sporulated as described above, and the resulting segregants phenotyped for decision threshold in duplicate.

We assumed a model $V_P = V_{GAL3} + V_{BG} + \varepsilon^2$, where phenotypic variance V_P is a sum of contributions from the variance due to *GAL3* V_{GAL3}, the variance due to strain background V_{BG}, and measurement error ε^2. We estimated measurement error by assuming a Gaussian form $\mathcal{N}(\mu, \sigma)$ and fitting it to the differences between replicate measurements across all segregants. The variance in inter-replicate differences should be twice the measurement variance, and thus $\varepsilon = \frac{\sigma}{\sqrt{2}}$. To filter out poor-quality data, we removed segregants where half the inter-replicate difference was greater than 1.5 (S2.3 Fig.). We calculated the mean of each allele population (μ_a or μ_{-a}), where the two allelic variants of *GAL3* are denoted by a and $-a$. To estimate the effect of the *GAL3* allele E_{GAL3}, we divided the difference of the mean of the two populations by 2. The variance due to *GAL3* is the square of E_{GAL3}.

Finally, the phenotypic variance of a segregant population (V_P) is composed of the measurement noise (ε^2) and the genotypic variance (V_G). V_P was calculated for the YJM978 x BC187 segregants and for both of the hybrid conversion segregants. Since *GAL3* is a major driver of the decision threshold phenotype, V_G was partitioned into two components: the contribution to variance of the background (V_{BG}) and the contribution to variance of *GAL3* (V_{GAL3}). The background variance was estimated by subtracting ε^2 and *GAL3* variance from the variance of the segregant population. The *GAL3* contribution ($\frac{V_{GAL3}}{V_G}$) was reported as the ratio of the variance in *GAL3* and the genotypic variance (V_G).

Growth curves and diauxic lag time metric

Growth curves were obtained as described in Wang et al. [6]. In short, growth curves were obtained by manually measuring the absorbance at 600 nm (OD_{600}) on a plate reader (PerkinElmer EnVision) for each plate approximately every 15 min for up to 20 h in a room maintained at 30°C and 75% humidity. Strains to be assayed were pinned into 500 µl of liquid YPD and incubated for 16 h, then diluted 1:200 into 500 µl of synthetic minimal medium + 0.5% glucose and grown for 6-8 h, and

finally diluted 1:150 into synthetic minimal medium + 0.25% glucose + 0.25% galactose or synthetic minimal medium + 0.25% glucose + 0.25% maltose for growth curve measurements. The final inoculation was performed into two different plates (with 2 replicates per plate); these replicate growth curves were nearly indistinguishable for all strains. Analysis of growth curve data was performed in MATLAB using custom-written code [6].

To obtain growth rates in glucose or galactose, additional growth curves were performed as above, except the final culture medium contained 0.5% glucose alone or 0.5% galactose alone. The exponential growth rate was extracted from these data as the mean growth rate between when $OD_{600} = 2^{-6}$ and $OD_{600} = 2^{-4}$ (or, equivalently, when culture density was approximately 1/16 and 1/4 of saturation, respectively).

Bioinformatic analysis

Sequences for the SGRP strains were downloaded from SGRP website. Sequences for the strains in the Liti library [95] were downloaded from https://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html. For the remaining strains with multiple distinct isolates reporter in the literature, a single genetic distance that matched the strain in our collection was selected. Using these sequencing databases, we extracted the *GAL3* region and aligned sequences using MUSCLE (Table S2.3, S2.12 Fig.). Based on the identified SNPs, we used mutfunc (http://mutfunc.com/) to predict the consequences of nonsynonymous SNPs in the *GAL3* variants (Table S2.4). These sequences were used for the McDonald Kreitman analysis using DnaSP [52] (Table S2.5). A neighbor-joining phylogenetic tree was generated using the seqneighjoin function on MATLAB (S2.7 Fig.) and genetic distances [37].

Orchestra High Performance Compute Cluster

Portions of this research were conducted on the Orchestra High Performance Compute Cluster at Harvard Medical School. This NIH supported shared facility consists of thousands of

processing cores and terabytes of associated storage and is partially provided through grant NCRR 1S10RR028832-01. See http://rc.hms.harvard.edu for more information.

Acknowledgments

Yarden Katz and Sarah Boswell for critical reading of the manuscript; Andrew Murray and Chiara Ricci-Tam for helpful discussions; and Shervin Javadi and Stratedigm for flow cytometry assistance.

References

1. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Micro. 2008;6: 613–624. doi:10.1038/nrmicro1932

2. Gancedo JM. Yeast carbon catabolite repression. Microbiology and Molecular Biology Reviews. 1998.

3. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology. Academic Press Inc. (London) Ltd; 1961;3: 318–356. doi:10.1016/S0022-2836(61)80072-7

4. Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, et al. Trait Variation in Yeast Is Defined by Population History. Kruglyak L, editor. PLoS Genet. 2011;7: e1002111–15. doi:10.1371/journal.pgen.1002111

5. Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature. 2005;436: 588–592. doi:10.1038/nature03842

6. Wang J, Atolia E, Hua B, Savir Y, Escalante-Chong R, Springer M. Natural Variation in Preparation for Nutrient Depletion Reveals a Cost–Benefit Tradeoff. Siegal ML, editor. PLoS Biol. 2015;13: e1002041–31. doi:10.1371/journal.pbio.1002041

7. Lambert G, Kussell E. Memory and Fitness Optimization of Bacteria under Fluctuating Environments. Matic I, editor. PLoS Genet. 2014;10: e1004556–10. doi:10.1371/journal.pgen.1004556

8. New AM, Cerulus B, Govers SK, Perez-Samper G, Zhu B, Boogmans S, et al. Different Levels of Catabolite Repression Optimize Growth in Stable and Variable Environments. Doebeli M, editor. PLoS Biol. 2014;12: e1001764–22. doi:10.1371/journal.pbio.1001764

9. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population Diversification in a Yeast Metabolic Program Promotes Anticipation of Environmental Shifts. Siegal ML, editor. PLoS Biol. 2015;13: e1002042–24. doi:10.1371/journal.pbio.1002042

10. Sorrells TR, Booth LN, Tuch BB, Johnson AD. Intersecting transcription networks constrain gene regulatory evolution. Nature. 2015;523: 361–365. doi:10.1038/nature14613

11. Borneman AR, Chambers PJ, Pretorius IS. Yeast systems biology: modelling the winemaker's art. Trends Biotechnol. 2007;25: 349–355. doi:10.1016/j.tibtech.2007.05.006

12. Tuch BB, Li H, Johnson AD. Evolution of Eukaryotic Transcription Circuits. Science. American Association for the Advancement of Science; 2008;319: 1797–1799. doi:10.1126/science.1152398

13. Galgoczy DJ, Cassidy-Stone A, Llinas M, O'Rourke SM, Herskowitz I, DeRisi JL, et al.

Genomic dissection of the cell-type-specification circuit in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences. National Acad Sciences; 2004;101: 18069–18074. doi:10.1073/pnas.0407611102

14. Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, et al. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nat Genet. Nature Publishing Group; 2007;39: 730–732. doi:10.1038/ng2047

15. Bradley RK, Li X-Y, Trapnell C, Davidson S, Pachter L, Chu HC, et al. Binding Site Turnover Produces Pervasive Quantitative Changes in Transcription Factor Binding between Closely Related Drosophila Species. Wray GA, editor. PLoS Biol. Public Library of Science; 2010;8: e1000343. doi:10.1371/journal.pbio.1000343

16. Dalal CK, Zuleta IA, Mitchell KF, Andes DR. Transcriptional rewiring over evolutionary timescales changes quantitative and qualitative properties of gene expression. eLife. 2016. doi:10.7554/eLife.18981.001

17. Tsong AE, Miller MG, Raisner RM, Johnson AD. Evolution of a Combinatorial Transcriptional Circuit. Cell. 2003;115: 389–399. doi:10.1016/S0092-8674(03)00885-7

18. Tsong AE, Tuch BB, Li H, Johnson AD. Evolution of alternative transcriptional circuits with identical logic. Nature. Nature Publishing Group; 2006;443: 415–420. doi:10.1038/nature05099

19. Lavoie H, Hogues H, Whiteway M. Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. Current Opinion in Microbiology. 2009;12: 655–663. doi:10.1016/j.mib.2009.09.015

20. Dalal CK, Zuleta IA, Mitchell KF, Andes DR. Transcriptional rewiring over evolutionary timescales changes quantitative and qualitative properties of gene expression. eLife. 2016. doi:10.7554/eLife.18981.001

21. Hittinger CT, Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch. Nature. 2007;449: 677–681. doi:10.1038/nature06151

22. Nehlin JO, Carlberg M, Ronne H. Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. The EMBO Journal. European Molecular Biology Organization; 1991;10: 3373–3377.

23. Johnston M, Flick JS, Pexton T. Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae. Molecular and Cellular Biology. American Society for Microbiology; 1994;14: 3834–3841. doi:10.1128/MCB.14.6.3834

24. Bennett MR, Pang WL, Ostroff NA, Baumgartner BL, Nayak S, Tsimring LS, et al. Metabolic gene regulation in a dynamically changing environment. Nature. 2008;454: 1119–1122. doi:10.1038/nature07211

25. Escalante-Chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ, et al. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. Proceedings of the

National Academy of Sciences. 2015;112: 1636–1641. doi:10.1073/pnas.1418058112

26. Bhat PJ. Galactose Regulon of Yeast. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. doi:10.1007/978-3-540-74015-5

27. Scott A, Timson DJ. Characterization of the Saccharomyces cerevisiae galactose mutarotase/UDP-galactose 4-epimerase protein, Gal10p. FEMS Yeast Research. The Oxford University Press; 2007;7: 366–371. doi:10.1111/j.1567-1364.2006.00204.x

28. Lohr D, Venkov P, Zlatanova J. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. Federation of American Societies for Experimental Biology; 1995;9: 777–787. doi:10.1096/fj.1530-6860

29. Acar M. A General Mechanism for Network-Dosage Compensation in Gene Circuits . Science. 2010;329: 1656–1660. doi:10.1126/science.1190544

30. Acar M. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435: 228–232. doi:10.1038/nature03524

31. Baganz F, Hayes A, Marren D, Gardner DCJ, Oliver SG. Suitability of replacement markers for functional analysis studies inSaccharomyces cerevisiae. Yeast. John Wiley & Sons, Ltd; 1997;13: 1563–1573. doi:10.1002/(SICI)1097-0061(199712)13:16<1563::AID-YEA240>3.3.CO;2-Y

32. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population genomics of domestic and wild yeasts. Nature. 2009;458: 337–341. doi:10.1038/nature07743

33. Fay JC, Benavides JA. Evidence for Domesticated and Wild Populations of Saccharomyces cerevisiae. PLoS Genet. 2005;1: e5–6. doi:10.1371/journal.pgen.0010005

34. Verma M, Bhat PJ, Venkatesh KV. Quantitative Analysis of GAL Genetic Switch of Saccharomyces cerevisiae Reveals That Nucleocytoplasmic Shuttling of Gal80p Results in a Highly Sensitive Response to Galactose. Journal of Biological Chemistry. 2003;278: 48764–48769. doi:10.1074/jbc.M303526200

35. Venturelli OS, El-Samad H, Murray RM. Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. Proceedings of the National Academy of Sciences. 2012;109: E3324–E3333. doi:10.1073/pnas.1211902109

36. Peng W, Liu P, Xue Y, Acar M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nature Communications. Nature Publishing Group; 1AD;6: 1–9. doi:10.1038/ncomms7226

37. Cromie GA, Hyma KE, Ludlow CL, Garmendia-Torres C, Gilbert TL, May P, et al. Genomic Sequence Diversity and Population Structure of Saccharomyces cerevisiae Assessed by RAD-seq. G3: Genes|Genomes|Genetics. 2013;3: 2163–2171. doi:10.1534/g3.113.007492

38. Treusch S, Albert FW, Bloom JS, Kotenko IE, Kruglyak L. Genetic Mapping of MAPK-Mediated Complex Traits Across S. cerevisiae. Copenhaver GP, editor. PLoS Genet.

2015;11: e1004913–16. doi:10.1371/journal.pgen.1004913

39. Albert FW, Treusch S, Shockley AH, Bloom JS, Kruglyak L. Genetics of single-cell protein abundance variation in large yeast populations. Nature. Nature Publishing Group; 2014;506: 494–497. doi:10.1038/nature12904

40. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. Nature Publishing Group; 2010;464: 1039–1042. doi:10.1038/nature08923

41. Edwards MD, Gifford DK. High-resolution genetic mapping with pooled sequencing. BMC Bioinformatics. BioMed Central Ltd; 2012;13: S8. doi:10.1186/1471-2105-13-S6-S8

42. Lavy T, Kumar PR, He H, Joshua-Tor L. The Gal3p transducer of the GAL regulon interacts with the Gal80p repressor in its ligand-induced closed conformation. Genes & Development. 2012;26: 294–303. doi:10.1101/gad.182691.111

43. Lorenz K, Cohen BA. Small- and Large-Effect Quantitative Trait Locus Interactions Underlie Variation in Yeast Sporulation Efficiency. Genetics. 2012;192: 1123–1132. doi:10.1534/genetics.112.143107

44. Duveau F, Metzger B, Gruber JD, Mack K. Mapping small effect mutations in Saccharomyces cerevisiae: impacts of experimental design and mutational properties. G3: Genes| Genomes| …. 2014. doi:10.1534/g3.114.011783/-/DC1

45. Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, et al. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell Systems. The Authors; 2015;1: 88–96. doi:10.1016/j.cels.2015.02.001

46. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. Nature. 2009;458: 342–345. doi:10.1038/nature07670

47. Bergstrom A, Simpson JT, Salinas F, Barre B, Parts L, Zia A, et al. A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. Molecular Biology and Evolution. 2014;31: 872–888. doi:10.1093/molbev/msu037

48. Lavy T, Kumar PR, He H, Joshua-Tor L. The Gal3p transducer of the GAL regulon interacts with the Gal80p repressor in its ligand-induced closed conformation. Genes & Development. 2012;26: 294–303. doi:10.1101/gad.182691.111

49. Melcher K, Xu HE. Gal80-Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression. The EMBO Journal. 2001.

50. Egriboz O, Goswami S, Tao X, Dotts K, Schaeffer C, Pilauri V, et al. Self-Association of the Gal4 Inhibitor Protein Gal80 Is Impaired by Gal3: Evidence for a New Mechanism in the GAL Gene Switch. Molecular and Cellular Biology. 2013;33: 3667–3674. doi:10.1128/MCB.00646-12

51. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila.

Nature. 1991.

52. Rozas J. DNA Sequence Polymorphism Analysis Using DnaSP. In: Mapelli V, editor. Yeast Metabolic Engineering. Totowa, NJ: Humana Press; 2009. pp. 337–350. doi:10.1007/978-1-59745-251-9_17

53. Elyashiv E, Bullaughey K, Sattath S, Rinott Y, Przeworski M, Sella G. Shifts in the intensity of purifying selection: An analysis of genome-wide polymorphism data from two closely related yeast species. Genome Research. 2010;20: 1558–1573. doi:10.1101/gr.108993.110

54. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang S-P, et al. A Catalog of Neutral and Deleterious Polymorphism in Yeast. Pritchard JK, editor. PLoS Genet. 2008;4: e1000183–15. doi:10.1371/journal.pgen.1000183

55. Roop JI, Chang KC, Brem RB. Polygenic evolution of a sugar specialization trade-off in yeast. Nature. Nature Publishing Group; 2016;530: 336–339. doi:10.1038/nature16938

56. Stern DL, Orgogozo V. THE LOCI OF EVOLUTION: HOW PREDICTABLE IS GENETIC EVOLUTION? Evolution. 2008;62: 2155–2177. doi:10.1111/j.1558-5646.2008.00450.x

57. Stern DL. Is Genetic Evolution Predictable? Science. 2009;323: 741–746. doi:10.1126/science.1158997

58. Peng W, Liu P, Xue Y, Acar M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nature Communications. Nature Publishing Group; 1AD;6: 1–9. doi:10.1038/ncomms7226

59. Lashkari DA, DeRisi JL, McCusker JH. Yeast microarrays for genome wide parallel genetic and gene expression analysis. 1997.

60. Monod J. The growth of bacterial cultures. Annual Reviews in Microbiology. 1949.

61. Orr HA. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. Genetics. 1998.

62. Friesen ML, Saxer G, Travisano M, Doebeli M. EXPERIMENTAL EVIDENCE FOR SYMPATRIC ECOLOGICAL DIVERSIFICATION DUE TO FREQUENCY-DEPENDENT COMPETITION IN ESCHERICHIA COLI. Evolution. 2004;58: 245–17. doi:10.1554/03-369

63. Spencer CC, Saxer G. Seasonal resource oscillations maintain diversity in bacterial microcosms. Evolutionary Ecology …. 2007.

64. Smith EN, Kruglyak L. Gene–Environment Interaction in Yeast Gene Expression. Mackay T, editor. PLoS Biol. 2008;6: e83–15. doi:10.1371/journal.pbio.0060083

65. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. Nature Publishing Group; 2010;464: 1039–1042. doi:10.1038/nature08923

66. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, et al. Dissecting the architecture of a quantitative trait locus in yeast. Nature. 2002;416: 326–330. doi:10.1038/416326a

67. Brem RB. Genetic Dissection of Transcriptional Regulation in Budding Yeast. Science. 2002;296: 752–755. doi:10.1126/science.1069516

68. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, et al. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet. 2003;35: 57–64. doi:10.1038/ng1222

69. Lorenz K, Cohen BA. Causal Variation in Yeast Sporulation Tends to Reside in a Pathway Bottleneck. Gibson G, editor. PLoS Genet. 2014;10: e1004634–12. doi:10.1371/journal.pgen.1004634

70. Gerke JP, Chen CTL, Cohen BA. Natural Isolates of Saccharomyces cerevisiae Display Complex Genetic Variation in Sporulation Efficiency. Genetics. 2006;174: 985–997. doi:10.1534/genetics.106.058453

71. Gerke J. Genetic interactions between transcription factors cause natural variation in yeast. Science. 2009;323: 495–498. doi:10.1126/science.1166426

72. Ruusuvuori P, Lin J, Scott AC, Tan Z, Sorsa S, Kallio A, et al. Quantitative analysis of colony morphology in yeast. BioTechniques. 2014;56: 1–14. doi:10.2144/000114123

73. Taylor MB, Ehrenreich IM. Genetic Interactions Involving Five or More Genes Contribute to a Complex Trait in Yeast. Fay JC, editor. PLoS Genet. 2014;10: e1004324–8. doi:10.1371/journal.pgen.1004324

74. Voordeckers K, De Maeyer D, van der Zande E, Vinces MD, Meert W, Cloots L, et al. Identification of a complex genetic network underlying Saccharomyces cerevisiaecolony morphology. Molecular Microbiology. 2012;86: 225–239. doi:10.1111/j.1365-2958.2012.08192.x

75. Granek JA, Magwene PM. Environmental and genetic determinants of colony morphology in yeast. PLoS Genet. 2010. doi:10.1371/journal.pgen.1000823.g001

76. Sanchez MR, Miller AW, Liachko I, Sunshine AB, Lynch B, Huang M, et al. Differential paralog divergence modulates genome evolution across yeast species. Fay JC, editor. PLoS Genet. 2017;13: e1006585–27. doi:10.1371/journal.pgen.1006585

77. Fay JC. The molecular basis of phenotypic variation in yeast. Current Opinion in Genetics & Development. Elsevier Ltd; 2013;23: 672–677. doi:10.1016/j.gde.2013.10.005

78. Ehrenreich IM, Bloom J, Torabi N, Wang X, Jia Y, Kruglyak L. Genetic Architecture of Highly Complex Chemical Resistance Traits across Four Yeast Strains. Gasch AP, editor. PLoS Genet. 2012;8: e1002570–9. doi:10.1371/journal.pgen.1002570

79. Bloom JS, Kotenko I, Sadhu MJ, Treusch S, Albert FW, Kruglyak L. Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. Nature Communications. Nature Publishing Group; 2015;6: 1–6. doi:10.1038/ncomms9712

80. Chandler CH, Chari S, Dworkin I. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends in Genetics. Elsevier Ltd; 2013;29: 358–366. doi:10.1016/j.tig.2013.01.009

81. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. Nature Publishing Group; 2013;: 1–6. doi:10.1038/nature11867

82. Phillips PC. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9: 855–867. doi:10.1038/nrg2452

83. Taylor MB, Ehrenreich IM. Genetic Interactions Involving Five or More Genes Contribute to a Complex Trait in Yeast. Fay JC, editor. PLoS Genet. 2014;10: e1004324–8. doi:10.1371/journal.pgen.1004324

84. Lee JT, Taylor MB, Shen A, Ehrenreich IM. Multi-locus Genotypes Underlying Temperature Sensitivity in a Mutationally Induced Trait. Siegal ML, editor. PLoS Genet. 2016;12: e1005929–18. doi:10.1371/journal.pgen.1005929

85. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science. 2014.

86. Plucain J, Hindré T, Le Gac M, Tenaillon O. Epistasis and allele specificity in the emergence of a stable polymorphism in Escherichia coli. 2014. doi:10.1126/science.1248688

87. Keren L, Hausser J, Lotan-Pompan M, Slutskin IV, Alisar H, Kaminski S, et al. Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. Cell. Elsevier Inc; 2016;: 1–32. doi:10.1016/j.cell.2016.07.024

88. Hashida-Okado T, Ogawa A, Kato I, Takesako K. Transformation system for prototrophic industrial yeasts using the AUR1gene as a dominant selection marker. FEBS Letters. 1998;425: 117–122. doi:10.1016/S0014-5793(98)00211-7

89. Cubillos FA, Parts L, Salinas F, Bergstrom A, Scovacricchi E, Zia A, et al. High-Resolution Mapping of Complex Traits with a Four-Parent Advanced Intercross Yeast Population. Genetics. 2013;195: 1141–1155. doi:10.1534/genetics.113.155515

90. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. Green SJ, editor. PLoS ONE. 2015;10: e0128036–15. doi:10.1371/journal.pone.0128036

91. Bergstrom A, Simpson JT, Salinas F, Barre B, Parts L, Zia A, et al. A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. Molecular Biology and Evolution. 2014;31: 872–888. doi:10.1093/molbev/msu037

92. Strope PK, Skelly DA, Kozmin SG, Mahadevan G, Stone EA, Magwene PM, et al. The 100-genomes strains, an S. cerevisiaeresource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. Genome Research. 2015;25: 762–774. doi:10.1101/gr.185538.114

93. Hou J, Friedrich A, Gounot J-S, Schacherer J. Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. Nature Communications. Nature Publishing Group; 2015;6: 7214. doi:10.1038/ncomms8214

94. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Research. 2013;41: 4336–4343. doi:10.1093/nar/gkt135

95. Liti G, Louis EJ. Advances in Quantitative Trait Analysis in Yeast. Fay JC, editor. PLoS Genet. 2012;8: e1002912–7. doi:10.1371/journal.pgen.1002912

Chapter 3: Nature uses specific and non-specific GAL genes to quantitatively tune bimodal signaling response of the glucose-galactose circuit

Jue Wang, Julius Palme, Kayla B. Lee, Michael Springer

Adapted from: https://www.biorxiv.org/content/early/2017/04/28/131938

Abstract

Bimodal gene expression by an isogenic population of cells is a pervasive feature of signaling networks, but the mechanisms modulating bimodality are poorly understood. We found that natural yeast strains induce the galactose-utilization (GAL) pathway with a variety of bimodal phenotypes in mixtures of glucose and galactose. This bimodality can be described in terms of two features of the induction profile, representing the fraction of cells that are uninduced and the expression level of the induced subpopulation. We mapped genomic loci underlying these two traits using bulk-segregant analysis, identified causal genes in three loci, and phenotyped allele-replacement strains containing all allelic combinations of these genes. One gene affected only the induced fraction of the GAL response, another affected only the level of induction, and a third gene affected both traits. Our results show that different quantitative features of a bimodal signaling response can be tuned independently by genetic perturbations and that this tuning can change the response from unimodal to bimodal. This modularity may help cells adapt to complex natural environments on physiological as well as evolutionary timescales.

Introduction

Organisms live in complex environments and must respond accordingly in order to survive. These responses can produce phenotypic heterogeneity within a population of isogenic cells, which may in turn have fitness consequences. Examples of this phenotypic variability include the persistence of bacterial cells in the presence of antibiotics or a cell fate switch to terminal differentiation [1]. An extreme example of this heterogeneity is when cells occupy two distinct metabolic states (ON or OFF). However, cells can also display a graded response where the mean metabolic state is modulated by the concentration of the input.

The galactose (GAL) signaling pathway of *S. cerevisiae* is an excellent model for studying phenotypic population heterogeneity, where coupled feedback loops have been shown to tune cellular responses system [2-4]. In response to galactose most cells activate their GAL metabolic genes and occupy an ON state; on the other hand, remaining cells in the OFF state are inhibited by glucose, a repressor of the GAL system. Bimodality of GAL gene expression is attributed to bistability arising from positive feedback through the Gal1p kinase and Gal3p transducer [2,3]. However, perturbing other pathway components such as Gal2p permease, Gal4p activator, and Gal80p repressor also affect quantitative features of the GAL response [3,5-7]. It has also been shown that the modality of the GAL response is affected by the metabolic conditions prior to encountering galactose [8]. Additionally, there is evidence that this switch between states can be modulated as a function of glucose concentration [9]. While much has been learned about heterogeneity in the GAL system [2-4], the vast majority of studies have focused on a small number of related background strains and growth conditions in galactose with non (or minimally) repressive carbon sources. How the bimodal response of the pathway is controlled and vary across perturbations and carbon sources is poorly understood.

In previous work, we found that natural yeast isolates differed widely in the inducibility of

GAL genes in glucose and galactose mixtures [10,11]. In particular, some strains displayed bimodal activation of GAL genes, while other strains were unimodal in the same conditions. Similar population heterogeneity has been seen in yeast maltose utilization [12] and bacterial utilization of various sugar mixtures [13]. This natural variation provides an opportunity to dissect the genetic variants modulating bimodality in nature and expand our knowledge of the repertoire of quantitative behaviors that can be achieved by this model circuit.

In this work, we showed that natural yeast isolates induce the GAL pathway with a diverse array of bimodal and unimodal expression patterns that vary with sugar conditions. We analyzed this variation in terms of two features of the response, representing the fraction of cells that remain uninduced and the expression level the induced subpopulation. Using bulk segregant analysis and CRISPR/cas9 allele replacement, we identified genetic variants underlying these two traits. Additionally, we found that the metabolic history of cells before inducing GAL genes and the carbon sources present also affects the bimodal response. The independent tuning of these two quantitative features of the GAL response can account for the diversity of unimodal and bimodal phenotypes observed in our natural isolates. This genetic flexibility may be advantageous for cells adapting to complex natural nutrient environments.

Results

Natural yeast isolates vary in the degree of bimodality of GAL induction

To study natural variation in the population behavior of the GAL pathway response, we measured the expression of a GAL1 promoter driving YFP (GAL1pr-YFP) in 34 geographically and ecologically diverse yeast strains [11,14,15] grown in a titration of glucose plus a constant level of galactose. We measured the GAL reporter response in a titration of glucose concentrations from 2% to 0.004% w/v in a background of constant 0.25% galactose. We previously reported a metric called the "decision threshold" to describe the glucose concentration where 50% of the cells have greater-

than-basal expression of the GAL reporter [10]. This metric is similar to those used in previous work [6,16], and focuses on *when* a cell decides to induce a pathway while differentiating it from *how strongly* a cell responds once induced.

As expected, we found that all strains are uninduced in high glucose and fully induced in low or no glucose (Fig. 3.1A). However, at intermediate glucose concentrations, some strains display a unimodal population with intermediate (i.e. sub- maximal) GAL expression (Fig. 3.1A, 3.1B), while other strains display a bimodal mixture of uninduced and partially induced cells (Fig. 3.1C, 3.1D). Additionally, strains with the same modality still have quantitatively different GAL induction profiles (S3.1 Fig.), raising the question of what mechanisms can give rise to these diverse signaling phenotypes.

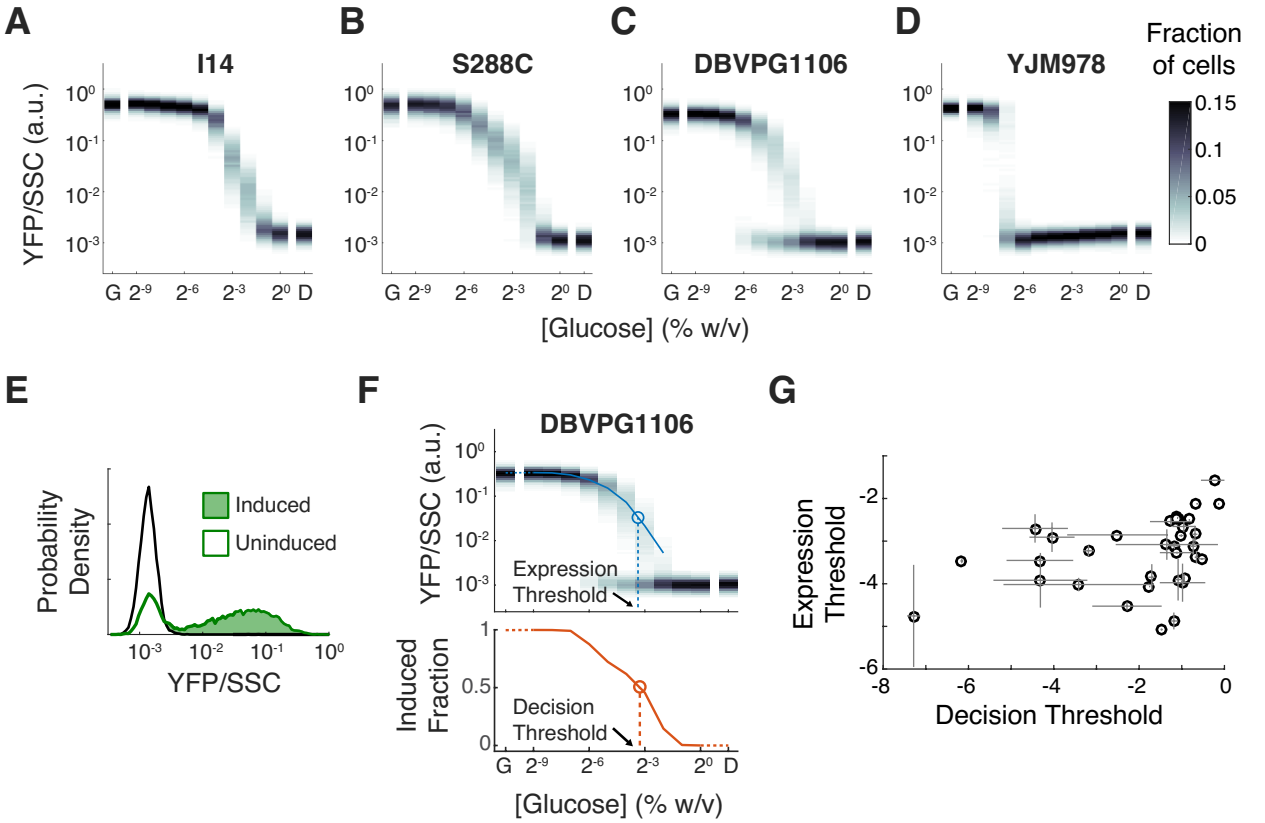


Figure 3.1. Natural variation in GAL induction can be mathematically decomposed into two uncorrelated features

Figure 3.1 (continued)

Each plot is a series of YFP fluorescence (normalized to side scatter "SSC") histograms from 12 sugar conditions for strains (A) I14, (B) S288C, (C) DBVPG1106, and (D) YJM978. Other phenotyped strains are shown in Figure S3.1. Darker regions represent more frequently observed YFP values. The middle 10 conditions in each plot are 0.25% galactose + the indicated concentrations of glucose. The first and last conditions contain only one sugar: "D", 2% glucose; "G", 2% galactose. (E) Identification of induced cell subpopulation (green shading) using a reference distribution from 2% glucose (black histogram) (Materials and Methods). (F) Calculation of induced level (blue line) and induced fraction (orange line) from a GAL induction profile of strain DBVPG1106. Expression threshold and decision threshold are calculated from induced level and fraction as described in the text. (G) Scatterplot of expression threshold versus decision threshold across 36 *S. cerevisiae* natural isolates. Plotted are the mean and S.D. of 3-10 replicates.

Variation in GAL bimodality phenotypes can be parameterized by two metrics

Upon close inspection, the GAL induction phenotypes generally seem to be a mixture of two components: an induced subpopulation that decreases in YFP level as glucose increases, and an uninduced subpopulation that remains at the same YFP level regardless of glucose concentration (S3.2A Fig.). The mixing of these components can be quantified as an induced fraction that decreases as a function of glucose concentration (S3.2B Fig.). Simply by varying the glucose-dependence of the induced fraction and of the induced subpopulation expression level, we can simulate many bimodal phenotypes, as well as unimodal phenotypes, reminiscent of the observed data (S3.1, S3.2B-C Fig.). In this framework, a strain which is unimodal in a particular condition has an induced fraction of one (but a sub-maximal induced level), while a strain that is bimodal in this condition has an induced fraction of less than one.

Applying this population decomposition framework to our data, we computationally separated induced and uninduced cells from each GAL reporter distribution (Fig. 3.1E) and calculated

two summary metrics for each strain's phenotype: E_{10}, the glucose concentration where the induced subpopulation reaches 10% of its maximal GAL expression level, and F_{50}, the glucose concentration where 50% of cells in the population are induced (Fig. 3.1F). For convenience, these metrics are in log2-transformed units, so a strain with $E_{10} = -1$ has an induced subpopulation that reaches 10% of maximal induction at $2^{-1} = 0.5\%$ w/v glucose. We find that the E_{10} and F_{50} are uncorrelated across natural isolates, suggesting the possible existence of genetic changes that can decouple them (Fig. 3.1G).

Presence of background carbon sources can tune bimodality

We measured galactose induction in the presence of glucose, and in intermediate mixtures of glucose and galactose, we see that some natural isolates have bimodal induction. Since the repression of glucose and activation of galactose have opposing effects on the strains, it can be difficult to differentiate the impact of either on the bimodal phenotype. Other studies have used galactose gradients with low concentrations of background sugars that have a minimally repressive (glucose or mannose) or non-repressive (raffinose) to observe bimodality in the GAL pathway. To better understand how other sugars impact bimodality, we grew strains in 2% raffinose for 14-16 hours followed by an 8-hour induction in a galactose gradient with 0.1% mannose [16], 0.1% glucose [6], 2% raffinose [2], or no background carbon source (S3.3 Fig.). Notably, the background sugars, glucose and mannose, had similar induction profiles, and raffinose and no background sugar had similar induction profiles. This presents the possibility that similarities in induction are a result of the role that the repression plays on the GAL pathway. Bimodal GAL induction was only present in repressive or minimally repressive conditions, which contradicts other studies that showed bimodality in GAL induction independent of the background carbon [2,3,6,16]. These differences are likely caused by pre-growth conditions, the metabolic state of the cell in the sugar gradients, a combination of the two, or even the specific strain that was used in the study.

Bulk segregant analysis identifies genetic loci associated with GAL induction variation

To analyze the genetic basis of E_{10} and F_{50}, we crossed strains S288C and DBVPG1106 and phenotyped random haploid segregants from their hybrid. These parent strains differ in both traits, and their segregants display semi-continuous, correlated variation in these traits with a small number of outliers. Therefore, E_{10} and F_{50} are likely modulated by multiple genes, at least some of which affect both traits.

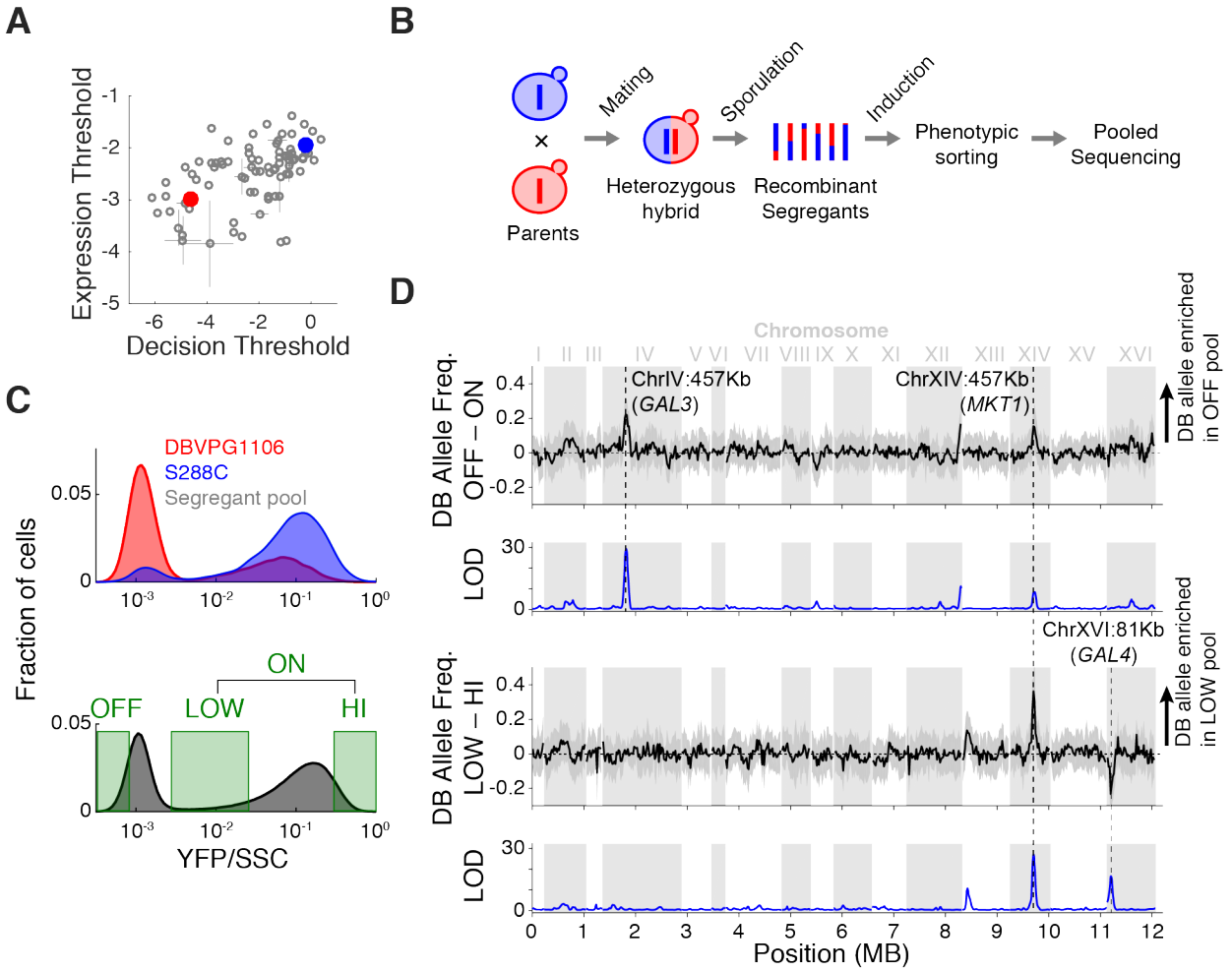


Figure 3.2. Bulk segregant analysis of expression threshold and decision threshold

(A) Expression threshold versus decision threshold across 90?? haploid segregants of the DBVPG1106 x S288C cross. Parent phenotypes are shown as filled circles: DBVPG1106 (red), S288C (blue). (B) Schematic of bulk segregant analysis strategy. (C) GAL reporter histograms of parent strains DBVPG1106 (red) and S288C (blue) and a pool of haploid segregants (black, bottom) in the

Figure 3.2 (continued)

sorting conditions, 0.25% glucose + 0.25% galactose. Green boxes are a schematic of the gates used to sort segregant cells into 3 phenotyped pools for sequencing (Gates used in actual sorting experiment are shown in Figure S3.4). ON pool allele counts are a computational sum of the LOW and HI pool allele counts (Materials and methods). (D) Genome-wide plots of differential allele frequency and log-odds-ratio (LOD) as computed by the MULTIPOOL algorithm (see Main Text, Materials and Methods). Top two panels show the OFF/ON comparison; bottom two panels show the LOW/HIGH comparison.

To identify these genes, we performed bulk-segregant linkage mapping using a pooled sorting strategy (Fig. 3.2B). We chose a glucose+galactose condition where the parental GAL1pr-YFP distributions were maximally different and used it to induce a pooled mixture of haploid (MATa) segregants (Fig. 3.2C). We then used FACS to sort the segregants into three pools: (1)uninduced ("OFF"), (2) induced and low-expression ("LOW"), and (3) induced and high-expression ("HI") cells (Fig. 3.2C). Each pool was sequenced to 15-33x coverage. We expected that a genomic locus affecting the induced level (and thus E_{10}) will differ in allele frequency between the LOW and HI pools, while any locus affecting the induced fraction (and thus F_{50}) would differ in parental allele frequency between the OFF pool and a computationally merged LOW+HI pool ("ON") (Materials and Methods).

We found 5 loci with significantly different allele frequencies between OFF/ON pools or between LOW/HI pools, defined as genomic regions with a peak log-odds (LOD) score > 5 calculated by MULTIPOOL [17] (Fig. 3.3D; Materials and Methods). To look for causal variants, we inspected gene annotations in a region of 2-LOD decrease around each LOD peak. The three most significant loci are centered at chrIV:457Kb, chrXIV:457Kb, and chrXVI:81Kb and contain the genes *GAL3*, *MKT1*, and *GAL4*, respectively (Fig. 3.3D). *GAL3* and *GAL4* are direct regulators of galactose sensing, while *MKT1* is known to have pleiotropic effects in crosses between S288C and natural isolates [18-21]. The *GAL3*-associated locus was significant only in the OFF/ON comparison, while the *GAL4*-

associated locus was only significant in the LOW/HI comparison, suggesting that the effect of these loci are specific to either E_{10} or F_{50}. The *MKT1*-associated locus was significant in both comparisons but had a higher LOD score in the LOW/HI than in the OFF/ON comparison. Unlike the other loci, the *GAL4*-associated locus was enriched for the S288C allele in the DBVPG1106-like segregant pool, suggesting the possibility of transgressive segregation. Two other loci, at chrXII:1053Kb and chrXIII:105Kb, were also significant and seemed to have phenotype-specific effects, but did not contain any obvious genes for follow-up. Therefore, we focused on the chrIV (*GAL3*-associated), chrXIV (*MKT1*-associated), and chrXVI (*GAL4*-associated) loci for further investigation.

GAL3 and GAL4 alleles specifically affect F_{50} and E_{10}, while MKT1 alleles affect both traits

To test if *GAL3*, *MKT1*, and *GAL4* alleles are causal variants in the chrIV, chrXIV, and chrXVI loci, we used CRISPR/cas9 to replace the coding region and flanking regions of each gene (Materials and Methods) in both DBVPG1106 and S288C with the allele from the other parent (S3.4 Fig.). In DBVPG1106, replacing the endogenous *GAL3* allele with $GAL3^{S288C}$ shifted F_{50} in the direction of the S288C parent (S3.4A Fig.). *MKT1* replacement also shifted F_{50} and had a small effect on E_{10} as well. *GAL4* replacement had a small but clear effect on E_{10} and no detectable effect on F_{50}. In the S288C background, allele replacements had similar trait-specificity but much smaller effects (S3.4B-C Fig.). In both parental backgrounds, the *GAL4* allele replacement resulted in a change in E_{10} away from the value of the other parent. This is consistent with the sign of allele-frequency differences of the *GAL4*-containing locus between LOW and HI pools (Fig. 3.2D) and confirms that *GAL4* is a transgressive allele in this cross. Overall, these results show that *GAL3*, *MKT1*, and *GAL4* are causal variants in their respective loci and corroborate the allelic effects inferred from our bulk segregant analysis. The single allele replacements only modestly altered the phenotype of the parent strains, suggesting that other genes make substantial contributions to the total phenotypic difference. Alternatively, there may be genetic interactions between our mapped genes such that allele replacement of 2 or 3 of them is sufficient to achieve conversion of one parental phenotype to the other. To assess these

possibilities, we constructed all 16 combinations of strain background, *GAL3* allele, *MKT1* allele, and *GAL4* allele from either the DBVPG1106 or S288C parent; and measured E_{10} and F_{50} of 2 independent isolates of each of the 16 genotypes. We examined the resulting phenotypic landscape (Fig. 3.3A) in terms of pairs of strains differing in the allelic status of one gene (or strain background) while other genetic factors are held constant. The effect of switching from the DBVPG1106 genetic variant to the S288C variant can be visualized as a vector in E_{10} versus F_{50} space (Fig. 3.3B-E) or as a trait difference (Fig. 3.3F).

This analysis reveals that the trait-specificity of single genetic changes are broadly consistent across different genotypic backgrounds (i.e. combinations of strain background and alleles at the other loci). This can be seen in the fact that effect vectors from DBVPG1106 to S288C variants in E_{10}-F_{50} space are parallel (Fig. 3.3B-E), or equivalently, that differential effects cluster by angle to the origin (Fig. 3.3F). Across the combinatorial allele replacement strains, it is clear that *GAL3* allele predominantly affects F_{50}, *GAL4* mostly affects E_{10}, and *MKT1* affects both traits. For example, a strain with $GAL4^{S288C}$ has a lower E_{10} than the congenic strain with $GAL4^{DBVPG1106}$ for all such strain pairs. These results show that E_{10} and F_{50} can be tuned independently in this cross.

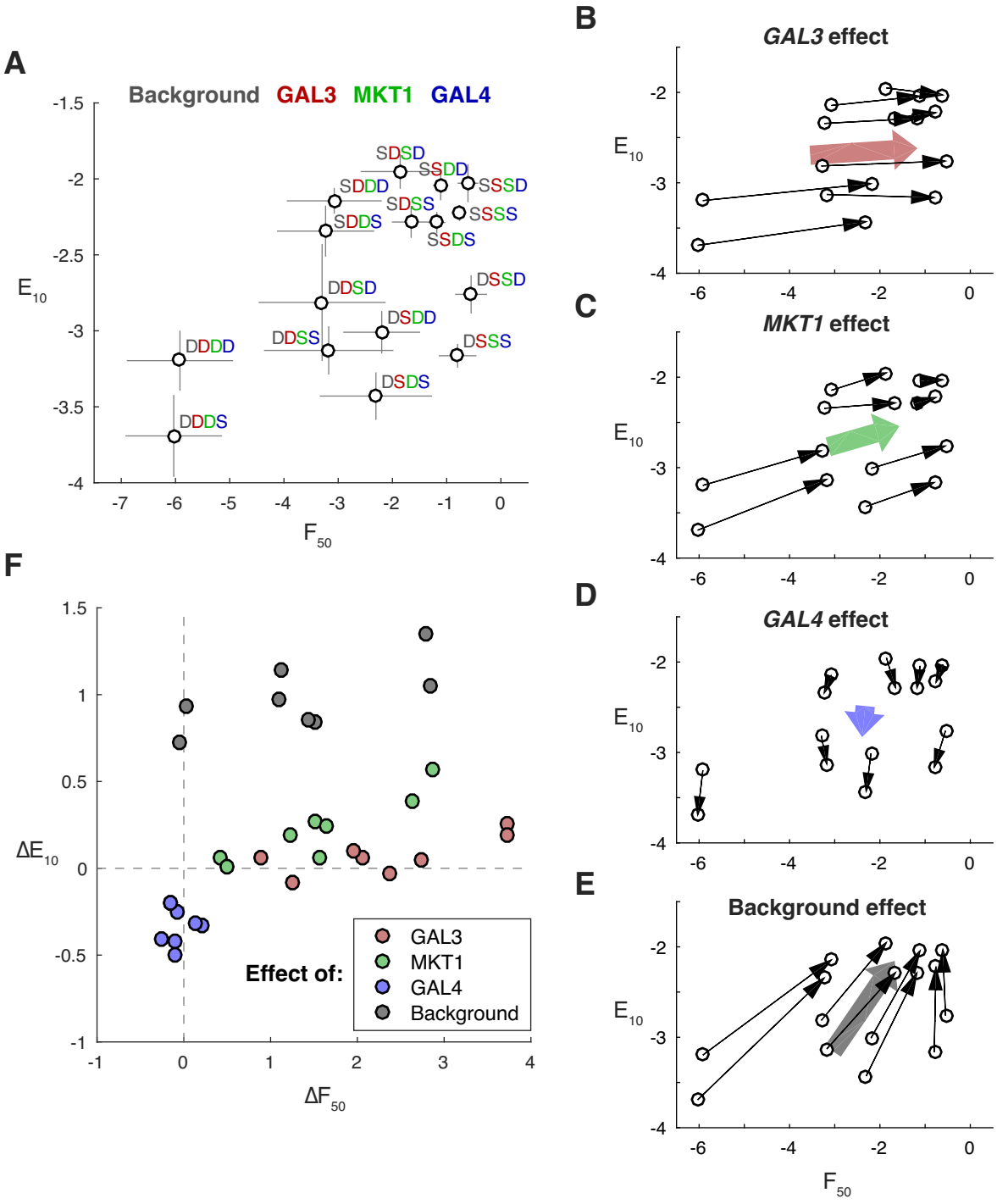


Figure 3.3. Combinatorial effects of strain background and *GAL3*, *MKT1*, and *GAL4* alleles.

(*A*) E_{10} versus F_{50} for all 16 combinations of S288C ("S") or DBVPG1106 ("D") strain background (gray letters), *GAL3* allele (red), *MKT1* allele (green), and *GAL4* allele (blue). Effects of switching from

Figure 3.3 (continued)

DBVPG1106 to S288C variant while holding other genetic variables constant are shown as arrows for switching (B) *GAL3* allele, (C) *MKT1* allele, (D) *GAL4* allele, or (E) strain background. (F) The effects shown in (B)-(E) are plotted as differences in E_{10} versus differences in F_{50}.

In contrast to trait-specificity, the effect sizes of single genetic changes across the combinatorial landscape are more complicated. In general, *GAL3* allele effects on F_{50} are large, spanning up to half the phenotypic distance between the parents. *MKT1* allele effects on F_{50} are almost as great, and combined with *GAL3* allele replacement, can essentially phenoconvert DBVPG1106 to S288C, but only along the F_{50} axis (DSSD versus in DDDD in Fig. 3.3A). However, the reciprocal replacement in the S288C background has a more modest effect (SDDS versus SSSS in Fig. 3.3A). Consistent with these findings, the strain background effect on F_{50} (which can be interpreted as the residual variation after allele replacements) varies widely, from negligible to almost as large as that of *GAL3* or *MKT1*. For the E_{10} trait, *GAL4* allele effects span between a third and half the phenotypic distance between the parents, but in the opposite direction required for phenoconversion. Therefore, triple allele replacement strains DSSS (DBVPG1106 $GAL3^{S288C}$ $MKT1^{S288C}$ $GAL4^{S288C}$) and SSSD still differ substantially in E_{10} from their respective wildtype SSSS and DDDD strains. Overall, strain background has effects on both E_{10} and F_{50} in most genotype backgrounds, indicating substantial variation in both traits not accessed by our allele swaps.

Independent tuning of F_{50} and E_{10} modulates the modality of the GAL response

Our results above show that F_{50} can be tuned independently of E_{10} by some genetic variants in the S288C x DBVPG1106 cross. To see if this is true over a larger range of F_{50}, we analyzed phenotypic data on S288C, BC187, and DBVPG1106 strains whose *GAL3* loci have been replaced with a panel of natural *GAL3* alleles that we previously showed to underlie a spectrum of GAL inducibility phenotypes [10]. As expected, F_{50} varied widely as the *GAL3* allele is changed (Fig. 3.4A). However,

variation in E_{10} with *GAL3* allele was minimal and driven almost entirely by the strain background (Fig. 3.4B).

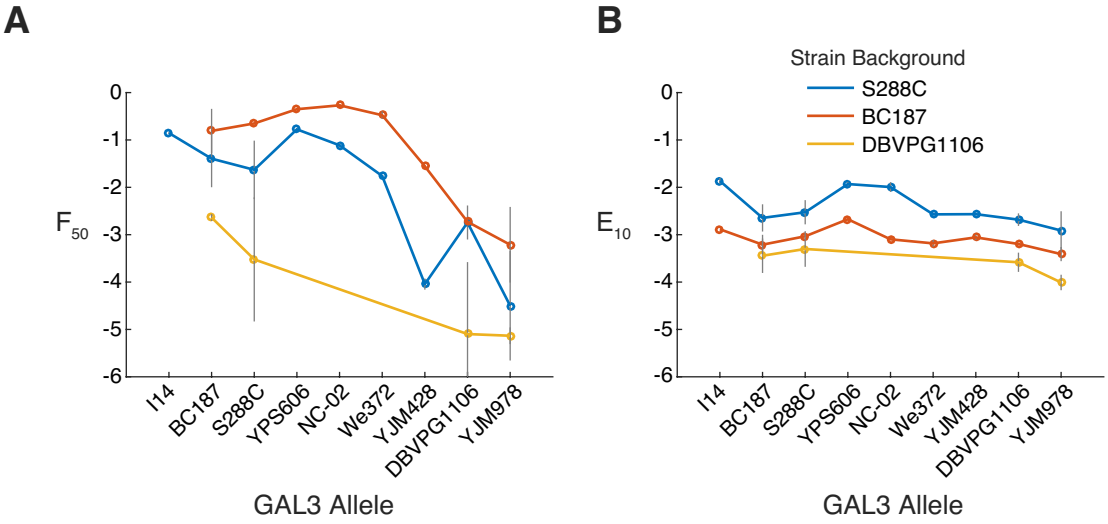


Figure 3.4. Perturbations that affect F_{50} do not affect E_{10}

(A) F_{50} and E_{10} for allele-replacement strains with S288C, BC187, or DBVPG1106 genetic backgrounds but containing alleles of GAL3 from various other natural isolates.

The definitions of the E_{10} and F_{50} metrics (Fig. 3.1F, S3.2 Fig.) imply that tuning either parameter independently should alter the apparent number of modes in GAL expression distributions. Since F_{50} varies over a wider range of glucose concentrations than E_{10} does under the perturbations we tested, we asked if independently tuning F_{50} affects modality. Indeed, plotting GAL reporter distributions shows that a number of allele replacements are able to convert strains from being bimodal to unimodal and vice versa. For example, DBVPG1106 is bimodal, but replacing alleles with $GAL3^{S288C}$ and $MKT1^{S288C}$ increases its F_{50} and makes it unimodal (Fig. 3.5A-B). Conversely, BC187 is unimodal, but decreasing its F_{50} by introducing $GAL3^{YJM978}$ makes it bimodal (Fig. 3.5C-D). Finally, Y12-WashU, one of the most obviously bimodal strains, is rendered unimodal when pre-conditioned in acetate rather than raffinose before galactose induction (Fig. 3.5E-F), suggesting that the environment can also perturb the circuits modality. This parallels the effect of *GAL3* alleles and

suggests that the independent tuning of E_{10} and F_{50} is a consequence of how the GAL circuit is integrated with carbon metabolism more broadly.

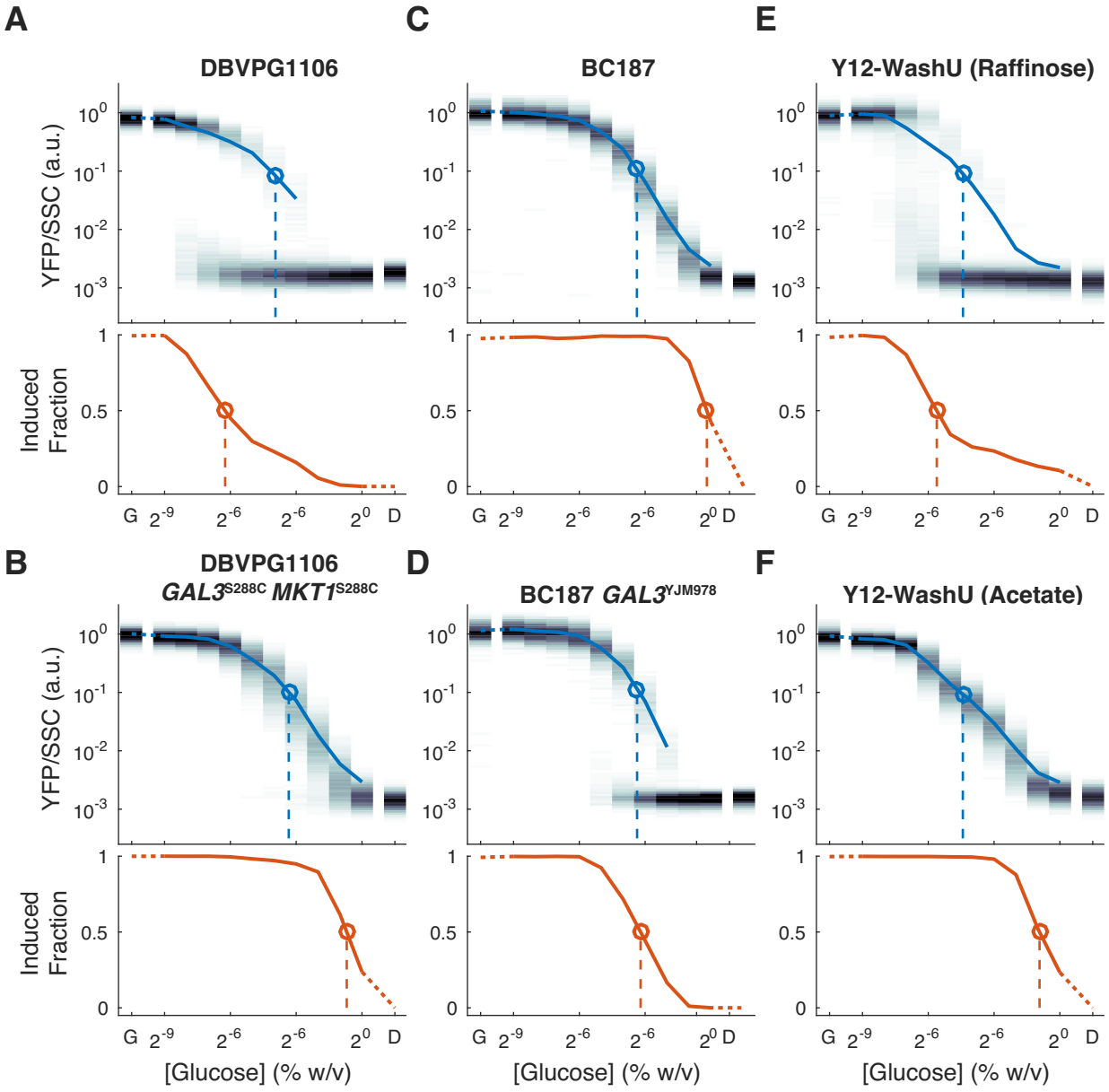


Figure 3.5. Changing F_{50} changes the number of modes of the GAL response

Plotted are GAL reporter histogram series on a glucose gradient + galactose, as in Figure 4.1F, with induced expression level (blue line), E_{10} (dotted vertical blue line), induced fraction (orange line), and F_{50} (dotted vertical orange line). These plots show one representative experiment for each strain/condition, out of the 3-12 replicates plotted in Figures 4.3 and 4.4. Strains: (A) DBVPG1106

Figure 3.5 (continued)

with all endogenous alleles; (B) DBVPG1106 with replacements by *GAL3* and *MKT1* alleles from S288C. (C) BC187 with endogenous alleles; (D) BC187 with replacement by *GAL3* allele from YJM978; (E) Y12-WashU cultured in raffinose prior to glucose + galactose (standard protocol); (F) Y12-WashU pre-cultured in acetate.

Discussion

Independent tuning and the molecular mechanism of bimodality

It is known that positive feedback on GAL gene expression through *GAL3* tunes the switching rate of cells between uninduced and induced states [3] and is a key contributor to the bistability of the pathway [2]. Changes in *GAL3* dosage affects the induced fraction of GAL genes [6], and a panel of natural *GAL3* variants confers a spectrum of GAL induction phenotypes [10]. We put these previous observations in context by showing that natural *GAL3* alleles specifically affect the sugar threshold where individual cells to switch to an induced state, while the level of induction in that state is set by *GAL4* and other unknown genes. Both these features combine to yield the population level behavior of the circuit, including apparent patterns of bimodality. Underscoring this point, we found that *GAL3* allele replacement is sufficient to convert a unimodal response to bimodal, and vice versa, while the level of the induced subpopulation remains unchanged. This degree of modularity in the quantitative behavior of the GAL circuit was previously unappreciated.

GAL4 is the transcription factor activating all inducible GAL genes [22,23]. Previously, changes in dosage of *GAL4* was found to have no effect on the GAL induced fraction [6]. The S288C variant of *GAL4* contains a non-conservative R95G mutation, as well as a conservative R508K mutation, relative to DBVPG1106 and other natural isolates. These mutations suggest that the S288C and DBVPG1106 *GAL4* alleles might differ in their ability to activate transcription of GAL genes. This effect could be specific to induced level if differences in *GAL4* activity only affect GAL promoters that

are in an active state and the latter variable is separately dictated by feedback loops such as *GAL3*. An important question for future work is whether this scenario is quantitatively plausible in a mathematical model of the GAL circuit, and what general features of this and other circuits allow for independent tuning.

Modularity of the GAL pathway and genetic background

We also find that *MKT1* alleles affect the GAL response and can play almost as large a role as *GAL3*. *MKT1* is involved in maintaining killer toxin [24], regulating translation [25], and affects numerous traits in crosses between S288C and natural isolates [19,26-30]. The S288C allele of *MKT1* is a loss-of-function variant relative to natural alleles and causes lower expression of mitochondrial genes [31,32]. In turn, deletion mutants of mitochondrial genes tend to exhibit aberrant GAL induction; this effect is more pronounced on the induced fraction than on induced level [33], echoing our observations. Therefore, it is likely that the effect of *MKT1* allele on GAL induction is due to perturbations to mitochondrial function.

We found that much of the variation in E_{10}, and to a lesser extent F_{50}, must be attributed to unknown alleles in the genetic background. This dovetails with other recent reports that many traits in yeast are dominated by large effects from one or a few loci but can be tuned quantitatively by many small-effect loci [10,33,34]. Moreover, *MKT1* is not a member of the canonical GAL pathway, and nor are any genes in 2 other loci that reached significance in our linkage mapping. Combined with observations that deletion mutants of up to quarter of all yeast genes have quantitatively perturbed GAL signaling [33], our results indicate that decision-making circuits are not as modular with respect to genetic variation as is often assumed.

Our results indicate that memory of metabolic state is encoded by the GAL circuit and persists even after the cells have reached steady-state in inducing conditions (Fig. 3.5B, Materials and Methods). This appears to be a distinct phenomenon from the "memory" of glucose or galactose pre-induction conditions previously attributed to bistability of the GAL network [3,35]. However, the fact

that pre-induction carbon source specifically affects F_{50}, just as *GAL3* allele does, suggests that this positive feedback loop may be a nexus of regulation of GAL genes by multiple signals in the cell. Indeed, recently it was shown that NAD(P) can directly bind Gal80p and thereby impact downstream GAL pathway expression [36,37]. Since these studies relied on bulk measurements, it will be interesting to revisit these investigations using quantitative, single-cell readouts of pathway behavior.

Physiological and ecological role of independent tuning

Our results raise the question of why independent tuning of induced fraction and induced level would exist in nature. Previously, we showed that natural variation in the timing of GAL induction during diauxic growth leads to a fitness tradeoff—some strains prepare for glucose exhaustion at an upfront cost while others maximize their growth rate on glucose but suffer a diauxic lag [11]. Related work showed that both strategies could be implemented by the same strain as part of a bimodal response [38], and that this may be an evolutionarily stable strategy [39]. Under this framework, tuning E_{10} and F_{50} separately would allow the timing of the inducing population, and its level of induction, to evolve separately. This could provide fitness benefits in certain conditions, although exactly what these conditions are would depend on the quantitative details of the costs and benefits of induction, an interesting issue to be explored in future work.

Materials and Methods

Strains and media

Strains were obtained as described in [11]. An initial set of 42 strains were assayed in glucose gradients + galactose. Strains CLIB324, L-1528, M22, W303, YIIC17-E5, YJM975, YJM981 were excluded from downstream analysis due to poor growth in our media conditions. Strain 378604X

was also excluded due to a high basal expression phenotype that was an outlier in our collection. The genetic basis of this behavior is likely an interesting topic for follow-up studies.

All experiments were performed in synthetic minimal medium, which contains 1.7g/L Yeast Nitrogen Base (YNB) (BD Difco) and 5g/L ammonium sulfate (EMD), plus D-glucose (EMD), D-galactose (Sigma), or raffinose (Sigma). Cultures were grown in a humidified incubator (Infors Multitron) at 30°C with rotary shaking at 230rpm (tubes and flasks) or 999rpm (500uL cultures in 1.2mL 96-well plates).

Flow cytometry assay in glucose gradient

GAL induction experiments were performed in a 2-fold dilution series of glucose concentration, from 2^{0}% to 2^{-9}% w/v, with constant 0.25% galactose. 2% glucose and 2% galactose conditions were also included with each glucose titration experiment. To assess and control for well-to-well variation, experiments were performed as a co-culture of a "query" strain to be phenotyped and as a "reference" strain that was always SLYB93 (natural isolate YJM978 with constitutive mCherry segmentation marker).

To start an experiment, cells were struck onto YPD agar from -80C glycerol stocks, grown to colonies, and then inoculated from colony into YPD liquid and cultured for 16-24 hours. Then, query and reference strain cultures were mixed 9:1 by volume and inoculated in a dilution series (1:200 to 1:6400) in S + 2% raffinose medium. The raffinose outgrowths were incubated for 16-20 hours, and then their optical density (OD_{600}) was measured on a plate reader (PerkinElmer Envision). One outgrowth culture with OD_{600} closest to 0.1 was selected for each strain, and then washed once in S (with no carbon sources). Washed cells were diluted 1:200 into glucose + galactose gradients in 96-well plates (600uL cultures in each well) and incubated for 8 hours. Then, cells were harvested and fixed by washing twice in Tris-EDTA pH 7.5 (TE) and resuspended in TE + 0.1% sodium azide before transferring to a shallow microtiter plate (CELLTREAT) for measurement. Flow cytometry was performed using a Stratedigm S1000EX with A700 automated plate handling system. Data analysis

was performed using custom MATLAB scripts, including Flow-Cytometry-Toolkit (https://github.com/springerlab/Flow-Cytometry-Toolkit).

Experiments using acetate as pre-induction carbon sources were done as above, except S + 2% potassium acetate were used instead of raffinose medium for the outgrowth step.

Crossing and generating segregants

To prepare parent strains for crossing and sporulation, we sporulated diploid natural isolates bearing the hoΔ::GAL1pr-YFP-hphNT1 reporter cassette and isolated random spores that displayed MAT**a** or MATα phenotypes in test crosses. We then introduced a constitutive fluorescent marker in tandem with the GAL reporter, to obtain MAT**a**; hoΔ::GAL1pr-YFP-mTagBFP2-kanMX4 or MATα; hoΔ::GAL1pr-YFP-mCherry-natMX4 parent strains. To the MAT**a** parent we also introduced a pRS413-derived plasmid bearing STE2pr-AUR1-C and hphNT1. This plasmid is maintained by hygromycin selection but also allows selection for MAT**a** cells by Aureobasidin A [40]. This plasmid design is inspired by a similar mating-type selection plasmid used in a recent study [41].

To isolate segregants for phenotyping, we crossed a parent with BFP-kanMX + MAT-selection plasmid to a parent with mCherry-natMX and isolated a $G418^{R}Nat^{R}Hyg^{R}$ diploid hybrid with the plasmid. We sporulated the hybrid by culturing it to saturation in YPD, diluting 1:10 in YP+2% potassium acetate and incubating at 30C for 8 hours, and washing and resuspending into 2% potassium acetate and incubating at 30C until >20% of cells were tetrads, or about 3 days. We incubated ~$5x10^{6}$ tetrads in 100uL water with 50U of zymolyase 100T (Zymo Research) for 5 hours at 30C, and then resuspended tetrads in 1mL of 1.5% NP-40 and sonicated for 10 seconds at power setting 3 on a probe sonicator (Fisher Scientific). The resulting segregants were plated on YPD + 0.5ug/mL Aureobasidin A ("AbA"; Clontech) and random colonies were picked into YPD liquid and saved as glycerol stocks. Haploidy was confirmed by mating to tester strains with known mating type. 90 segregants were phenotyped for GAL induction as described above.

Sorting-based bulk-segregant analysis

To generate segregant pools, we prepared a diploid hybrid and sporulated it as described above. To reduce the size of recombination blocks and improve the resolution of linkage mapping [42], we then performed the following "intercross" protocol 4 times: from spore suspension, used Sony SH800 Cell Sorter to sort $4x10^6$ BFP+ or mCherry+ (but not +/+ or -/-) cells into 100uL YPD + 40ug/mL tetracycline; incubate for 16 hours at 30C without shaking; add 5mL YPD + 200ug/mL G418 + 100ug/mL ClonNat + 200ug/mL Hygromycin B and incubate 48 hours at 30C with shaking; sporulate cultures and prepare sonicated spore suspension. After the 4^{th} sporulation cycle, the sonicated spores were resuspended in YPD + 0.5ug/mL AbA and incubated at 30C for 16 hours. This culture was frozen as a glycerol stock, as well as used to inoculate the galactose-induction sorting experiment.

To sort segregant pools for bulk genotyping, we inoculated the intercrossed, MAT**a**-selected segregants from a saturated YPD culture into S + 2% raffinose + AbA at dilutions of 1:200, 1:400, 1:800, and 1:1600, and incubated at 30C for 16-24 hours. We chose the raffinose culture with OD closest to 0.1, washed once in S (0.17% Yeast Nitrogen Base + 0.5% Ammonium Sulfate), and diluted 1:200 into S + 0.25% glucose + 0.25% galactose + AbA. We incubated the glucose-galactose culture at 30C for 8 hours, and then used a Sony SH800 sorter to isolate pools of 30,000 cells with the 5% lowest ("OFF") and highest ("HI") YFP expression, among cells whose Back Scatter (BSC) signal was between 10^5 and $3x10^5$. The "LOW" pool was similarly obtained, but from the 5% of cells with lowest non-basal expression (S3.4 Fig.). The sorted cells were resuspended in YPD + AbA and incubated at 30C until saturation, about 16-24 hours. An aliquot of this culture was saved for -80C glycerol stocks, and another was used to prepare sequencing libraries.

To sequence the segregant pools, we extracted genomic DNA from 0.5mL of saturated YPD culture of each segregant pool using the PureLink Pro 96 kit (Thermo Fisher K182104A). From these genomic preps, we made sequencing libraries using Nextera reagents (Illumina FC-121-1030)

following a low-volume protocol [43]. We adjusted the input DNA concentration so that resulting libraries had mean fragment sizes of 200-300bp as measured on a BioAnalyzer. Libraries were multiplexed and sequenced in an Illumina NextSeq flow cell to a depth of 16-33x.

Reads from the Illumina sequencing were aligned to the sacCer3 reference genome using bwa mem, and SNP counts were generated using samtools mpileup, on the Harvard Medical School Orchestra cluster. These outputs were processed in MATLAB using custom code as follows: SNPs with coverage less than 2 or more than 1000 were removed. The LOW and HI pools were computationally merged into an ON pool. To make sure the two pools contributed equally to the merged pool, at each SNP, allele counts in the pool with higher coverage were randomly subsampled to the coverage of the other pool. The final allele counts in each pool were output to text files by chromosome and given as inputs to the MULTIPOOL algorithm (mp_inference.py version 0.10.2) [17] to compute LOD scores. Loci with maximal LOD>5 were considered significant; previous work showed that this corresponded to an FDR of 5% [41,44]. This correspondence may differ under our experimental conditions; therefore, the 2 loci that we did not validate experimentally should be interpreted with caution.

CRISPR/Cas9 allele replacement

Allele replacement strains were constructed using 3 rounds of gene knockout followed by CRISPR/Cas9-mediated markerless integration of heterologous allele. Initially, strains were prepared by introducing Cas9 on a CEN/ARS plasmid (SLVF11); this plasmid is derived from a previous one [45], but we replaced the auxotrophic *URA3* marker with AUR1-C to allow Aureobasidin A selection on prototrophic natural isolates. In each round of allele replacement, a gene plus upstream and downstream flanking sequences (-784bp to +815bp for *GAL3*, -449bp to +372bp for *MKT1*, -191bp to +139bp for *GAL4*) was deleted by integration of a kanMX6 marker with 40bp flanking homology. Then, a donor DNA, a guide RNA insert, and a guide RNA backbone were simultaneously transformed into the strain [46]. The donor DNA contains the new allele, its flanking

sequences, and an additional 40bp of homology to target it to the correct genomic locus. The guide RNA insert was a linear DNA containing a SNR52 promoter driving a guide RNA gene containing a 20bp CRISPR/Cas recognition sequence linked to a crRNA scaffold sequence, plus 40bp of flanking homology on both ends to a guide RNA backbone. The guide RNA backbone was a 2u plasmid containing natMX4 (pRS420). This was linearized by NotI + XhoI digestion before transformation. Allele re-integration transformations were plated on cloNAT to select for in vivo assembly of the guide RNA into a maintainable plasmid, and Aureobasidin A to select for presence of Cas9. Successful re-integration was verified by colony PCR and Sanger sequencing was performed on a subset of strains and on all donor DNAs to verify the sequence of allelic variants.

References

1. Nature MF, 2000. Feedback control of intercellular signalling in development. naturecom.

2. Venturelli OS, El-Samad H, Murray RM. Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. Proceedings of the National Academy of Sciences. 2012;109: E3324–E3333. doi:10.1073/pnas.1211902109

3. Acar M. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435: 228–232. doi:10.1038/nature03524

4. o MSAN, Leidy C, Pedraza JM. Tuning the range and stability of multiple phenotypic states with coupled positive–negative feedback loops. Nature Communications. Nature Publishing Group; 2013;4: 1–8. doi:10.1038/ncomms3605

5. Ramsey SA, Smith JJ, Orrell D, Marelli M, Petersen TW, de Atauri P, et al. Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. Nat Genet. 2006;38: 1082–1087. doi:10.1038/ng1869

6. Acar M. A General Mechanism for Network-Dosage Compensation in Gene Circuits . Science. 2010;329: 1656–1660. doi:10.1126/science.1190544

7. Hawkins KM, Smolke CD. The Regulatory Roles of the Galactose Permease and Kinase in the Induction Response of the GAL Network in Saccharomyces cerevisiae. Journal of Biological Chemistry. 2006;281: 13485–13492. doi:10.1074/jbc.M512317200

8. Biggar SR, journal GCTE, 2001. Cell signaling can direct either binary or graded transcriptional responses. embojembopressorg.

9. VERMA M, BHAT PJ, Venkatesh KV. Steady-state analysis of glucose repression reveals hierarchical expression of proteins under Mig1p control in Saccharomyces cerevisiae. Biochem J. 2005;388: 843–849. doi:10.1042/BJ20041883

10. Lee KB, Wang J, Palme J, Escalante-Chong R, Hua B, Springer M. Polymorphisms in the yeast galactose sensor underlie a natural continuum of nutrient-decision phenotypes. Fay JC, editor. PLoS Genet. 2017;13: e1006766–27. doi:10.1371/journal.pgen.1006766

11. Wang J, Atolia E, Hua B, Savir Y, Escalante-Chong R, Springer M. Natural Variation in Preparation for Nutrient Depletion Reveals a Cost–Benefit Tradeoff. Siegal ML, editor. PLoS Biol. 2015;13: e1002041–31. doi:10.1371/journal.pbio.1002041

12. New AM, Cerulus B, Govers SK, Perez-Samper G, Zhu B, Boogmans S, et al. Different Levels of Catabolite Repression Optimize Growth in Stable and Variable Environments. Doebeli M, editor. PLoS Biol. 2014;12: e1001764–22. doi:10.1371/journal.pbio.1001764

13. Solopova A, van Gestel J, Weissing FJ, Bachmann H, Teusink B, Kok J, et al. Bet-hedging

during bacterial diauxic shift. Proceedings of the National Academy of Sciences. 2014;111: 7427–7432. doi:10.1073/pnas.1320063111

14. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population genomics of domestic and wild yeasts. Nature. 2009;458: 337–341. doi:10.1038/nature07743

15. Cromie GA, Hyma KE, Ludlow CL, Garmendia-Torres C, Gilbert TL, May P, et al. Genomic Sequence Diversity and Population Structure of Saccharomyces cerevisiae Assessed by RAD-seq. G3: Genes|Genomes|Genetics. 2013;3: 2163–2171. doi:10.1534/g3.113.007492

16. Peng W, Liu P, Xue Y, Acar M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nature Communications. Nature Publishing Group; 1AD;6: 1–9. doi:10.1038/ncomms7226

17. Edwards MD, Gifford DK. High-resolution genetic mapping with pooled sequencing. BMC Bioinformatics. BioMed Central Ltd; 2012;: 1–11. doi:10.1186/1471-2105-13-S6-S8

18. Ehrenreich IM, Bloom J, Torabi N, Wang X, Jia Y, Kruglyak L. Genetic Architecture of Highly Complex Chemical Resistance Traits across Four Yeast Strains. Gasch AP, editor. PLoS Genet. 2012;8: e1002570–9. doi:10.1371/journal.pgen.1002570

19. Kim HS, Huh J, Fay JC. Dissecting the pleiotropic consequences of a quantitative trait nucleotide. FEMS Yeast Research. 2009;9: 713–722. doi:10.1111/j.1567-1364.2009.00516.x

20. Sinha H, Nicholson BP, Steinmetz LM, McCusker JH. Complex Genetic Interactions in a Quantitative Trait Locus. PLoS Genet. 2006;2: e13–8. doi:10.1371/journal.pgen.0020013

21. Duveau F, Metzger B, Gruber JD, Mack K. Mapping small effect mutations in Saccharomyces cerevisiae: impacts of experimental design and mutational properties. G3: Genes| Genomes| …. 2014. doi:10.1534/g3.114.011783/-/DC1

22. Lohr D, Venkov P, Zlatanova J. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. Federation of American Societies for Experimental Biology; 1995;9: 777–787. doi:10.1096/fj.1530-6860

23. Ren B, Robert F, Wyrick JJ, Aparicio O, 2000. Genome-wide location and function of DNA binding proteins. Science.

24. Cell RW, 1980. Plasmids controlling exclusion of the K2 killer double-stranded RNA plasmid of yeast. Elsevier.

25. Tadauchi T, Inada T, Matsumoto K, Irie K. Posttranscriptional Regulation of HO Expression by the Mkt1-Pbp1 Complex. Molecular and Cellular Biology. 2004;24: 3670–3681. doi:10.1128/MCB.24.9.3670-3681.2004

26. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, et al. Dissecting the architecture of a quantitative trait locus in yeast. Nature. 2002;416:

326–330. doi:10.1038/416326a

27. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. Nature Publishing Group; 2010;464: 1039–1042. doi:10.1038/nature08923

28. Dimitrov LN, Brem RB, Kruglyak L, Gottschling DE. Polymorphisms in Multiple Genes Contribute to the Spontaneous Mitochondrial Genome Instability of Saccharomyces cerevisiae S288C Strains. Genetics. 2009;183: 365–383. doi:10.1534/genetics.109.104497

29. Deutschbauer AM, Davis RW. Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat Genet. 2005;37: 1333–1340. doi:10.1038/ng1674

30. Demogines A, Smith E, Kruglyak L, Alani E. Identification and Dissection of a Complex DNA Repair Sensitivity Phenotype in Baker's Yeast. Frankel WN, editor. PLoS Genet. 2008;4: e1000123–14. doi:10.1371/journal.pgen.1000123

31. Gerber AP, Herschlag D, Brown PO. Extensive Association of Functionally and Cytotopically Related mRNAs with Puf Family RNA-Binding Proteins in Yeast. Sean Eddy, editor. PLoS Biol. 2004;2: e79–14. doi:10.1371/journal.pbio.0020079

32. Lee S-I, Dudley AM, Drubin D, Silver PA, Krogan NJ, Pe'er D, et al. Learning a Prior on Regulatory Potential from eQTL Data. Moses AM, editor. PLoS Genet. 2009;5: e1000358–25. doi:10.1371/journal.pgen.1000358

33. Hua B, Springer M. Small effect-size mutations cumulatively affect yeast quantitative traits. 2017. doi:10.1101/126409

34. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. Nature Publishing Group; 2013;: 1–6. doi:10.1038/nature11867

35. Stockwell SR, Rifkin SA. A living vector field reveals constraints on galactose network induction in yeast. 2016 Sep pp. 1–16. doi:10.1101/073817

36. Li Y, Chen G, Liu W. Multiple metabolic signals influence GAL gene activation by modulating the interaction of Gal80p with the transcriptional activator Gal4p. Molecular Microbiology. 2010;78: 414–428. doi:10.1111/j.1365-2958.2010.07343.x

37. Kumar PR, Yu Y, Sternglanz R, Johnston SA, 2008. NADP regulates the yeast GAL induction system. Science.

38. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population Diversification in a Yeast Metabolic Program Promotes Anticipation of Environmental Shifts. Siegal ML, editor. PLoS Biol. 2015;13: e1002042–24. doi:10.1371/journal.pbio.1002042

39. Healey DW. Phenotypic heterogeneity and evolutionary games in microbial populations. 2015. pp. 1–106.

40. Hashida-Okado T, Ogawa A, Kato I, Takesako K. Transformation system for

prototrophic industrial yeasts using the AUR1gene as a dominant selection marker. FEBS Letters. 1998;425: 117–122. doi:10.1016/S0014-5793(98)00211-7

41. Treusch S, Albert FW, Bloom JS, Kotenko IE, Kruglyak L. Genetic Mapping of MAPK-Mediated Complex Traits Across S. cerevisiae. Copenhaver GP, editor. PLoS Genet. 2015;11: e1004913–16. doi:10.1371/journal.pgen.1004913

42. Cubillos FA, Parts L, Salinas F, Bergstrom A, Scovacricchi E, Zia A, et al. High-Resolution Mapping of Complex Traits with a Four-Parent Advanced Intercross Yeast Population. Genetics. 2013;195: 1141–1155. doi:10.1534/genetics.113.155515

43. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. Green SJ, editor. PLoS ONE. 2015;10: e0128036–15. doi:10.1371/journal.pone.0128036

44. Albert FW, Treusch S, Shockley AH, Bloom JS, Kruglyak L. Genetics of single-cell protein abundance variation in large yeast populations. Nature. Nature Publishing Group; 2014;506: 494–497. doi:10.1038/nature12904

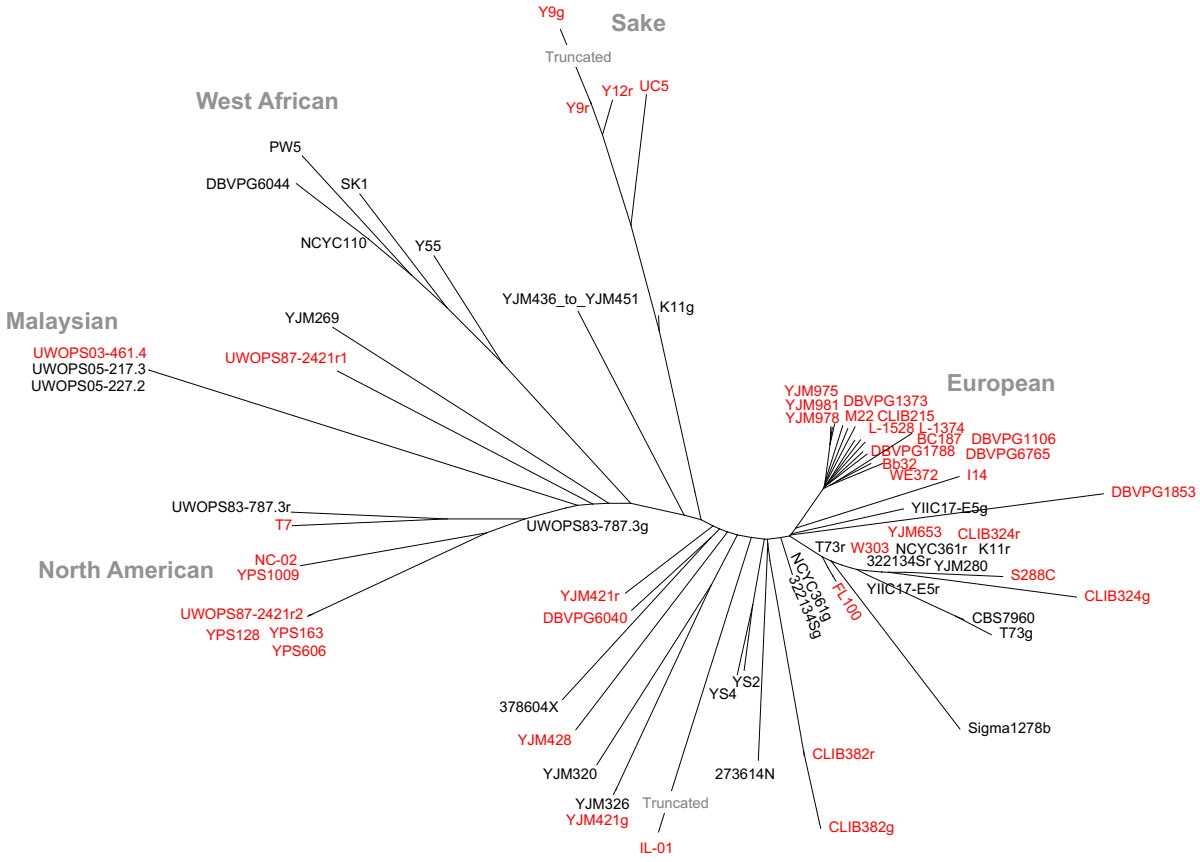
45. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Research. 2013;41: 4336–4343. doi:10.1093/nar/gkt135

46. Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, et al. Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell Systems. The Authors; 2015;1: 88–96. doi:10.1016/j.cels.2015.02.001

Appendix I.

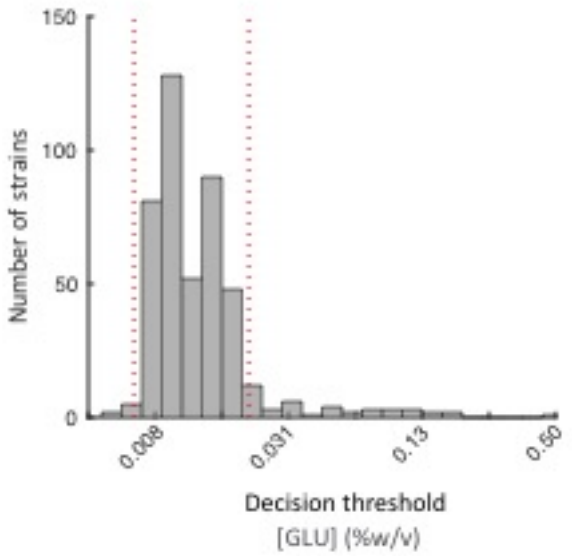
Supporting information for Chapter 2

Polymorphisms in the yeast galactose sensor underlie a natural continuum of nutrient-decision phenotypes



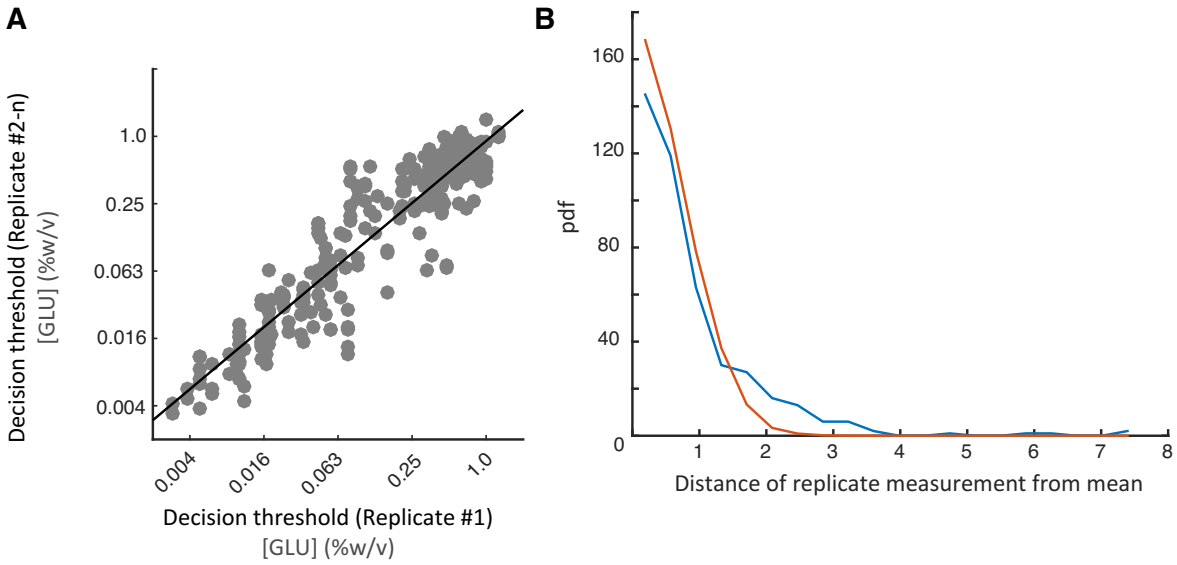
S2.1. Phylogenetic tree of *S. cerevisiae* used in this study.

Phylogenetic tree of common natural isolates of *S. cerevisiae* constructed based on sequencing data from Cromie et al. 2013 [37]. Strains highlighted in red were used in this study, while strains in black were not.



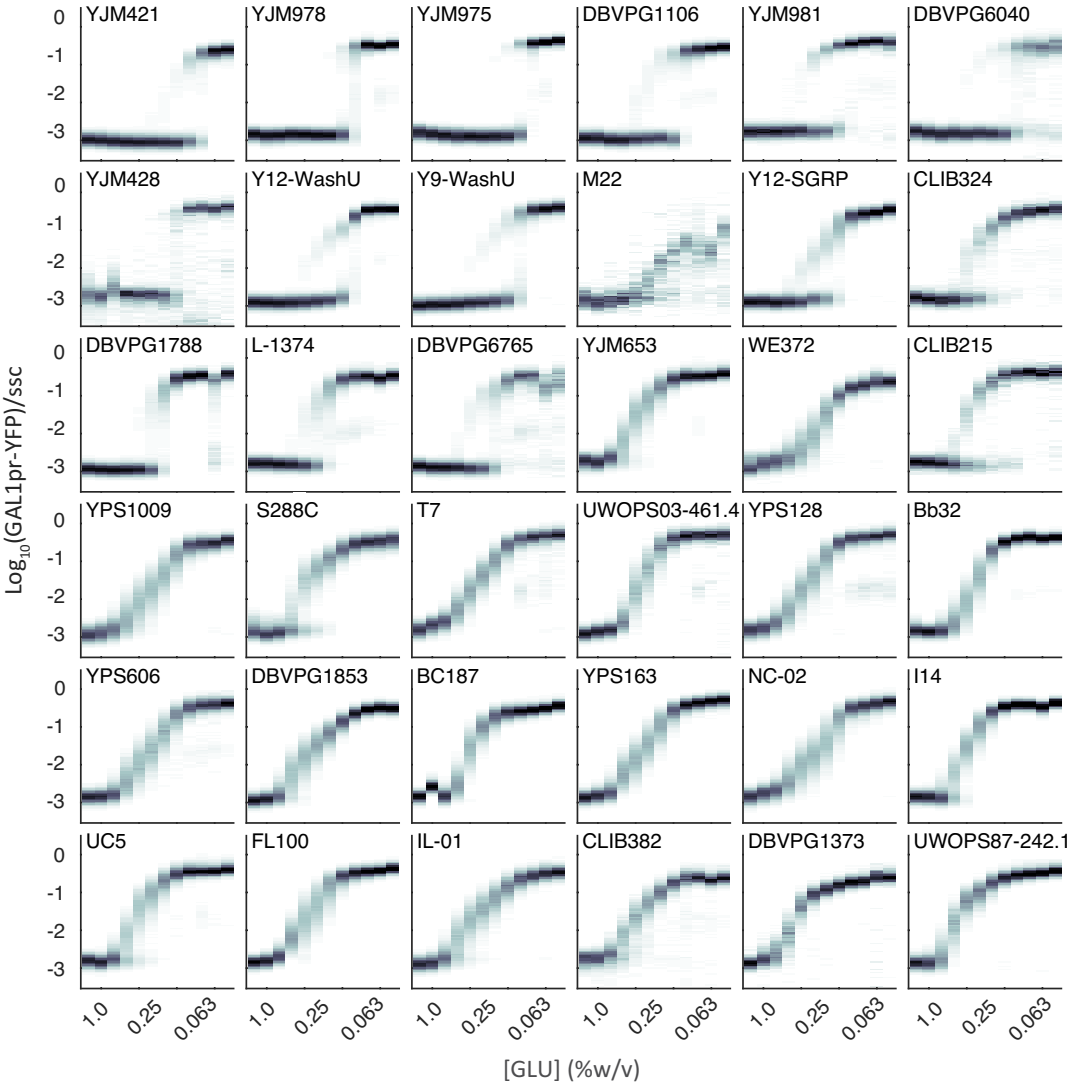
S2..2. Quality control for query strains based on the reference strain.

Each experiment contained a reference strain. The decision threshold of the reference strain was roughly normally distributed, with a long tail. Based on technical measurements, the tails are due to unintended variation in the assay, e.g. cells grown at too high of an OD, as opposed to biological variation. To eliminate this variation, we truncated the 5% highest and lowest values (red dashed lines). The standard deviation of the remaining, roughly normal, distribution was calculated and used to eliminate samples.



S2.3. Determining a cut-off for query outliers.

(A) Remaining strains were plotted, replicate #1 vs. replicate #2-n, where n is the total number of replicates. A total of 68 strains out of 480 experiments were removed in our quality control. (B) The absolute value of the difference between each distinct measurement of a sample and the mean of all other replicate for that sample is plotted (blue). The same technique was used on simulated data derived from a normal distribution with a standard deviation of 0.75 (red). Based on this a 1.5 standard deviation was chosen to eliminate samples that were likely due to some unintended source of bias.

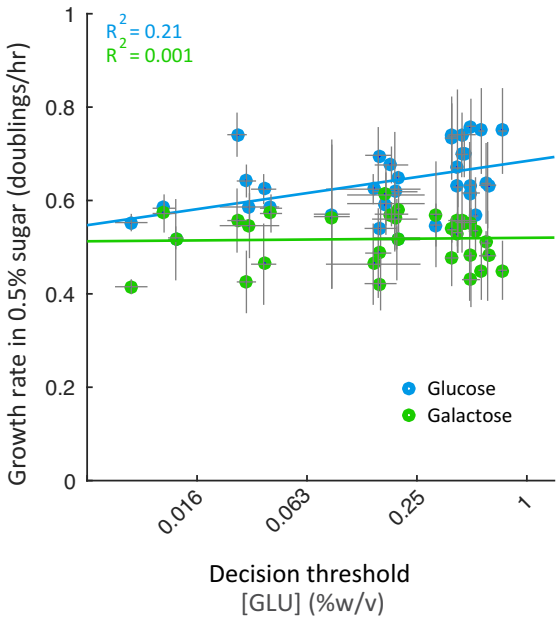


S2.4. Steady-state expression of *GAL1pr-YFP* from a panel of natural isolates in mixtures of glucose and galactose.

Representative YFP induction profiles of the diploid natural isolates assayed in Fig. 1. Cells were grown for 8 hours, a time previously determined to be sufficient for expression to reach steady-state [25], in a titration from 1% to 0% glucose (two-fold dilution series) in constant background 0.25% galactose. Flow cytometry profiles are plotted for each glucose concentration. Each panel contains

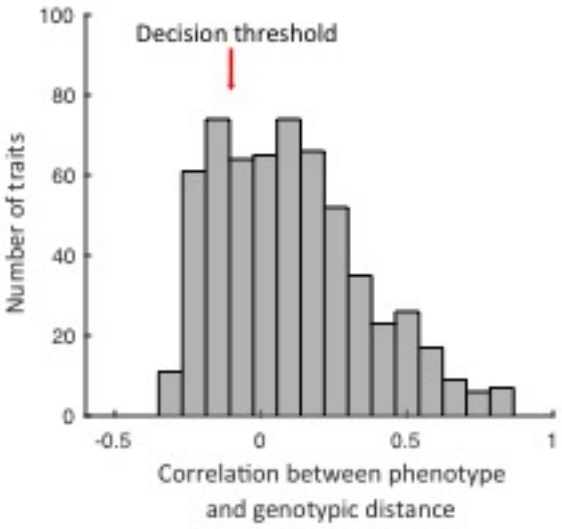
S2.4 (continued)

10 distinct glucose and galactose concentrations and 2% pure glucose or galactose. The color density represents the probability density function across of cells for different fluorescent intensity levels. Strains are ordered by increasing decision threshold.



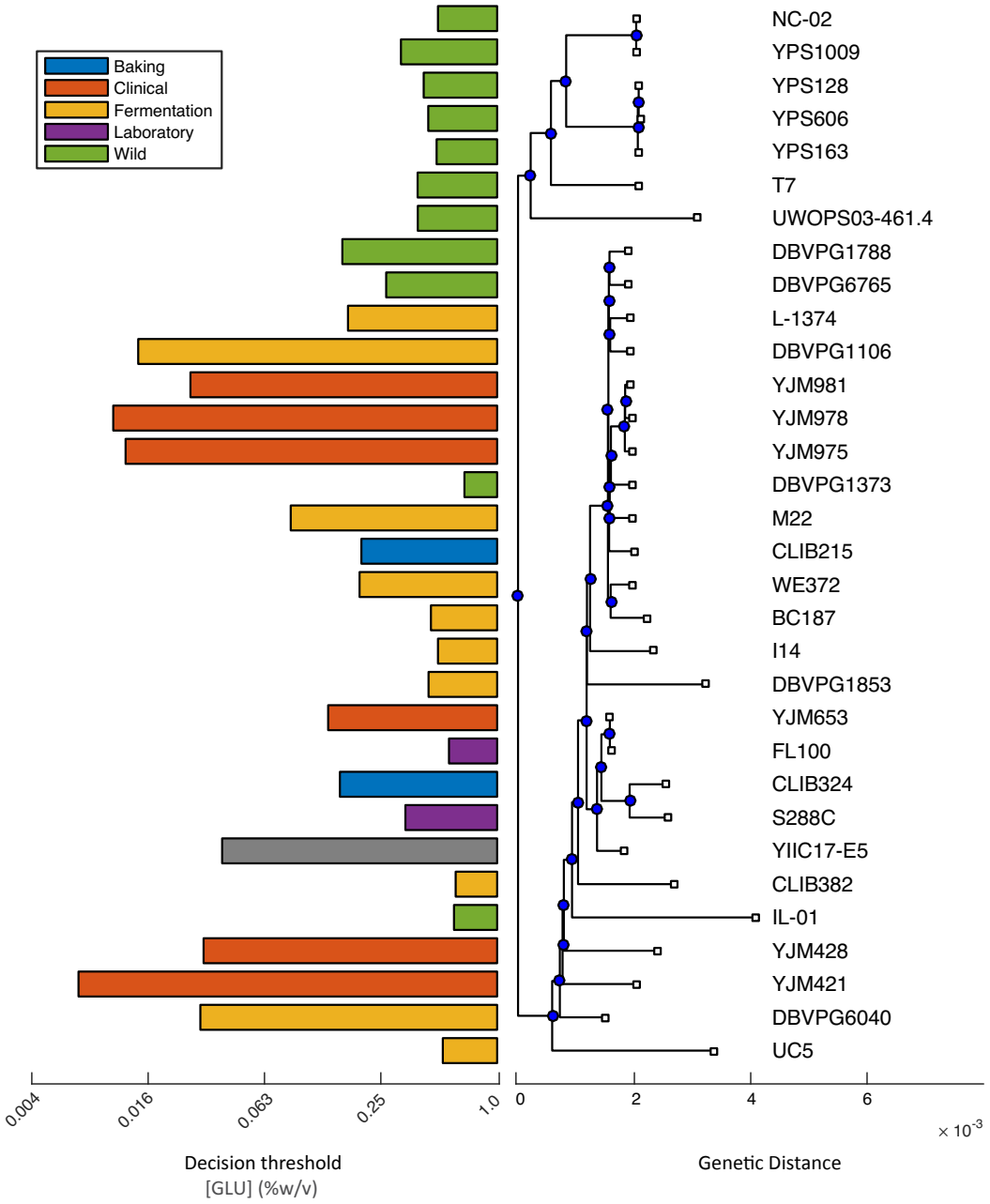
S2.5. Growth rate in 0.5% glucose or 0.5% galactose is not strongly correlated with decision threshold.

Cells were grown in medium containing 0.5% glucose or 0.5% galactose and the OD_{600} was measured every 15 minutes by plate reader (Materials and methods). The growth rate was then calculated for each strain and condition (Materials and methods). The growth rate in glucose (blue) or galactose (green) of natural isolates is plotted versus the decision threshold (from Fig. 1). Error represents S.E.M. of three replicates for growth rate and at least two replicates for decision threshold. The line is a linear least squared fit.



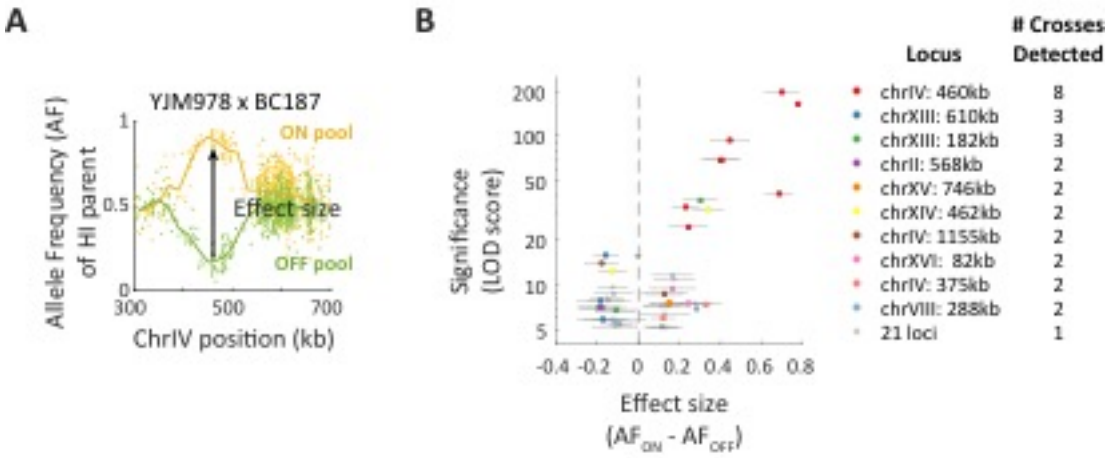
S2.6. Correlation between genetic distance and phenotypic distance for decision threshold and traits from literature.

Genetic distance [37] and phenotypic distance for a number of traits [4] had been previously measured and determined to be weakly correlated [4]. The histogram of correlation between genetic and phenotypic distance is plotted. The correlation between genetic distance and decision threshold is denoted with the red arrow.



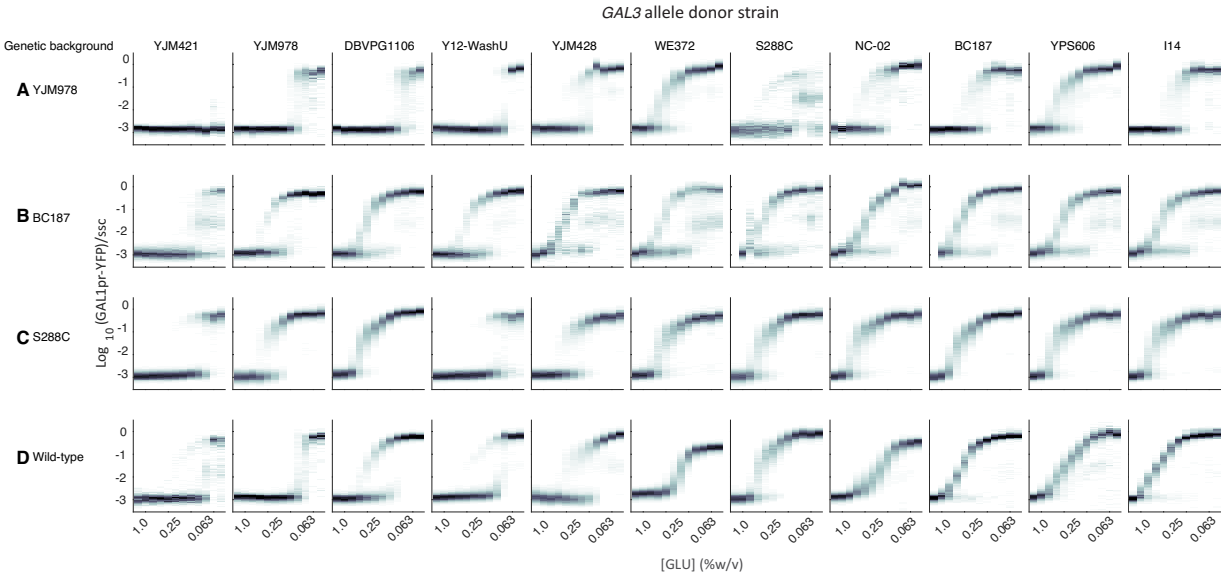
S2.7. Relationship of decision threshold with phylogeny and ecological niche.

Phylogenetic tree was constructed based on the Cromie et al. distance matrix (Materials and methods) with the bar plot indicating decision threshold (from Fig. 1). Color of bars indicate the ecological niche of strain.



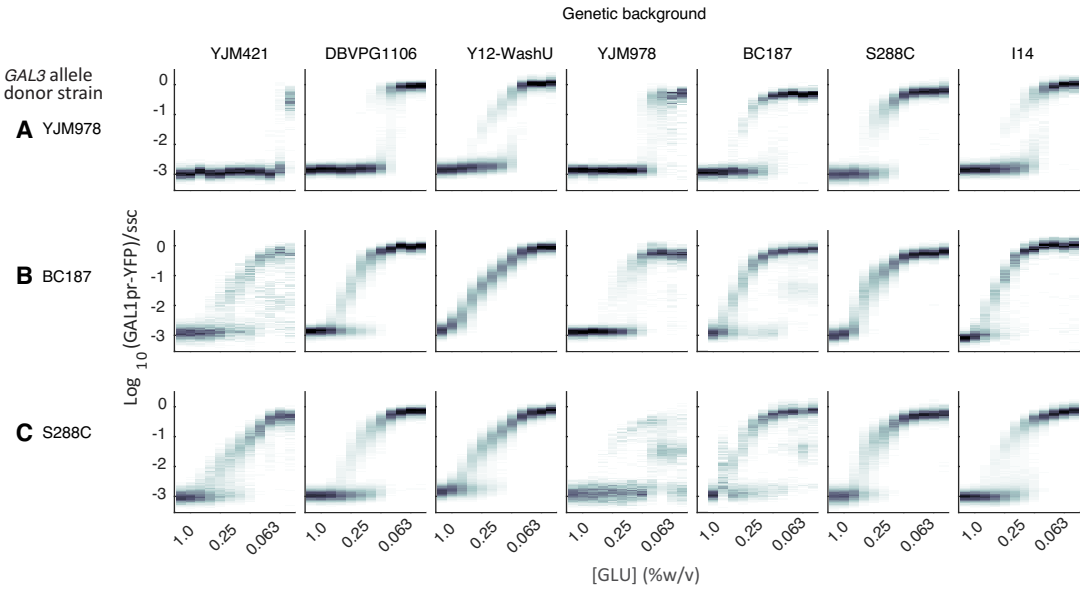
S2.8. Significance and effect size of detected loci.

(A) Allele frequency of the ON parent (BC187) in the YJM978xBC187 cross across a region of chromosome IV spanning the chrIV:460Kb locus. The difference in allele frequency between ON and OFF pools at the locus can be used as a proxy for its effect size on the GAL induction phenotype. (B) Scatterplot of significance (LOD score) versus effect size (allele frequency difference) for all 49 LOD peaks where LOD > 5. Significant LOD peaks from different crosses were "clustered" into a single locus if they lay within 20kb of each other. Dots representing LOD peaks are colored by clustered locus.



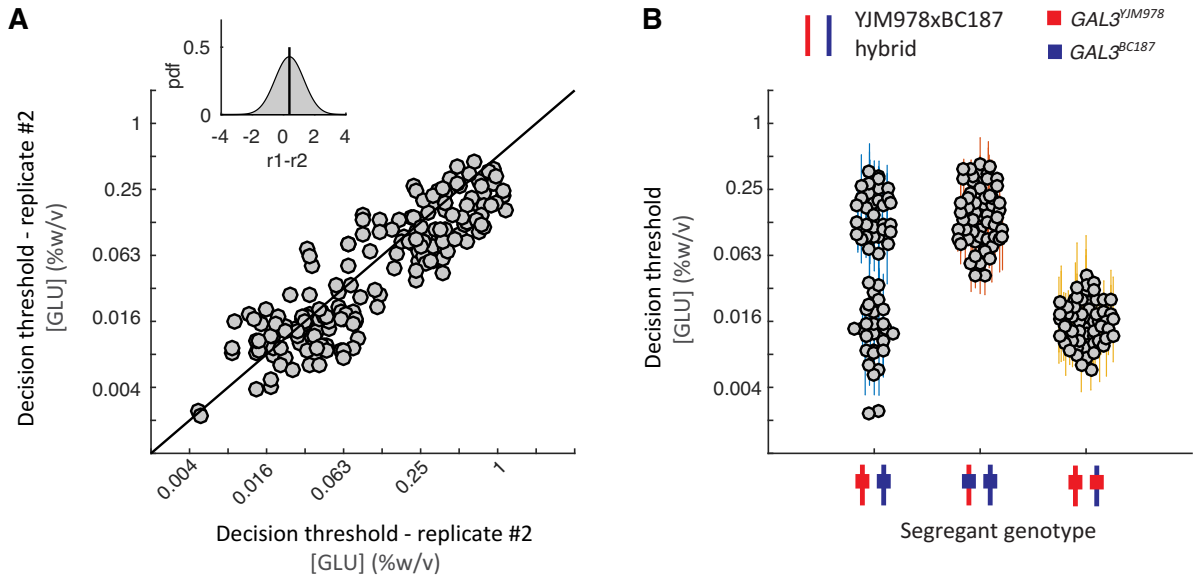
S2.9. Representative YFP induction profiles of *GAL3* allele replacements.

Homologous *GAL3* allele replacement strains were assayed in a gradient of glucose in a background of 0.25% galactose (Fig. 3A-C). The alleles were assayed in three backgrounds (A) YJM978, (B) BC187, and (C) S288C. (D) The parental strain is shown for comparison.



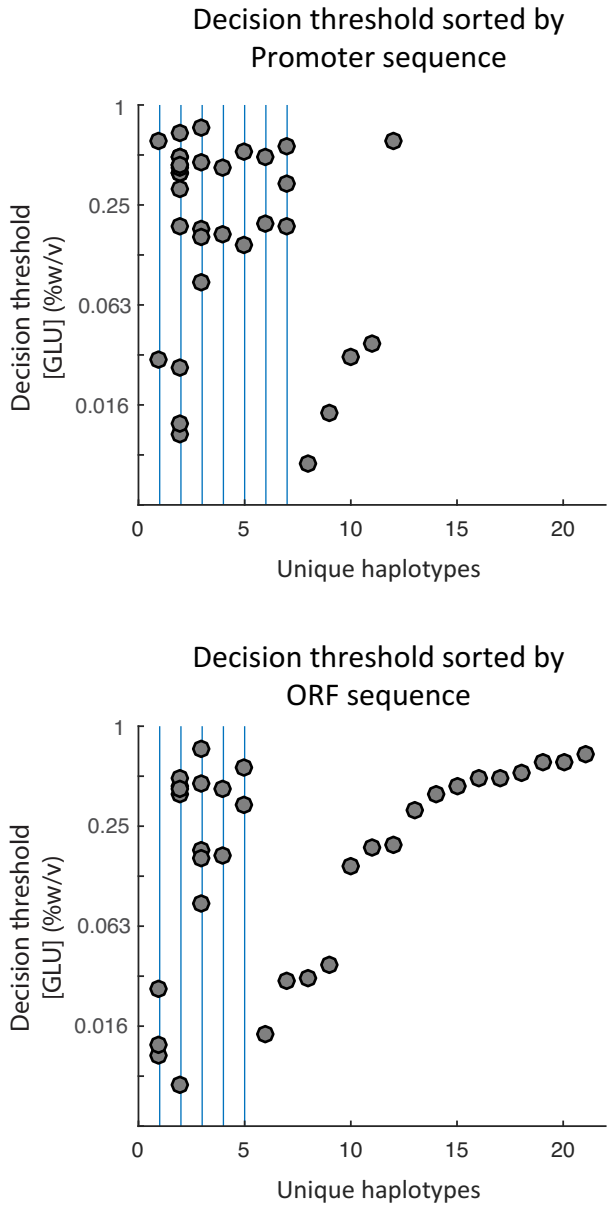
S2.10. Hemizygous hybrids YFP density plots.

Homologous *GAL3* allele replacement strains were assayed in a gradient of glucose in a background of 0.25% galactose (Fig. 3D). Three different alleles (A) YJM978, (B) BC187, and (C) S288C were assayed in seven genetic backgrounds.



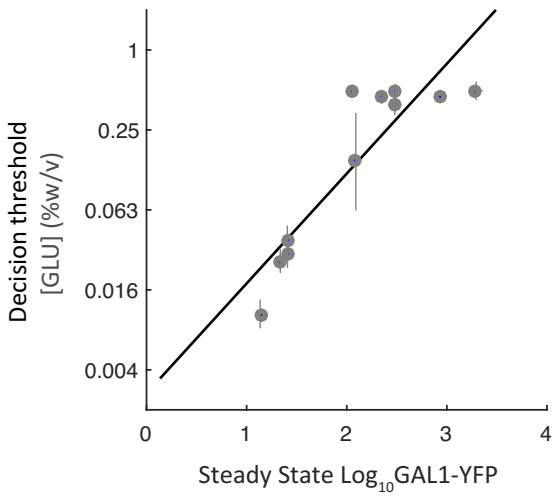
S2.11. Phenotypic variation of hybrid (and hybrid conversion) segregants.

(A) Plot of the decision threshold for replicate 1 and replicate 2 from each segregant assayed. Inset: probability density function of the difference of replicate 1 and replicate 2. The variance from this distribution was used to determine the measurement error. (B) Decision threshold of segregants produced from hybrid conversion (Error represents range of the two segregants).



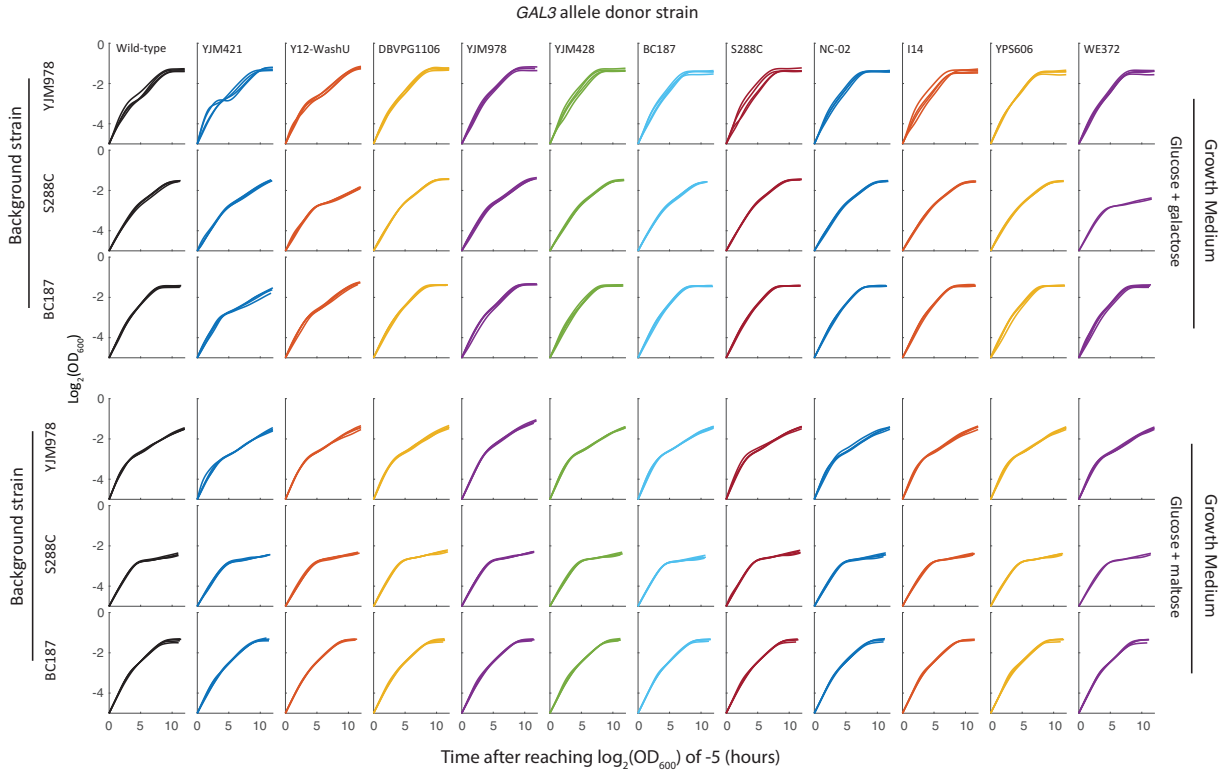
S2.12. Decision threshold versus haplotype of promoter or ORF region.

Decision threshold plotted versus unique haplotypes of the (A) promoter or (B) ORF region. Haplotype clusters that contain at least two strains are shown on the left side of the graph. The blue line is a guide to show how strains cluster.



S2.13. Scatter plot of decision threshold versus *GAL1*-YFP steady state expression [6].

Scatter plot of steady state *GAL1* expression levels versus decision threshold of a subset of strains from Fig. 1. We previously showed that *GAL1* expression levels before the diauxic lag are inversely correlated with the diauxic lag length [6]. We extend that show that the decision threshold is correlated to these *GAL1* expression levels.



S2.14. Growth curves of *GAL3* allele replacement strains.

Replicate data of growth curves of *GAL3* allele replacement strains in the YJM978, BC187, and S288C background in glucose+galactose (top) or glucose+maltose (bottom). Wild-type growth curves are shown for each background strain in black. Each color represents a different color *GAL3* donor allele. Time is shown relative to $Log_2(OD_{600})$ reaching -5.

Appendix II.

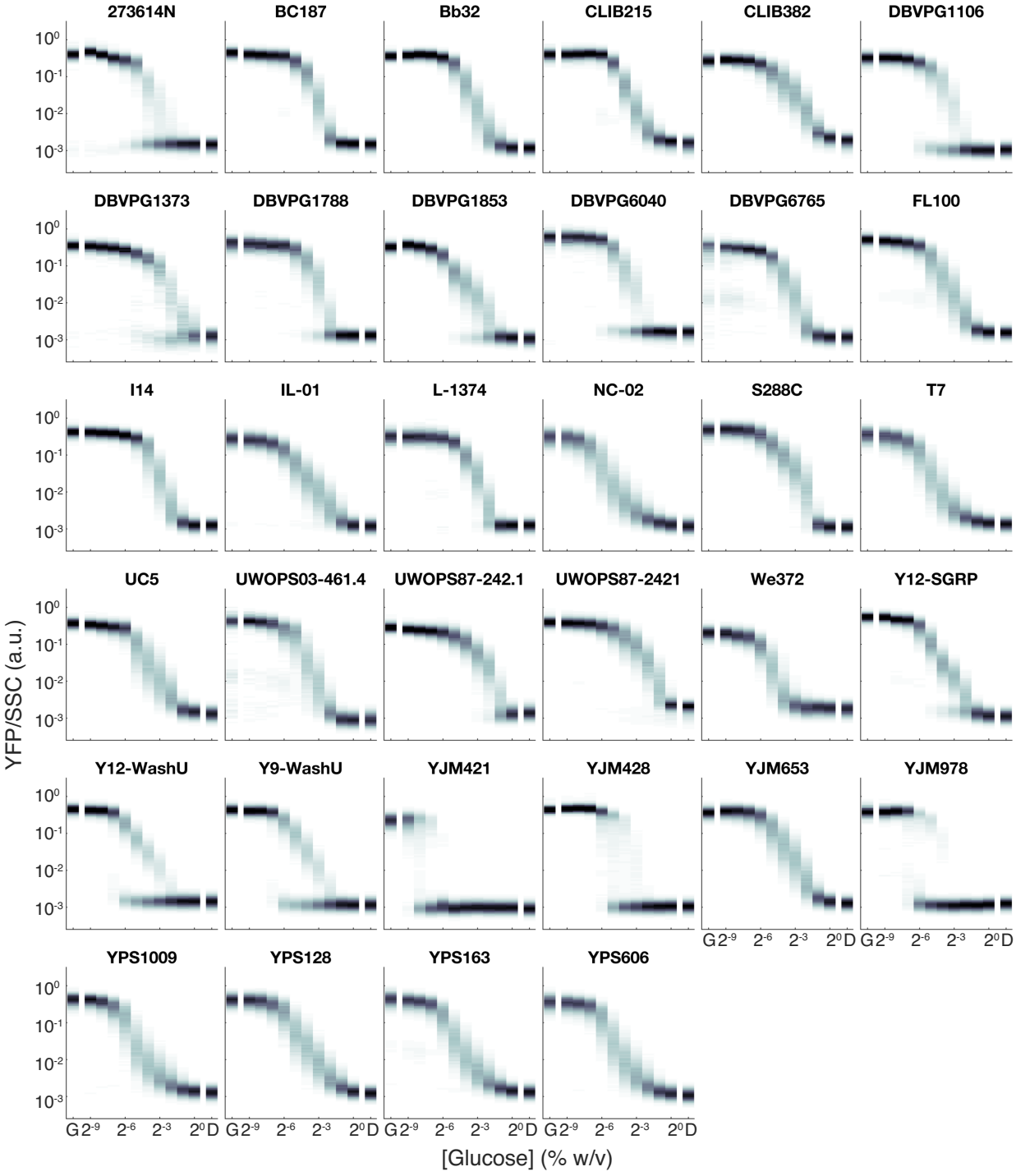
Supporting information for Chapter 3

Chapter 3: Nature uses specific and non-specific GAL genes to quantitatively tune bimodal signaling response of the glucose-galactose circuit

Figure S3.1. GAL response phenotypes for 34 natural isolates

Plotted are series of GAL1pr-YFP fluorescence (normalized to side scatter "SSC") histograms from 12 sugar conditions for 34 strains. One replicate experiment (out of 3-10 replicates) is shown for each strain. Data from all replicates is used to calculate E_{10} and F_{50} for the scatterplot in Figure 3.1G.

Figure S3.1 (Continued)



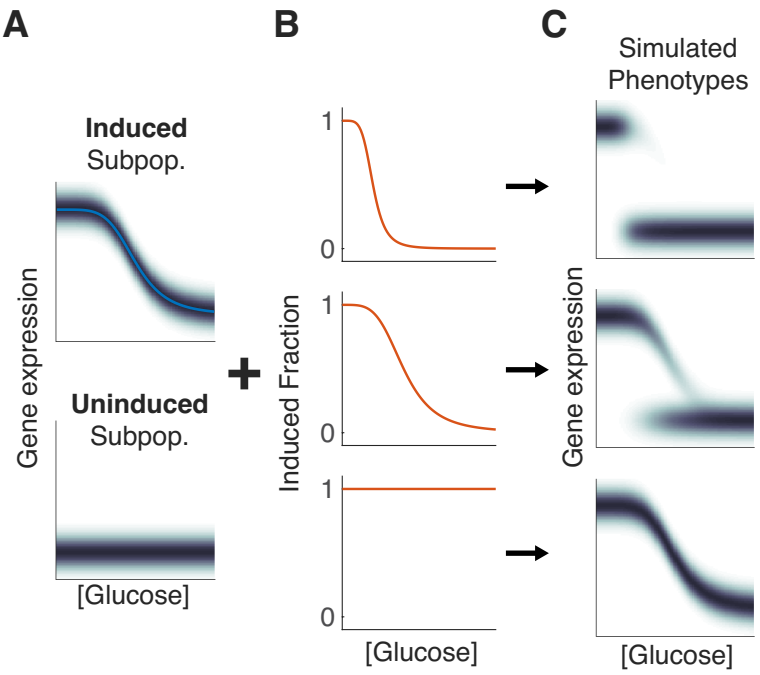


Figure S3.5. Bimodal phenotypes simulated using a subpopulation decomposition framework

(A) Two simulated subpopulations, where the mean of the induced population is shown in blue. (B) 3 possible functions for the dependence of induced fraction on glucose. (C) Simulated population behaviors using the 3 induced fraction functions.

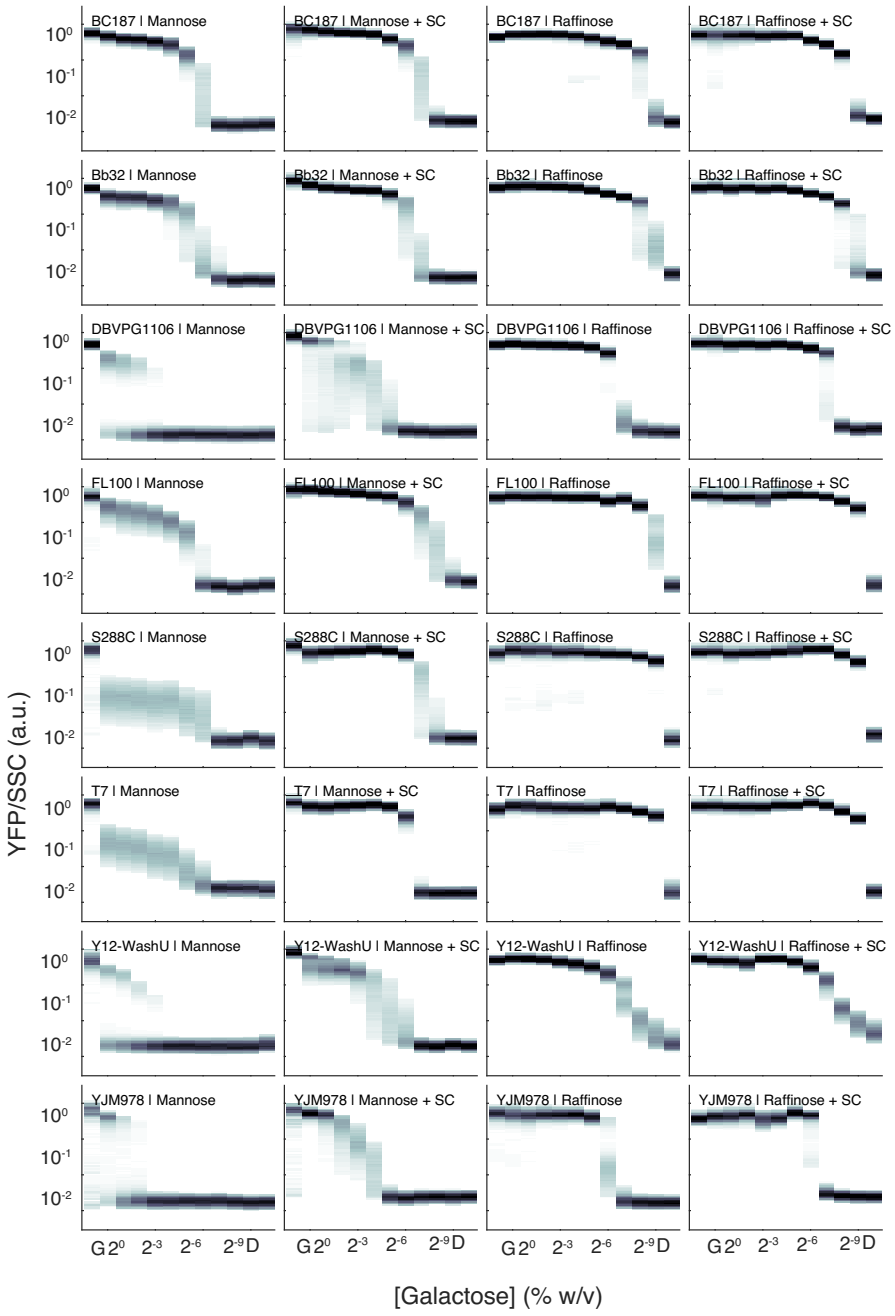


Figure S3.3. GAL response phenotypes for 8 natural isolates in different background sugars

Plotted are series of GAL1pr-YFP fluorescence (normalized to side scatter "SSC") histograms from 12 sugar conditions for 8 different strains in 0.1% mannose or 2% Raffinose with or without amino acids to compare our phenotypes in growth conditions similar to other studies

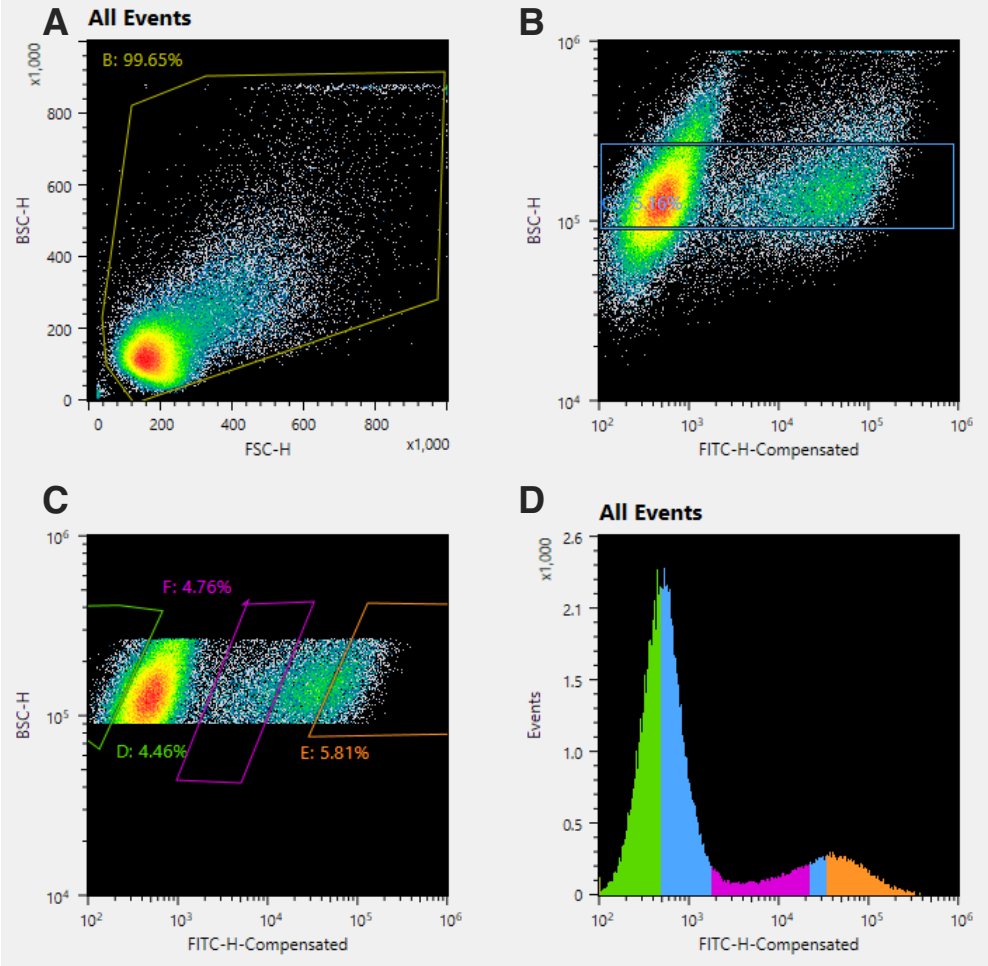


Figure S3.4. Sorting strategy for bulk-segregant analysis

(A) Backscatter versus forward scatter of unsorted segregant pool, obtained on Sony SH800 cell sorter. (B) Backscatter versus FITC (YFP), showing mixture of uninduced and induced cells. Backscatter was used as a proxy for cell size; therefore, it is correlated with fluorescence. Gating on backscatter (rectangle) isolates differences in GAL1pr-YFP reporter among the cells. (C) Gates for OFF, LOW, and HI cells were drawn after gating on backscatter and shaped to follow the backscatter-FITC correlation. (D) View of gated populations as histogram on FITC axis.

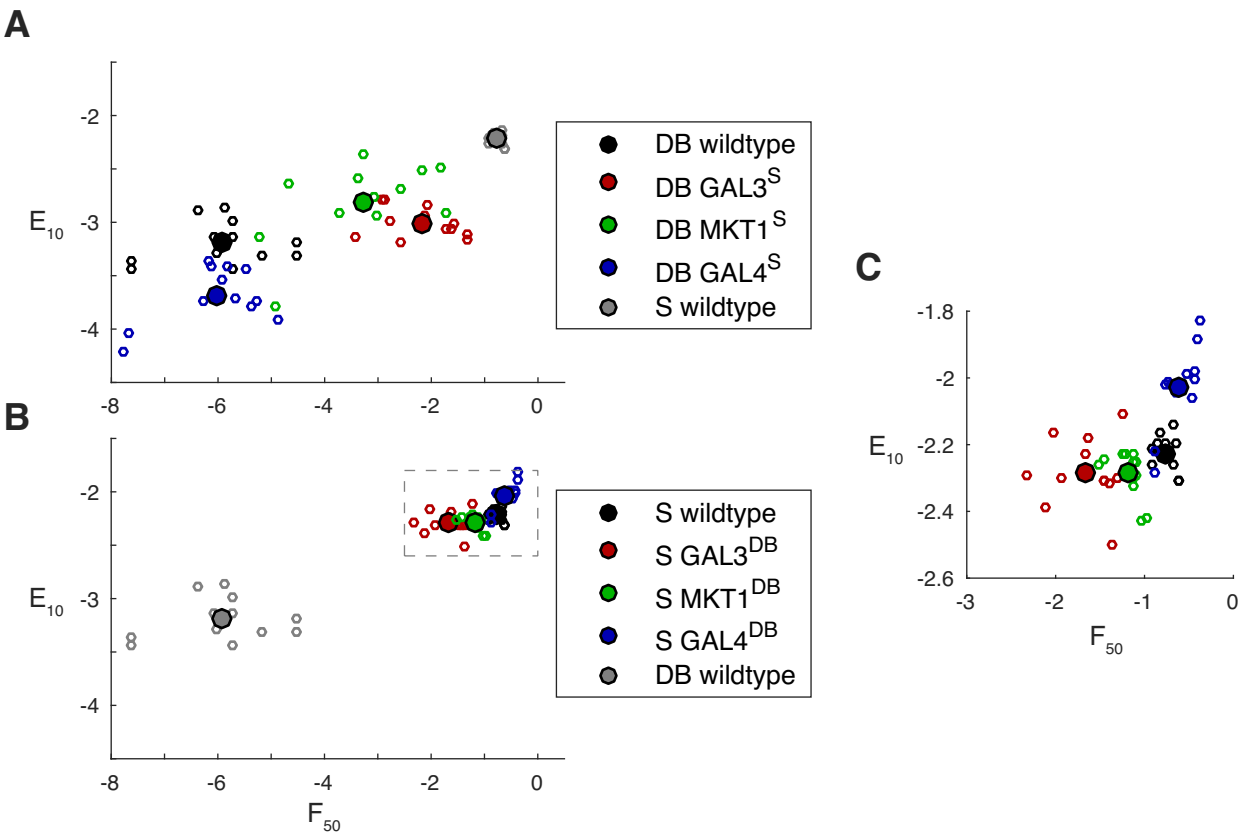


Figure S3.5. Effect of allele replacement of *GAL3*, *MKT1*, or *GAL4* in DBVPG1106 and S288C backgrounds.

Scatterplots of E_{10} versus F_{50} for (A) DBVPG1106 strains where the indicated genes (and flanking regions) have been replaced by their S288C alleles; (B) S288C strains containing replacements by DBVPG1106 alleles. (C) Enlargement of region in (B) outlined by dotted rectangle. Small circles are individual replicates (12 replicates per genotype, comprising 6 replicates each for 2 independently constructed isolates – see Materials and Methods); large circles indicate the mean. These plots show a subset of the same data as in Figure 4.3.