Fitness Tradeoffs, Bimodality, and the Genetics of Natural Variation in Glucose-Galactose Sensing Across Saccharomyces Cerevisiae

Fitness tradeoffs, bimodality, and the genetics of natural variation in glucose-galactose sensing across *Saccharomyces cerevisiae*

A dissertation presented

by

Jue Wang

to

The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Systems Biology

Harvard University

Cambridge, Massachusetts

September 2016

Dissertation Advisor: Dr. Michael Springer Jue Wang

Fitness tradeoffs, bimodality, and the genetics of natural variation in glucose-galactose sensing across *Saccharomyces cerevisiae*

Abstract

In nature, cells must constantly sense their environment and respond appropriately. These cellular "decisions" are implemented by molecular components that can mutate and evolve. How do mutations lead to changes in cellular decision-making? How do the decisions made by cells affect their fitness? This thesis explores these questions by examining the same metabolic decision made by geographically and ecologically diverse strains of *Saccharomyces cerevisiae* (budding yeast). When yeast encounter a mixture of the sugars glucose and galactose, they utilize glucose while repressing galactose-utilization (GAL) genes. When glucose is exhausted, cells undergo a "diauxic lag" while inducing GAL genes, and then resume growth on galactose. We found that some yeast strains can induce GAL genes before glucose is exhausted, which reduces their diauxic lag but imposes an initial growth cost. The degree of pre-induction depends on the sugar-sensing threshold of the GAL circuit, which we map to a natural allelic series of the signaling gene *GAL3*. However, because the GAL response is bimodal, we find that *GAL3* alleles only modulate the fraction of cells that induced in a given condition, while the expression level attained by those induced cells is tuned by alleles of other genes. Overall, our results reveal the repertoire of mutations that can quantitatively tune a cellular decision in nature, as well as the downstream effects of this tuning on physiology and fitness.

Table of Contents

Chapter 1. Introduction 1

Chapter 2. Natural Variation in Preparation for Nutrient Depletion Reveals a Cost–Benefit Tradeoff 9

Chapter 3. Polymorphisms in the yeast galactose sensor underlie a natural continuum of nutrient-decision phenotypes 59

Chapter 4. Natural Genetic Variation Can Independently Tune the Induced Fraction and Induction Level of a Bimodal Signaling Response 97

Appendix I. Supporting Information for Chapter 2 126

Appendix II. Supporting Information for Chapter 3 144

Appendix III. Supporting Information for Chapter 4 15

Acknowledgements

Despite the reputation of a PhD, I found mine to be mostly enjoyable. This is due in no small part to the support of my wife, Liz Adams, and her ability to bear my long, detailed rants about technical troubleshooting. Even though she is not a scientist herself, she somehow manages to make me more productive while also helping me to enjoy life more outside the lab.

I will always admire my advisor Mike Springer's optimism, keen eye, and wide-ranging interests. From working with him I've learned a great deal about microbial physiology and gained a quantitative perspective on signaling and regulatory circuits. Most importantly, he's shown me how to maintain enthusiasm for ideas and questions in the face of practical setbacks.

I owe a huge debt to the mentors who kindled my interest in research and biology. Cary James, Brian Frederick, and Francois Amar encouraged me to do research in high school and funded me to present my results. Roy Kishony took me on as an undergrad researcher and opened my eyes to the mysteries of how cells respond to their environments. In grad school, Chris Marx and Johan Paulsson were a valuable source of advice and fascinating scientific anecdotes. Daniel Segre, Michael Desai, and Jeff Gore generously served on my DAC.

My labmates were a crucial source of fun, support, and inspiration. I thank Bo Hua for his mathematical quickness and endless ideas for impractical side projects; Yoni Savir for his cutting critiques delivered with a smile and constant encouragement to think bigger; Christine DeGennaro for her yeast expertise; Sarah Boswell for her not-screwing-up-experiments expertise; Esha Atolia, Kayla Lee, Chiara Ricci-Tam, and Julius Palme for tolerating my attempts to mentor them and probably teaching me more in the process; and Renan Escalante-Chong for starting the work on the GAL pathway.

I also thank Shervin Javadi and Jodi Moore for patiently guiding me as I went from clueless novice to proper flow cytometrist; Sam Keough, Barb Grant, and the rest of lab ops for their sheer competence; Mason Miranda and IT for never making me feel like computer problems were my fault even though they often were; Becky Ward for her writing advice; Sam Reed for her bottomless savvy in the logistics of academia and perhaps all other life matters.

One of the most enjoyable aspects of my PhD experience was being embedded in the HMS SysBio community. Even those whom I barely know have contributed in some way to my happiness and intellectual development, by giving a good talk or piquing my thoughts with a happy hour chat. In particular, I thank my classmates Ylaine Gerardin, Thomas Graham, and Keisuke Ishihara for their excellent scientific cocktail-party-conversation skills; Nick Leiby, Hallie Kuhn, and Anna Chen for their camaraderie; Rishi Jajoo for his finely honed sense for interesting questions; Nate Lord, Jake Wintermute, Max Staller, and Zeba Wunderlich for educating and inspiring me during and after my rotations; and Daniel Schultz and Michael Baym for teaching me the thankless art of robot-wrangling.

My friends are also partly to blame for my positive experience of grad school. I thank Reed Gochberg and Lucas Toffoli for sharing my love of bacon and pastries; Zach Taxin, Pedro Teixeira, Derek Jones, and John Harrison for their antics; Emily Boehm and Jill Bunting for their laser wits; Hannah Taber and Larson Hogstrom for being ideal roommates; and Dave Kleinschmidt for being an old kindred spirit and an inimitable scholarly peer.

Finally, I thank my parents Qin and Binke Wang, who have given me everything I have. I can only hope that I live up to their selflessness, curiosity, and sense of adventure.

Chapter 1.

Introduction

In nature, cells must constantly sense their environment and respond appropriately [1–3]. These cellular "decisions" are implemented by molecular components that can mutate and evolve. How do mutations lead to changes in cellular decision-making? How do the decisions made by cells affect their fitness? This thesis explores these questions by examining a metabolic decision made by *Saccharomyces cerevisiae* (budding yeast): whether to consume the sugars glucose, galactose, or both.

The yeast galactose-utilization (GAL) pathway is a compelling model for cellular decision-making because it is well-characterized and exhibits a deceptively simple logic: when galactose is present and glucose is not, GAL genes are expressed; in all other cases, they stay uninduced [4–6]. In this sense, yeast "prefer" glucose over galactose. However, this traditional picture ignores the fact that sugar concentration, like most environmental signals, is a continuous quantity, and decision circuits must in general map quantitative inputs to quantitative outputs. One of the most detailed measurements of such a mapping was recently made in the GAL pathway by Escalante-Chong, Savir, et al [7]. They found that mere presence of glucose is not sufficient to repress GAL genes, as previously assumed. Instead, repression requires a certain ratio of glucose to galactose, a behavior they call "ratio-sensing". They also found that different natural isolates of yeast induced GAL genes at different ratios of galactose to glucose. These observations motivated the project in Chapter 2 by raising two questions: 1) What are the physiological consequences of ratio-sensing by the GAL pathway? 2) Why do different strains of yeast have different induction thresholds?

To investigate these questions, we decided to study diauxic growth, a well-known physiological manifestation of microbial nutrient preferences [8,9]. When cells grow in a mix of preferred and alternative nutrients, they first consume the preferred nutrient to exhaustion. Then, growth slows while cells induce alternative nutrient utilization genes; this period is called the "diauxic lag" and can last up to several hours. Finally, cells resume growth on the alternative nutrient. The findings of Escalante-Chong, Savir, et al. imply that cells could potentially activate GAL genes before glucose exhaustion and thereby avoid a diauxic lag. Additionally, strain variation in the threshold of ratio-sensing should result in a spectrum of diauxic lag across natural yeast isolates. In fact, variation in diauxic lag had already been seen across natural yeast strains in glucose-maltose mixtures [10] and in experimentally evolved bacteria [11] and yeast [10]. We confirmed the connection between GAL regulatory variation and diauxic lag, and unexpectedly found that GAL gene expression has a considerable fitness cost, suggesting a rationale for why diverse regulatory phenotypes exist. These conclusions have been echoed by other studies published alongside and after ours [12,13].

The differences in diauxic lag seemed to be caused by differences in how the GAL pathway integrates signals from glucose and galactose. This is done by the 7 genes of the canonical GAL pathway [6] and a much larger glucose signaling network [14,15]. In the GAL pathway, the Gal1p, Gal7p, and Gal10p enzymes help catabolize galactose [6]. Gal2p permease transports galactose into the cell [16]. Intracellular galactose binds Gal3p and converts it to an active form [17]. Activated Gal3p sequesters the Gal80p repressor [18], thus preventing it from sequestering Gal4p [19]. Free Gal4p binds to promoters of *GAL1, GAL2, GAL7, GAL10,* and *GAL80* and activates their expression [20]. These genes are unexpressed in the absence of galactose and actively repressed in the presence of glucose by the transcriptional repressor Mig1p [21,22]. Additional glucose-mediated mechanisms trigger degradation of GAL pathway

transcripts and Gal2p transporters. GAL pathway components are obvious candidates for mutations driving the natural variation; indeed, multiple related yeast species have lost function of some GAL genes, suggesting rapid evolution in this pathway [23,24]. On the other hand, polymorphisms in glucose signaling are also plausible. Experimentally evolved yeast strains with shortened glucose-maltose diauxic lag were found to have mutations in the glucose signaling genes *HXK2* and *STD1* [10].

The strain differences in GAL regulation we observed are particularly interesting in light of numerous recent studies dissecting the genetics of nutrient utilization and stress resistance across natural yeast isolates [25–29]. These studies showed that budding yeast harbor extensive phenotypic variation as a species, based on a changing repertoire of alleles required for growth in various environments. At a practical level, these studies demonstrated that meiotic recombination and linkage analysis, long staples of yeast genetics, are also powerful tools for revealing the molecular mechanisms underlying natural variation [30–33]. Despite these advances, the effects of natural polymorphisms on quantitative cellular decisions are mostly unexplored. In chapter 3, we address this gap by mapping the mutations underlying strain differences in glucose-galactose signaling.

In chapters 2 and 3, we were intrigued to see an array of bimodal and unimodal GAL induction patterns across our strains. In the most obvious cases, a subpopulation of cells was uninduced while another subpopulation induced at near-maximal levels. The range of sugar conditions that elicited bimodality varied across strains, and some strains produced a unimodal response in all conditions tested. Bimodality, and non-genetic heterogeneity more broadly, is widely observed across microbes and multicellular organisms, and can play important roles in environmental sensing, differentiation, and disease [34–37]. A variety of mechanisms, typically

involving positive feedback, can give rise to bimodality [35,38–40]. In the GAL pathway, bimodality is attributed to bistability [41] and requires positive feedback through Gal1p and Gal3p [42]. Additionally, the sugar-dosage-response of GAL expression distributions is quantitatively altered by perturbing *GAL1*, *GAL2*, *GAL3*, *GAL80* [41,43,44], or *MIG1* [22]. Under certain environmental conditions and genetic perturbations, GAL genes can also convert to a unimodal, graded behavior [22]. This latter phenomenon is poorly understood and seems to be phenocopied by the differences between our strains, which we know can be dissected genetically. In chapter 4, we took this opportunity to delve into the genetic factors affecting the modality of the GAL response.

Chapter 4 was also motivated by a methodological challenge when studying bimodality or any other distributed phenotype: high dimensionality. The behaviors of mutants often vary in multiple ways compared to the wildtype, and usually the analysis must focus on a summary metric that seems most salient. For example, most recent works, including ours in chapter 3, compute the "induced fraction" to quantify the degree of induction of GAL genes [44,45]. However, there are many ways a distribution can contain X% induced cells. This means that most of the quantitative behavior encoded in the GAL response is currently being ignored by the literature (with a few recent exceptions [12,42]). Given that the GAL pathway is one of the best-characterized models for quantitative gene regulation, many new insights are likely if the full distribution of the GAL response can be taken into account or dimensionally reduced in an unbiased way. Our work in chapter 4 makes a step toward this goal by considering a second dimension of the GAL response that is subject to natural variation. This allows us not only to uncover an additional set of genetic variants affecting GAL regulation, but obtain a deeper understanding of the dimension that is already widely studied.

References

1. Perkins TJ, Swain PS. Strategies for cellular decision-making. Mol Syst Biol. Nature Publishing Group; 2009;5: 326. doi:10.1038/msb.2009.83

2. Balázsi G, van Oudenaarden A, Collins JJ. Cellular decision making and biological noise: from microbes to mammals. Cell. 2011;144: 910–25. doi:10.1016/j.cell.2011.01.030

3. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. 2008;9: 583–93. doi:10.1038/nrg2398

4. Bhat PJ. Galactose Regulon of Yeast: From Genetics to Systems Biology. Berlin: Springer; 2008.

5. Johnston M. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. Microbiol Rev. 1987;51: 458.

6. Lohr D, Venkov P, Zlatanova J, Program B, Academy B. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. 1995;9: 777–787.

7. Escalante-chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. Proc Natl Acad Sci. 2015;112: 1636–1641. doi:10.1073/pnas.1418058112

8. Dienert F. Sur la fermentation du galactose et sur l'acoutumance des levures à ce sucre. Faculté des sciences de Paris. 1900.

9. Monod J. Recherches sur la croissance des cultures bacteriennes. Paris: Hermann; 1942.

10. 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. doi:10.1371/journal.pbio.1001764

11. 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: 245–60.

12. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. PLoS Biol. 2015;13: e1002042. doi:http://dx.doi.org/10.1101/002907

13. Roop JI, Chang KC, Brem RB. Polygenic evolution of a sugar specialization trade-off in yeast. Nature. Nature Publishing Group; 2016;

14. Zaman S, Lippman SI, Schneper L, Slonim N, Broach JR. Glucose regulates transcription in yeast through a network of signaling pathways. Mol Syst Biol. 2009;5: 245.

doi:10.1038/msb.2009.2

15. Broach JR. Nutritional Control of Growth and Development in Yeast. Genetics. 2012;192: 73–105. doi:10.1534/genetics.111.135731

16. Tschopp JF, Emr SD, Field C, Schekman R. GAL2 codes for a membrane-bound subunit of the galactose permease in Saccharomyces cerevisiae. J Bacteriol. 1986;166: 313–318.

17. 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 Dev. 2012;26: 294–303.

18. Peng G, Hopper JE. Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. Proc Natl Acad Sci U S A. 2002;99: 8548–53. doi:10.1073/pnas.142100099

19. 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. Mol Cell Biol. 2013;33: 3667–3674.

20. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, et al. Genome-wide location and function of DNA binding proteins. Science. 2000;290: 2306–9. doi:10.1126/science.290.5500.2306

21. Nehlin JO, Carlberg M, Ronne H. Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. EMBO J. 1991;10: 3373–7.

22. Biggar SR, Crabtree GR. Cell signaling can direct either binary or graded transcriptional responses. EMBO J. 2001;20: 3167–76. doi:10.1093/emboj/20.12.3167

23. Hittinger CT, Rokas A, Carroll SB. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. Proc Natl Acad Sci U S A. 2004;101: 14144–9. doi:10.1073/pnas.0404319101

24. Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A. High-resolution yeast phenomics resolves different physiological features in the saline response. Proc Natl Acad Sci U S A. 2003;100: 15724–9. doi:10.1073/pnas.2435976100

25. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James S a, et al. Population genomics of domestic and wild yeasts. Nature. Nature Publishing Group; 2009;458: 337–41. doi:10.1038/nature07743

26. Warringer J, Zorgo E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, et al. Trait Variation in Yeast Is Defined by Population History. PLoS Genet. 2011;7. doi:10.1371/journal.pgen.1002111

27. Schacherer J, Shapiro J a, Ruderfer DM, Kruglyak L. Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. Nature. Nature

Publishing Group; 2009;458: 342–5. doi:10.1038/nature07670

28. Liti G, Louis EJ. Advances in quantitative trait analysis in yeast. PLoS Genet. 2012;8: e1002912. doi:10.1371/journal.pgen.1002912

29. Ehrenreich IM, Bloom J, Torabi N, Wang X, Jia Y, Kruglyak L. Genetic architecture of highly complex chemical resistance traits across four yeast strains. PLoS Genet. 2012;8: e1002570.

30. Cubillos F a, Parts L, Salinas F, Bergström 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.

31. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. 2013;494: 234–237.

32. Treusch S, Albert FW, Bloom JS, Kotenko IE, Kruglyak L. Genetic Mapping of {MAPK-Mediated} Complex Traits Across S. cerevisiae. PLoS Genet. 2015;11: e1004913.

33. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro J a, et al. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. Nature Publishing Group; 2010;464: 1039–42. doi:10.1038/nature08923

34. Veening J-W, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol. 2008;62: 193–210. doi:10.1146/annurev.micro.62.081307.163002

35. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet. 2005;6: 451–464. doi:10.1038/nrg1615

36. Grimbergen AJ, Siebring J, Solopova A, Kuipers OP. Microbial bet-hedging: the power of being different. Curr Opin Microbiol. 2015;25: 67–72.

37. Macarthur BD, Ma'ayan A, Lemischka IR. Systems biology of stem cell fate and cellular reprogramming. Nat Rev Mol Cell Biol. Nature Publishing Group; 2009;10: 672–681. doi:10.1038/nrm2766

38. To T-L, Maheshri N. Noise can induce bimodality in positive transcriptional feedback loops without bistability. Science. 2010;327: 1142–5. doi:10.1126/science.1178962

39. Keller AD. Model genetic circuits encoding autoregulatory transcription factors. J Theor Biol. 1995;172: 169–185. doi:10.1006/jtbi.1995.0014

40. Ferrell JE. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Curr Opin Cell Biol. 2002;14: 140–148. doi:10.1016/S0955-0674(02)00314-9

41. Acar M, Becskei A, Oudenaarden A van. Enhancement of cellular memory by reducing

stochastic transitions. Nature. 2005;435: 228. doi:10.1038/nature03524

42. Venturelli OS, El-Samad H, Murray RM. Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. Proc Natl Acad Sci U S A. 2012;109: 1–10. doi:10.1073/pnas.1211902109

43. Hawkins KM, Smolke CD. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. J Biol Chem. 2006;281: 13485–92. doi:10.1074/jbc.M512317200

44. Acar M, Pando BF, Arnold FH, Elowitz MB, van Oudenaarden A. A general mechanism for network-dosage compensation in gene circuits. Science (80-). 2010;329: 1656–60. doi:10.1126/science.1190544

45. Peng W, Liu P, Xue Y, Acar M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nat Commun. 2015;6: 6226.

Chapter 2.

Natural Variation in Preparation for Nutrient Depletion Reveals a Cost–Benefit Tradeoff

Jue Wang, Esha Atolia, Bo Hua, Yonatan Savir, Renan Escalante-Chong, Michael Springer

Adapted from Wang et al. (2015) PLoS Biology. doi: 10.1371/journal.pbio.1002041

Maximizing growth and survival in the face of a complex, time-varying environment is a common problem for single-celled organisms in nature. When offered two different sugars as carbon sources, microorganisms first consume the preferred sugar, then undergo a transient growth delay, the "diauxic lag", while inducing genes to metabolize the less preferred sugar. This delay is assumed to be inevitable, due to selection to maximize use of the preferred sugar. Contrary to this view, we found that many natural isolates of *Saccharomyces cerevisiae* display short or non-existent diauxic lags when grown in mixtures of glucose (preferred) and galactose. These strains induce galactose-utilization (GAL) genes hours before glucose exhaustion, thereby "preparing" for the transition from glucose to galactose metabolism. The extent of preparation varies across strains, and seems to be determined by the steady-state response of GAL genes to mixtures of glucose and galactose rather than by induction kinetics. Although early GAL induction gives strains a competitive advantage once glucose runs out, it comes at a cost while glucose is still present. Costs and benefits correlate with the degree of preparation: strains with higher expression of GAL genes prior to glucose exhaustion experience a larger upfront growth cost but also a shorter diauxic lag. Our results show that classical diauxic growth is only one extreme on a continuum of growth strategies constrained by a cost-benefit tradeoff. This type of continuum is likely to be common in nature, as similar tradeoffs can arise whenever cells evolve to use mixtures of nutrients.

Introduction

Natural environments contain complex, time-varying mixtures of nutrients and stresses. Understanding how cells use external cues to maximize growth and survival is key to understanding the evolution and function of regulatory circuits. Gene regulation allows cells to express pathways for specific tasks only in conditions when they are needed, to maximize the benefit of these pathways while minimizing their metabolic cost [1–4]. Regulatory circuits have evolved elaborate behaviors such as bet-hedging, signal integration, and environmental anticipation in response to the complexity of natural environments [5].

A classic example of gene regulation occurs during microbial growth on mixtures of carbon sources. For example, when budding yeast or *E. coli* grow in the sugars glucose and galactose, they first consume glucose, while dedicated signaling mechanisms repress galactose-utilization (GAL) genes [6–11]. When glucose has been exhausted, cells temporarily stop growing, induce GAL genes, and start growing again. The transient pause in growth, called the diauxic lag, can last up to several hours.

The diauxic lag is commonly thought to be a consequence of selection to minimize expression of superfluous metabolic pathways when a nutrient that can be more efficiently utilized is available [12–14]. This is supported by the observation that cells growing in two sugars that support similar growth rates do not exhibit a diauxic lag [8]. However, recent studies have shown that even in the same nutrient mixture, the length of diauxic lag can vary among experimentally evolved bacterial strains [15,16] or yeast isolates [17]. In both cases, evolved strains lacking a diauxic shift possessed weaker catabolite repression of secondary carbon pathways than the ancestor. This leads to a fitness cost during growth in the preferred nutrient, but a fitness advantage when the environment shifts rapidly between preferred and alternative

nutrients. These results raise the question of whether similar mechanisms and fitness tradeoffs underlie the diauxic lag variation seen in natural yeast isolates [17].

To address this question, we monitored culture density and gene expression in ecologically diverse *Saccharomyces cerevisiae* natural isolates growing in mixtures of glucose and galactose. As expected, we found a spectrum of diauxic lag phenotypes, from strains with non-existent lags to those with more classical lag times of many hours. Strikingly, the variation in lag time is not due to differences in how fast strains can execute induction of GAL genes, but rather the timing of when they begin to induce. Short-lag strains induce GAL genes up to 4 hours before glucose is exhausted, in effect "preparing" for the transition to galactose metabolism. The degree of preparation correlates with the strength of glucose repression; strains that induce GAL genes at higher glucose levels also induce them earlier during diauxic growth. These results suggest that natural variation in catabolite repression is not only a key determinant of microbial fitness during sudden nutrient shifts [17], but also gradually changing nutrient conditions. Finally, we show that the observed phenotypic variation follows a tradeoff: early GAL induction benefits cells by preparing them for glucose exhaustion, but the cost of expressing GAL genes reduces growth rate while glucose is still present. This tradeoff is likely a general constraint on microbial growth strategies in mixed-nutrient environments.

Results

Natural yeast strains vary in length of diauxic lag

We grew 43 *Saccharomyces cerevisiae* strains in a carbon-limited medium containing 0.25% glucose and 0.25% galactose, the preferred and alternative carbon source respectively (Figure 2.1A). The *S. cerevisiae* strains come from a range of geographical locations and environments [18,19] and are all prototrophic, allowing us to omit amino acids from the media and avoid

potential complications from their role as alternative carbon sources [20]. Bulk growth of the cells was measured by recording the optical density of each culture every 10 minutes for 44 hours using an automated plate reader (Materials and methods).

The growth curves generally display an initial phase of fast growth followed by a second phase of relatively slower growth, as expected in a 2-sugar mixture (Figure 2.1B-C, S2.1). However, the strains varied in the extent of growth lag, or a local minimum in growth rate, between the two growth phases (top versus bottom strains in Figure 2.1B-C). Some strains (e.g. YJM978) had a long diauxic lag during which growth rate almost reaches zero, whereas some strains (e.g. BC187) had a brief lag period during which even the minimum growth rate was relatively high. Strain SLYG78, a derivative of the commonly used laboratory strain S288C, exhibits a prominent lag phase (Figure S2.1), consistent with previous studies and the traditional understanding of *S. cerevisiae* as having a diauxic-growth phenotype [6,17]

To quantify the variation in diauxic lag, we defined a "diauxic lag time" metric as the time required to reach a strain's smoothed maximal growth rate in galactose after having dropped below this growth rate during glucose depletion (horizontal black lines in Figure 2.1C, S2.2B, Materials and Methods). In growth curves that do not have a local growth-rate minimum, we defined the lag time as zero (Figure 2.1C, S2.2B). This lag time metric was robust to small differences in culture behavior ($R^2 = 0.96$; Figure S2.2C) and to the method of calculation (Figure S2.2D, Materials and Methods).

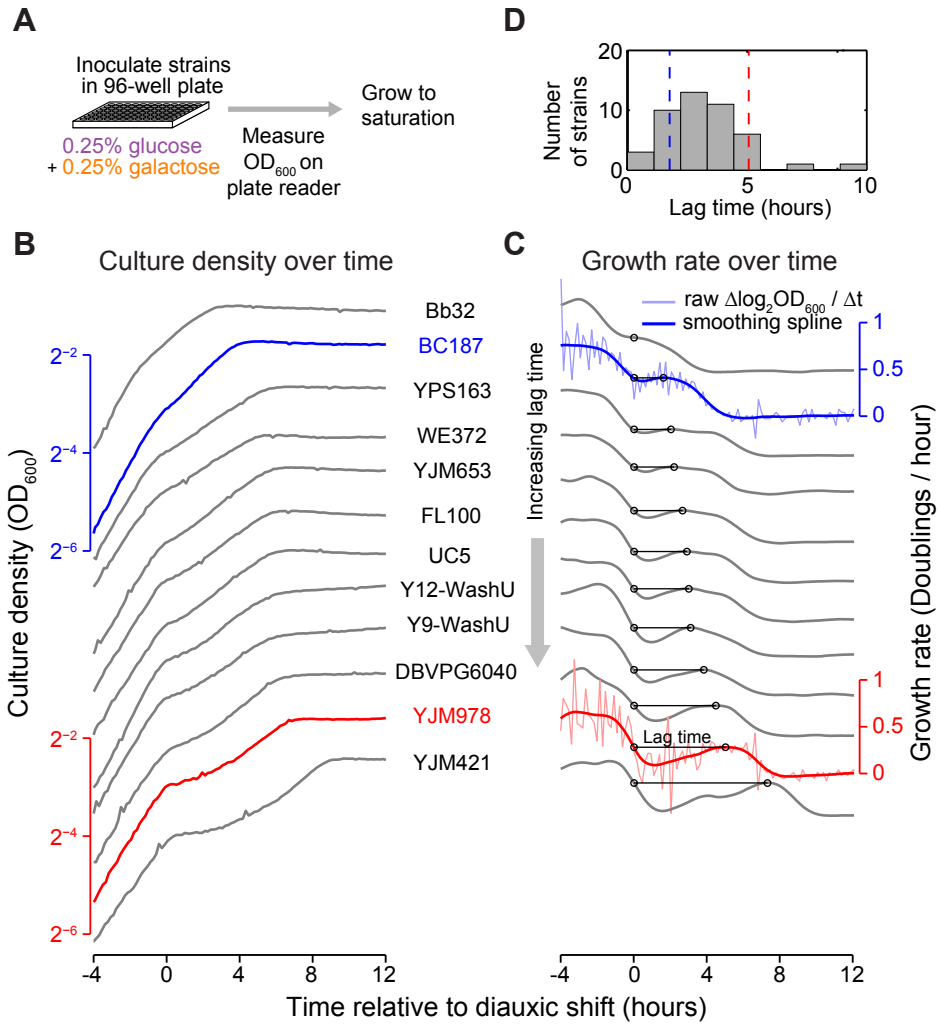


Figure 2.1. Natural yeast strains vary in length of diauxic lag.

(A) Schematic of growth curve experiment in "diauxic growth conditions", defined as batch culture in synthetic minimal media with 0.25% glucose and 0.25% galactose. **(B)** Growth curves (OD_{600} versus time relative to diauxic shift) plotted top-to-bottom in order of increasing diauxic lag. A single replicate growth curve is shown for each of 11 strains with similar growth rates in galactose-only media. (Growth curves for both replicates of all 43 strains assayed are shown in Figure S2.1). Strains BC187 and YJM978 are highlighted in red and blue, respectively. **(C)** Smoothed growth rate versus time relative to diauxic shift for the same strains as in (B). Example plots of raw OD differentials (light blue, light red) used to obtain the smoothed growth rate are shown for BC187 and YJM978. Diauxic lag time metric is denoted by horizontal black line with circles (see also Figure S2.2). (D) Histogram of diauxic lag time across all natural isolates assayed. Data used for histogram are the mean of 2 replicates (Materials and methods).

We found that diauxic lag time varies continuously in our strains from 0 to 9 hours, with a mean of 3.2 hours and a standard deviation of 1.6 hours (Figure 2.1D). The continuous nature of the observed variation was robust to the choice of metric, as a related but distinct growth-curve feature, the minimum mid-diauxic growth rate, also varies continuously and correlates strongly with lag time ($R^2 = 0.71$, Figure S2.2C). Lag time was not correlated with growth rate in pure glucose or galactose, and even among a subset of strains with similarly high growth rates in galactose-only media (subset shown in Figure 2.1B-C) we saw wide variability in the diauxic lag time (Figure S2.3). This suggests that the observed variation is due to differences in metabolic regulation rather than in maximal sugar utilization rates.

Several strains displayed no measurable diauxic lag and seem to transition instantly from glucose consumption to galactose consumption. This implies either that these strains can induce GAL genes quickly upon glucose exhaustion, induce GAL genes before glucose exhaustion, or both. To examine these possibilities, we characterized strains YJM978 and BC187, which represent long-lag and short-lag phenotypes, respectively (red and blue curves in Figure 2.1).

Strain BC187 induces galactose-responsive genes before glucose is exhausted

We cultured BC187 and YJM978 in 0.25% glucose plus 0.25% galactose and monitored GAL pathway expression and glucose and galactose concentrations until saturation, when both sugars were depleted (Figure 2.2A, Materials and methods). We refer to this as a "diauxic growth experiment." To enable single-cell measurement of GAL gene induction, we integrated a cassette containing yellow fluorescent protein driven by the GAL1 promoter (GAL1pr-YFP), which has been shown to be a faithful proxy for GAL pathway expression [21–23], at a neutral chromosomal locus (Figure 2.2A, top, Materials and Methods). We measured GAL1pr-YFP expression and extracellular sugar concentration by flow cytometry and enzymatic assay,

respectively, over the entire diauxic growth cycle (Figure 2.2A, bottom). To quantify the timing of GAL gene induction, we defined t_{low} and t_{high}, respectively, as the time when GAL1pr-YFP expression reaches 2-fold above basal levels and 1/4 of maximal levels, relative to the moment of glucose exhaustion (Figure 2.2B).

Strain YJM978, which has a long diauxic lag, does not induce galactose-responsive genes until after glucose is exhausted, consistent with the classical understanding of diauxic growth (Figure 2.2C; t_{low} = 1.7 ± 0.1 hours, t_{high} = 2.7 ± 0.1 hours). In contrast, BC187, which has a short diauxic lag, begins GAL induction significantly before glucose exhaustion (Figure 2.2D; t_{low} = -3.0 ± 0.1 hours, p = 0.02 by t-test on n = 2 replicates). Even using the more conservative t_{high} metric, BC187 reaches near-maximal induction before glucose exhaustion (t_{high} = -0.5 ± 0.1 hours). Pre-induction of GAL genes by BC187 leads to significant galactose consumption, even before glucose is fully exhausted (Figure S2.5). Both strains use glucose and galactose to completion and reach a similar yield (Figure S2.1), indicating that differences in induction time are not due to drastic differences in carbon utilization efficiency. Both strains have undetectable GAL1pr-YFP expression in glucose-only media (Figure S2.6), ruling out the possibility that galactose metabolism is constitutively active in BC187. In effect, BC187 "prepares" for the diauxic shift by inducing GAL genes before glucose exhaustion.

Figure 2.2. A short-lag strain induces galactose-utilization (GAL) genes hours before the diauxic shift. **(A)** *Top*: Schematic of GAL1pr-YFP transcriptional reporter and cartoon of fluorescence distribution as measured by flow cytometry. *Bottom*: Schematic of diauxic growth GAL induction experiment. **(B)** Definitions of induction metrics, t_{low} and t_{high}, when reporter expression is at low but above-basal or near-maximal levels, respectively. Diauxic growth for strains **(C)** YJM978 and **(D)** BC187. GAL reporter expression distributions (gray shading), GAL reporter median (red line), glucose concentration (purple circles), and galactose concentration (orange circles). Time is defined relative to the moment when culture achieves a density of 10^6 cells / mL (Figure S2.4). Purple and orange lines are smoothing spline fits to glucose and galactose measurements. Dotted purple line indicates time of glucose exhaustion, calculated using a local linear fit (Materials and methods). Data shown in (B) and (C) represent two replicate experiment. GAL reporter expression distribution is shown for only one of the two replicates. **(E)** Comparison of induction start time, t_{low}, and near-maximal induction time, t_{high}, for YJM978 (red bars) and BC187 (blue bars) cultures. Bars and error bars represent the mean and range, respectively, of two replicates.

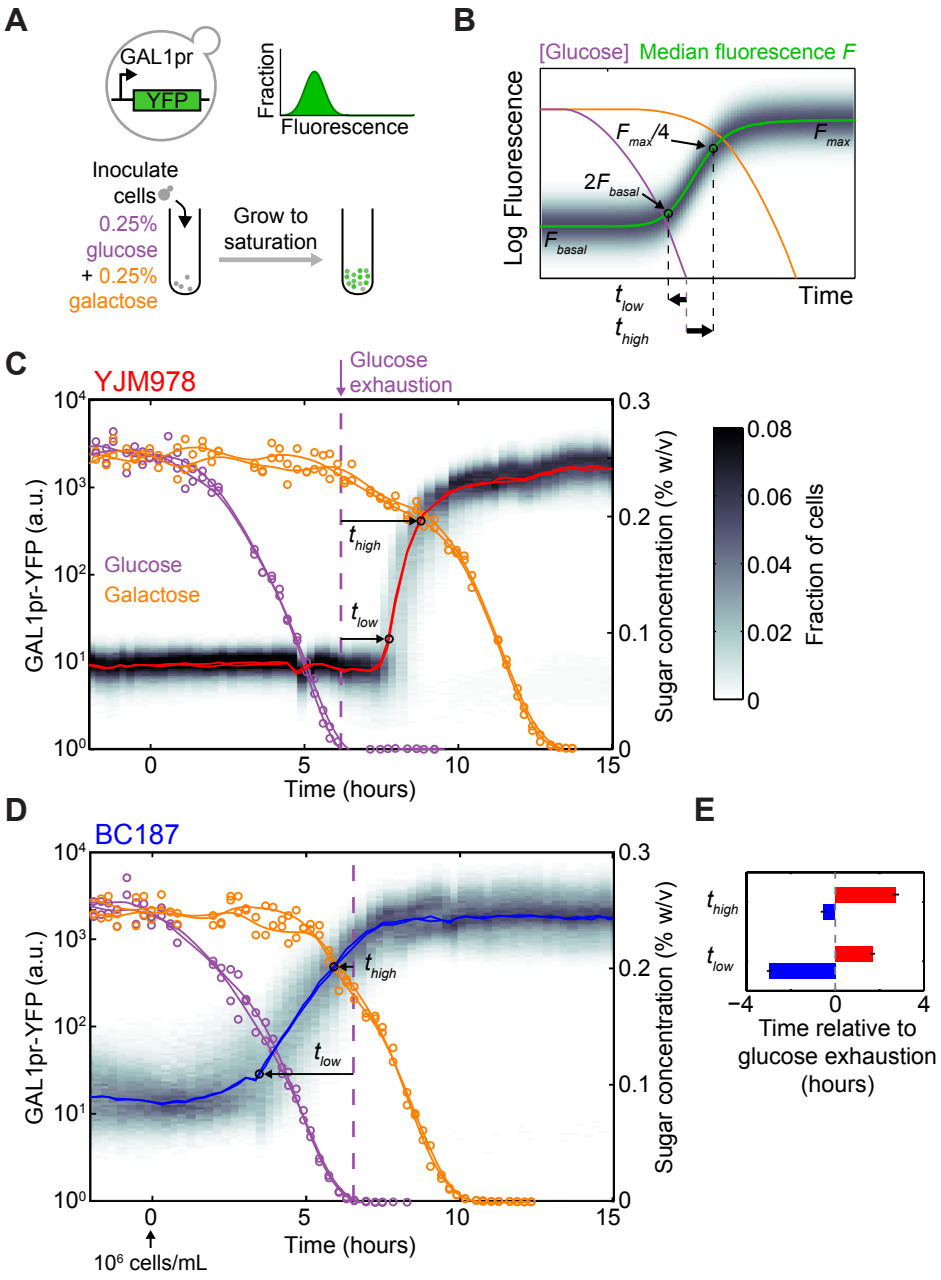


Figure 2.2 (Continued). A short-lag strain induces galactose-utilization (GAL) genes hours before the diauxic shift.

Preparation is a continuous trait among natural yeast isolates

To determine if GAL induction prior to glucose exhaustion is a typical behavior of natural isolates, we integrated a GAL1pr-YFP reporter into 13 additional strains (for a total of 15 strains; see Table S1B) and monitored their expression during diauxic growth (same conditions as in Figure 2.2, Materials and Methods). Directly measuring sugar concentrations is laborious and less precise than measuring YFP fluorescence by flow cytometry, so we used YJM978 as a "reference" strain to signal the exhaustion of glucose, and co-cultured it with a "query" strain whose GAL induction kinetics we wanted to assay (Figure 2.3A). The reference strain was modified to express a fluorescent marker to distinguish it from the query strain (Figure S2.7A, Materials and Methods). To quantify differences in GAL induction time, we defined the "preparation time" as the difference in time between when the query and reference strains reach 1/16 of their maximal median GAL1pr-YFP expression (Figure 2.3B-C). Preparation time ranged from -3.8 to 0.04 hours relative to YJM978 with a mean of -1.3 hours, indicating that most strains induce GAL genes earlier than YJM978. The preparation time measured by this method is highly reproducible and robust to the query-to-reference mixing ratio (Figure S2.7B,C,E,F, Materials and Methods).If the degree of preparation determines the extent to which a strain has a diauxic shift, then strains that begin inducing GAL genes earlier should also have a shorter diauxic lag. We find a strong correlation ($R^2 = 0.83$, $p = 9.2 \times 10^{-7}$) between preparation time and the diauxic lag time (Figure 2.3D). However, earlier-inducing strains appeared to take longer to reach full induction, or "execute" induction more slowly. We defined the "execution time" as the time required for median GAL1pr-YFP expression to increase from 1/64 to 1/4 of its maximal level (Figure 2.3B-C). The execution time anticorrelated with preparation time (Figure 2.3E inset) and lag time (Figure 2.3E), contradicting the naive expectation that a strain with a shorter diauxic lag will induce GAL genes more quickly. Taken together, our data show that the

length of diauxic lag correlates to *when* strains begin to transition to galactose metabolism, not *how fast* they can execute the transition once they begin.

Related studies have observed population heterogeneity of growth rates and gene expression during sudden media shifts and diauxic growth [17,23,24]. In our experimental conditions, strains BC187 and YJM978 do not display bimodality in GAL1pr-YFP expression during diauxic growth (Figure 2.2C,D). A small number of strains do display bimodal expression at steady-state in glucose + galactose (Figure S2.6) and possibly also during diauxic growth (Figure S2.7D, G-I). However, the time window of any bimodality is short compared to induction time differences between strains (Figure S2.7G-I, Materials and methods). Therefore, although single-cell variation is likely an important dimension of regulatory behavior in some strains [17,23,24], our analysis of population-level dynamics already captures a major regulatory mode of diauxic growth.

Figure 2.3. Diauxic lag time is correlated with the start time of GAL induction.

(A) Schematic of co-culture GAL induction experiment. Each of 15 query strains (gray) are co-cultured with reference strain YJM978 expressing constitutive mCherry marker (red), and sampled for flow cytometry every 15 minutes from mid-exponential phase until saturation. **(B)** Illustration of how preparation time and execution time metrics are defined. **(C)** Median GAL1pr-YFP expression versus time for query (gray) and reference (red) strain in 3 co-cultures selected to illustrate a range of preparation times. Strain I14 had above-basal reporter expression at the start of sampling, so its execution time was computed by linear extrapolation. **(D)** Scatterplot of diauxic lag time (from Figure 2.1) versus preparation time. **(E)** Scatterplot of diauxic lag time versus execution time. **(E, inset)** Scatterplot of preparation time versus execution time. Dotted gray lines in (D) and (E) indicate least-squares linear fits used to calculate coefficients of determination R^2 and p-values. Data for diauxic lag time are the mean and range of two replicates, and for preparation time and execution time are the mean and s.e.m. of three replicates.

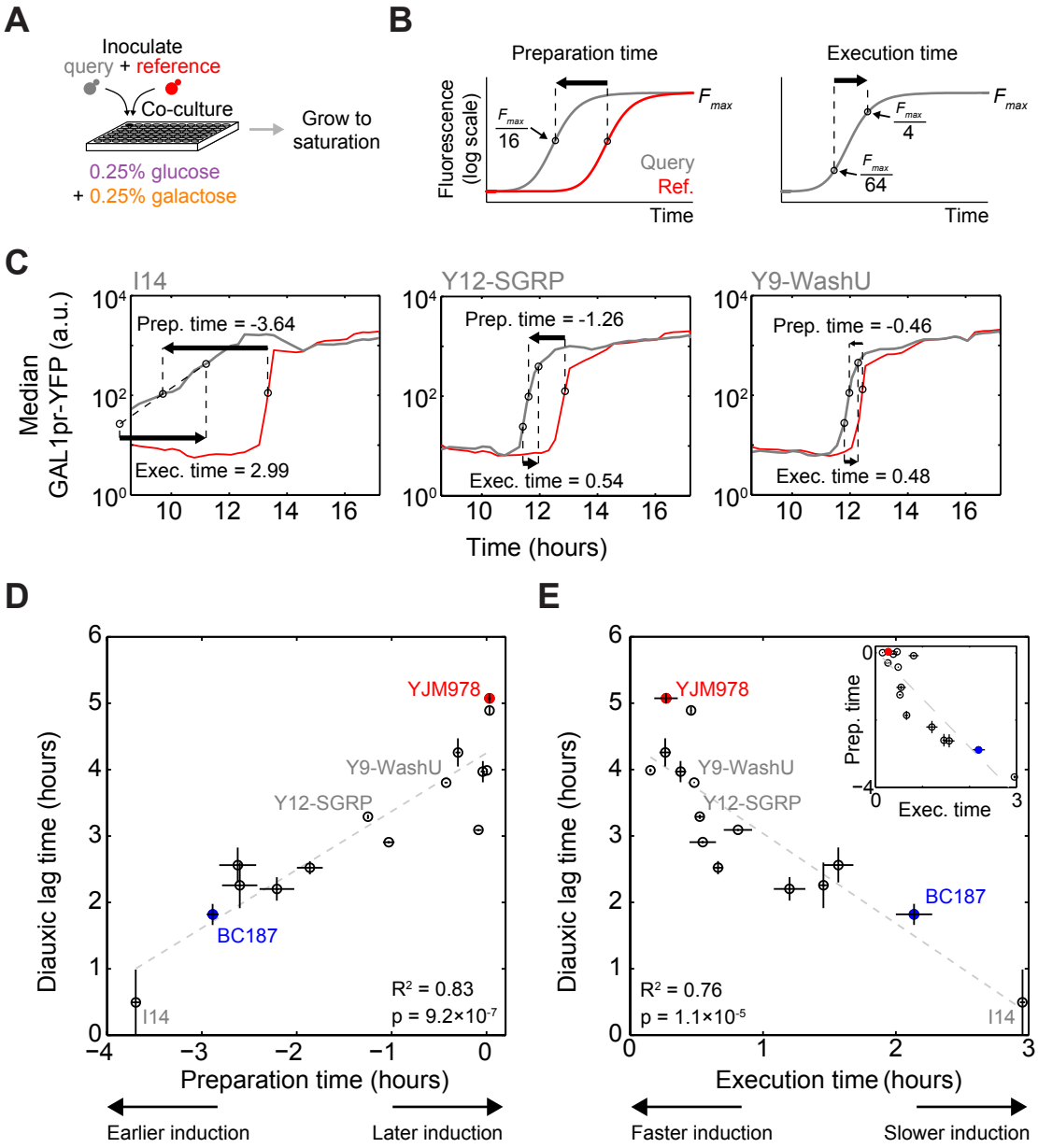


Figure 2.3 (Continued). Diauxic lag time is correlated with the start time of GAL induction.

GAL induction kinetics after sudden media shift are poorly correlated with diauxic lag time

Our observations above rule out a model of the diauxic lag in which all strains begin inducing upon glucose exhaustion and vary only in how quickly they can reach maximal induction. However, it is possible that instead of inducing at glucose exhaustion, all strains induce when glucose is depleted below a certain threshold and vary in the delay before displaying observable GAL1pr-YFP expression. In this scenario, strains with a short delay between the start of induction and observable GAL1pr-YFP expression would appear to be preparing whereas strains with a long delay would appear to be inducing only after glucose exhaustion.

When cells are grown in glucose, the GAL pathway is repressed [7,25]. To ask whether differences in glucose de-repression kinetics could explain diauxic lag variation in our natural isolates, we grew strains in 2% glucose and transferred them into 2% galactose. We found significant variation in induction delay, defined as the time until median GAL1pr-YFP expression has increased 2-fold after transfer into galactose (Figure 2.4A). Some strains began to induce 5 hours after media switch, while one strain did not induce even after 18 hours. In contrast, the execution time of induction varied only between 0.6 to 1.6 hours (Table S2), suggesting that once glucose repression is relieved, GAL expression quickly induces from basal to maximal in all strains. Strikingly, induction delay after glucose-to-galactose shift was a poor predictor of both preparation time (Figure 2.4B; $R^2 = 0.16$) and diauxic lag time (Figure 2.4C; $R^2 = 0.13$). In particular, strains BC187 and I14 have short diauxic lags and early preparation times but very long induction delays after glucose-to-galactose media shift. Strain I14 had a similar behavior. When these two strains were omitted from the data, a weak correlation emerged ($R^2 = 0.56$; $p = 0.005$), suggesting that glucose de-repression kinetics may play a role in the diauxic lag in our strains, but potentially convolved with a second response such as cell stress.

Figure 2.4. Diauxic lag time is correlated poorly with GAL induction kinetics but strongly with steady-state GAL expression in a glucose-galactose mixture.

(A) Median GAL1pr-YFP expression versus time for BC187 (blue line), YJM978 (red line), and 13 other strains (gray lines) after transfer from 2% glucose into 2% galactose. **(B)** Scatterplot of preparation time (from Figure 2.3) versus induction delay after glucose-to-galactose shift, defined as the time until median GAL expression reaches 2-fold above basal expression. Black triangle indicates strain YJM981, which did not induce above background during the entire 18-hour experiment; this strain was omitted from the R^2 calculation. (C) Scatterplot of diauxic lag time (from Figure 2.1) versus induction delay after glucose-to-galactose shift. **(D)** *Top:* Schematic of how sugar concentrations for steady-state measurements were chosen from the diauxic growth experiment. *Bottom:* Measured steady-state GAL1pr-YFP expression distributions for BC187, YJM978, and 13 other strains in 0.0625% glucose + 0.25% galactose. **(E)** Scatterplot of preparation time versus mean steady-state GAL1pr-YFP expression from (D). **(F)** Scatterplot of diauxic lag time versus mean steady-state GAL1pr-YFP expression from (D). Dotted gray lines in (B), (C), (E), and (F) indicate least-squares linear fits used to calculate coefficients of determination R^2 and p-values.

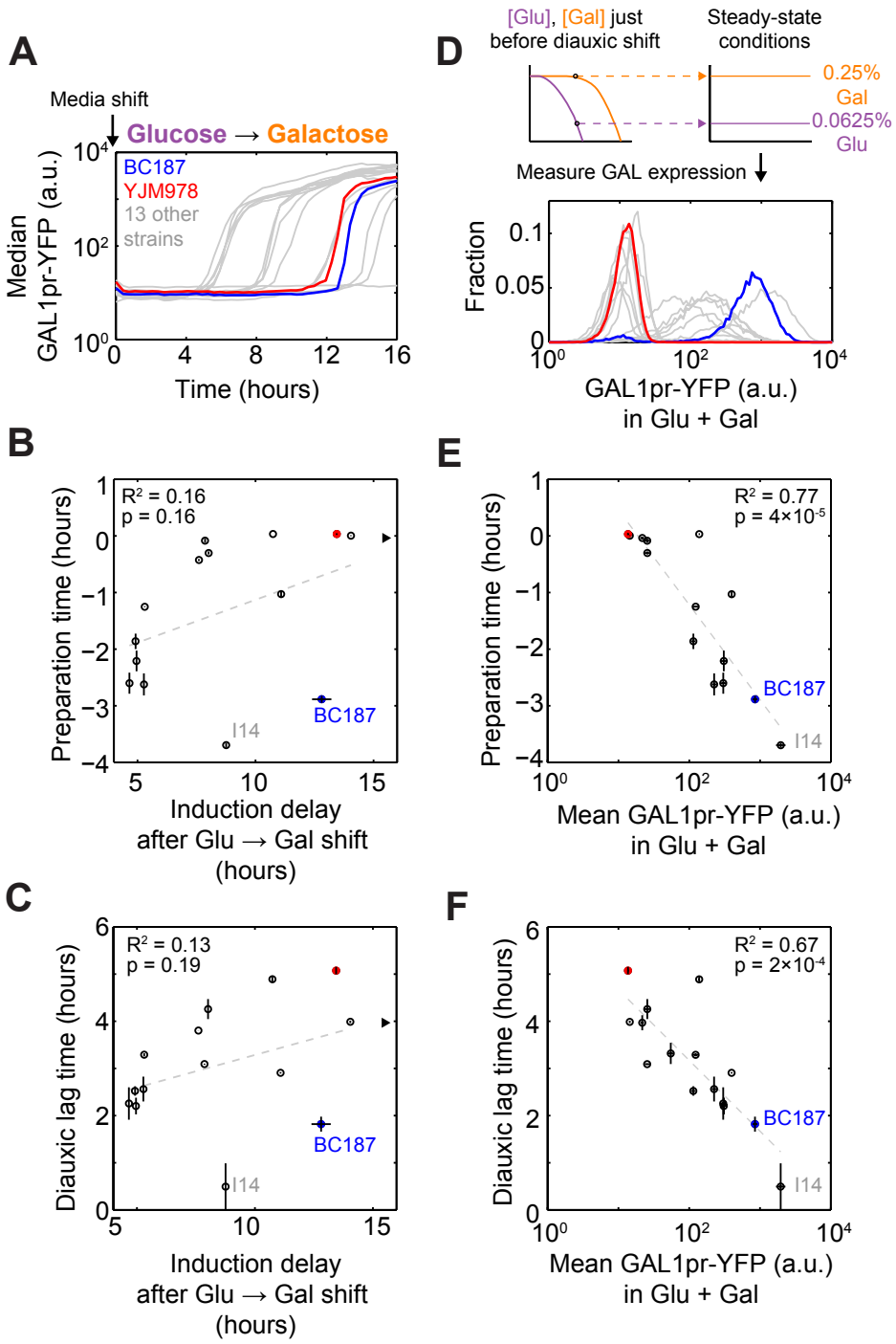


Figure 2.4 (Continued). Diauxic lag time is correlated poorly with GAL induction kinetics but strongly with steady-state GAL expression in a glucose-galactose mixture.

Differences in steady-state sugar sensing explains variation in preparation and diauxic lag time

Given that some strains can induce GAL genes in the presence of glucose, we hypothesized that differences in steady-state GAL expression in glucose + galactose may underlie differences in preparation. We measured GAL expression of our natural isolates in 0.0625% glucose + 0.25% galactose to simulate the conditions of a diauxic batch culture just before glucose exhaustion (Figure 2.4D). To ensure that we observed the steady-state response of the GAL reporter, we measured induction after cultures reached steady-state but before appreciable glucose had been depleted (Figure S2.8, Materials and Methods). We found that steady-state GAL expression in glucose + galactose varied from uninduced to almost maximal (Figure 2.4D, S6). On the other hand, all strains were uninduced in glucose-only media and maximally induced in galactose-only media (Figure S2.6), suggesting that strains vary not in overall glucose-repressibility or inducibility of GAL genes, but in how they integrate signals from both sugars in the mixed environment prior to diauxic shift.

We found that steady-state GAL expression in the glucose-galactose mixture correlated significantly with preparation time (Figure 2.4E; $R^2 = 0.77$, $p = 4x10^{-5}$) and diauxic lag time (Figure 2.4F; $R^2 = 0.67$, $p = 2x10^{-4}$). In other words, the strains that induce earlier during diauxic growth are those with higher steady-state GAL1pr-YFP expression in glucose + galactose. This suggests that short-lag strains do not suddenly switch GAL genes from "off" to "on" during diauxic growth, but instead express them at quasi-steady-state levels appropriate to the degree of glucose depletion. Consistent with this, the steady-state GAL1pr-YFP expression of these strains is proportional to their expression 3 hours before reference strain induction during diauxic growth (Fig S9A). Furthermore, BC187 grown in 3 sugar mixtures representing different moments during diauxic growth (0.25% galactose plus 0.25%, 0.125%, or 0.0625% glucose)

displayed intermediate GAL1pr-YFP expression even after reaching steady-state, not basal or maximal expression as would be expected for a switch-like response (Figure S2.8, S2.9B). Taken together, our data strongly suggest that differences in preparation, and therefore diauxic lag time, are due to differences in the steady-state response of GAL genes to glucose-galactose mixtures.

All strains prepare for glucose exhaustion during diauxic growth

Comparing the timing of GAL gene induction between diauxic growth and sudden media shift conditions offers a more sensitive measure of preparation for glucose exhaustion than the diauxic growth experiment alone. Even a long-lag strain like YJM978, which does not show observable GAL1pr-YFP expression until after glucose is exhausted (Figure 2.2 and 2.3), displays a much shorter induction delay during diauxic growth (t_{low} = 1.7 ± 0.1 hours; Figure 2.2C) than after media shift from glucose to galactose (induction delay = 12.2 hours, Figure 2.4A). To directly test if YJM978 could prepare for glucose exhaustion, we grew it in 0.125% glucose with or without 0.25% galactose and suddenly transferred the cells to galactose. We found that pre-growth in medium containing both galactose and glucose leads to an induction delay approximately one hour shorter than pre-growth in glucose alone, even though GAL1pr-YFP expression is indistinguishable from basal levels in both pre-growth media (Figure S2.10). As YJM978 has one of the longest diauxic lags in our set of strains, these data indicate that all strains prepare for glucose exhaustion to some degree.

Preparation for glucose exhaustion has an immediate cost but delayed benefit

The fact that all of our strains prepared for glucose exhaustion by pre-inducing GAL genes suggests that preparation provides a fitness benefit. Consistent with this, strains with shorter diauxic lag times take less time after the diauxic shift to reach saturation (Figure 2.1B-C, S2.11A-

B). But if preparation is always advantageous, then why don't all strains display this phenotype? In the diauxic growth experiment of Figure 2.2, we noted that the YJM978 culture exhausts glucose 23 ± 4 minutes before BC187 does (Figure S2.11C), even though BC187 eventually exhausts both sugars first. Since BC187 and YJM978 grow at similar rates in glucose-only media (Figure S2.3), this suggests that BC187 is paying a cost for expressing GAL genes before glucose is exhausted.

To directly measure potential costs and benefits experienced by BC187 during diauxic growth, we performed a competitive fitness assay by co-culturing BC187 and YJM978 under diauxic growth conditions. In addition to GAL1pr-YFP reporter expression, we also monitored the relative abundance of the two strains by tagging them with constitutive fluorophores (Figure 2.5A, Materials and Methods). We plotted the log-ratio of BC187 to YJM978 cell counts versus time and found four different phases of relative fitness during a diauxic growth cycle (Figure 2.5A). Initially, when both sugar concentrations are high, both strains exhibit low GAL1pr-YFP expression (Figure 2.5B, Phase I) and grow at comparable rates (growth rate difference less than 0.062 doublings/hr at 95% confidence). When glucose is depleted below 0.1%, BC187 displays increased GAL1pr-YFP expression while YJM978 does not (Figure 2.5B, Phase II). During this phase, BC187 has a significant fitness disadvantage of -0.17 doublings/hr relative to YJM978 (Figure 2.5A, pink-shaded point in 5C; $p = 0.0025$ for non-zero slope by t-test). After glucose exhaustion, YJM978 begins to induce GAL1pr-YFP (Figure 2.5B, Phase III), and here BC187 has a significant fitness advantage of 0.38 doublings/hr relative to YJM978 (Figure 2.5A, light-blue-shaded point in 5C; $p = 7.7 \times 10^{-5}$ for non-zero slope by t-test). Once GAL1pr-YFP is fully induced in both strains the relative fitness again is comparable (Figure 2.5A, Phase IV; fitness difference less than 0.06 doublings/hour at 95% confidence).

Figure 2.5. Preparation for glucose exhaustion has upfront cost and delayed benefit.

(A) Log2-ratio of BC187 cell number versus YJM978 cell number versus time during diauxic growth in two replicate co-cultures. A positive value on the vertical axis at any given moment indicates that BC187 has divided more times than YJM978 since time = 0, and therefore has a net fitness advantage. Raw data (black circles) and smoothing splines (gray curves) are shown for two replicates. **(B)** Median GAL1pr-YFP expression of BC187 (blue lines) and YJM978 (red lines), glucose concentration (purple circles, lines), and galactose concentration (orange circles, lines) from (A). Definitions of "high", "medium" and "low" sugar concentrations are indicated for reference. Vertical dotted gray lines in (A) and (B) demarcate 4 phases of relative fitness and GAL1pr-YFP expression during the experiment (see *Results*). **(C)** Comparison of growth rate differences during the diauxic growth versus steady-state sugar conditions. Data points with shaded backgrounds and labeled "diauxic growth" represent the slope of the data in (A) during Phase II (pink background) and Phase III (light blue background). Positive values indicate that BC187 grows faster than YJM978. Data are the mean and s.e.m. of n=6 (phase II) or n=14 (phase III) discrete derivatives in the shaded regions from (A). Points with white background and labeled "steady-state" are computed from the same data as in Figure S2.12C, and represent the mean and s.e.m. of 3-6 replicates. P-values are computed by 2-sample t-test.

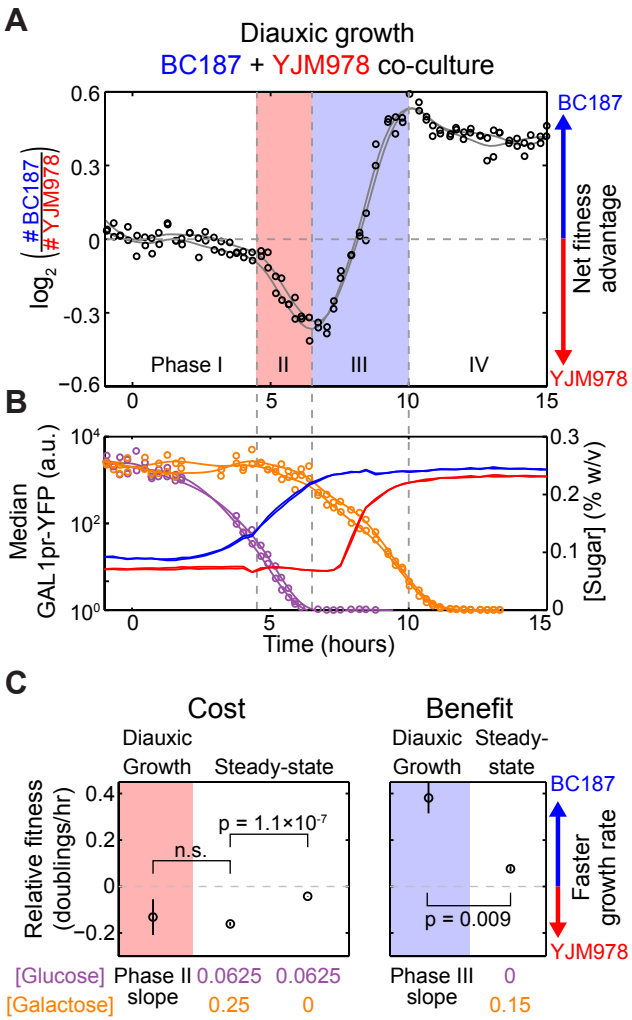


Figure 2.5 (Continued). Preparation for glucose exhaustion has upfront cost and delayed benefit.

This experiment shows that BC187 grows more slowly than YJM978 just before glucose exhaustion (Figure 2.5A, Phase II). To rule out that this is due to differences in utilization of low glucose concentrations unrelated to GAL regulation, we measured the absolute growth rates of the two strains in 0.0625% glucose with or without an additional 0.25% galactose,where sugar concentrations were held constant by frequent dilution (Figure S2.12, Materials and Methods). We found that BC187 grows at 0.62 doublings/hour in glucose alone, but significantly slower, at 0.51 doublings/hour, in glucose plus galactose (Figure S2.12C; $p = 3.2x10^{-4}$ by t-test on n = 3-6 replicates per condition). YJM978 has the same growth rate of 0.67 doublings/hour in both conditions. Neither strain shows GAL1pr-YFP expression in glucose alone, but in glucose plus galactose, BC187 displays near-maximal induction while YJM978 remains at background (Figure S2.12D). These results correspond to a relative fitness of BC187 to YJM978 of -0.043 doublings/hour in glucose alone and -0.16 doublings/hour in glucose plus galactose. Only the latter is comparable to the fitness difference of -0.13 doublings/hour just prior to glucose exhaustion during diauxic growth (Figure 2.5C, left panel). Therefore, the fitness difference prior to glucose exhaustion is due to a *steady-state* cost of BC187's early response to galactose.

In principle, the fitness difference after glucose exhaustion (Figure 2.5A, phase III) could be due to differences in galactose utilization rather than a benefit from pre-induction of GAL genes. To rule this out, we measured the steady-state relative fitness of the strains in 0.15% galactose (Figure S2.12C), corresponding to the carbon conditions just after glucose exhaustion when BC187 has its largest fitness advantage, 0.38 doublings/hour, over YJM978 (Figure 2.5A-B, Phase III),. In contrast, when galactose is held constant at 0.15%, BC187 has only a 0.076 doublings/hour advantage over YJM978 (Figure 2.5C, S2.12C). This steady-state relative fitness is significantly lower than the fitness difference during Phase III of diauxic growth (p = 0.009 by

t-test; Figure 2.5C, right panel), showing that the majority of the fitness benefit after glucose exhaustion during diauxic growth is *kinetic*, not steady-state.

These results indicate that GAL pathway expression has a strong influence on growth rate in both constant and time-varying sugar environments. If this is a direct result of GAL gene activity, then cells from the same population with non-genetic variation in GAL expression should also exhibit different growth rates. To test this, we performed time-lapse microscopy to measure the growth rate and GAL expression of BC187 cells growing in 0.125% glucose + 0.25% galactose mixture, a partially inducing condition (Figure S2.13, Materials and Methods). To maximize the dynamic range of GAL expression of the observed cells, we performed three experiments, pre-induced cells to low, medium, and high GAL1pr-YFP expression by culturing in 0.125% glucose, 0.125% glucose + 0.25% galactose, and 0.25% galactose respectively. We found that growth rate and GAL1pr-YFP expression displayed a significant negative correlation across cells of the same population, regardless of the pre-culture medium (Figure S2.13B). Furthermore, cell populations pre-induced to higher GAL1pr-YFP levels displayed lower growth rates than populations pre-induced to lower GAL levels (Figure S2.13C). Therefore, the fitness differences between bulk cultures of different strains may be due to effects of GAL expression at the single-cell level.

Synthetic expression of GAL genes recapitulates costs and benefits

Given the long-established role of GAL genes in performing and regulating galactose metabolism [10], our findings strongly suggest that GAL expression causes the observed costs and benefits. Nevertheless, it is possible that unknown genes outside of the GAL pathway can also mediate cellular responses to the environments we studied. To show that expression of GAL pathway genes alone is sufficient to produce a fitness cost and a benefit, we introduced the

chimeric transcription factor GEV into the S288C lab strain background (Figure 2.6A, "S288C-GEV") [26,27]. The presence of β-estradiol, an otherwise inert compound in yeast, triggers the GEV protein to activate genes responsive to the GAL pathway activator GAL4p [28,29]. Therefore, S288C-GEV cells grown in glucose + β-estradiol will express all the inducible genes in the GAL pathway, as well as a GAL1pr-YFP reporter we integrated into this strain (Figure 2.6B, Materials and Methods). As expected, we find that S288C-GEV has a fitness cost relative to an unmodified S288C strain when grown in glucose + β-estradiol (Figure 2.6C, top panel, black line). This cost is absent in glucose-only media (Figure 2.6C, top panel, purple line), where S288C does not express GAL genes (Figure 2.6C, bottom panel).

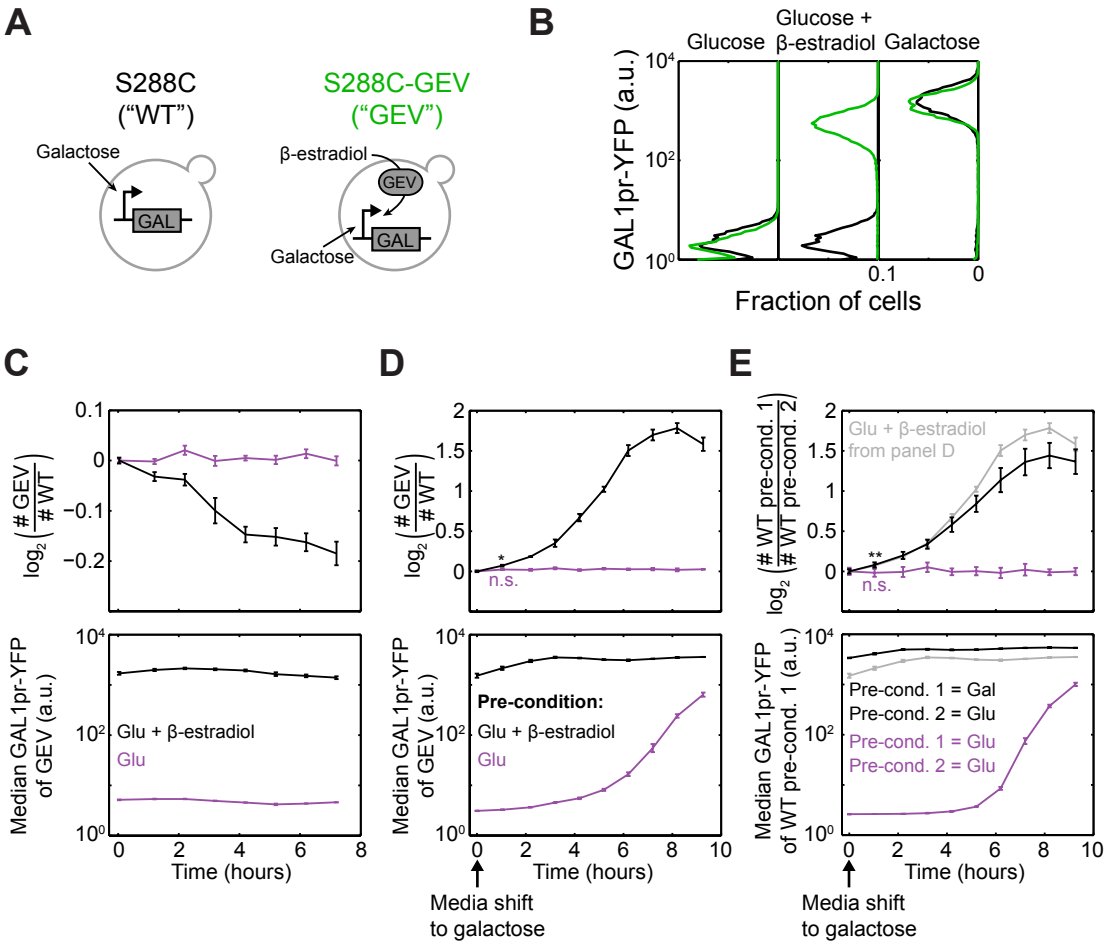


Figure 2.6 (Continued). Synthetic induction of GAL genes is costly in glucose but beneficial during transition to galactose.

(A) Strains S288C ("WT") and S288C-GEV, a congenic strain expressing the GEV protein ("GEV") were used. Both WT and GEV strains induce GAL genes ("GAL") in response to galactose; strain GEV also induces GAL genes in response to β-estradiol. **(B)** GAL1pr-YFP expression histograms of strains WT (black) and GEV (green) at steady-state in 2% glucose, 2% glucose + 30nM β-estradiol, or 2% galactose. The same concentrations were used in the following experiments. **(C)** *Top*: log2-ratio of GEV to WT cell counts during steady-state co-culture in glucose (purple) or glucose + β-estradiol (black). *Bottom*: median GAL1pr-YFP expression of strain WT during this experiment. **(D)** *Top*: log2-ratio of GEV to WT cell counts upon sudden shift to galactose, after pre-growth in glucose (purple) or glucose + β-estradiol (black). Asterisk "*": $p = 0.008$ for change in log2-strain ratio by 2-sample t-test. "n.s.": not significant, or $p > 0.05$. *Bottom*: median GAL1pr-YFP expression of strain GEV during this experiment. **(E)** *Top*: log2-ratio of cell counts of two WT strains pre-grown in different conditions and shifted to galactose. The strains were either both pre-conditioned in glucose (purple) or the query strain (numerator of log-ratio) was pre-conditioned in galactose while the reference strain (denominator of log-ratio) was pre-conditioned in glucose (black). The black line from (D) is reproduced in gray in (E) to compare synthetic and "natural" GAL pre-induction. "**": $p = 0.01$ for change in log2-strain ratio by 2-sample t-test. *Bottom*: median GAL1pr-YFP expression of the strain from pre-condition 1 during this experiment. The WT strain from pre-condition 1 contains a GAL1pr-YFP reporter, whereas the WT strain from pre-condition 2 expresses constitutive mCherry to distinguish the cells. Data in (C-E) are mean and s.e.m. of 3 replicates.

We find that S288C-GEV pre-induced in glucose + β-estradiol has an advantage over uninduced S288C when transferred suddenly to galactose medium (Figure 2.6D). We see a similar advantage when strain S288C is "naturally" pre-induced by growing in galactose, and then mixed with uninduced S288C and shifted together to galactose (Figure 2.6E). Therefore, induction of GAL genes recapitulates the benefits of galactose pre-growth in preparing cells for a transition to galactose. Surprisingly, the advantage of pre-induction (Figure 2.6D-E, slope of black line) is largest 3-6 hours after medium shift rather than immediately. However, this delay is seen for both synthetic and "natural" pre-induction, suggesting that it is due to stresses of the

medium shift unrelated to sugar metabolism (Materials and Methods). In fact, even the immediate advantage of pre-induction is significant; by one hour after shift to galactose, the synthetically pre-induced strain has made 0.068 more doublings than the non-pre-induced strain ($p = 0.008$; "*" in Figure 2.6D). This is almost identical to the immediate advantage conferred by "natural" pre-induction (Figure 2.6E, gray and black lines). Therefore, expression of GAL genes alone is sufficient to cause a fitness cost in glucose-containing environments and a fitness benefit during transitions to galactose.

Tradeoff between costs and benefits of preparation is a general constraint

Our data indicate that BC187 pre-induces GAL genes at a cost before the diauxic shift but reaps a benefit afterward, whereas YJM978 minimizes its preparation cost at the expense of experiencing a growth lag. To see if this tradeoff also constrains our other natural isolates, we assayed 15 strains to determine the cost they incur by responding to galactose while glucose is present. We defined the "galactose cost" of each strain as the relative difference in its steady-state growth rate in glucose + galactose versus glucose only, or specifically, as $(R_{glu+gal} - R_{glu}) / R_{glu}$, where $R_{glu+gal}$ represents growth rate in 0.03125% glucose + 0.25% galactose and R_{glu} represents growth rate in 0.03125% glucose. Galactose cost ranged from 0 to -0.6, meaning that a strain may grow up to 60% slower simply because galactose is present in addition to glucose. The cost experienced by a given strain increased with its GAL1pr-YFP expression in glucose + galactose (Figure 2.7B), suggesting that the growth rate reduction is due to expression or activity of GAL genes. Additionally, when the cost measurement was repeated in 0.125% glucose + 0.25% galactose, a condition which elicits lower GAL1pr-YFP expression from most strains, the magnitude of galactose cost also decreased (Figure S2.14). These results confirm the presence of a tradeoff: no strain can partially induce GAL genes without also experiencing a decrease in growth rate.

To further illustrate this tradeoff, we used the minimum mid-diauxic growth rate ("minimum rate") as a direct metric for the benefit of preparation (Figure 2.6A bottom). This metric is correlated with lag time and intuitively captures why preparation is beneficial: the more prepared a strain is, the higher its growth rate will be just after glucose exhaustion (Figure S2.2C). Furthermore, the minimum rate is not correlated with growth rate in glucose or galactose alone, and therefore is not convolved with steady-state metabolic differences (Figure S2.3). As expected, we found a negative correlation between preparation cost and minimum rate (Figure 2.7C). Our model strains for short-lag and long-lag phenotypes, BC187 and YJM978, appeared near the extremes of this tradeoff, with the phenotypes of most other strains in between.

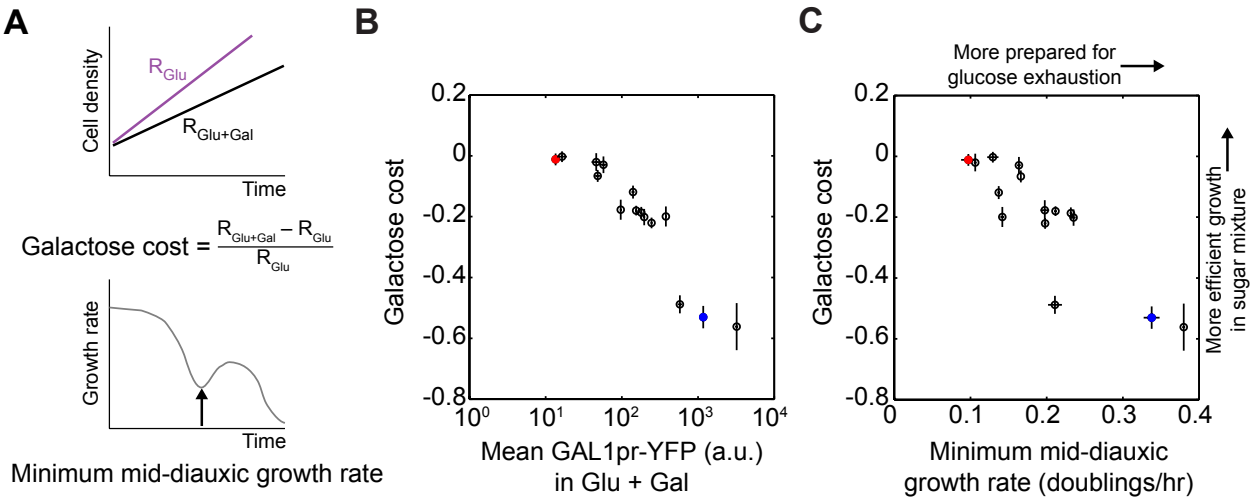


Figure 2.7. Tradeoff between costs and benefits of preparation underlies natural variation in GAL pathway expression.

Figure 2.7 (Continued). Tradeoff between costs and benefits of preparation underlies natural variation in GAL pathway expression.

(A) Illustration of how (*Top*) galactose cost and (*Bottom*) the minimum mid-diauxic growth rate are defined (see also Figure S2.2, S14, and Materials and methods). Glucose and glucose + galactose conditions indicate 0.03125% glucose and 0.3125% glucose + 0.25% galactose media, respectively. **(B)** Scatterplot of galactose cost versus mean GAL1pr-YFP expression at steady-state in glucose + galactose. Data points are mean and s.e.m. of n=3 replicates. **(C)** Scatterplot of galactose cost versus minimum mid-diauxic growth rate. The latter is computed from the growth curves shown in Figure 2.1 and S1. Data points are the mean and s.e.m. of n=3 replicates (galactose cost) or mean and range of n=2 replicates (minimum rate).

Discussion

"Why no lag phase?" An old problem revisited again

A recent study by New et al. found that yeast strains evolved to respond quickly to sudden glucose-to-maltose (i.e. preferred-to-alternative sugar) transitions tended to also have shorter diauxic lags [17]. These evolved isolates acquired mutations that weakened carbon catabolite repression, so that maltose-utilization (MAL) genes are partially induced in otherwise repressing glucose levels. New et al. found that partial MAL expression is costly when glucose is available, but enables cells to resume growth more quickly when the environment changes suddenly from glucose to maltose.

Here we confirm the link between diauxic lag duration and glucose repression found by New et al., and observe an analogous expression cost in the GAL pathway, consistent with other reports [14,30]. Additionally, we extend the previous results in two ways. First, we show that variation in glucose repression leads to a spectrum of GAL pre-induction phenotypes during diauxic growth, and that this "preparation" is mediated by steady-state sugar-sensing rather than induction or de-repression kinetics. Secondly, we demonstrate that the same cost-benefit

tradeoff that constrains lab evolution in sudden nutrient shifts also applies to natural isolates in gradually depleting nutrient mixtures. Overall, our results suggest that the mechanisms and selective forces that New et al. found in evolved strains are very likely also relevant in nature.

Preparation arises from weak catabolite repression during gradual glucose depletion

Preparation for the diauxic shift can be attributed to two key features of the yeast GAL pathway. First, some strains express GAL genes at relatively high levels in glucose-galactose mixtures [31]. This partial induction has the effect of allowing cells to anticipate sudden nutrient shifts, which New et al. also hypothesized to underlie differences in diauxic lag duration [17]. However, it is not obvious *a priori* whether partial induction of GAL genes happens fast enough relative to glucose depletion to be physiologically relevant. Our experiments show that cells are indeed able to induce GAL genes before glucose exhaustion. For example, in our culture conditions, strain BC187 takes 4.1 hours and YJM978 3.3 hours to deplete glucose from 0.2% to 0%, while both strains can execute induction from 1/64 to 1/4 of maximal expression in less than 2 hours. Even long-lag strains, which do not display observable induction prior to the diauxic shift, still begin to induce sooner during diauxic growth than after a sudden nutrient shift, suggesting that all strains can prepare for glucose depletion. These findings contribute to growing evidence that batch culture is a continuous dynamical process and that this feature plays an important role in cellular regulation [32,33].

Induction timing, not speed, underlies variation in diauxic lag

Previous studies have described differences in diauxic lag in terms of how quickly strains can transition from preferred to non-preferred nutrient metabolism [15–17]. We find that in a gradually depleting glucose-galactose mixture, “fast” or “slow” changes in growth rate are not due to “fast” or “slow” induction of GAL genes from basal to maximal, nor high basal

induction, but rather "early" or "late" initiation of induction relative to glucose exhaustion. This clarifies a distinction between induction "speed" and "timing" that has not been addressed explicitly in previous work on diauxic growth.

New et al. observed a correlation between diauxic growth phenotype and growth delay after glucose-to-maltose shift [17], suggesting that common mechanisms underlie the behavior of cells in sudden and gradual nutrient shifts. We observe that diauxic lag duration is only weakly correlated to induction delay after a sudden glucose-to-galactose shift (Figure 2.4A-C), and instead that diauxic lag is more strongly correlated to preparation time and partial GAL expression (Figure 2.3, 2.4D-F). This discrepancy may be due to differences in our experimental systems, and suggest that our strains may experience stress after the glucose-to-galactose shift incurred by sudden loss of a metabolizable carbon source (Materials and Methods).

Preparation as a widespread regulatory strategy

Other examples of preparation have recently been described in microbes encountering specific sequences of nutrients or stresses. For example, when *E. coli* encounter either heat shock or low oxygen, they induce both heat-responsive and low-oxygen-responsive genes, presumably an adaptive response when entering the warm, oxygen-deprived mammalian gut [34]. The co-regulation was decoupled by lab evolution under repeated heat shock in constant high oxygen, suggesting that the secondary response was neutral or costly when not needed. Anticipatory responses can also be asymmetric. When domesticated yeast encounter stresses typical of early stages of fermentation, they acquire resistance to later stresses; however, later stresses do not trigger resistance to early ones [35]. These results demonstrate that simple biochemical circuits can evolve the ability to anticipate environmental changes when the environmental cues occur in a predictable temporal sequence [36]. We now show that low or decreasing levels of a

preferred nutrient can serve as a predictive cue for eventual depletion. Since this is inevitable when cells deplete a mixture of nutrients at unequal rates, and mixed-nutrient environments are ubiquitous in nature, environmental anticipation may be a more widespread regulatory strategy than previously recognized.

Natural variation in diauxic lag may result from a tradeoff between costs and benefits

To be considered a meaningful example of preparation, a response must be beneficial in the future but neutral or costly in the present [35,36]. We showed that anticipatory GAL induction is costly--specifically, that many strains grow faster in glucose-only media than in media containing the same concentration of glucose plus an inducing concentration of galactose. The magnitude of cost is correlated to the degree of GAL expression across genetically diverse natural isolates, as well as across cells of the same strain with non-genetic expression variation. This cost can likely be attributed to the expression or activity of GAL pathway genes, because a strain that synthetically induces GAL genes in an otherwise non-inducing environment also exhibits a growth defect. These results rule out the possibility that strains induce GAL genes in glucose + galactose because it provides additional energy and thus a selective advantage.

The cost of GAL induction confirms part of the traditional rationale for the diauxic lag: strains that maintain stringent repression of alternative sugar pathways gain an advantage by maximizing their growth rate on glucose. On the other hand, we show that pre-induction also has a benefit that can sometimes more than compensate for its cost. Simply by being able to grow when glucose runs out, BC187 is able to double its population size over 3 hours while YJM978 undergoes a lag phase. This benefit is recapitulated when synthetically pre-induced cells are shifted from glucose to galactose media. The prevalence of short-lag phenotypes

among natural strains shows that diauxic lag is by no means an inevitable phenotype in nature, and may be selectively advantageous only in certain conditions.

We find that strains seem to face a tradeoff between fast growth on glucose and readiness to grow on galactose when glucose runs out. In principle, these goals need not be in conflict, and given the countless ways that genetic variation can tune growth rates and gene expression, perhaps evolution can optimize multiple traits simultaneously. In fact, a naïve analysis reveals no tradeoff between our diauxic growth metrics and unnormalized growth rates in glucose or galactose (Figure S2.3), consistent with a similar observation by New et. al. [17]. Therefore, although the correlations that we observe across natural isolates suggest there can be a causal relationship between GAL gene regulation and fitness consequences during diauxic growth, definitive proof of this idea requires future work incorporating genetic and mechanistic analyses.

Given these caveats, it is nevertheless striking that we do observe a tradeoff between minimum diauxic growth rate and a galactose cost metric normalized for baseline growth rate differences in glucose. Like other examples of biological tradeoffs [2,3,37], our observation suggests the presence of underlying constraints despite substantial variation in other traits. In our strains, this constraint is likely the combination of an upper "speed limit" on how quickly GAL induction can be executed and an unavoidable cost of pre-induction.

Bet hedging, mixed strategies, and optimal growth

In this study, we have focused on the timing of induction of entire cell populations during diauxic growth. Some of our natural isolates display bimodal GAL induction, similar to lab-evolved isolates, suggesting that the core phenomenon of preparation may be further modulated by heterogeneity across single cells. In fact, a different lab strain W303 has been

found to implement both early and late induction strategies simultaneously in subpopulations of the same culture [23], reminiscent of microbial "bet-hedging" observed in other contexts [38–40]. This "mixed strategy" can be evolutionarily stable, as mutants with unimodal GAL induction are unable to invade the bimodal wildtype strain in glucose-galactose mixtures [41]. Similar population diversification during diauxic growth has been observed in bacteria [24,42,43]. Additionally, cellular decisions in nutrient mixtures can be influenced by epigenetic memory [22,44,45] and inter-species signalling [46,47]. An important goal of future investigation will be determining the relative importance of these different contributions to cellular decision-making in complex natural environments.

Materials and Methods

Strains and Media

Natural isolate yeast strains were obtained from multiple sources: 23 strains were part of the *Saccharomyces* Genome Resequencing Project (SGRP) and obtained from the National Collection of Yeast Cultures [18]; 18 strains were obtained from the Fay lab at Washington University [19]; strain Bb32 was obtained from the Broad Institute [48]; strain SLYG78 was constructed for this study. Some strains were obtained in duplicate, which we indicate by affixing "-SGRP" or "-WashU" to the strain name. One of these, Y12, displayed reproducibly different diauxic growth phenotypes depending on the source collection—this may be due to strain mislabeling (Table S1, personal communication, Justin Fay) [49]. All strains are homozygous diploid and prototrophic.

Growth curves were performed on 43 strains, and a subset of 15 natural isolates were chosen for subsequent analyses. A full strain list, as well as detailed genotypes of the 15-strain subset, can be found in Table S1. With the exception of SLYG78, the subset strains were transformed with vector SLVA63 or SLVD02 digested with NotI, which replaces the chromosomal HO locus with GAL1pr-YFP linked to the kanMX4 or hphNT1 selection marker respectively. Deletion of HO does not affect growth rate [50]. Strain SLYG78 was constructed by transforming S288C-lineage haploid strains FY4 and FY5 [51] with GAL1pr-YFP and TDH3pr-mTagBFP2 (vectors SLVD02 and SLVD13), respectively, and mating them to obtain a diploid. Strains BC187 and YJM978 were transformed a second time with SLVA19 or SLVA06, which replace the 2nd HO locus with TDH3pr-mTagBFP or TDH3pr-mCherry linked to natMX4, respectively. These strains are designated BC187yb and YJM978ym in this section and in the supporting materials, but simply 'BC187' and 'YJM978' in the main text for clarity. Strain BC187ym was used for time-lapse microscopy experiments (Figure S2.13) instead of BC187yb (see "Single-cell measurements by

time-lapse microscopy"); the two strains are identical other than the fluorescent protein they express. Strains for synthetic GAL induction via GEV are described below. All yeast transformations were done by the standard lithium acetate procedure [52].

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 carbon sources. YNB contains no amino acids and extremely small amounts of other carbon-containing compounds, and therefore the added sugars comprise the sole carbon source. For diauxic growth experiments (Figures 2.1-2.3), the synthetic minimal media base was supplemented with 2.5g/L glucose (EMD) and 2.5g/L galactose (Sigma) to obtain 0.25% glucose plus 0.25% galactose w/v. We chose a 1:1 mixture of sugars to maximize the amount of growth curve data in both diauxic growth phases, and a total carbon concentration of 0.5% w/v because it is the highest that can be completely exhausted in synthetic minimal medium before non-carbon nutrients become yield-limiting. Unless noted otherwise, cultures were grown in a humidified incubator (Infors Multitron) at 30°C with rotary shaking at 230rpm (tubes and flasks) or 999rpm (deep 96-well plates).

Growth curves and diauxic lag time metric

Growth curves (Figure 2.1) were obtained using an automated robotic workcell in a room maintained at 30°C and 75% humidity. Strains were cultured in 150uL of medium in optical-bottom 96-well plates (CellTreat). Plates were cycled between a shaker (Liconic) and a plate reader (Perkin Elmer Envision) using a robotic arm (Caliper Life Sciences Twister II), and absorbance at 600nm (OD_{600}) was measured for each plate approximately every 10 minutes for up to 48 hours. In the humidity-controlled room, evaporation of medium was negligible within this time. Strains to be assayed were pinned from glycerol stock onto YPD agar and incubated

for 16 hours, and then pinned into 600uL of liquid YPD and incubated another 16 hours. These cultures were diluted 1:200 into 600uL of synthetic minimal + 0.5% glucose and grown for 8 hours, and finally diluted 1:300 into synthetic minimal + 0.25% glucose + 0.25% galactose for growth curve measurements. The final inoculation was performed into 2 different plates; these replicate growth curves were nearly indistinguishable for all strains (Figure S2.1).

Analysis of growth curve data was performed in MATLAB using custom-written code. Raw OD_{600} readings were background-corrected by subtracting the median OD of 5-10 media-only wells on each plate. OD increased linearly with culture density in the density range of our cultures (Figure S2.2A). The OD of a typical saturated culture in our experiment was approximately 0.3, which corresponds to 5×10^7 cells/mL. To analyze the diauxic lag, a smoothed growth rate was obtained by log2-transforming the data, computing discrete derivatives between consecutive data points as $(\underline{OD}_i - OD_{i-1}) / (t_i - t_{i-1})$ and fitting the derivatives to a cubic spline using the MATLAB function `csaps` with a smoothing parameter of 0.75. This smoothed derivative represents the instantaneous growth rate in units of doublings/hour. The diauxic lag time metric was computed as the difference in time between the last local maximum in the smoothed growth rate and the previous point where the culture had the same growth rate; the earlier point was also used as the time of diauxic shift (Figure 2.1, S2.2B). The minimum mid-diauxic growth rate was computed as the minimum value of the smoothed growth rate between these two times (Figure S2.2B). In strains that did not have a local minimum in smoothed growth rate, we defined the diauxic lag as zero and the minimum mid-diauxic growth rate as the value of the smoothed growth rate at its inflection point between the two growth phases; this inflection point was also used as the time of diauxic shift (Figure S2.2B, strain Bb32). Similar results were obtained if the 2 metrics were calculated using a sliding-window average on the discrete derivatives instead of a smoothing spline (Figure S2.2D). We chose the smoothing-

spline method because it facilitated calculation of a second derivative to allow identification of inflection points in the growth rate (Figure 2.2B, red lines).

To obtain growth rates in glucose or galactose (Figure S2.3), additional growth curves were performed as above, except the final culture medium contained 0.5% glucose alone or 0.5% galactose alone. 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 when culture density is approximately 1/16 and 1/4 of saturation.

Flow cytometry and sugar assays on diauxic batch cultures

We assayed the gene expression and sugar consumption of BC187yb, YJM978yb, or a co-culture of the 2 during diauxic growth (Figure 2.2, 2.5) by inoculating them from single colonies into liquid YPD, incubating for 16 hours, mixing 1:1 by volume if co-culturing, and then diluting 1:100 – 1:500 into 2% raffinose and growing for 20 hours to ~$1.5x10^6$ cells/mL. The raffinose cultures were pelleted by centrifugation, washed once, and then resuspended in 0.25% glucose + 0.25% galactose medium in 2 replicate cultures of 50mL each. The cultures were incubated in flasks at 30°C with shaking, and a sample was removed every 15 minutes until saturation, about 18 hours. Some sample was placed on ice and diluted 1:2 – 1:100 in Tris-EDTA pH8.0 and read immediately on a Stratedigm S1000EX cytometer. The flow cytometer injected a defined volume, so we can estimate the absolute culture density (Figure S2.4A). The remaining sample was filter-sterilized using a Pall 0.2um filter plate and the flow-through stored at -20°C. Media flow-throughs were later thawed and assayed for glucose and galactose concentrations by mixing with a sugar-specific oxidase (Megazyme) and measuring the absorbance of the reaction at 340nm. A standard curve of known sugar concentrations was also assayed and used to infer concentration from absorbance. We expect YFP signal to change one hour slower than GAL1

protein levels, due to fluorophore maturation time [53]. This may be why galactose decreases slightly in the YJM978 culture before GAL1pr-YFP increases (Figure 2.2D). However, since all strains have the same reporter, this should not affect induction time differences between strains.

Analysis of flow cytometry and sugar time course data

Flow cytometry data was analyzed using custom MATLAB code. In co-culture experiments, a 2D Gaussian mixture model (using the `gmdistribution` class) was fit to mCherry and side-scatter signal to segment the nonfluorescent and mCherry-expressing populations. When BC187yb was co-cultured with YJM978ym, segmentation was applied to both mCherry and BFP signal to exclude debris and doublets. We optimized flow cytometry conditions to minimize the occurrence of doublets (<1%), and therefore segmentation with 1 or 2 fluorescent markers gave equivalent results. GAL1pr-YFP expression histograms were computed on the log10-transformed YFP signal of each segmented subpopulation.

Results of diauxic growth experiments (Figure 2.2B-C, 2.5A-B) are plotted so that time zero corresponds to when culture density is 10^6 cells/mL rather than inoculation time (Figure S2.4B-D). This allows the glucose consumption rate of each strain to be compared by looking at the glucose exhaustion time (Figure S2.11). To determine the glucose exhaustion time for each dataset in Figure 2.2, a line was fit to all glucose data points whose values lay between 0.01% and 0.05%, and the x-intercept of this line was taken as the time of glucose exhaustion. This method is more robust to measurement noise at low sugar concentrations than simply finding the time when concentration reaches some low threshold.

Diauxic growth time-course measurements on multiple strains

To determine the timing of GAL pathway induction in multiple natural isolates (Figure 2.3), we co-cultured GAL1pr-YFP-labeled versions of each "query" strain with a "reference" strain,

YJM978ym, which contains a constitutive fluorescent protein, TDH3pr-mCherry, as well as a GAL1pr-YFP reporter (Table S1; also see "Strains and media"). Query strains were grown in liquid YPD for 16 hours and then mixed with the reference strain YJM978ym at ratios of 1:4, 1:1, and 4:1 by volume. The mixed cultures were diluted 1:20 into YPD and grown for 4 hours, and then diluted 1:200 in 2% raffinose and grown for 12 hours. The raffinose cultures were then diluted 1:200 into 0.25% glucose + 0.25% galactose cultures split across 40 96-well plates. These were placed in a shaking incubator and allowed to grow for 8 hours before beginning sampling. Every 15 minutes a plate was removed from the incubator and its contents were mixed 1:1 with Tris-EDTA pH8.0 + 0.2% sodium azide to stop growth and protein synthesis, and incubated for 1 hour at room temperature to allow fluorophore maturation. The 40 plates were then measured on the flow cytometer with the aid of a robotic arm.

We confirmed that the constitutive fluorophore does not affect the time of induction by co-culturing two YJM978 strains, with and without the TDH3pr-mCherry (Figure S2.7B). We also compared the GAL induction start time (t_{low}) of BC187 and YJM978 when they are cultured separately and when co-cultured, and saw no significant difference for either strain (Figure S2.7C). To check that growth rate differences between strains do not affect how quickly glucose is depleted, and therefore the timing of GAL induction, we performed each co-culture experiment at 3 different initial ratios of query to reference strain, and obtained almost identical results (Figure S2.7D-E). Therefore, this assay is robust to the presence and amount of reference strain, and we used the 3 inoculating ratios as replicates for data analysis.

To analyze population heterogeneity in GAL induction (Figure S2.7G-I), we computed the "ON fraction" as the fraction of cells with YFP signal greater than 1/32 of maximal median YFP. This threshold is just above the uninduced YFP level (Figure S2.7G). The ON fraction increases

monotonically in most of our strains. Some strains have a small pre-induced population at the start of sampling (Figure S2.7H), consistent with the steady-state bimodality we have seen. Some strains do not seem to reach complete induction (ON fraction = 1), and in fact decrease in ON fraction due to an increasing YFP-off population toward the end of the timecourse (also see Figure S2.7D). This is unlikely for biological reasons (all glucose and most galactose has been depleted at that point) and may reflect the presence of contaminants in the fixative. Our metrics are computed on data before this potential contaminant reaches appreciable concentrations and do not affect the reported results. Likewise, before this point at least 90% of cells induce as one coherent population in all our strains (Figure 2.7H-I) rather than as two-subpopulations as seen by Venturelli et al. in strain W303 [23], which we did not assay here. The environmental and genetic determinants of induction time heterogeneity are potentially interesting to dissect in future experiments.

Sudden medium shift experiments

The medium shift experiment in Figure 2.4A was performed by inoculating strains from colony into liquid YPD, incubating for 16 hours, and then diluting 1:500-8000 into 2% glucose so that cell density was approximately 1×10^6 after 12 hours of further incubation. At this point, cultures were pelleted by centrifugation at 1000g for 2 minutes and washed once in 2% galactose. The cultures were pelleted again and resuspended in 2% galactose, and a sample of cells was removed from each culture and measured on the flow cytometer every 20 minutes for 18 hours. The same protocol was used when shifting strain YJM978ym from 0.125% glucose + 0.25% galactose to 0.125% glucose (Figure S2.10).

A similar experiment by New et al. using time-lapse microscopy after a glucose-to-maltose shift found that the average single-cell growth lag correlated with a metric similar to our diauxic lag

time [17]. The apparent discrepancy between New et al. and our observations in Figure 2.4A-C is likely explained by differences in our metrics, the circuit studied (GAL versus MAL), and/or growth media. In particular, we used Yeast Nitrogen Base, which contains no carbon sources other than glucose or galactose, whereas New et al. used YP, which contains peptone and yeast extract. We speculate that auxiliary carbon sources may modulate the response of cells to sudden primary carbon shifts, a potentially interesting effect for future investigation.

Calculation of induction metrics

For both the diauxic growth (Figure 2.2-2.3) and sudden medium shift experiments (Figure 2.4A), we analyzed GAL1pr-YFP expression kinetics by calculating the time that a certain threshold value of median YFP signal was reached, and using these induction times to define other metrics (e.g. preparation time). These induction time calculations were always done by linear interpolation between two data points that bracketed the threshold YFP value. The threshold values of YFP signal were chosen to reflect the meaning of a given metric—for example, we considered the "start" of induction to be when YFP signal reached 2-fold above the basal expression of that strain (usually the initial value in a timecourse), and the "end" of induction to be when YFP signal was 4-fold below maximal expression. If the same metric was used in different experimental designs (for example, execution time during diauxic growth or after medium shift), we occasionally chose different YFP thresholds to define the metric due to variation in the range of observed data. In general, however, our results were robust to the choice of threshold. For example, preparation time can be computed using a different definition of "mid-induction time" with almost identical results (Figure S2.7F). For a detailed description of each metric used in this study, and when they can be compared across experiments, see Table S2.

Steady-state GAL expression and growth rate measurements

To measure the steady-state behavior of cells in defined glucose and galactose concentrations, we inoculated cells from colony into liquid YPD for 16 hours, diluted in 2% raffinose and grew for 20 hours, and then inoculated into glucose and/or galactose media and grew for at least 8 hours before sampling. To maintain steady-state conditions, we diluted the cultures 1:3 – 1:10 with fresh media every 2 hours so that the culture density stayed below 10^6 cells/mL (Figure S2.12A, light-colored lines). Based on the observed glucose consumption rates, this ensures that less than 10% of the glucose in a 0.0625% glucose medium is depleted. As a further check, we continued the experiment up to 48 hours and found that GAL expression reached steady-state at 8 hours and stayed constant (Figure S2.8), indicating that our protocol was sufficient to prevent physiologically relevant changes in sugar concentrations.

To measure the steady-state relative and absolute growth rates of strains BC187yb and YJM978ym (Figure 2.5C,E), we co-cultured them in various glucose and/or galactose media and sampled and diluted the cultures every 2 hours for 12 hours. We determined the growth rate difference (a.k.a. selection rate) by fitting a line to the log2-ratio of cell counts for each strain over time (Figure 2.5C, S2.12B). We determined absolute growth rates from the same data by fitting a line to the log2-dilution-adjusted-cell-concentration (Figure S2.12A, C; see also Figure S2.4). We obtained precise dilution factors by weighing culture tubes when empty and during each dilution. These experiments were done with n=3-6 replicates. To compare steady-state growth rate differences to those from diauxic growth (Figure 2.5A), we computed discrete derivative of the log2-strain-ratio at all consecutive data points in Phase I or Phase II, and compared their distribution with our steady-state measurements by a 2-sample t-test (Figure 2.5C).

Single-cell measurements by time-lapse microscopy

To prepare cells for time-lapse microscopy (Figure S2.13), we inoculated BC187ym from a colony into liquid YPD and grew for 16 hours, diluted in 2% raffinose and grew for 16 hours, and then diluted into 0.125% glucose, 0.125% glucose + 0.25% galactose, or 0.25% galactose for 8 hours to a density of 5×10^5 cells/mL. Cells are then diluted 1:300 into 0.125% glucose + 0.25% galactose medium into wells (~1000 cells / well) on a glass-bottom 96-well plate pre-coated with concanavalin A (Sigma) and left to settle for 1 hour. BC187ym contained a GAL1pr-YFP promoter and a TDH3pr-mCherry marker for image segmentation. Imaging was performed on a Nikon Eclipse Ti inverted microscope through a 20x objective lens. Exposures were taken every hour for 4 hours in bright field, YFP (ex. 500/24, em. 542/27), and mCherry (ex. 562/40, em. 641/75) channels, from 30 camera positions across 2 wells per pre-media condition, for a total of 90 camera positions. Image acquisition was controlled using custom MATLAB code using Micromanager/ImageJ.

Microscopy data were analyzed using custom MATLAB code. Microcolonies (clumps of 1-10 cells) were segmented in each mCherry image by applying a Gaussian blur to smooth cell boundaries, followed by a tophat filter to even out background, and thresholding to identify cells. Microcolonies were tracked across each time series by identifying overlapping areas. Colonies that split up, merged, entered, or exited the image during the acquisition time period were omitted from downstream analysis. Growth rate was computed as the change in log2 of a microccolony's pixel area between first and last time points, divided by elapsed time (4 hours). YFP concentration was computed as the final average background-subtracted YFP signal per pixel area of a microcolony, where background YFP was taken as the median pixel intensity.

Synthetic GAL induction using the GEV system

Synthetic induction experiments (Figure 2.6) were performed using 3 strains derived from FY5, a MATα S288C derivative (Table S1) [51]. Strain SLYA32 ("wt" reference strain in Figure 2.6C-E) was transformed with a constitutive TDH3pr-mCherry-natMX4 (vector SLVA06) to allow flow cytometry segmentation. Strain SLYA39 ("WT" in Figure 2.6B, query strain in 2.6E) was transformed with a GAL1pr-YFP-natMX4 reporter (vector SLVA64). Strain SLYH71 ("GEV" in Figure 2.6) was transformed with a tandem GAL1pr-YFP- ACT1pr-GEV-hphNT1 replacing the HO locus (vector SLVD04). The GEV sequence was subcloned from vector pAGL, a generous gift from the Botstein lab [26]. To perform competitive growth experiments (Figure 2.6C-D), query and reference strains were inoculated from single colonies into YPD, grown overnight, mixed 1:1 by volume, and then diluted 1:100 into YPD and grown 6 hours to OD~0.3. Then the cultures were concentrated 5x by centrifugation and diluted in triplicate 1:300 (1:60 dilution of cells) into 2% glucose or 2% glucose + 30nM β-estradiol and grown 12 hours to pre-induce. If needed (Figure 2.6D), cells were shifted to 2% galactose by centrifugation at 3000g for 2min, washing in new medium, pelleting again, and resuspending. For the experiment in Figure 2.6E, the above protocol was used, except query and reference strains were kept in separate cultures until the time of medium shift, and then mixed and resuspended together into new medium. The cultures were sampled immediately after medium shift, and then every 30 minutes for 9 hours, to measure the strain ratio by flow cytometry. The query strain in Figure 2.6E, black line, is shifted from galactose medium back to the same medium, so the apparent delay in fitness advantage it exhibits may reflect a stress response to centrifugation and resuspension.

Measuring galactose cost

To obtain the galactose cost (Figure 2.7, S2.14), we measured the growth rates of multiple strains in glucose and glucose + galactose. We co-cultured strains with the YJM978ym reference in

0.03125% glucose alone or 0.03125% glucose + 0.25% galactose (0.125% glucose in Figure S2.14), allowed them to grow for 8 hours, and then measured the cell count ratio at 2 timepoints 4 hours apart. To minimize glucose depletion, we inoculated cells so that their density at the end of the experiment did not exceed 3×10^6 cells / mL. We computed the growth rate difference between query and reference strain as

$$\Delta R = \left[\log_2 \frac{N_{query,final}}{N_{ref,final}} - \log_2 \frac{N_{query,initial}}{N_{ref,initial}}\right] / 4\ hours,$$

where N refers to the number of cells of a particular strain at a particular timepoint. We computed the absolute growth rate of the reference strain in each well as

$$R_{ref} = \log_2 \frac{N_{ref,final}}{N_{ref,initial}} / 4\ hours,$$

and then found the average and s.e.m. of reference strain growth rates across all replicates of each condition as $\langle R_{ref,glu} \rangle$ and $\langle R_{ref,glu+gal} \rangle$ (see Table S2). We computed the absolute growth rates of query strains as $R_{query} = \langle R_{ref} \rangle + \Delta R$ in each of the two conditions. Then we computed the galactose cost metric as $(R_{glu+gal} - R_{glu}) / R_{glu}$, where R denotes query strain growth rates in each condition. Error bars are the s.e.m. of galactose cost, computed from the s.e.m. of measured ΔR and $\langle R_{ref} \rangle$ values using standard uncertainty propagation formulas [54].

Raw data and MATLAB code

Raw data and MATLAB analysis code used to generate all figures in this paper are deposited in the Dryad repository and are openly available via: http://dx.doi.org/10.5061/dryad.39h5m [55].

References

1. Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature. 2005;436: 588–92. doi:10.1038/nature03842

2. Poelwijk FJ, de Vos MGJ, Tans SJ. Tradeoffs and Optimality in the Evolution of Gene Regulation. Cell. Elsevier Inc.; 2011;146: 462–470. doi:10.1016/j.cell.2011.06.035

3. Lang GI, Murray AW, Botstein D. The cost of gene expression underlies a fitness trade-off in yeast. Proc Natl Acad Sci U S A. 2009;106: 5755–60. doi:10.1073/pnas.0901620106

4. Scott M, Mateescu EM, Zhang Z, Hwa T, Gunderson CW. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. Science (80-). 2010;330: 1099–102. doi:10.1126/science.1192588

5. Perkins TJ, Swain PS. Strategies for cellular decision-making. Mol Syst Biol. Nature Publishing Group; 2009;5: 326. doi:10.1038/msb.2009.83

6. Dienert F. Sur la fermentation du galactose et sur l'acoutumance des levures à ce sucre. Faculté des sciences de Paris. 1900.

7. Gancedo JM. Yeast Carbon Catabolite Repression. Microbiol Mol Biol Rev. 1998;62.

8. Monod J. Recherches sur la croissance des cultures bacteriennes. Paris: Hermann; 1942.

9. Johnston M, Flick JS, Pextont T. Multiple Mechanisms Provide Rapid and Stringent Glucose Repression of GAL Gene Expression in Saccharomyces cerevisiae. 1994;14. doi:10.1128/MCB.14.6.3834.Updated

10. Lohr D, Venkov P, Zlatanova J, Program B, Academy B. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. 1995;9: 777–787.

11. Bhat PJ. Galactose Regulon of Yeast: From Genetics to Systems Biology. Berlin: Springer; 2008.

12. Magasanik B. Catabolite Repression. Cold Spring Harb Symp Quant Biol. 1961;26: 249–256. doi:10.1101/SQB.1961.026.01.031

13. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol. 2008;6: 613–24. doi:10.1038/nrmicro1932

14. MacLean RC. Pleiotropy and GAL pathway degeneration in yeast. J Evol Biol. 2007;20: 1333–8. doi:10.1111/j.1420-9101.2007.01351.x

15. 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: 245–60.

16. Spencer CC, Bertrand M, Travisano M, Doebeli M. Adaptive diversification in genes that regulate resource use in Escherichia coli. PLoS Genet. 2007;3: e15. doi:10.1371/journal.pgen.0030015

17. 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. doi:10.1371/journal.pbio.1001764

18. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James S a, et al. Population genomics of domestic and wild yeasts. Nature. Nature Publishing Group; 2009;458: 337–41. doi:10.1038/nature07743

19. Cromie G a, 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 (Bethesda). 2013;3: 2163–71. doi:10.1534/g3.113.007492

20. Hinnebusch AG. Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol. 2005;59: 407–50. doi:10.1146/annurev.micro.59.031805.133833

21. Acar M, Pando BF, Arnold FH, Elowitz MB, van Oudenaarden A. A general mechanism for network-dosage compensation in gene circuits. Science (80-). 2010;329: 1656–60. doi:10.1126/science.1190544

22. Acar M, Becskei A, Oudenaarden A van. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435: 228. doi:10.1038/nature03524

23. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. PLoS Biol. 2015;13: e1002042. doi:http://dx.doi.org/10.1101/002907

24. Solopova A, van Gestel J, Weissing FJ, Bachmann H, Teusink B, Kok J, et al. Bet-hedging during bacterial diauxic shift. Proc Natl Acad Sci U S A. 2014;111: 7427–32. doi:10.1073/pnas.1320063111

25. Nehlin JO, Carlberg M, Ronne H. Control of yeast GAL genes by MIG1 repressor: a transcriptional cascade in the glucose response. EMBO J. 1991;10: 3373–7.

26. McIsaac RS, Silverman SJ, McClean MN, Gibney P a, Macinskas J, Hickman MJ, et al. Fast-acting and nearly gratuitous induction of gene expression and protein depletion in Saccharomyces cerevisiae. Mol Biol Cell. 2011;22: 4447–59. doi:10.1091/mbc.E11-05-0466

27. Quintero MJ, Maya D, Arévalo-Rodríguez M, Cebolla A, Chávez S. An improved system for estradiol-dependent regulation of gene expression in yeast. Microb Cell Fact. 2007;6: 10. doi:10.1186/1475-2859-6-10

28. Parthun R, Jaehnings JA. Purification and Characterization of the Yeast Transcriptional

activator GAL4. J Biol Chem. 1990;265: 209–213.

29. Hashimoto H, Kikuchi Y, Nogi Y, Fukasawa T. Regulation of expression of the galactose gene cluster in Saccharomyces cerevisiae. Isolation and characterization of the regulatory gene GAL4. Mol Gen Genet. 1983;191: 31–8.

30. Malakar P, Venkatesh K V. GAL regulon of Saccharomyces cerevisiae performs optimally to maximize growth on galactose. FEMS Yeast Res. 2013; 1–11. doi:10.1111/1567-1364.12109

31. Biggar SR, Crabtree GR. Cell signaling can direct either binary or graded transcriptional responses. EMBO J. 2001;20: 3167–76. doi:10.1093/emboj/20.12.3167

32. DeRisi JL. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. Science (80-). 1997;278: 680–686. doi:10.1126/science.278.5338.680

33. Slavov N, Budnik BA, Schwab D, Airoldi EM, van Oudenaarden A. Constant Growth Rate Can Be Supported by Decreasing Energy Flux and Increasing Aerobic Glycolysis. Cell Rep. Elsevier; 2014;7: 705–714. doi:10.1016/j.celrep.2014.03.057

34. Tagkopoulos I, Liu Y-C, Tavazoie S. Predictive behavior within microbial genetic networks. Science (80-). 2008;320: 1313–7. doi:10.1126/science.1154456

35. Mitchell A, Romano GH, Groisman B, Yona A, Dekel E, Kupiec M, et al. Adaptive prediction of environmental changes by microorganisms. Nature. Nature Publishing Group; 2009;460: 220–4. doi:10.1038/nature08112

36. Mitchell A, Pilpel Y. A mathematical model for adaptive prediction of environmental changes by microorganisms. Proc Natl Acad Sci U S A. 2011;108: 7271–6. doi:10.1073/pnas.1019754108

37. Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo a, et al. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. Science. 2012;336: 1157–60. doi:10.1126/science.1217405

38. Arnoldini M, Vizcarra IA, Peña-Miller R, Stocker N, Diard M, Vogel V, et al. Bistable Expression of Virulence Genes in Salmonella Leads to the Formation of an Antibiotic-Tolerant Subpopulation. PLoS Biol. 2014;12: e1001928. doi:10.1371/journal.pbio.1001928

39. Levy SF, Ziv N, Siegal ML. Bet Hedging in Yeast by Heterogeneous, Age-Correlated Expression of a Stress Protectant. Hurst LD, editor. PLoS Biol. 2012;10: e1001325. doi:10.1371/journal.pbio.1001325

40. Fridman O, Goldberg A, Ronin I, Shoresh N, Balaban NQ. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. Nature. Nature Publishing Group; 2014;99. doi:10.1038/nature13469

41. Healey D, Gore J. Phenotypic heterogeneity implements a game theoretic mixed strategy in a clonal microbial population. bioRxiv. 2014; doi:10.1101/011049

42. Boulineau S, Tostevin F, Kiviet DJ, ten Wolde PR, Nghe P, Tans SJ. Single-cell dynamics reveals sustained growth during diauxic shifts. PLoS One. 2013;8: e61686. doi:10.1371/journal.pone.0061686

43. Kotte O, Volkmer B, Radzikowski JL, Heinemann M. Phenotypic bistability in Escherichia coli's central carbon metabolism. Mol Syst Biol. 2014;10: 736.

44. Brown JCS, Lindquist S. A heritable switch in carbon source utilization driven by an unusual yeast prion. Genes Dev. 2009;23: 2320–32. doi:10.1101/gad.1839109

45. Zacharioudakis I, Gligoris T, Tzamarias D. A yeast catabolic enzyme controls transcriptional memory. Curr Biol. 2007;17: 2041–6. doi:10.1016/j.cub.2007.10.044

46. Jarosz DF, Lancaster AK, Brown JCS, Lindquist S. An Evolutionarily Conserved Prion-like Element Converts Wild Fungi from Metabolic Specialists to Generalists. Cell. Elsevier Inc.; 2014;158: 1072–1082. doi:10.1016/j.cell.2014.07.024

47. Jarosz DF, Brown JCS, Walker GA, Datta MS, Ung WL, Lancaster AK, et al. Cross-Kingdom Chemical Communication Drives a Heritable, Mutually Beneficial Prion-Based Transformation of Metabolism. Cell. Elsevier Inc.; 2014;158: 1083–1093. doi:10.1016/j.cell.2014.07.025

48. Mortimer RK, Romano P, Suzzi G, Polsinelli M. Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of Saccharomyces cerevisiae derived from natural fermentation of grape musts. Yeast. 1994;10: 1543–52. doi:10.1002/yea.320101203

49. Fay JC, Benavides J a. Evidence for domesticated and wild populations of Saccharomyces cerevisiae. PLoS Genet. 2005;1: 66–71. doi:10.1371/journal.pgen.0010005

50. Baganz F, Hayes a, Marren D, Gardner DC, Oliver SG. Suitability of replacement markers for functional analysis studies in Saccharomyces cerevisiae. Yeast. 1997;13: 1563–73. doi:10.1002/(SICI)1097-0061(199712)13:16<1563::AID-YEA240>3.0.CO;2-6

51. Brachmann CB, Davies a, Cost GJ, Caputo E, Li J, Hieter P, et al. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast. 1998;14: 115–32. doi:10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2

52. Gietz RD, Woods R a. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol. 2002;350: 87–96.

53. Gordon A, Colman-lerner A, Chin TE, Benjamin KR, Yu RC, Brent R. Single-cell quantification of molecules and rates using open-source microscope-based cytometry.

2007;4: 175–181. doi:10.1038/NMETH1008

54. Ku HH. Notes on the Use of Propagation of Error Formulas. 1966;79: 75–79.

55. Wang J, Atolia E, Hua B, Savir Y, Escalante-Chong R, Springer M. Data from: Natural Variation in Preparation for Nutrient Depletion Reveals a Cost-Benefit Tradeoff. Dryad Digit Repos. 2014; doi:10.5061/dryad.39h5m

Chapter 3.

Polymorphisms in the Yeast Galactose Sensor Underlie a Natural Continuum of Nutrient-Decision Phenotypes

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

*Equal contribution

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 identified a locus containing the galactose sensor, *GAL3*, which 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 time, a trait that has been previously demonstrated to be selectable. These results suggest that ecological constraints on the galactose pathway may have led to variation in a single protein, which allows 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, are often composed of mixtures of multiple nutrients. This is likely to select for cells that not only utilize nutrients efficiently, but also best decide which subsets of nutrients to utilize.

Comparative studies have highlighted the plasticity of transcriptional regulatory networks. While transcription factor binding sites are typically conserved, the location of the binding sites in the genome can rapidly evolve [6]. Chromatin immunoprecipitation followed by sequencing in yeast [7-9], mice and human [10], and flies [11] 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 [12-15]. The development of genomic tools has greatly aided the interspecific comparison of regulatory binding sites.

The comparative paucity of examples of evolution of other components of regulatory networks should not be taken as a sign of their relative absence, but instead is likely due to the lack of high throughput tools to identify these changes. 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 [16]. 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 [17]; 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, and given the previously observed changes in the GAL regulatory network, we characterized differences in the induction of the GAL pathway in natural isolates of the budding yeast, *S. cerevisiae*. In the presence of high concentrations of glucose, the preferred carbon source, yeast cells repress the GAL pathway [18,19]. 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 [20] where the induction of the pathway is dependent on the ratio of glucose and galactose [21]. These observations, combined with the deep molecular understanding in the literature [22,23], make the GAL pathway an excellent model for studying natural variation in cellular decision-making.

Here, we use single-cell measurements to quantify the extent of variation in GAL induction, followed by bulk segregant analysis and synthetic allele replacements to find the determinants of this variation. We found that the glucose concentration needed to induce GAL genes varies ~100-fold across closely related natural isolates of *S. cerevisiae.* 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 [24]. 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*) (Figure 3.1A). *GAL1* is the first metabolic gene in the galactose utilization pathway [25] and this promoter has been used by numerous studies as a faithful readout of pathway activity [21,26-28]. The reporter construct was integrated into the neutral HO locus [29] in 42 different *S. cerevisiae* strain backgrounds (Figure S3.1) [30,31]. These 42 strains span a range of phylogenetic and ecological diversity [30,31]. 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 (Table S3.1).

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 (Figure 3.1B, 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 [21]. 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).

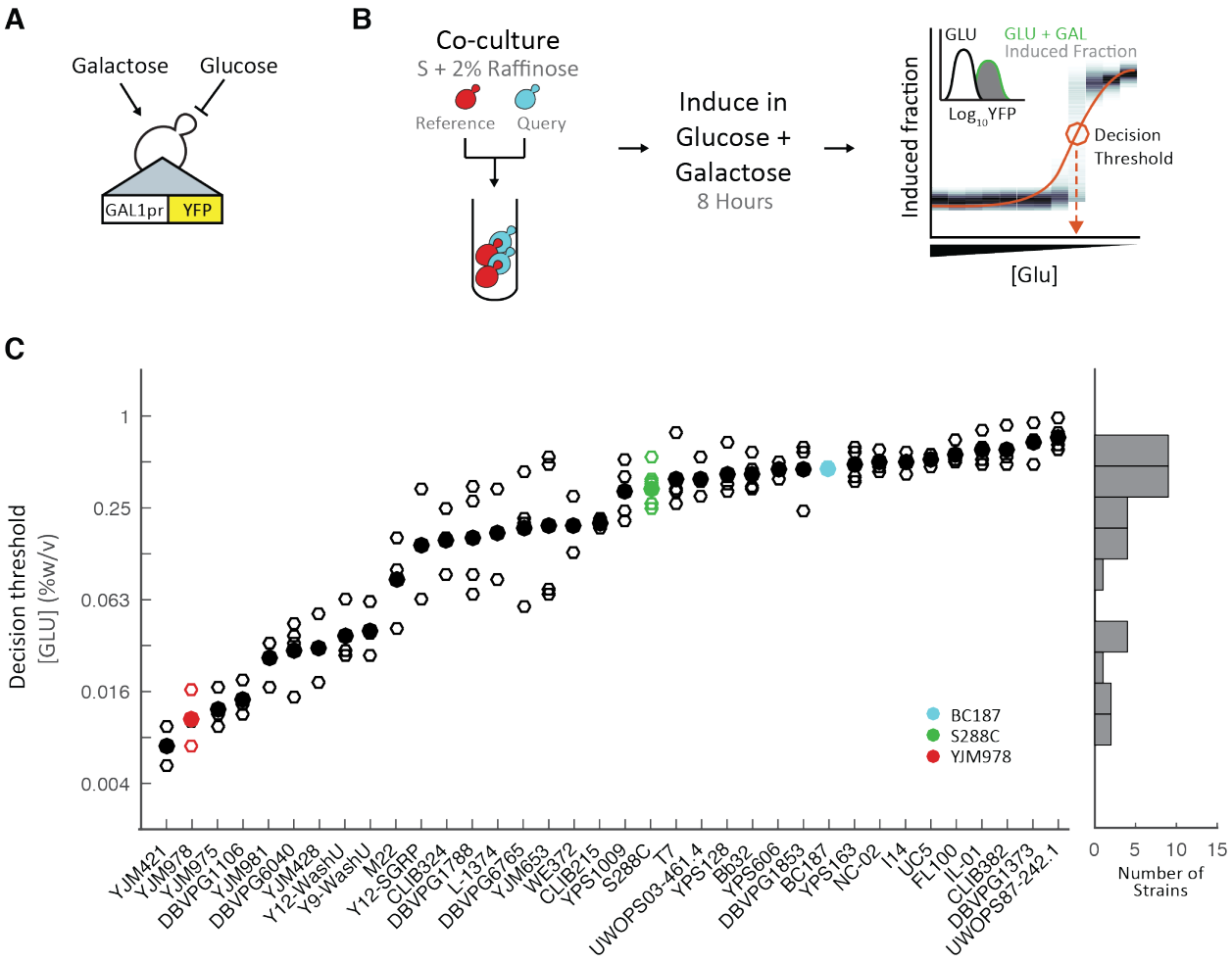


Figure 3.1. Natural isolates of *S. cerevisiae* vary in the decision to induce the GAL pathway.
(A) Schematic of the reporter construct (*GAL1pr-YFP*). (B) Schematic of co-culture pre-growth, glucose and galactose induction, and flow cytometry measurement for a glucose titration. The decision threshold, 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 was a large amount of variability in the concentration of glucose at which cells induced the GAL pathway (Figure 3.1, S3.4). We also observed bimodal expression in some strains and conditions, a likely consequence of cellular heterogeneity and ultrasensitivity in the GAL circuit [28,32,33]. 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" or F_{50}, 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 [27,34], 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 was highly reproducible across replicate measurements for all of our natural isolates (Figure S3.3).

Quantitatively, the decision threshold varies over a range of 108 ± 0.7-fold glucose concentrations across our strains (Figure 3.1, S3.4). 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 (Figure 3.1C). This glucose concentration corresponds to a galactose+glucose ratio of ~1:1. By eye the decision threshold appear to form 2 clusters, but given the reproducibility of our measurements there are significantly more than 2 (Materials and Methods). Therefore, there is a reproducible continuum of a phenotypes centered around 2 typical behaviors.

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 (Figure S3.5, [35]). 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 11 closely related strains of the wine/European clean lineage. Despite the close phylogenetic relationship of these strains, there is significant variation in the decision threshold (ANOVA, 11 strains, p-value=$3.8*10^{-9}$). The two most phenotypically distinct strains in this lineage, YJM975 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 [36]) to pairwise phenotypic distance (Materials and Methods). We did not find a significant correlation (R^2=-0.08) between genetic distance and decision threshold. When we compared the correlation between individual traits tested in Warringer et al. with the genetic distance, the decision threshold fell within one standard deviation of the mean of each individual correlation [4] (Figure S3.6). This suggests that the relationship between phylogenetic distance and phenotype

for the decision threshold is similar to previously analyzed phenotypes. Similarly, we found that differences in the decision threshold are not correlated with ecological source of the isolate (Figure S3.7). Thus, while the decision threshold varies, the cause of this variation and even whether the variation is selected or neutral is unclear.

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 (Figure 3.2A) [37-39]. We crossed eight strains that span the phenotypic and phylogenetic diversity of *S. cerevisiae* in a round-robin design (Figure 3.2). This design is known to efficiently sample parental genetic variation and allow downstream linkage analyses to detect loci with a range of effect sizes [37]. 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 at 20-50x coverage in bulk to determine the parental allele frequencies in each pool. We used the MULTIPOOL software [40] 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 (Figure 3.2B). 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 [37] (Figure S3.8).

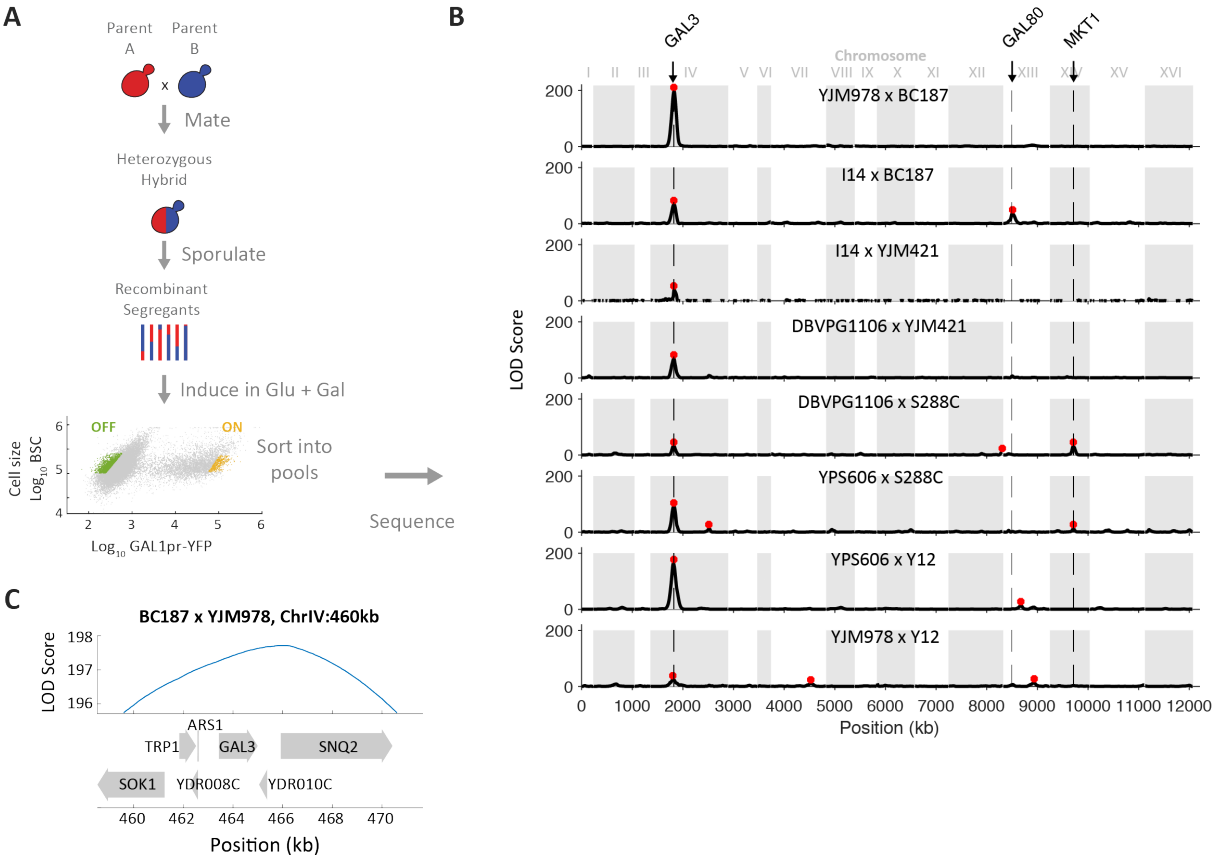


Figure 3.2. Bulk segregant analysis identifies one major-effect locus underlying natural variation in decision threshold.

(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 S3.2. (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.

Over all 8 crosses, we found 16 loci where segregant pools differ in allele frequency at LOD > 10 (Figure 3.2B). 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 (Figure 3.2B). 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 (Figure 3.2C). This includes *GAL3*, whose product directly binds galactose and positively regulates the GAL pathway [41]. The support interval for chr:460 looked similar in other crosses (Table S3.2). One other locus, at chrXIV:462, reached LOD > 10 in two crosses; the remaining significant loci were confined to a single 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 [37]. In our experiments we suspect this may arise from the different conditions used for sorting each cross (Materials and Methods), i.e. gene-by-environment effects. 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 (Figure S3.8, Table S3.2, Materials and Methods). 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 (Figure 3.1). 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 [42] (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 (Figure 3.1, 3.3A-C, S3.9).

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 (Figure 3.3D, S3.10). 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 (Figure 3.3), 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.

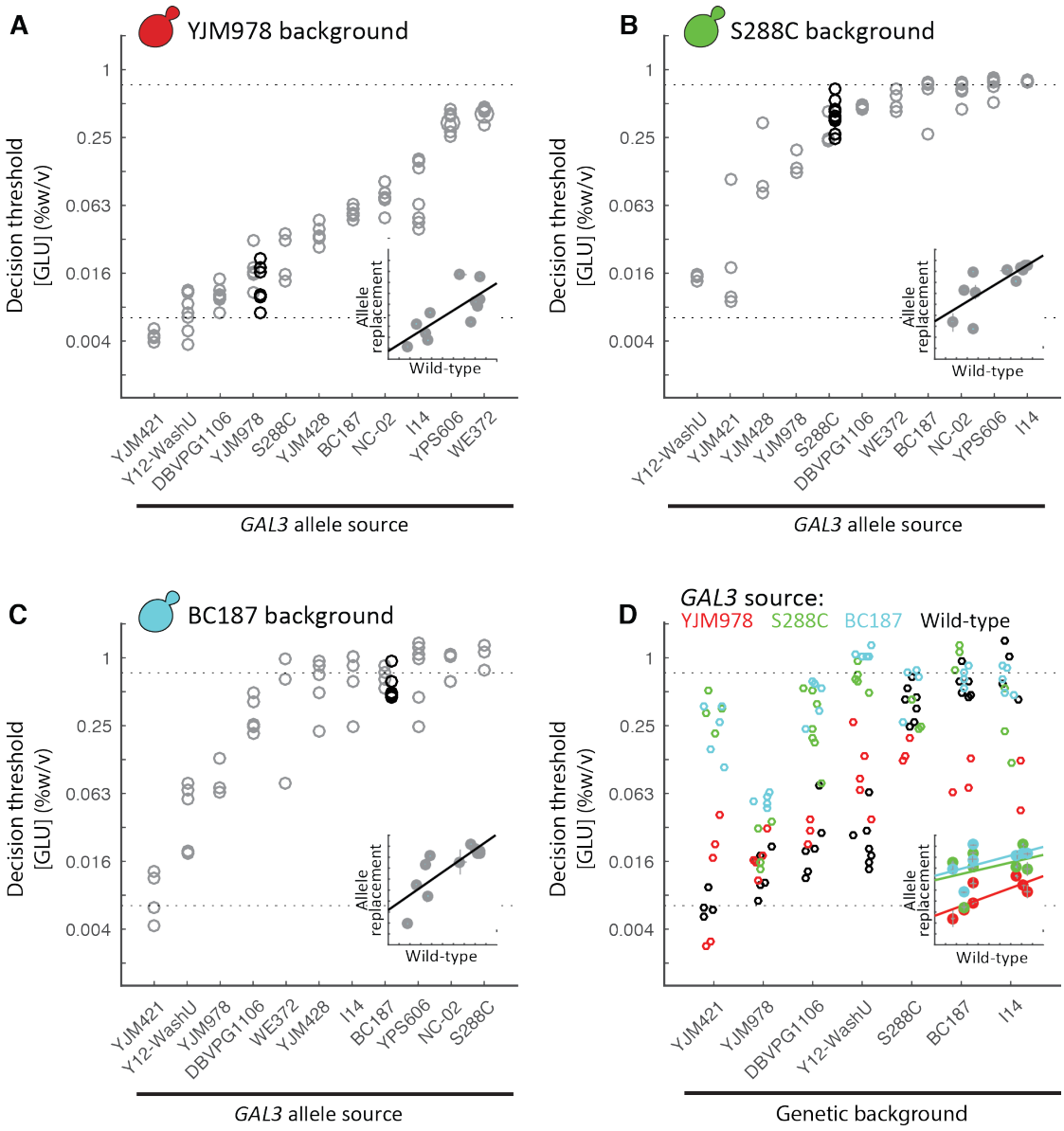


Figure 3.3. *GAL3* allele largely sets the decision threshold.

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: Scatterplot of natural isolate versus allele replacement decision threshold; error bar represents S.E.M. (D) Decision threshold of three allelic variants of *GAL3* (Mean +/- S.E.M., n≥2) inserted into various genetic backgrounds: $GAL3^{YJM978}$ (red), $GAL3^{S288C}$ (green), and $GAL3^{BC187}$ (blue), haploid wild-type strain (black). Strains are ordered based on wild-type decision threshold. Inset: Scatterplot of natural isolate (decision threshold of background strain) versus the decision threshold of the allele replacement, error bar represents S.E.M.

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 quantitatively separate the contributions of *GAL3* allele versus other genes to variation in decision threshold, we calculated the contribution of *GAL3* to the total variance of decision threshold in meiotic segregants from a YJM978 and BC187 hybrid. We selected these two strains because the *GAL3* locus was the only significant locus from our BSA on this cross, and hence this should serve as a rough upper bound on the *GAL3* contribution in other strains. We constructed three YJM978 x BC187 hybrid strains: 1) a 'wild-type' hybrid (YJM978 x BC187), 2) a hybrid with *GAL3* only from YJM978 (YJM978 x BC187 *gal3Δ::GAL3*^{YJM978}) and 3) 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 (Figure 3.4, S3.11). Strikingly, by converting a single allele in each hybrid, we were able to dramatically reduce the phenotypic variation of the segregant populations.

The total variance of each 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 ($V_{Background}$) and the variance due to *GAL3* (V_{GAL3}). Hence the total variance could be simplified to $V_P = \varepsilon^2 + V_{GAL3} + V_{Background}$ (Figure 3.4, S3.11). By definition, in the allele swap segregants (hybrids 2 and 3) $V_{GAL3} = 0$.

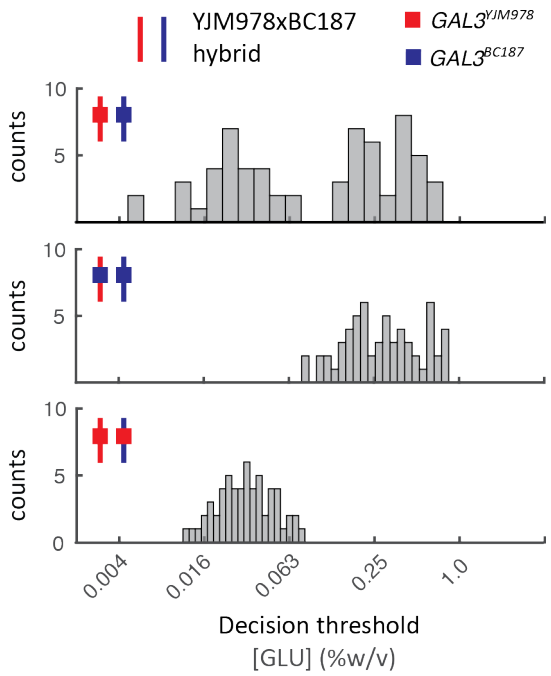


Figure 3.4. The *GAL3* allele accounts for 70-90% of the decision threshold.

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

The contribution of measurement noise (ε^2), V_{GAL3}, and $V_{Background}$ to the total variance (3.5) of was 0.5, 2.6, and 0.4 respectively (Materials and Methods). Hence 86% of the genetic variance between YJM978 and BC187 is controlled by *GAL3*. With the knowledge of V_{GAL3} from comparing segregants from hybrid 2 and 3, $V_{Background}$ can also be determined from the 'wild-type' hybrid. In this case, the background variance was estimated by subtracting the measurement noise and *GAL3* variance from the variance of the hybrid 1 segregant population. Two segregants from hybrid 1 have a decision threshold lower than what we would have expected from segregants of hybrid 2 (Figure S3.11). These two strains increase the $V_{Background}$ yielding a value of 1.4. This would correspond to *GAL3* explaining 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 undersampled the distribution from hybrid 2. These calculations suggest that *GAL3* could contribute anywhere from 70-90% to the variance of the decision threshold phenotype.

GAL3 tunes the GAL diauxic lag length

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]. This behavior is characterized by two phases of growth separated by a period with little growth, during which cells induce the genes required to metabolize the second sugar. This period of little growth is referred to as the diauxic lag. Previously, our lab has shown that BC187 and YJM978 have different diauxic lag lengths [35]. Furthermore, higher *GAL1* expression levels before the diauxic lag is inversely correlated with diauxic lag length [35]. In addition to this, changes in the decision threshold are correlated to *GAL1* expression levels (Figure S3.12), and likely because of this, diauxic lag length is negatively correlated with the decision threshold (Figure 3.5B). Hence, *GAL3* alleles would be expected to lead to changes in the length of the diauxic lag.

To determine if changing the *GAL3* allele is sufficient to change the diauxic lag length in natural isolates, we preformed diauxic shift experiments on our allele replacement strains (Figure 3.5A). Diauxic shift experiments with six *GAL3* alleles (I14, YJM421, Y12-WashU, BC187, and S288c) in three strain backgrounds (YJM978, S288C and BC187) allele replacements are shown in Figure 3.5. Diauxic shift experiments were performed by measuring the OD_{600} every 15 mins for strains inoculated into medium containing 0.25% glucose and 0.25% galactose for approximately 20 hours (Materials and methods). If the *GAL3* allele is sufficient to change the diauxic lag, the expectation would be that the diauxic lag length in glucose+galactose medium would be strongly affected by *GAL3* variants. Specifically, *GAL3* alleles from S288C, BC187, and I14

would have short lags and *GAL3* alleles from YJM978, DBVPG1106, and YJM421 would have long lags. Indeed, simply changing the *GAL3* allele in either the YJM978, BC187, or S288C background was enough to change the diauxic lag (Figure 3.5A, S3.13). Previous work had selected for strains that had a different diauxic lag in glucose+maltose, the mutants they found also changed the glucose+galactose diauxic lag [24]. To determine if the *GAL3* alleles we identified had a specific effect on the GAL diauxic lag or also affected the maltose diauxic lag, we grew cells in 0.25% glucose+0.25% maltose medium. The *GAL3* allele only affects the diauxic lag in glucose+galactose, not in glucose+maltose (Figure 3.5A, inset).

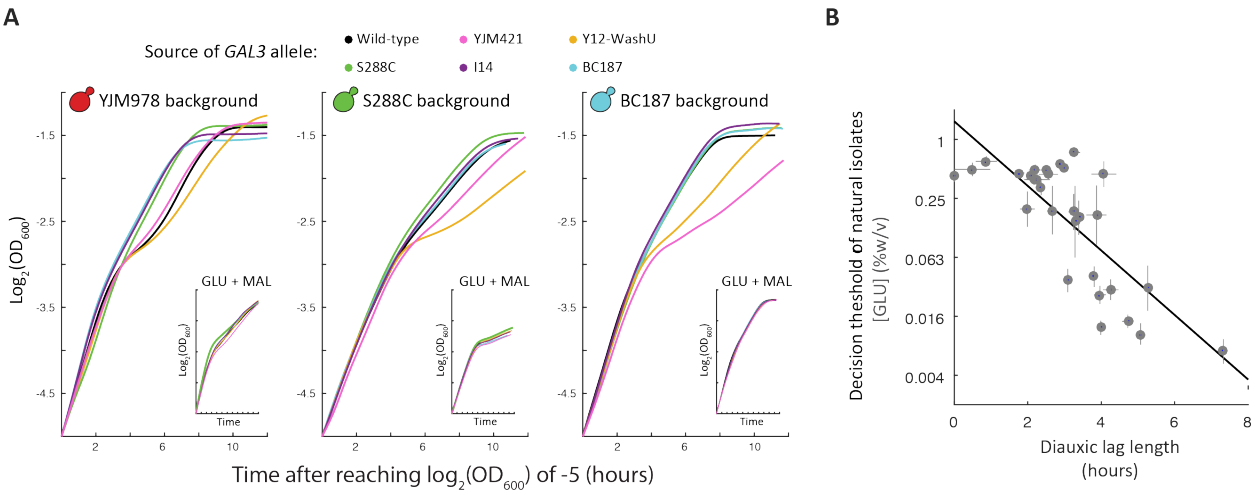


Figure 3.5. Changing *GAL3* alleles specifically affects the glucose-galactose diauxic lag.
(A) Growth curves (OD_{600} versus time) of allele replacement strains in three genetic backgrounds: YJM978 (red), S288C (green), BC187 (blue) of cells grown in a mixture 0.25% glucose and 0.25% galactose. Cultures grew for 6 to 8 hours before entering the diauxic lag. A single replicate is shown (additional replicates are shown in Figure S3.13). 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 Figure 3.1) is inversely correlated with the diauxic lag length (as measured in [35]).

Discussion

Natural genetic variation in the GAL pathway

Determining the genetic source of phenotypic variation is a major challenge of modern biology. Two recent studies have examined phenotypic variation in the GAL pathway, each at different genetic distances. We examine variation between natural isolates of *S. cerevisiae*, Roop et al. between *S. cerevisiae* and *S. bayanus*, and Peng et al. between *S. cerevisiae* and *S. paradoxus*. Interestingly, the conclusions from each genetic distance have been distinct.

We found that there is significant variation in the glucose and galactose concentrations at which natural isolates of *S. cerevisiae* induce the GAL pathway. Using BSA, we identified the galactose sensor, *GAL3*, as a major driver of this phenotypic variation, which can account for 70-90% of the variation in a single cross. Additionally, there was a strong correlation between the decision thresholds of homologous *GAL3* allele-replacements in three distinct backgrounds and their corresponding *GAL3* donor strains. This work demonstrates that there can be significant variation in the signaling pathways that interpret the environment even in closely related strains.

Our results from examining variance in *S. cerevisiae* might be surprising as one might expect other members of the GAL pathway to be key drivers of the decision threshold. By changing the copy number of the GAL regulatory genes, Acar et al. showed that solely changing the level of Gal3p or Gal80p could change the 'inducibility profile' of *S. cerevisiae* [27]. Surprisingly, while we found evidence for a role of *GAL3* in all our crosses, we only obtained evidence for a potential role for *GAL80* in a single cross, I14 x BC187. Additionally, in all our crosses we found that the *GAL3* allele had the largest effect size. Interestingly, work from Peng et al. showed that between *S. cerevisiae* and *S. paradoxus*, promoter variation in *GAL80* is a major factor in the

variation in response profiles [43]. One possibility is that *GAL80* is under additional selective constraint in *S. cerevisiae* and hence is not found as a source of variation between *S. cerevisiae* strains. For example, in *S. uvarum* there are two alleles of *GAL80* and the deletion of one of the alleles leads to rapid galactose catabolism and metabolic overload [44]. The single copy of *GAL80* in *S. cerevisiae* and *S. paradoxus* might similarly have a distinct function in each species and hence could affect the mutational plasticity. Alternatively, it could be possible that a significant amount of the variation between *S. cerevisiae* and *S. paradoxus* is driven by changes in the coding sequence of *GAL3* or that the strain of *S. paradoxus* used by Peng et al. was an outlier.

Synthetic perturbations of feedback in the GAL pathway by Venturelli et al. also highlighted the role of *GAL1* [33]. Roop et al. examined the variation between *S. cerevisiae* and *S. bayanus* by swapping coding and promoter sequences. They found *GAL1*, *GAL3*, *GAL4*, and *GAL10* affected diauxic lag length, while *GAL2*, *GAL7*, and *GAL80* did not. While the phenotype they examined was distinct, this still highlights the potential different genetic sources of variation that are occurring between the different strains. More work will be required to understand the relative importance of different members of the GAL pathway variation and the potential similarities and differences caused by variation in each.

Genetic complexity of signaling traits

Our limited understanding of complex traits has been underscored by the shortcomings of genome-wide association studies (GWAS). While GWAS has helped to identify players in multiple diseases and traits, much of the expected heritability is missing for many traits [45]. There have been multiple potential explanations including a large number of small effect loci and epistasis [46]. Using model organisms with quantitative and tractable genetic systems have the potential to elucidate these issues.

The potential role of small effect genes is highlighted by the fact that *GAL3* was the only one significant locus identified by BSA in a single cross, but there was still 10-30% of unaccounted variance. BSA places a maximum effect size on the remaining causative alleles, and hence suggests that small levels of variation in multiple genes influence the decision threshold in the GAL pathway. Given the number of genes that are known to be involved in the GAL pathway, many of the small effect alleles must be in other cellular processes. Indeed, by examining the effect of deletion mutants upwards of 40% of genes in the yeast genome have the potential to influence the GAL response (Bo et al., under review). This result might also explain why swapping the main regulators of the GAL pathway between *S. cerevisiae* and *S. bayanus* was only able to partially interconvert the phenotypes [47].

This could also explain the difference in the number of causative alleles that we identified. Previous X-QTL analyses in yeast that have identified multiple loci, typically average more than 10 per trait [37,38,48]. One possible explanation for this is that our trait is 'less' complex and is therefore driven by fewer genes. But, at least two alternative hypotheses exist. One, while the trait might be equally complex, variation of this trait might be under more selective constraint than previously observed phenotypes, and hence there is less standing genetic variation. Two, there could be a similar number of genes involved, but if the largest effect size genes drive majority of the variance, so the power to identify other small effect size genes is diminished.

While many of the *GAL3* allele swaps had effects that are consistent with a linear additive model, there were two types of epistasis that were evident. First, while the difference between the S288C, BC187, and YJM978 *GAL3* alleles are largely preserved, in two of the strain backgrounds (YJM978 and S288C) the differences between the alleles are highly compressed (Figure 3.3D). This suggests a direct epistatic interaction between the *GAL3* allele and another

gene. Second, *GAL3* alleles that had decision thresholds above 0.25% glucose in 0.25% galactose placed in the S288C and BC187 backgrounds have a smaller spread than that for the YJM978 background, suggesting that some feature of the circuit limits the effect of the *GAL3* allele appears to saturate at high decision thresholds. This could be explained by epistasis between *GAL3* and an emergent property of the system, i.e. pathway saturation. The underlying genetic interactions could be quite complex; any gene that is involved in the linearity of the pathway could have an epistatic interaction with *GAL3*. By looking at the coherent behavior of whole pathways, the interaction of genes, and emergent properties we can gain more insight than by limiting our observations to a strict gene by gene comparison. Given these results, the GAL system will likely be a good system for testing the principles of quantitative genetics.

Evolution and selection of the GAL pathway

When grown in multiple sugars, cultures often go through two distinct growth phases separated by a diauxic lag [3]. This phenomenon results from cells first consuming the preferred sugar, followed by a "lag" where cells must induce the genes necessary to metabolize the second, less preferred sugar, in order to achieve a higher growth rate. Several recent works have highlighted that even for the same mixture of sugars, related strains can have a continuum of diauxic lag times [24,35]. The diauxic lag length is selectable [24] and correlated to the expression level of the GAL pathway before glucose depletion [35]. Here we extend this to show that this variation in 'preparation' of the GAL pathway is driven by variation in *GAL3* and directly correlated with the decision threshold (Figure 3.5B). We confirm this by showing that allelic swaps of *GAL3* are sufficient to alter the diauxic lag length. Hence, differences in the diauxic lag lengths result from differences in how cells 'perceive' their environment.

Interestingly, one laboratory evolution study produced distinct variation from what we observed in natural isolates [24]. New et al. selected for strains that had a shorter diauxic lag during the switch from utilizing glucose to maltose, which in principle this could be achieved by changing the response to either maltose or glucose. The evolved strains had mutations in genes that caused weakened catabolite repression, i.e. mutations in the glucose sensing and metabolic genes, *HXK2* and *STD1*, instead of, by analogy to our system, maltose regulators. One potential explanation for this difference is that New et al. selected for an altered diauxic lag in maltose; nature might counter select against mutants that affect the diauxic lag in other sugars. Supporting this, the New et al.'s evolved strains having shorter diauxic lags in both maltose and galactose, while our *GAL3* allele replacements specifically tune the diauxic length in galactose and do not affect the lag length in maltose (Figure 3.5A, inset). The difference between these classes of mutants also raise the possibility that populations with different lifestyles, e.g., generalists versus specialist [49], might be enriched for different classes of mutants.

More broadly, certain types of fluctuating environments might select for strains that are able to quickly adapt along a specific phenotypic axes of variation. For example, in environments where there was a selective advantage in being able to rapidly tune diauxic lag time, strains where a single protein specifically regulated the GAL decision threshold could have a selective advantage. Hence, it is interesting to consider whether the variation we see in *GAL3* alleles is under selection as a result of environmental pressures, which vary between different strains depending on their ecological niche, for altered diauxic lag length. This raises the interesting hypothesis that variation in signaling pathways is prevalent in nature and that this variation might be an indicator of the type of environment the organism evolved in. Given these results, the GAL system will likely serve as a rich system for understanding the relationship between

environmental variation, quantitative traits and the underlying signaling pathways that give rise to such traits.

Materials and methods

Strains and media

Strains were obtained as described in [35]. Strains used in this study can be found in Table S3.1. All strains from the collection and those assayed in Figure 3.1 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). 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. [21]. 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) (Figure S3.2). 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) (Figure S3.3). 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 (Figure S3.3). All strains were measured at least twice; replicates were performed on different days.

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 [50]. This plasmid design is inspired by a similar mating-type selection plasmid used in a recent study [37].

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 [51], 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 [52]. 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 to a depth of 16-33x.

Genome sequences of round-robin parents

Non-S288C parental genomes for the bulk segregant analysis were obtained from the literature: I14 from [37]; BC187, YJM978, DBVPG1106, and Y12 from [53]; YPS606 from [54]. 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 [36] 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) 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. 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; [40]) with the options `-m contrast -r 100 -c 2200 -n 1000`, following previous practice [37]. 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 [55]. Thus we used n=200 for this cross.

We defined significant loci as LOD peaks where LOD > 10 (Figure 3.2B). Previous bulk segregant analyses using MULTIPOOL used a less stringent cutoff of LOD > 5 [37,38]. This corresponded to a false discovery rate of 5% in one study [38], but led to a much higher number of unreplicated locus calls in another study [37]. 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 (Table S3.2). 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 (Figure S3.8, Table S3.2).

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 [56], 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 [42]. 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.

Hybrid conversions and determining *GAL3* allelic contribution

Hybrid conversions were constructed by mating CRISPR/Cas9 generated allele replacement strains (YJM978 *GAL3*^{BC187} or BC187 *GAL3*^{YJM978}) to either BC187 or YJM978 (haploid, wild-type). Segregants were generated as mentioned above (see crossing and generating segregants). Segregants from the original hybrid (BC187 x YM978) were also phenotyped. Since the hybrid conversion strains had a known *GAL3* allele, the allelic contribution of the GAL3 alleles was estimated by subtracting the background effect, and measurement noise from the total phenotypic variance of the two hybrid conversion segregant populations. A range of allelic contribution was estimated by using the variance of the original hybrid (YJM978 x BC187) and the two hybrid conversion populations.

Estimating measurement error

$$x_i{}^1 \sim x_i + \varepsilon^1$$

$$x_i{}^2 \sim x_i + \varepsilon^2$$

$$\therefore x_i{}^1 - x_i{}^2 = \varepsilon^1 + \varepsilon^2$$

$$x_i{}^1 - x_i{}^2 \sim \mathcal{N}(\mu, \sqrt{2}\sigma)$$

$$\varepsilon = \frac{\sigma}{\sqrt{2}}$$

The measurement noise (ε) was estimated by determining the difference of two replicate measurements, which was normally distributed. The variance of the noise was estimated as the noise term. Superscripts were used to denote the two different replicate measurements. If the deviation from the mean of the two replicate measurements was greater than a threshold value (1.5 as calculated above for detecting outliers), both measurements were removed.

Estimating the *GAL3* effect

$$\mu_a = \frac{\sum x_i{}^a}{N}, \mu_{-a} = \frac{\sum x_i{}^{-a}}{N}$$

$$E_{GAL3} = \frac{\mu_a - \mu_{-a}}{2}$$

$$V_{GAL3} = (E_{GAL3})^2$$

The two allelic variants of GAL3 are denoted by *a* and *–a*. In order to estimate the effect of the *GAL3* allele (E_{GAL3}), the mean of each allele population (*a* or *–a*) was calculated, by taking the difference of the mean of the population and dividing by 2.

Determining allelic contribution

$$V_P = \varepsilon^2 + V_G$$

$$\text{where, } V_P = \frac{\sum(x_i - \bar{x})^2}{N}$$

$$\text{where, } V_G = V_{background} + V_{GAL3}$$

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 ($V_{background}$) and the contribution to variance of *GAL3* (V_{GAL3}). The background variance was estimated by subtracting the measurement noise and *GAL3* variance from the variance of the segregant population. The *GAL3* contribution was reported as the ratio of the variance in *GAL3* and the genotypic variance (V_G).

$$GAL3\ contribution = \frac{V_{GAL3}}{V_G}$$

Growth curves and diauxic lag time metric

Growth curves were obtained as described in Wang et al. [35]. 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 [35].

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 [57] 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. A neighbor-joining phylogenetic tree was generated using the seqneighjoin function on MATLAB.

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

7. 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

8. 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

9. 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

10. 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

11. 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

12. 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

13. 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

14. 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

15. 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

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

18. 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.

19. 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

20. 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

21. 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

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

23. 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

24. 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

25. 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

26. 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

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

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

29. 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

30. 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

31. 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

32. 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

33. 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

34. 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

35. 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

36. 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

37. 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

38. 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

39. 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

40. 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

41. 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

42. 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

43. 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

44. Kuang MC, Hutchins PD, Russell JD, Coon JJ, Hittinger CT. Ongoing resolution of duplicate gene functions shapes the diversification of a metabolic network. eLife. 2016;5: 859–28. doi:10.7554/eLife.19027

45. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461: 747–753. doi:10.1038/nature08494

46. Zuk O, Hechter E, Sunyaev SR. The mystery of missing heritability: Genetic interactions create phantom heritability. 2012. doi:10.1073/pnas.1119675109/-/DCSupplemental

47. 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

48. 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

49. Jarosz DF, Lancaster AK, Brown JCS, Lindquist S. An Evolutionarily Conserved Prion-like Element Converts Wild Fungi from Metabolic Specialists to Generalists. Cell. Elsevier Inc; 2014;158: 1072–1082. doi:10.1016/j.cell.2014.07.024

50. 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

51. 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

52. 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

53. 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

54. 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

55. 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

56. 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

57. 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 4.

Natural Genetic Variation Can Independently Tune the Induced Fraction and Induction Level of a Bimodal Signaling Response

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

Bimodal gene expression by genetically identical 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. The phenotypic variation can be described in terms of two uncorrelated features representing the fraction of cells that are induced and the expression level of the induced subpopulation. We mapped genomic loci underlying these two traits using bulk-segregant analysis, identified causal genes in 3 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. Additionally, the genetic effect on induced fraction could be phenocopied by varying the growth conditions prior to galactose induction. Our results show that different quantitative features of a bimodal signaling response can be tuned independently by genetic and environmental 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

Non-genetic heterogeneity is a pervasive feature of gene expression and cellular signaling [1–3]. Bimodal responses, where cells in an isogenic population adopt one of two distinct states, are particularly important for microbes coping with fluctuating environments [4,5] and cells of multicellular organisms differentiating into discrete types [6,7]. The galactose-utilization (GAL) pathway in *Saccharomyces cerevisiae* (budding yeast) is a well-characterized bimodal response and a classic model of microbial decision-making [8,9]. GAL enzymes are tightly repressed in glucose and activated almost 1000-fold in galactose [10]. In mixtures of glucose and galactose, GAL genes induce as a function of the galactose-to-glucose ratio [11] and display complex patterns of bimodal expression [12].

Bimodality of GAL gene expression is attributed to bistability arising from positive feedback through the Gal1p kinase and Gal3p transducer [13,14]. However, perturbing other pathway components such as Gal2p permease, Gal4p activator, and Gal80p repressor also affect quantitative features of the GAL response [14–17]. Additionally, the modality of the GAL response is affected by the metabolic conditions prior to encountering galactose [12]. Despite the complex response of GAL expression distributions to genetic and environmental perturbations, most studies of the pathway have focused on one quantitative feature such as the induced fraction [16,18,19], with a few recent exceptions [13,20]. How multiple quantitative features of the pathway are controlled and vary across perturbations is poorly understood.

In previous work, we found that natural yeast isolates differed widely in the inducibility of GAL genes in glucose + galactose mixtures [19,21]. 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 [22] and bacterial

utilization of various sugar mixtures [23]. 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 traits representing the induced fraction of cells and the expression level the induced subpopulation, which vary in an uncorrelated way across natural isolates. Using bulk segregant analysis and CRISPR/cas9 allele replacement, we identified genetic variants underlying these two traits and showed that the variants can affect the traits independently of each other. Additionally, we found that the metabolic history of cells before inducing GAL genes also affects the bimodal response in a trait-specific way. 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 underlying 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 [21,24,25] grown in a titration of glucose plus a constant level of galactose. As expected, we found that all strains are uninduced in high glucose and fully induced in low or no glucose (Figure 4.1). However, at intermediate glucose concentrations, some strains display a unimodal population with intermediate (i.e. sub-maximal) GAL expression (Figure 4.1B) while other strains display a bimodal mixture of

uninduced and partially induced cells (Figure 4.1C). Additionally, strains with the same modality still have quantitatively different GAL induction profiles (Figure S4.1), raising the question of what mechanisms can give rise to these diverse signaling phenotypes.

Variation in GAL bimodality phenotypes can be parameterized by two uncorrelated 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 (Figure S4.2A). The mixing of these components can be quantified as an induced fraction that decreases as a function of glucose concentration (Figure S4.2B). 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 (Figure S4.1, S4.2B-C). 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.

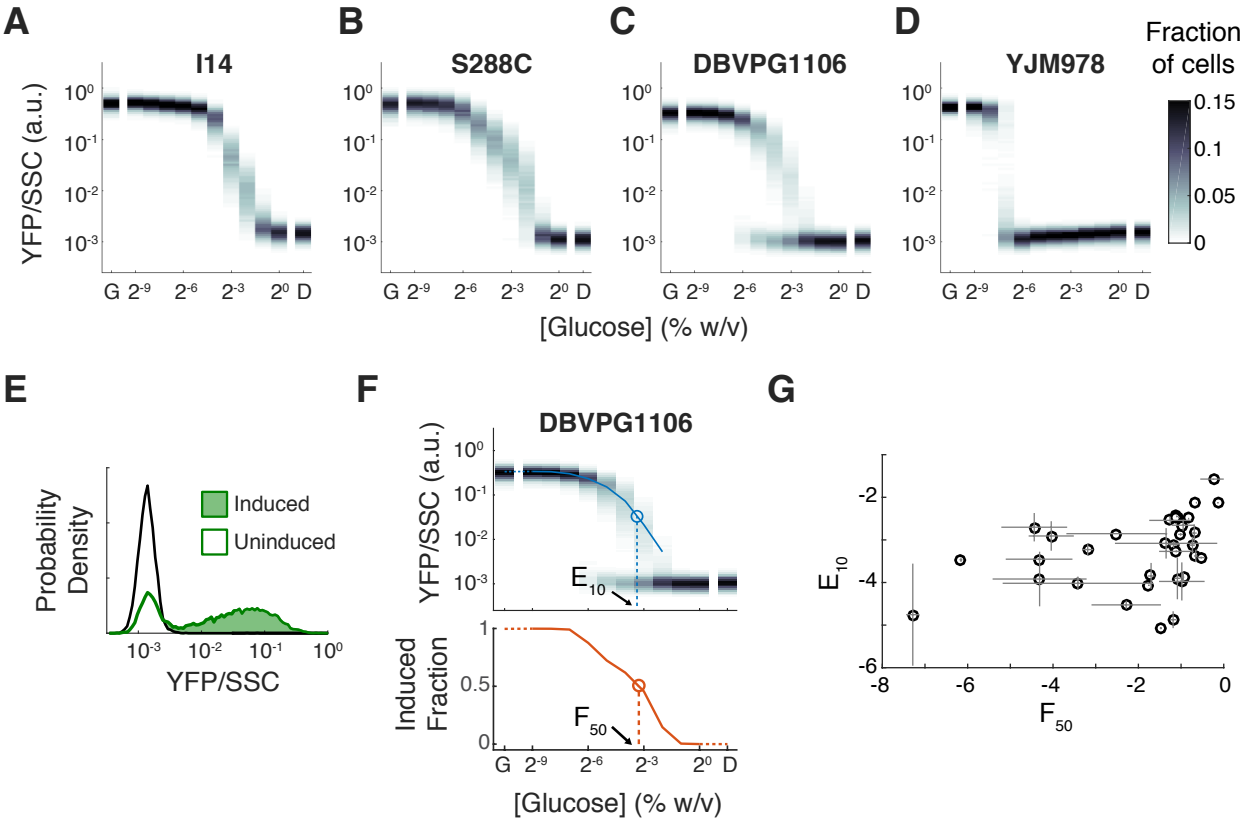


Figure 4.1. Natural variation in GAL induction can be analyzed in terms of two uncorrelated features
Each plot is a series of YFP fluorescence (normalized to side scatter "SSC") histograms from 12 sugar conditions for strains (A) I14, (B) S288C, (C) DBVPG1106, and (D) YJM978. Other phenotyped strains are shown in Figure S4.1. 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) Induced level (blue line) and induced fraction (orange line), and the corresponding E_{10} and F_{50} metrics, for strain DBVPG1106. (G) Scatterplot of E_{10} versus F_{50} across 34 *S. cerevisiae* natural isolates (mean and S.D.; n=3-10).

Applying this population decomposition framework to our data, we computationally separated induced and uninduced cells from each GAL reporter distribution (Figure 4.2E) 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 (Figure 4.2F). For convenience, we express these metrics in units of log2-transformed glucose concentration, 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 (Figure 4.2G).

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.

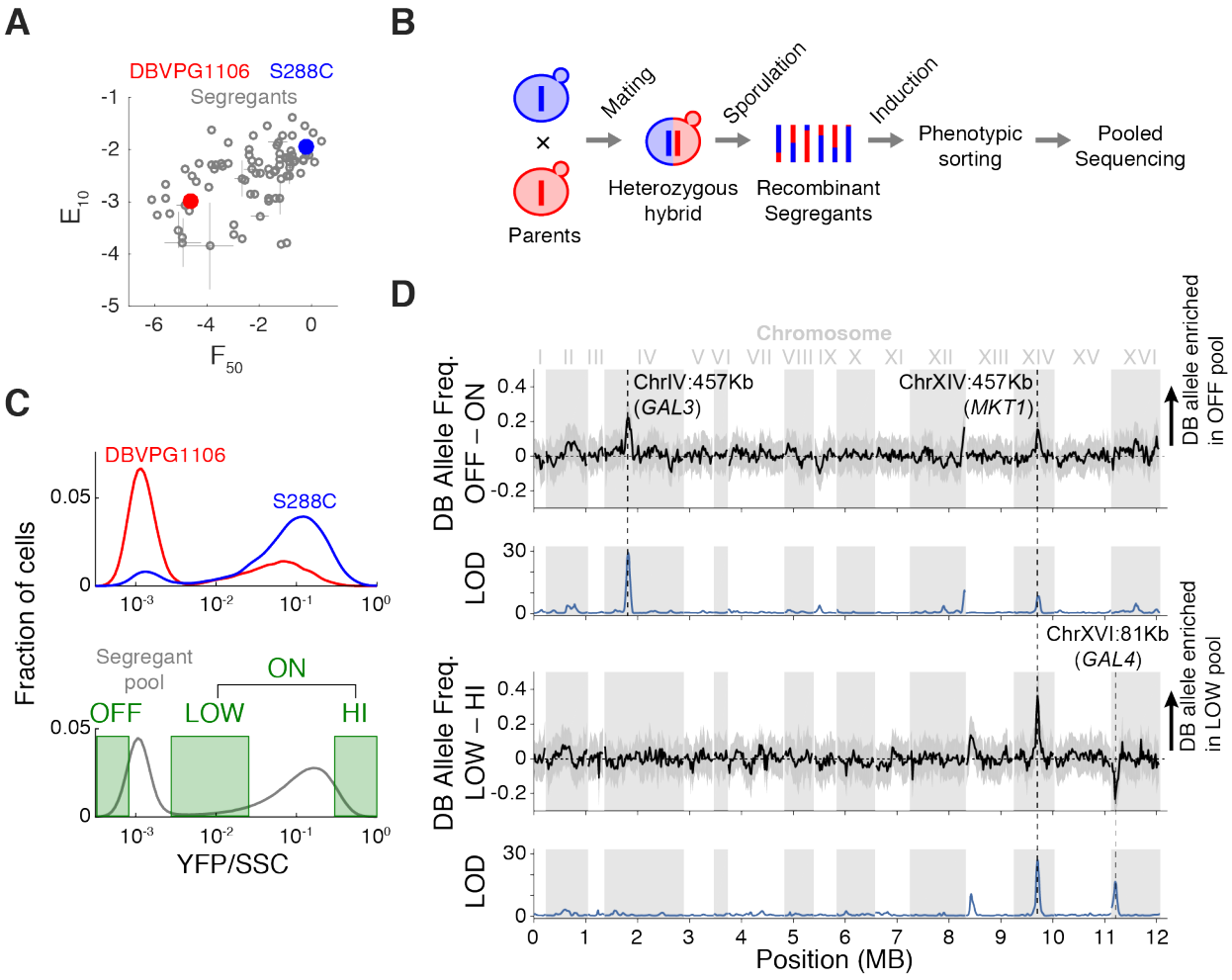


Figure 4.2. Bulk segregant analysis of E_{10} and F_{50}

(A) E_{10} versus F_{50} 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 (gray, bottom) in the 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 S4.3). 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 (Figure 4.2B). 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 (Figure 4.2C). We then used FACS to sort the segregants into pools of uninduced ("OFF"), induced and low-expression ("LOW"), and induced and high-expression ("HI") cells (Figure 4.2C), and sequenced each pool 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 [26] (Figure 4.3D; 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 (Figure 4.3D). *GAL3* and *GAL4* are direct regulators of galactose sensing, while *MKT1* is known to have pleiotropic effects in crosses between S288C and natural isolates . 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 (Figure S4.4). In DBVPG1106, replacing the endogenous *GAL3* allele with $GAL3^{S288C}$ shifted F_{50} in the direction of the S288C parent (Figure S4.4A). *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 (Figure S4.4B-C). 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 (Figure 4.2D) 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.

Figure 4.3. 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 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}.

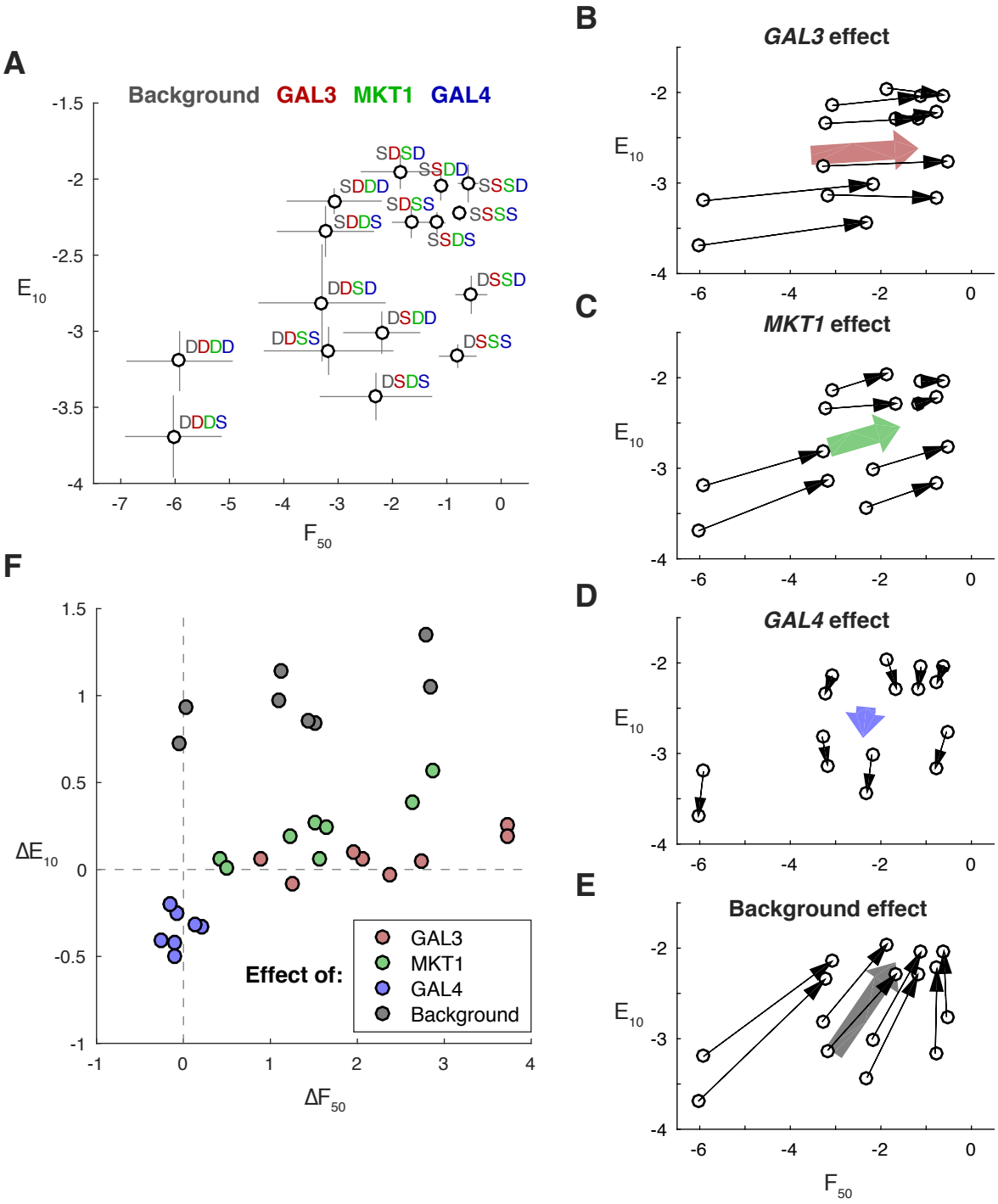


Figure 4.3 (Continued). Combinatorial effects of strain background and *GAL3*, *MKT1*, and *GAL4* alleles.

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 (Figure 4.3A) 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 the DBVPG1106 genetic variant to the S288C variant can be visualized as a vector in E_{10} versus F_{50} space (Figure 4.3B-E) or as a trait difference (Figure 4.3F).

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 (Figure 4.3B-E), or equivalently, that differential effects cluster by angle from the origin (Figure 4.3F). 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.

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 Figure 4.3A). However, the reciprocal replacement in the S288C background has a more modest effect (SDDS versus SSSS in Figure 4.3A). 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.

Genetic and environmental perturbations that affect F_{50} do not affect E_{10}

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 [19]. As expected, F_{50} varied widely as the *GAL3* allele is changed (Figure 4.4A). However, variation in E_{10} with *GAL3* allele was minimal and driven almost entirely by the strain background (Figure 4.4B).

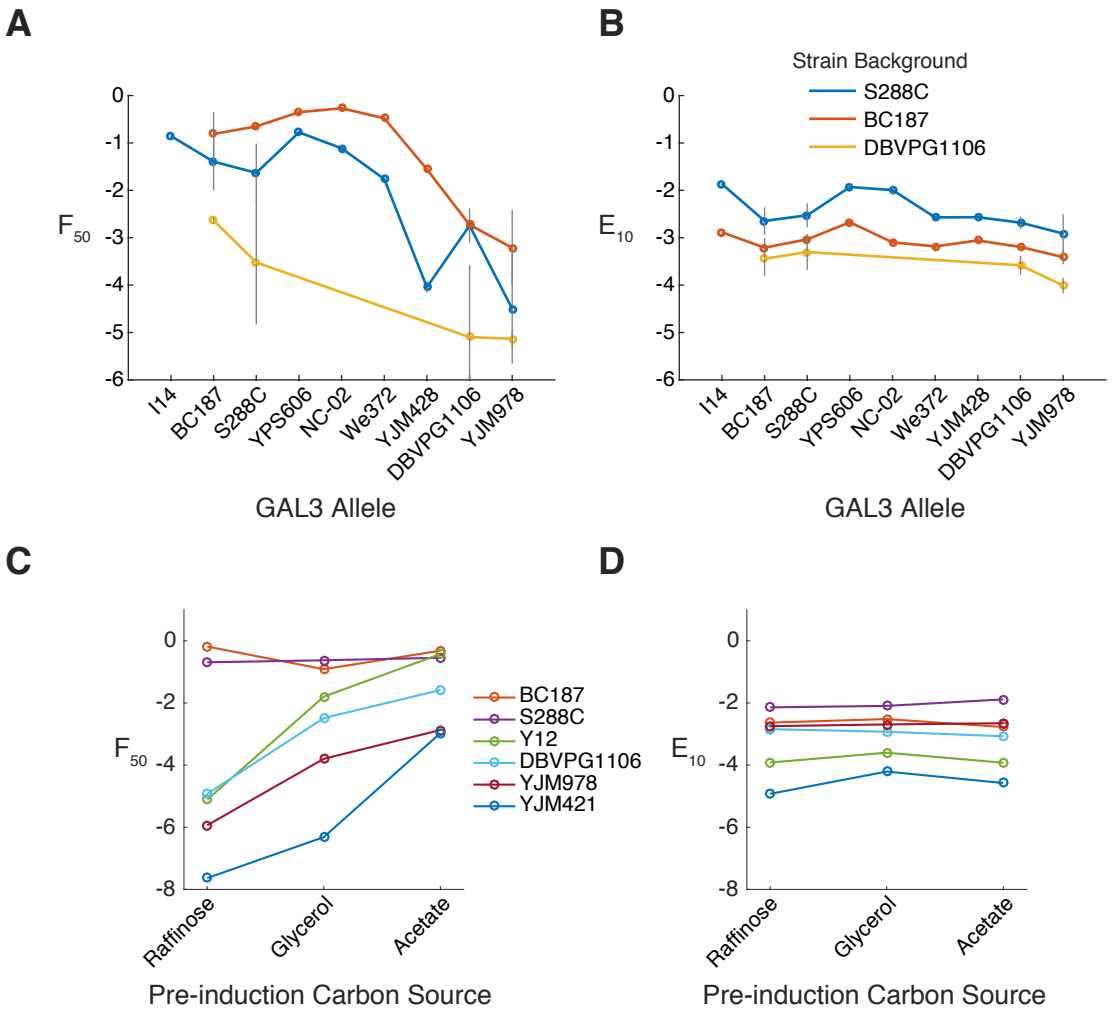


Figure 4.4. Perturbations that affect F_{50} do not affect E_{10}

(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. (B) F_{50} and E_{10} for 8 natural isolate strains induced in glucose+galactose after being cultured in raffinose, glycerol, or acetate. Raffinose pre-culture is the standard condition used for the other experiments in this paper.

Previously, a laboratory yeast strain was found to induce GAL genes bimodally or unimodally depending on the carbon source prior to galactose induction [12]. To see how metabolic memory affects E_{10} and F_{50} across our natural isolates, we pre-grew six strains in raffinose, acetate, or glycerol prior to induction in glucose + galactose. These carbon sources are neither inducers of GAL genes nor signals for glucose catabolite repression [27]. Nevertheless, they

caused the yeast strains to exhibit different F_{50} upon subsequent induction in glucose + galactose (Figure 4.4C). We did not observe a carbon source pre-conditioning effect on E_{10} (Figure 4.4D). 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.

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

The definitions of the E_{10} and F_{50} metrics (Figure 4.1F, S4.2) 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 (Figure 4.5A-B). Conversely, BC187 is unimodal, but decreasing its F_{50} by introducing $GAL3^{YJM978}$ makes it bimodal (Figure 4.5C-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 (Figure 4.5E-F).

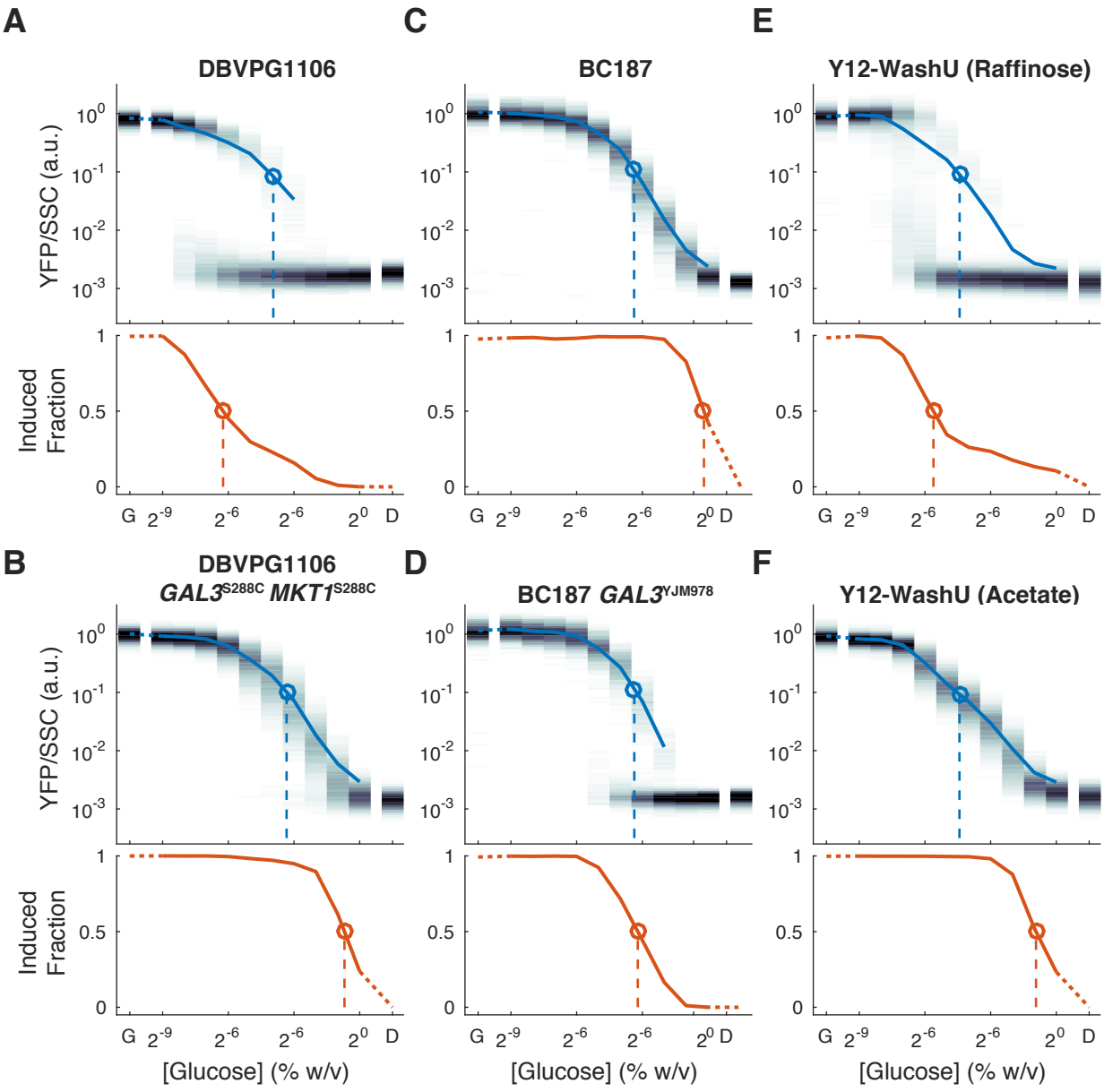


Figure 4.5. Changing F_{50} changes the number of modes of the GAL response

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 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

Our results show that the apparently complex bimodal GAL responses of natural yeast isolates can be simplified by being analyzed in terms of the dose-response of the induced fraction, F_{50}, and of the induced expression level, E_{10}. We show that these traits vary independently across natural isolates and can be tuned independently by alleles of *GAL3* and *GAL4* in the DBVPG1106 x S288C cross.

It is known that positive feedback on GAL gene expression through *GAL3* tunes the switching rate of cells between uninduced and induced states [14] and is a key contributor to the bistability of the pathway [13]. Additionally, changes in GAL3 dosage affects the induced fraction of GAL genes [16], and a panel of naturally occurring *GAL3* variants confers a spectrum of GAL induction phenotypes across yeast strains [19]. However, these studies only analyzed the effect of *GAL3* on induced fraction [14,16,19] or only considered whether a response was unimodal or bimodal [20]. Indeed, heterozygous deletion of *GAL4* was previously found to have no effect on the GAL induced fraction [16]. 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 [10,28]. 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. Residue 95 is on a loop linking the DNA-binding and regulatory domains of *GAL4* and directly participates in interactions with Gal11p [29,30], a component of the RNA polymerase II mediator complex that enhances expression of *GAL* genes [31]. These observations 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, genetic background, and metabolic state

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 [32], regulating translation [33], and affects numerous traits in crosses between S288C and natural isolates [34–39]. The S288C allele of *MKT1* is a loss-of-function variant relative to natural alleles and causes lower expression of mitochondrial genes [40,41]. 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 [42], 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 [19,42–44]. 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 [42], our results indicate that decision-making circuits are not as modular with respect to genetic variation as is often assumed.

In addition to genetic effects on GAL induction, we found that culturing cells in raffinose, glycerol, or acetate prior to induction in a glucose + galactose led to very different GAL phenotypes. Raffinose is commonly used to pre-culture cells for GAL induction studies (including most experiments here) because unlike glucose, it does not visibly repress GAL genes [14,45]. We chose glycerol and acetate by the same criterion. Otherwise, however, these carbon sources elicit very different physiological responses. Raffinose is hydrolyzed to release fructose [46], which can then be fermented [47,48]. Glycerol and acetate, by contrast, must be utilized via respiration [49], which entails expression changes in many genes [50] as well as differences in ATP/ADP ratio and redox state [51]. Therefore, our results suggest that factors other than canonical glucose catabolite repression may be important in setting the inducibility of GAL genes.

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 (Figure 4.5B, 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 [14,52]. 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 [53,54]. 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 [21]. Related work showed that both strategies could be implemented by the same strain as part of a bimodal response [20], and that this may be an evolutionarily stable strategy [55]. 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 [21]. 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 (600uL cultures in 1mL 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 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 8.0 (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 glycerol and acetate as pre-induction carbon sources were done as above, except S + 3% glycerol or 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 [56]. A similar mating-type-selection plasmid was used in a recent study [57].

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 [58], we then performed the following "intercross" protocol 4 times: from spore suspension, use 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 (Figure S4.3). The sorted cells were resupended 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 [59]. 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) [60] to compute LOD scores. Loci with maximal LOD>5 were considered significant; previous work showed that this corresponded to an FDR of 5% [57,61]. 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 [63], 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 [62]. 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. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet. 2005;6: 451–464. doi:10.1038/nrg1615
2. Balázsi G, van Oudenaarden A, Collins JJ. Cellular decision making and biological noise: from microbes to mammals. Cell. 2011;144: 910–25. doi:10.1016/j.cell.2011.01.030

3. Raj A, van Oudenaarden A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell. 2008;135: 216–226. doi:10.1016/j.cell.2008.09.050

4. Grimbergen AJ, Siebring J, Solopova A, Kuipers OP. Microbial bet-hedging: the power of being different. Curr Opin Microbiol. 2015;25: 67–72.

5. Veening J-W, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol. 2008;62: 193–210. doi:10.1146/annurev.micro.62.081307.163002

6. Xiong W, Ferrell JEJ. A positive-feedback-based bistable "memory module" that governs a cell fate decision. Nature. 2003;426: 460–465. doi:10.1038/nature02119.1.

7. Macarthur BD, Ma'ayan A, Lemischka IR. Systems biology of stem cell fate and cellular reprogramming. Nat Rev Mol Cell Biol. Nature Publishing Group; 2009;10: 672–681. doi:10.1038/nrm2766

8. Johnston M. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. Microbiol Rev. 1987;51: 458.

9. Bhat PJ. Galactose Regulon of Yeast: From Genetics to Systems Biology. Berlin: Springer; 2008.

10. Lohr D, Venkov P, Zlatanova J, Program B, Academy B. Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. 1995;9: 777–787.

11. Escalante-chong R, Savir Y, Carroll SM, Ingraham JB, Wang J, Marx CJ. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. Proc Natl Acad Sci. 2015;112: 1636–1641. doi:10.1073/pnas.1418058112

12. Biggar SR, Crabtree GR. Cell signaling can direct either binary or graded transcriptional responses. EMBO J. 2001;20: 3167–76. doi:10.1093/emboj/20.12.3167

13. Venturelli OS, El-Samad H, Murray RM. Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. Proc Natl Acad Sci U S A. 2012;109: 1–10. doi:10.1073/pnas.1211902109

14. Acar M, Becskei A, Oudenaarden A van. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435: 228. doi:10.1038/nature03524

15. Hawkins KM, Smolke CD. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. J Biol Chem. 2006;281: 13485–92. doi:10.1074/jbc.M512317200

16. Acar M, Pando BF, Arnold FH, Elowitz MB, van Oudenaarden A. A general mechanism for network-dosage compensation in gene circuits. Science (80-). 2010;329: 1656–60. doi:10.1126/science.1190544

17. Ramsey S a, 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: 1082–7. doi:10.1038/ng1869

18. Peng W, Liu P, Xue Y, Acar M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nat Commun. 2015;6: 6226.

19. Lee KB, Wang J, Escalante-Chong R, Palme J, Springer M. Natural variation in S. cerevisiae glucose-galactose sensing is driven by polymorphisms in the signal transducer Gal3p. Unpublished. 2016;

20. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. PLoS Biol. 2015;13: e1002042. doi:http://dx.doi.org/10.1101/002907

21. Wang J, Atolia E, Hua B, Savir Y, Escalante-chong R. Natural Variation in Preparation for Nutrient Depletion Reveals a Cost – Benefit Tradeoff. PLoS Biol. 2015; 1–31. doi:10.1371/journal.pbio.1002041

22. 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. doi:10.1371/journal.pbio.1001764

23. Solopova A, van Gestel J, Weissing FJ, Bachmann H, Teusink B, Kok J, et al. Bet-hedging during bacterial diauxic shift. Proc Natl Acad Sci U S A. 2014;111: 7427–32. doi:10.1073/pnas.1320063111

24. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James S a, et al. Population genomics of domestic and wild yeasts. Nature. Nature Publishing Group; 2009;458: 337–41. doi:10.1038/nature07743

25. Cromie G a, 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 (Bethesda). 2013;3: 2163–71. doi:10.1534/g3.113.007492

26. Edwards MD, Gifford DK. High-resolution genetic mapping with pooled sequencing. BMC Bioinformatics. BioMed Central Ltd; 2012;13 Suppl 6: S8.

27. Gancedo JM. Yeast Carbon Catabolite Repression. Microbiol Mol Biol Rev. 1998;62.

28. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, et al. Genome-wide location and function of DNA binding proteins. Science. 2000;290: 2306–9. doi:10.1126/science.290.5500.2306

29. Hidalgo P, Ansari AZ, Schmidt P, Hare B, Simkovich N, Farrell S, et al. Recruitment of the transcriptional machinery through GAL11P: Structure and interactions of the GAL4

dimerization domain. Genes Dev. 2001;15: 1007–1020. doi:10.1101/gad.873901

30. Hong M, Fitzgerald MX, Harper S, Luo C, Speicher DW, Marmorstein R. Structural Basis for Dimerization in DNA Recognition by Gal4. Structure. 2008;16: 1019–1026. doi:10.1016/j.str.2008.03.015

31. Himmelfarb HJ, Pearlberg J, Last DH, Ptashne M. GAL11P : A Yeast Mutation That Potentiates the Effect of Weak GAL4-Derived Activators. Cell. 1990;63: 1299–1309.

32. Wickner RB. Plasmids Controlling Exclusion of the K2 Killer Double-Stranded RNA Plasmid of Yeast. Cell. 1980;21.

33. Tadauchi T, Inada T, Matsumoto K, Irie K. Posttranscriptional regulation of HO expression by the Mkt1-Pbp1 complex. Mol Cell Biol. 2004;24: 3670–3681.

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

35. Deutschbauer AM, Davis RW. Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat Genet. 2005;37: 1333–40. doi:10.1038/ng1674

36. Demogines A, Smith E, Kruglyak L, Alani E. Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. PLoS Genet. 2008;4: e1000123.

37. 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.

38. Kim HS, Fay JC. A combined-cross analysis reveals genes with drug-specific and background-dependent effects on drug sensitivity in Saccharomyces cerevisiae. Genetics. 2009;183: 1141–1151.

39. 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.

40. 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. PLoS Genet. 2009;5: e1000358.

41. 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. Public Library of Science; 2004;2: e79. doi:10.1371/journal.pbio.0020079

42. Hua B, Springer M. Hierarchical structure explains why a quarter of the yeast genome affects a quantitative trait.

43. Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. 2013;494: 234–7. doi:10.1038/nature11867

44. Hou J, Sigwalt A, Fournier T, Pflieger D, Peter J, de Montigny J, et al. The Hidden Complexity of Mendelian Traits across Natural Yeast Populations. Cell Rep. 2016;

45. Johnston M, Flick JS, Pextont T. Multiple Mechanisms Provide Rapid and Stringent Glucose Repression of GAL Gene Expression in Saccharomyces cerevisiae. 1994;14. doi:10.1128/MCB.14.6.3834.Updated

46. Carlson M, Osmond BC, Botstein D. Mutants of yeast defective in sucrose utilization. Genetics. 1981;98.

47. Zimmermann, Friedrich K Entian K-D. Yeast sugar metabolism. CRC Press; 1997.

48. Huberts DHEW, Niebel B, Heinemann M. A flux-sensing mechanism could regulate the switch between respiration and fermentation. FEMS Yeast Res. 2012;12: 118–128. doi:10.1111/j.1567-1364.2011.00767.x

49. Turcotte B, Liang XB, Robert F, Soontorngun N. Transcriptional regulation of nonfermentable carbon utilization in budding yeast. FEMS Yeast Res. 2010;10: 2–13. doi:10.1111/j.1567-1364.2009.00555.x

50. DeRisi JL. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. Science (80-). 1997;278: 680–686. doi:10.1126/science.278.5338.680

51. van Dijken JP, Scheffers WA. Redox balances in the metabolism of sugars by yeasts. FEMS Microbiol Lett. 1986;32: 199–224. doi:10.1016/0378-1097(86)90291-0

52. Stockwell SR, Landry CR, Rifkin S a. The yeast galactose network as a quantitative model for cellular memory. Mol Biosyst. Royal Society of Chemistry; 2014;

53. Kumar PR, Yu Y, Sternglanz R, Johnston SA, Joshua-Tor L, Rajesh Kumar P, et al. NADP Regulates the Yeast GAL Induction System. Science (80-). 2008;319: 1090–2. doi:10.1126/science.1151903

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

55. Healey D, Axelrod K, Gore J. Negative frequency-dependent interactions can underlie phenotypic heterogeneity in a clonal microbial population. Mol Syst Biol. 2016;12: 877.

56. Hashida-Okado T, Ogawa A, Kato I, Takesako K. Transformation system for prototrophic industrial yeasts using the AUR1 gene as a dominant selection marker. FEBS Lett. 1998;425: 117–122. doi:10.1016/S0014-5793(98)00211-7

57. Treusch S, Albert FW, Bloom JS, Kotenko IE, Kruglyak L. Genetic Mapping of {MAPK-Mediated} Complex Traits Across S. cerevisiae. PLoS Genet. 2015;11: e1004913.

58. Cubillos F a, Parts L, Salinas F, Bergström A, Scovacricchi E, Zia A, et al. High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. Genetics. 2013;195: 1141–55. doi:10.1534/genetics.113.155515

59. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLoS One. 2015;10: e0128036. doi:10.1371/journal.pone.0128036

60. Edwards MD, Gifford DK. High-resolution genetic mapping with pooled sequencing. BMC Bioinformatics. BioMed Central Ltd; 2012;13 Suppl 6: S8. doi:10.1186/1471-2105-13-S6-S8

61. 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: 494–7. doi:10.1038/nature12904

62. 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 Syst. Elsevier; 2015; 1–9. doi:10.1016/j.cels.2015.02.001

63. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res. 2013;41.

Appendix I.

Supporting Information for Chapter 2

Natural Variation in Preparation for Nutrient Depletion Reveals a Cost–Benefit Tradeoff

Figure S2.1. Growth curves of all 43 strains assayed.

Plots of $\log_2$ (OD_{600}) versus time for 43 strains, after subtracting background (0.03) from the raw OD_{600} readings. Two replicates are shown in each panel. Time axes have been adjusted so that OD = 2^{-6} at time zero, to exclude an initial interval of 0-12 hours during which data can be noisy due to low OD (examples shown in Figure S2.2B). The strains are shown sorted from shortest to longest diauxic lag time from top left to bottom right. Plots with an asterisk "*" in top-right corner are strains shown in Figure 2.1B-C based on their galactose growth rate (Figure S2.3). Plots outlined in green represent the 15-strain subset used for GAL induction measurements (Table S1; Figure 2.3-2.4).

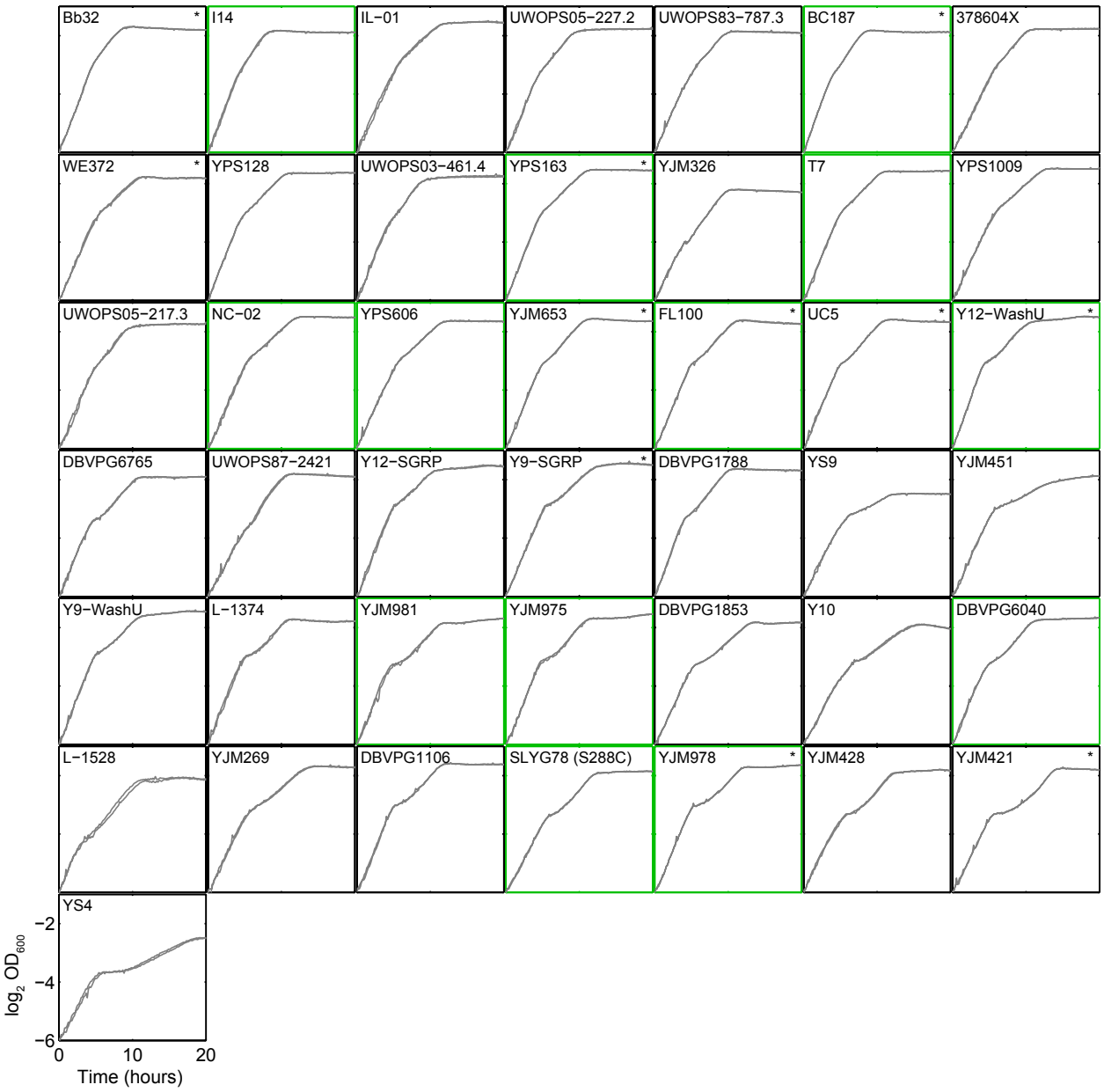


Figure S2.1 (Continued). Growth curves of all 43 strains assayed.

Figure S2.2. Diauxic lag and minimum mid-diauxic growth rate metrics correlate across replicates. **(A)** Measured optical density (i.e. absorbance at 600nm) versus actual culture density, obtained by serial dilution of a yeast culture saturated under growth curve assay conditions. Dilution series were prepared in triplicate. OD600 was linear with culture density in this range, and displayed a background (y-intercept) value of ~0.03. **(B)** Example growth curves (*top*) and growth rate plots (*bottom*) for 2 strains. Light gray lines show raw discrete derivatives computed from the growth curve data (Materials and Methods), blue lines show cubic spline fits to the discrete derivatives, and red lines show derivatives of the splines, or the smoothed 2nd derivative of the growth curves. Both replicates are shown for each strain. Strain Bb32 (*left*) did not have a local growth rate minimum, and therefore its diauxic lag duration was defined to be zero and its minimum mid-diauxic growth rate was defined to be the time of the inflection point in growth rate (Materials and Methods). More often, strains displayed a phenotype like SLYG78 (*right*), a S288C derivative, which had a clear minimum rate during diauxic shift. **(C)** Scatter plots of the diauxic lag time (*left*) and minimum mid-diauxic growth rate (*middle*) across 2 replicate experiments, and between diauxic lag time and minimum rate (*right*). All 3 plots are strongly correlated, showing that our metrics were robust to measurement noise and that the continuous phenotypic variation in diauxic growth is not an artifact of the lag time metric used in Figure 2.1. **(D)** Diauxic lag time and minimum mid-diauxic growth rate were calculated from a sliding-window average on the discrete derivatives of growth curves (as opposed to the cubic spline fit used for Figure 2.1). This method yielded almost identical results, showing that the metrics were not sensitive to the method of calculation.

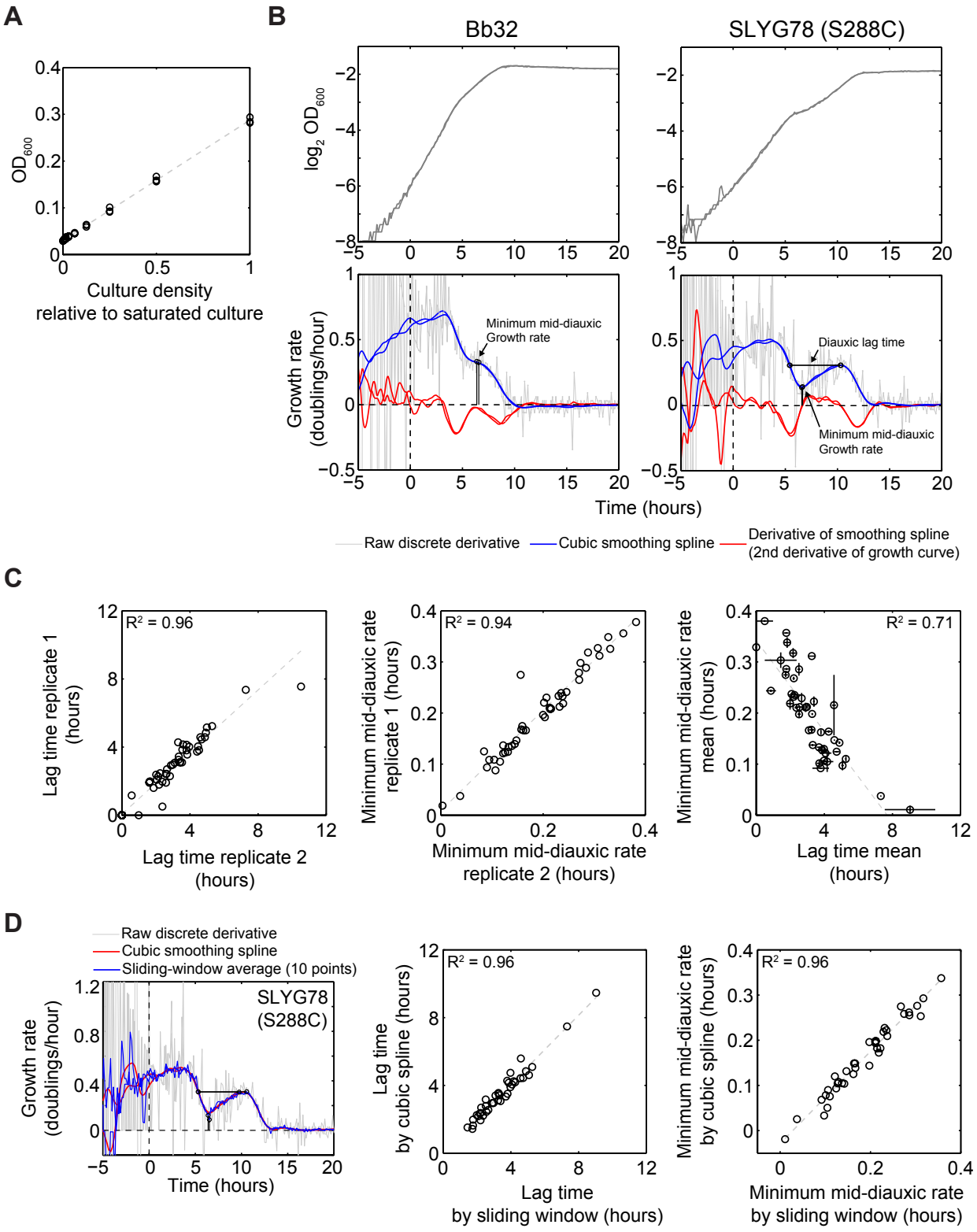


Figure S2.2 (Continued). Diauxic lag and minimum mid-diauxic growth rate metrics correlate across replicates.

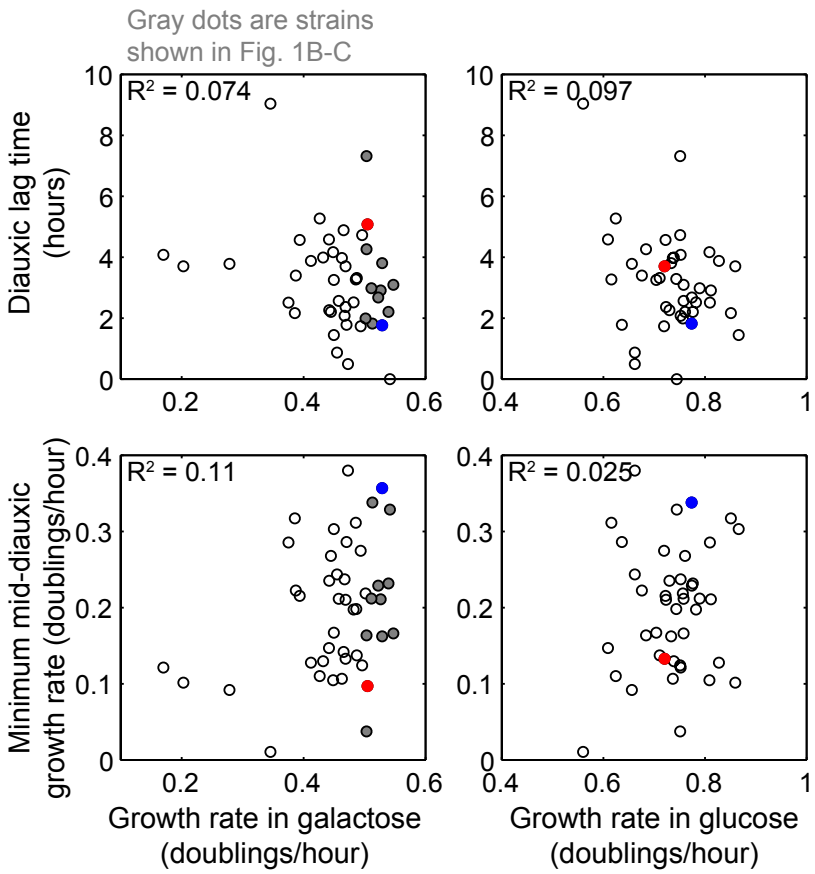


Figure S2.3. Diauxic lag time is not correlated with growth rate in glucose-only or galactose-only medium.

Scatterplots of diauxic lag time (*top*) and minimum mid-diauxic growth rate (*bottom*) versus steady-state growth rates in galactose alone (*left*) or in glucose alone (*right*). Steady-state growth rates were measured in a separate growth curve experiment (Materials and Methods). In general, the diauxic lag metrics correlated poorly with steady-state growth rates, suggesting that the phenotypic variation in diauxic growth cannot be solely explained by differences in glucose or galactose metabolism. Strains with growth rates between 0.5 and 0.6 doublings/hour in galactose are shown in Figure 2.1B-C (filled dots). This includes BC187 (blue) and YJM978 (red).

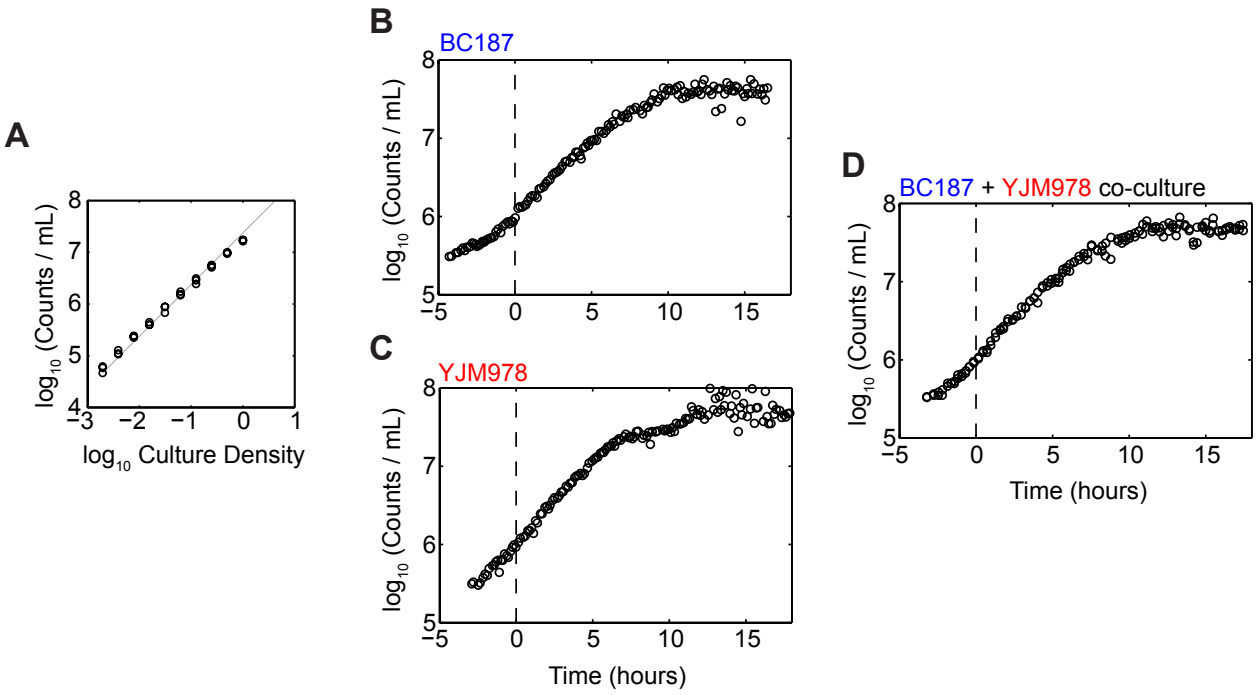


Figure S2.4. Determination of absolute cell concentration by flow cytometric counting.

(A) Absolute cell concentration as measured by flow cytometer versus actual relative culture density of a dilution series of a yeast culture, prepared in triplicate. Gray line shows predicted results extrapolated from the lowest density measurement, which agrees well with observed values. Absolute cell concentration was determined during the diauxic growth experiments in Figure 2.2B-C and 2.5A-B, and are plotted versus time for **(B)** BC187, **(C)** YJM978, and **(D)** a co-culture of BC187 and YJM978. Data for both replicates are shown—these are almost overlapping. The time axis is adjusted so that the culture is at 10^6 cells/mL at time zero. This time was determined by interpolation on a linear fit to four consecutive datapoints. The same adjustment was applied to the time axis for all plots in Figures 2.2 and 2.5.

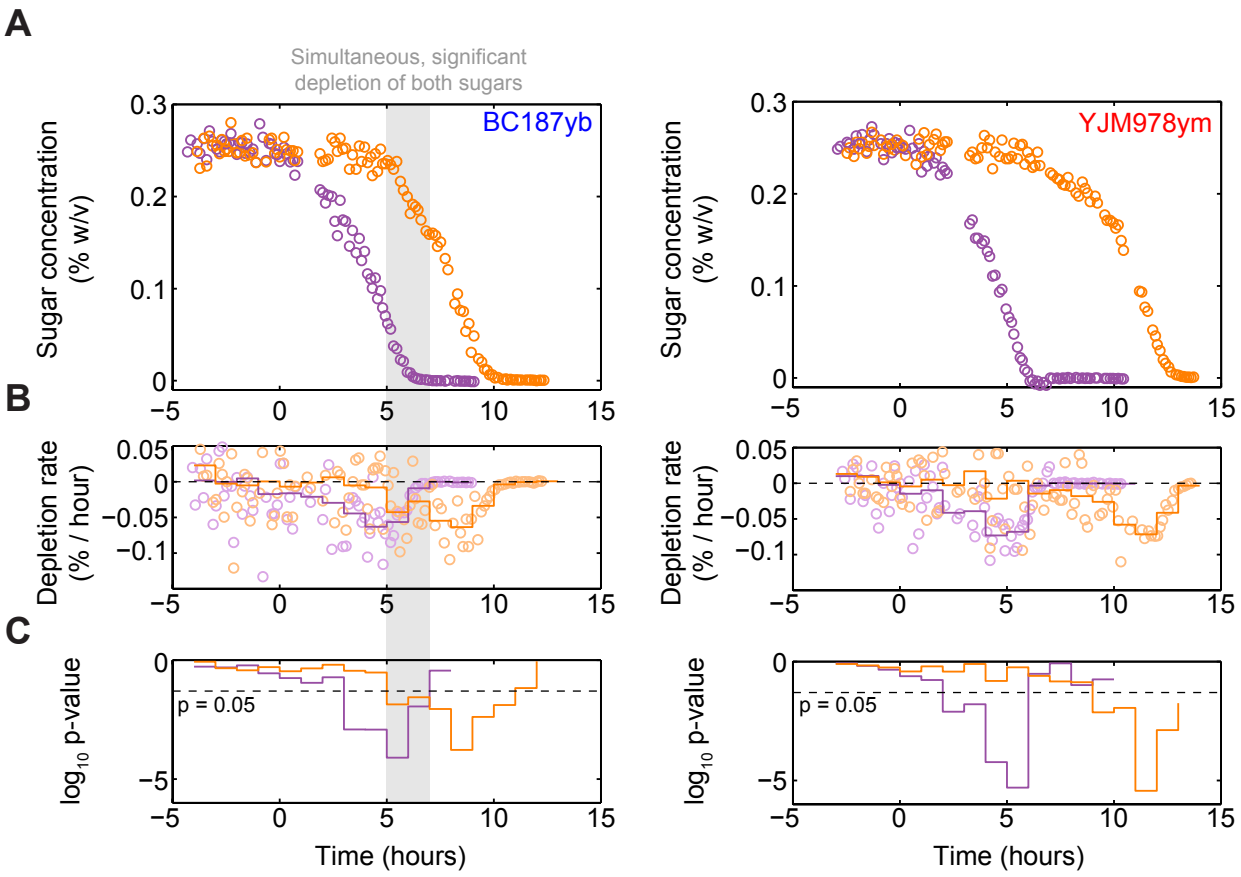


Figure S2.5. Strain BC187 can consume galactose and glucose simultaneously.

(A) Glucose and galactose concentrations versus time for (*left*) a culture of BC187, and (*right*) a culture of YJM978, from the same experiment as in Figure 2.2. Both replicates are plotted. Time zero corresponds to culture density of 10^6 cells/mL. To determine whether either strain begins to consume galactose prior to glucose exhaustion, we computed **(B)** the sugar depletion rate by taking discrete derivatives (circles) of the sugar concentrations, or slopes between consecutive data points, for both replicate datasets. We then binned the time axis into 1-hour intervals and computed, via a one-tailed t-test, the probability of observing the discrete derivatives in each time interval given a null hypothesis that the mean discrete derivative in that interval is 0 or positive. Mean sugar depletion rate for each bin is shown as lines in (B) and the $\log_{10}$ p-value for the significance test is shown in **(C)**. Dotted black lines in (C) indicate where p=0.05. For BC187, there is a 2-hour interval over which there is statistically significant depletion of both glucose and galactose, at a significance threshold of 0.05. By contrast, the time intervals of significant glucose and galactose depletion for YJM978 do not overlap temporally. These conclusions are robust to the width or position of time bins.

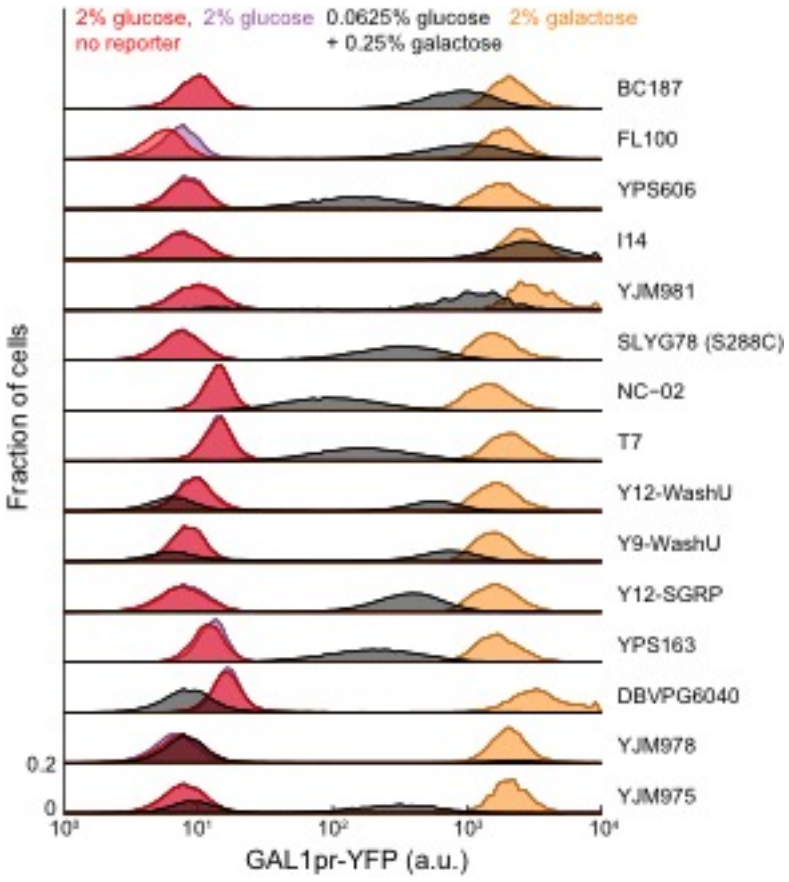


Figure S2.6. GAL1pr-YFP expression is highly variable across natural isolates in glucose + galactose but not in glucose alone.

Steady-state GAL1pr-YFP expression histograms for 15 strains showing partial expression in 0.0625% glucose + 0.25% galactose (black), basal expression in 2% glucose (purple), and maximal expression in 2% galactose (orange). Additionally, the parent strains without a GAL1pr-YFP reporter cassette were assayed in 2% glucose (red). Partial expression varies widely across strains in glucose + galactose, yet YFP signal above autofluorescence is undetectable from most strains in glucose-only medium (compare purple and red histograms). A number of strains (YJM981, Y12-WashU, Y9-WashU, YJM975) display bimodal expression in glucose+galactose. Measurements were taken at steady-state (as in Figure 2.4D; see Materials and Methods); distributions are unsmoothed histograms of 20,000 or more cells.

Figure S2.7. Co-culture method to determine timing of GAL induction relative to glucose depletion. **(A)** Example scatterplot of YFP versus mCherry signal by flow cytometry. Reference strain and query strain cells are distinguished by mCherry (red vs gray) and the YFP of each subpopulation is used to compute induction time. **(B)** Median GAL1pr-YFP expression of YJM978 with or without constitutive fluorophore in a co-culture of the two strains. Both strains contain the GAL reporter, which is unaffected by the constitutive fluorophore. **(C)** Start time of GAL induction t_{low} in BC187yb and YJM978ym cultured alone or in co-culture with each other, mean and range of 2 replicates. There is no significant difference in induction timing between separate and mixed cultures. **(D)** Median GAL1pr-YFP profiles for 15 strains from the co-culture experiment of Figure 2.3. Query and reference strain were mixed at three initial ratios. Density plot in background shows full YFP distributions of the query strain for the 1:4 query:reference condition (except for strain SLYG78, where 4:1 is shown). **(E)** Scatterplot of preparation time from different inoculating ratios. Preparation time was nearly identical across different inoculating ratios. The three ratios were used as replicates in Figure 2.3D-E. **(F)** Scatterplot of preparation time calculated as the time difference between query and reference strains at 1/32 or at 1/16 of maximal induction. The metric is robust to this difference (Spearman correlation = 0.97). **(G)** Definition of "ON fraction" as the fraction of cells with YFP signal higher than 1/32 of the maximal median YFP. **(H)** Possible ON fraction profiles. If a single population completely induces from basal to maximal ("Coherent induction"), the ON fraction will increase monotonically from 0 to 1. If the a culture splits into subpopulations with different induction times ("Early & late subpopulations"), the ON fraction will in two distinct phases, as in Venturelli et al. [23]. If a subset of cells never induce ("Non-inducing subpopulation"), the ON fraction will saturate below 1. **(I)** ON fraction versus time of the 15 strains in (D), from the 1:4 inoculation. Each strain is a different colored line, and strains BC187 and YJM978 are highlighted. Most profiles are consistent with "coherent induction", although in some strains, a small subpopulation consisting of less than 10% of all cells may have pre-induction before sampling. In some strains, the ON fraction decreases after saturating (see also panel D) – this is likely due to an experimental artifact (Materials and methods).

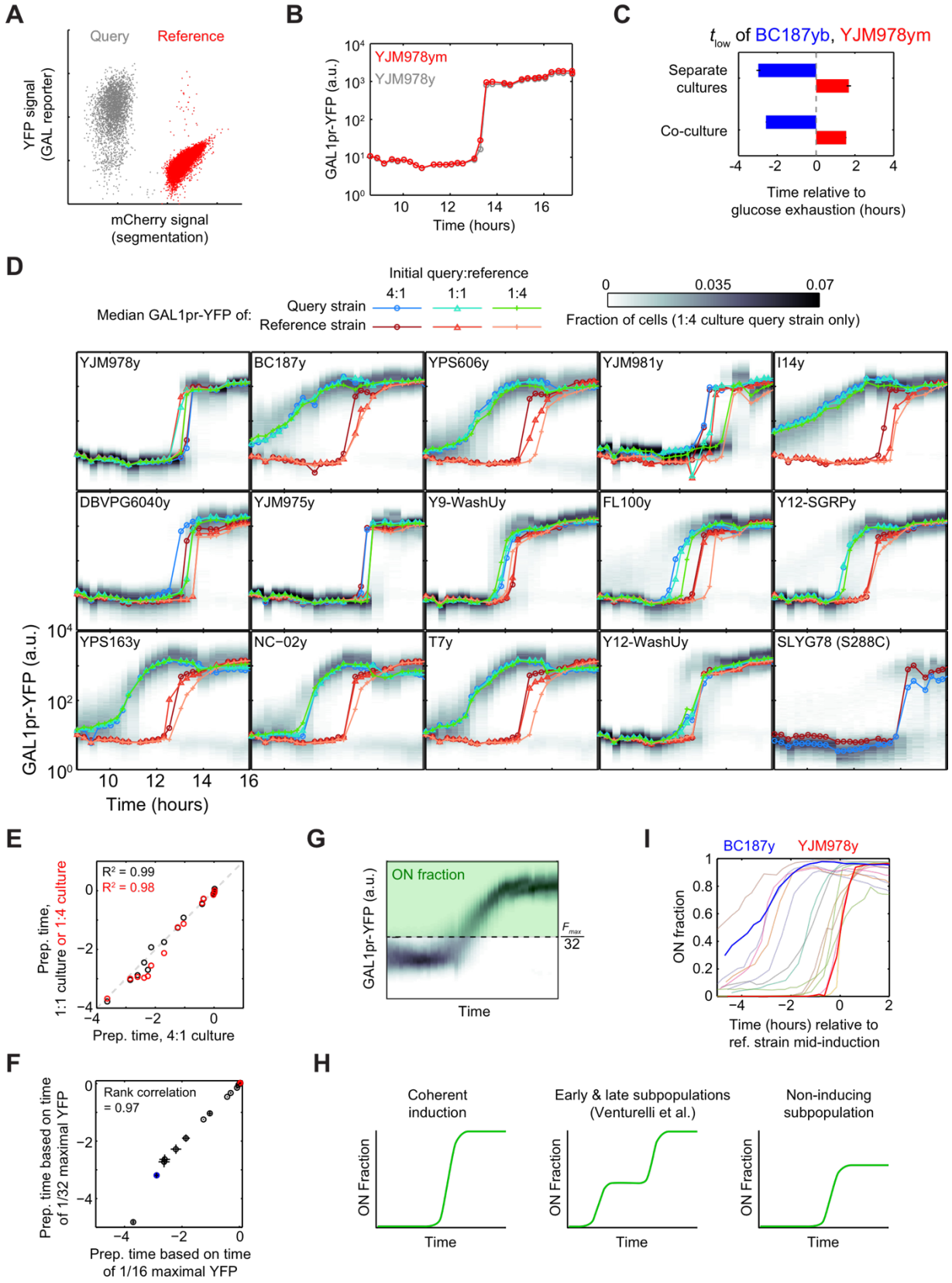


Figure S2.7 (Continued).

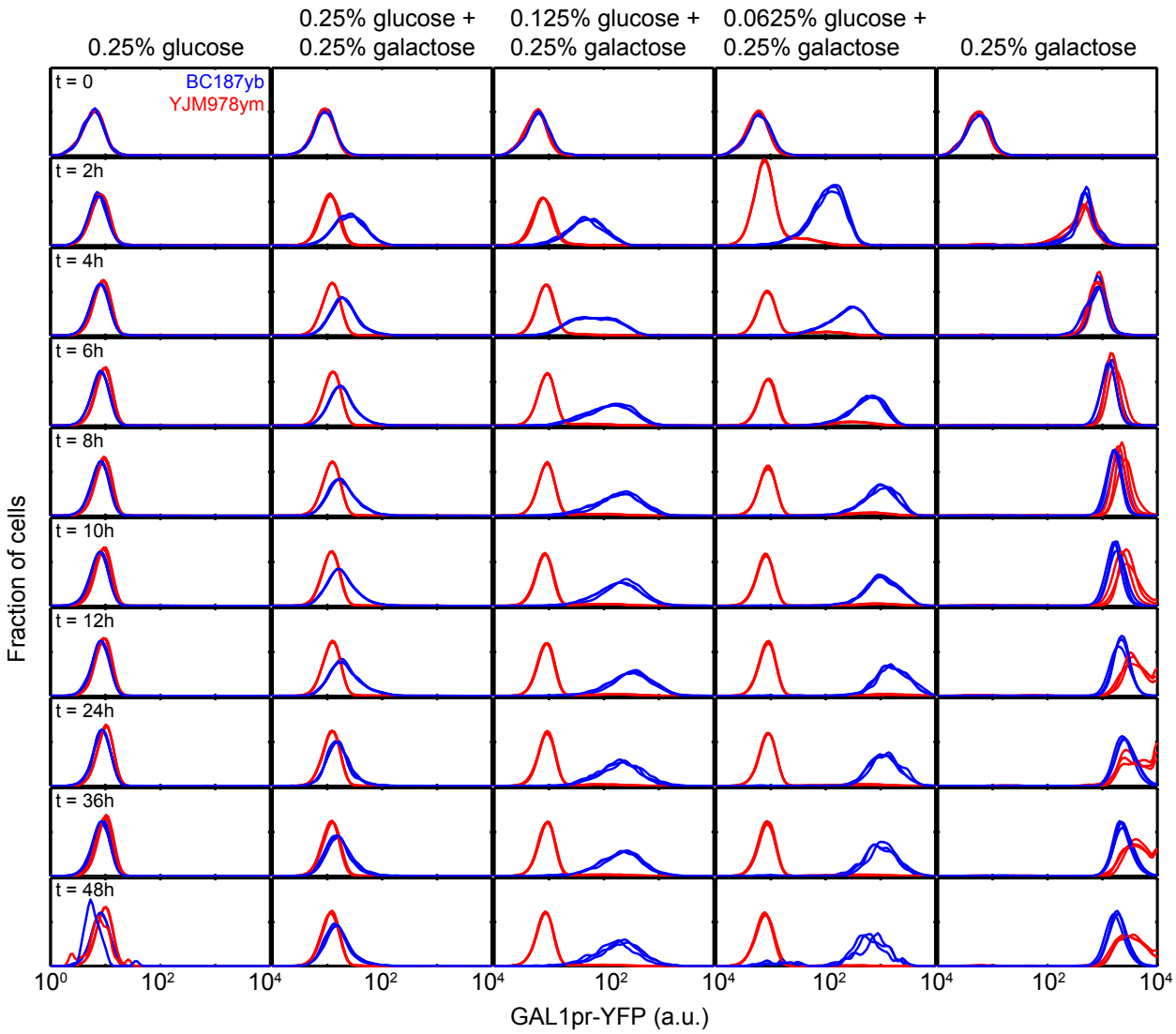


Figure S2.8. GAL1pr-YFP expression reaches steady-state after 8 hours of growth in galactose medium. GAL1pr-YFP expression distributions over time in repressing (0.25% glucose), inducing (0.25% galactose), and mixed-sugar (0.0625% glucose + 0.25% galactose) conditions for BC187yb (blue) and YJM978ym (red). Cultures were pre-grown in 2% raffinose to minimize the induction delay upon starting the experiment. Cells were diluted every two hours to maintain a density of less than 10^5 cells/mL. After 12 hours, the dilution factor was increased and dilution / sampling interval increased to 12 hours, and the cultures were monitored up to 48 hours. In conditions where either strain induces, expression stops increasing after eight hours.

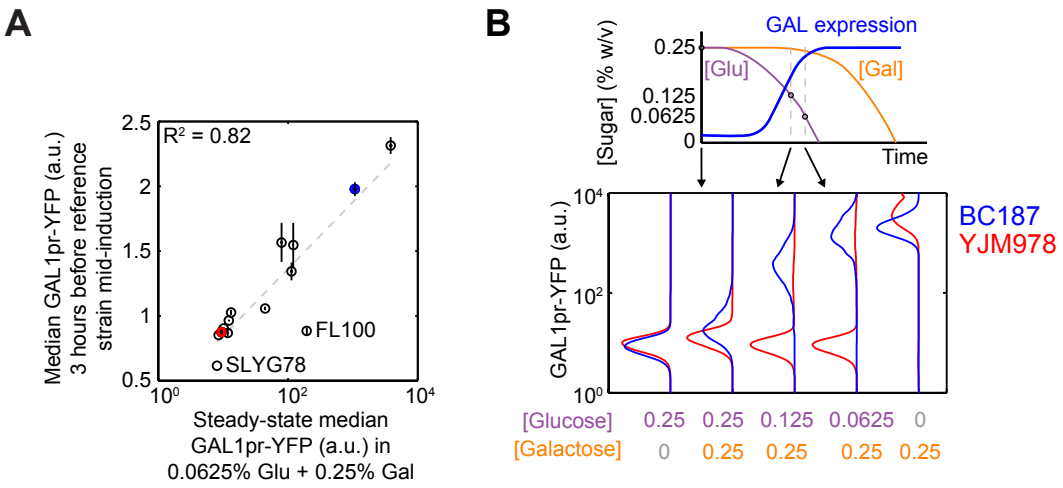


Figure S2.9. Strains induce GAL1pr-YFP at quasi-steady-state levels during gradual glucose depletion. **(A)** Scatterplot of median GAL1pr-YFP expression of query strains three hours before reference strain mid-induction time (computed from data in Figure 2.3) versus the median GAL1pr-YFP expression of the same strains growing at steady-state in 0.0625% glucose + 0.25% galactose. **(B)** Steady-state GAL1pr-YFP distributions for strains BC187 and YJM978 (*Bottom*) in glucose + galactose conditions chosen from different moments of diauxic growth (*Top schematic*). BC187 induces at intermediate levels at steady-state in glucose + galactose mixtures, rather than at basal or maximal levels, as would be expected if the level of GAL expression responds in a switch-like manner to decreasing glucose.

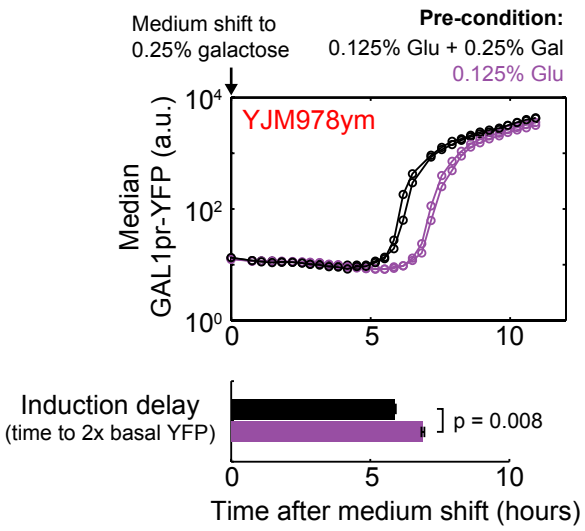


Figure S2.10. Pre-growth of YJM978 in a non-inducing galactose concentration accelerates GAL induction in subsequent medium shift.

Median GAL1pr-YFP expression versus time for YJM978ym cells suddenly transferred from glucose to galactose (purple), or from glucose + galactose to galactose (black). This strain induces GAL genes significantly earlier (p=0.008 by 2-sample t-test) in response to sudden galactose induction when pre-grown in the presence of some galactose.

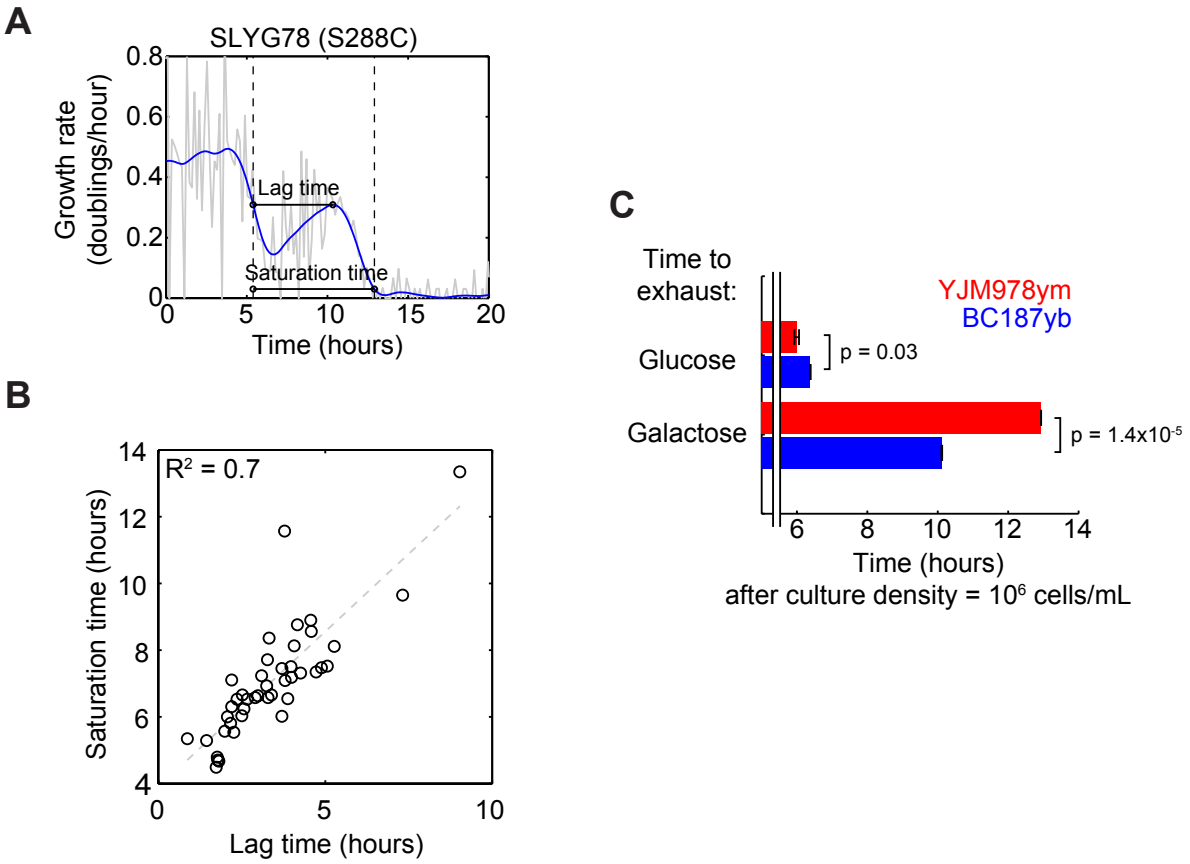


Figure S2.11. Short-lag strains reach saturation faster, but BC187 exhausts glucose more slowly than YJM978.

(A) Example calculation of saturation time, which is defined as the time for a strain to grow from the diauxic shift to saturation. **(B)** Scatterplot of saturation time versus diauxic lag time. The two metrics are strongly correlated, showing that strains that have a shorter diauxic lag also reach saturation sooner after diauxic shift. **(C)** Time to exhaustion of glucose or galactose in cultures of BC187yb (blue) or YJM978ym (red). YJM978 exhausts glucose significantly before BC187, even though BC187 exhausts galactose—and therefore both sugars—much faster than YJM978 under the assay conditions. Data are mean and range of n=2 replicates. P-value was calculated by 2 sample t-test.

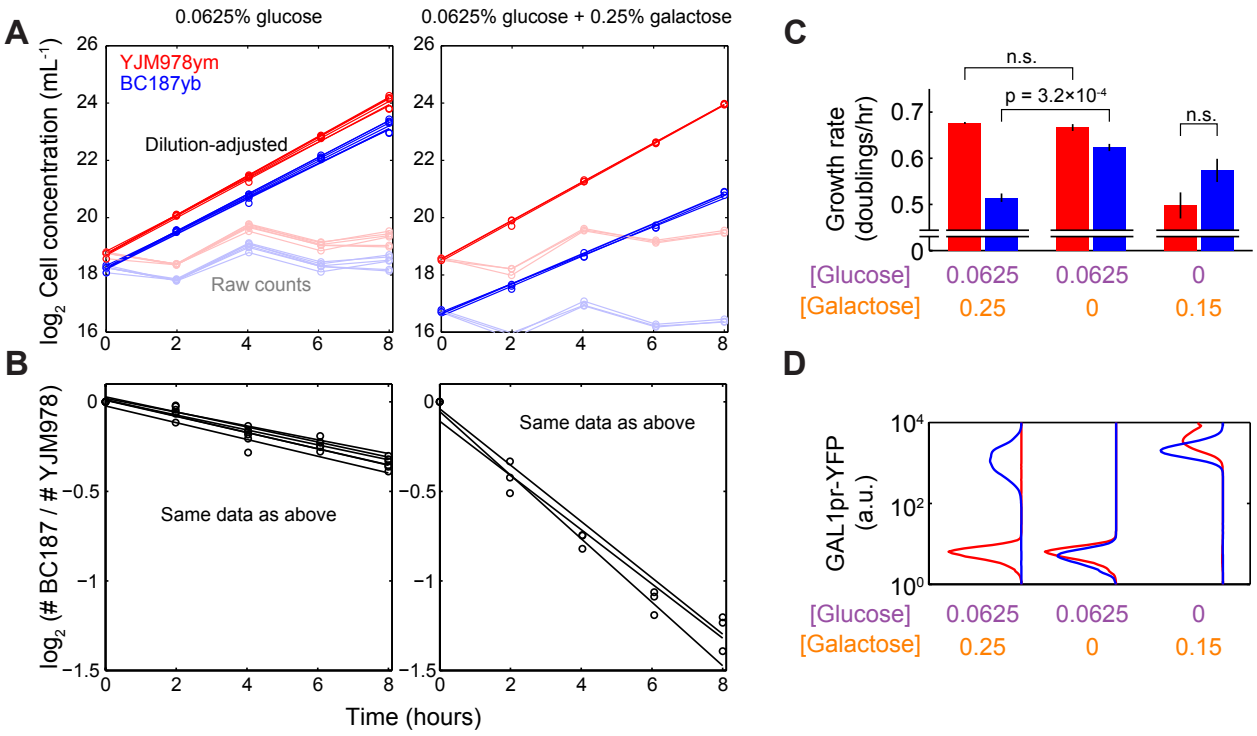


Figure S2.12. Absolute and relative fitness of BC187 and YJM978.

(A) Log2 absolute cell concentration versus time of strains BC187yb (blue) and YJM978ym (red) in 0.0625% glucose (left) and 0.0625% glucose + 0.25% galactose (right). Cultures were sampled every two hours after they had reached steady-state GAL1pr-YFP expression. Cultures were periodically diluted so that raw cell densities (light color) did not exceed 2^{20} or 10^6 cells/mL. Dilution-corrected data (dark color) were used to calculate growth rates. **(B)** Log2-ratio of BC187yb cell count to YJM978ym cell count in the same cultures as shown in (A). Relative fitnesses (i.e. growth rate differences) reported in Figure 2.5C are computed from line fits to these plots. **(C)** Steady-state growth rates of BC187yb (blue) and YJM978ym (red) in 0.0625% glucose + 0.25% galactose, 0.0625% glucose, or 0.15% galactose, as determined by linear fit to plots as in (A). Bar graphs are mean and s.e.m. of 3-6 replicates. P-values are computed by 2-sample t-test; "n.s." indicates $p > 0.05$. **(D)** Steady-state GAL1pr-YFP expression distributions of BC187 (blue lines) and YJM978 (red lines) in the conditions from (C). Only one timepoint and replicate is shown; others had identical fluorescence distributions.

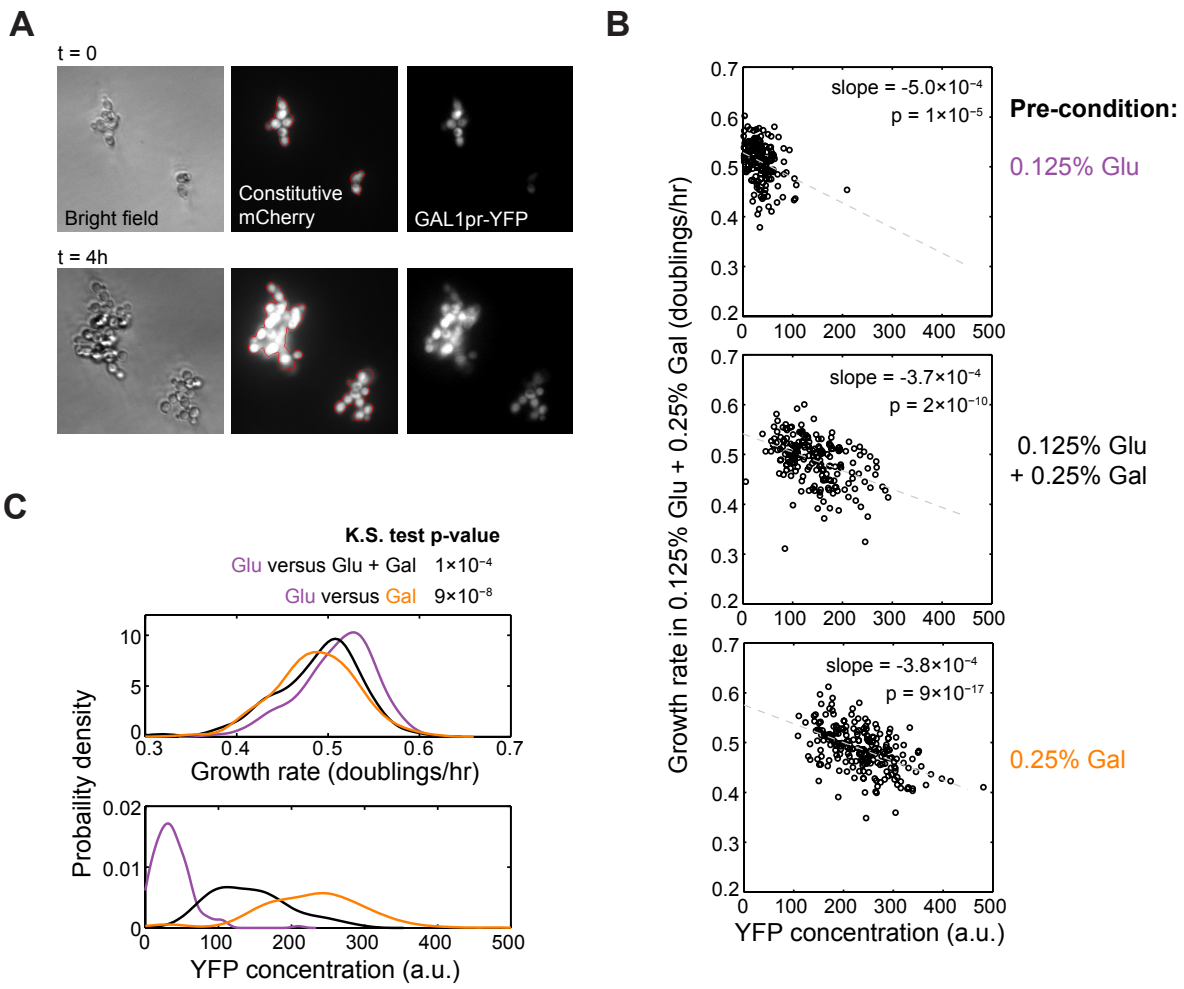


Figure S2.13. Single-cell growth rate correlates negatively with GAL1pr-YFP expression

(A) Example time-lapse microscopy images of BC187ym microcolonies (1-10 cells) at initial (top) and final (bottom) timepoints. Segmentation boundaries (red) were determined by analyzing mCherry images, and GAL1pr-YFP reporter concentration was determined as the final average YFP signal per pixel of each microcolony (Materials and methods). **(B)** Scatterplots of growth rate versus YFP concentration for microcolonies pre-induced in 0.125% glucose (n=165 microcolonies), 0.125% glucose + 0.25% galactose (n=196), or 0.25% galactose (n=223) prior to transferring to 0.125% glucose + 0.25% galactose for imaging. Growth rate displayed a significant negative correlation with GAL1pr-YFP concentration. **(C)** Distributions of growth rate (top) and YFP concentration (bottom) across microcolonies from the 3 pre-growth conditions. For clarity, plotted are smoothed probability densities estimated using a Gaussian kernel (Materials and methods). P-values are computed by a Kolmogorov-Smirnov test against the null hypothesis that growth rates of microcolonies from two pre-growth conditions have the same distribution.

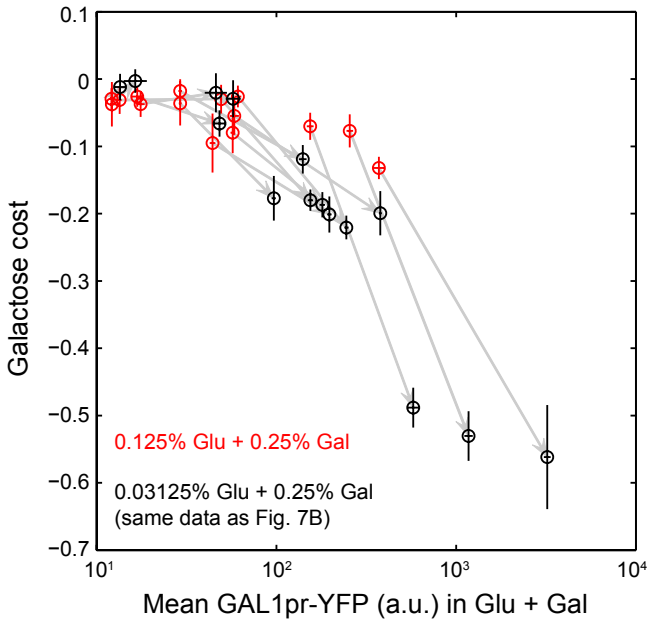


Figure S2.14. Galactose cost and GAL expression change in a correlated way between different media conditions.

Scatterplot of galactose cost versus mean GAL1pr-YFP expression at steady-state in two different sets of glucose or glucose + galactose media. Black circles are the same data as in Figure 2.7B, whereas red circles are data obtained from 0.125% glucose and 0.125% glucose + 0.25% galactose, which induces GAL genes to a lesser extent. Gray arrows connect strains between the 2 conditions. Most arrows point toward the lower left, indicating that as galactose stays constant and glucose decreases (such as during diauxic growth), GAL expression increases at the same time that the growth cost due to the presence of galactose increases.

Table S2.1. Strains used in this study.

Available at 10.1371/journal.pbio.1002041.s015.

This file contains three worksheets. Worksheet **(A)** lists the 43 natural isolates assayed by growth curves in Figure 2.1. Worksheet **(B)** lists the GAL1pr-YFP reporter strains constructed from a subset of 15 natural isolates. Worksheet **(C)** lists strains used in time-lapse microscopy and synthetic GAL induction experiments.

Table S2.2. Phenotypic measurements of natural isolates.

Available at 10.1371/journal.pbio.1002041.s016.

This file contains four worksheets. Worksheet **(A)** summarizes the metrics used in the paper and describes how they are defined and inter-related. Worksheet **(B)** contains values of the diauxic lag time and minimum mid-diauxic growth rate metrics from both replicates of the growth curve experiments in Figure 2.1. Worksheet **(C)** contains values of preparation time and other metrics measured on a subset of 15 natural isolates and used in Figures 2.3-2.4. Worksheet **(D)** contains growth rate measurements used to determine the galactose cost, as well as GAL expression data, plotted in Figures 2.7 and S2.14.

Appendix II.

Supporting Information for Chapter 3

Polymorphisms in the Yeast Galactose Sensor Underlie a Natural Continuum of Nutrient-Decision Phenotypes

Table S3.1. List of strains used in this study.

Strain ID	Genotype
SLL16-A01	Bb32 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-B01	BC187 MATa/x; hoΔ::GAL1pr-YFP-kanMX4
SLL16-C01	CLIB215 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-D01	CLIB324 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-E01	CLIB382 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-F01	DBVPG1106 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-G01	DBVPG1373 MATa/x; hoΔ::GAL1pr-YFP-kanMX4
SLL16-H01	DBVPG1788 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-A02	DBVPG1853 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-B02	DBVPG6040 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-C02	DBVPG6765 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-D02	FL100 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-E02	I14 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-F02	IL-01 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-G02	L-1374 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-H02	M22 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-A03	NC-02 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-B03	s288c MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-C03	T7 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-D03	UC5 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-E03	UWOPS03-461.4 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-F03	UWOPS87-2421 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-G03	WE372 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-H03	Y12-SGRP MATa/x; hoΔ::GAL1pr-YFP-kanMX4
SLL16-A04	Y12-WashU MATa/x; hoΔ::GAL1pr-YFP-hphNT1

SLL16-B04	Y9-WashU MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-C04	YJM421 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-D04	YJM428 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-E04	YJM653 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-F04	YJM975 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-G04	YJM978 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-H04	YJM981 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-A05	YPS1009 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-B05	YPS128 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-C05	YPS163 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-D05	YPS606 MATa/x; hoΔ::GAL1pr-YFP-hphNT1
SLL16-A06	I14 MATa; hoΔ::GAL1pr-YFP-TDH3pr-BFP-kanMX4
SLL16-B06	BC187 MATa; hoΔ::GAL1pr-YFP-TDH3pr-BFP-kanMX4
SLL16-C06	YPS606 MATa; hoΔ::GAL1pr-YFP-TDH3pr-BFP-kanMX4
SLL16-D06	S288C MATa; hoΔ::GAL1pr-YFP-TDH3pr-BFP-kanMX4
SLL16-E06	Y12-WashU MATa; hoΔ::GAL1pr-YFP-TDH3pr-BFP-kanMX4
SLL16-F06	DBVPG1106 MATa; hoΔ::GAL1pr-YFP-TDH3pr-BFP-kanMX4
SLL16-G06	YJM978 MATa; hoΔ::GAL1pr-YFP-TDH3pr-BFP-kanMX4
SLL16-H06	YJM421 MATa; hoΔ::GAL1pr-YFP-TDH3pr-BFP-kanMX4
SLL16-A07	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM421
SLL16-B07	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3Y12-WashU
SLL16-C07	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3DBVPG1106
SLL16-D07	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM978
SLL16-E07	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3S288C
SLL16-F07	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM428
SLL16-G07	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3BC187
SLL16-H07	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3NC-02
SLL16-A08	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3I14
SLL16-B08	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YPS606
SLL16-C08	YJM978 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3WE372
SLL16-D08	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM421
SLL16-E08	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3Y12-WashU
SLL16-F08	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3DBVPG1106

SLL16-G08	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM978
SLL16-H08	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3S288C
SLL16-A09	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM428
SLL16-B09	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3BC187
SLL16-C09	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3NC-02
SLL16-D09	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3I14
SLL16-E09	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YPS606
SLL16-F09	BC187 MATx; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3WE372
SLL16-G09	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM421
SLL16-H09	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3Y12-WashU
SLL16-A10	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3DBVPG1106
SLL16-B10	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM978
SLL16-C10	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3S288C
SLL16-D10	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM428
SLL16-E10	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3BC187
SLL16-F10	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3NC-02
SLL16-G10	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3I14
SLL16-H10	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YPS606
SLL16-A11	S288C MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3WE372
SLL16-B11	YJM421 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM978
SLL16-C11	YJM421 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3S288C
SLL16-D11	YJM421 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3BC187
SLL16-E11	DBVPG1106 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM978
SLL16-F11	DBVPG1106 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3S288C
SLL16-G11	DBVPG1106 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3BC187
SLL16-H11	Y12-WashU MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM978
SLL16-A12	Y12-WashU MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3S288C
SLL16-B12	Y12-WashU MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3BC187
SLL16-C12	I14 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3YJM978
SLL16-D12	I14 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3S288C
SLL16-E12	I14 MATa; hoΔ::GAL1pr-kanMX4; gal3Δ::GAL3BC187
SLYB93	YJM978 MATa/x; hoΔ::GAL1pr-YFP-kanMX4/ hoΔ::TDH3pr-mCherry-natMX4

Table S3.2. List of significant loci and associated genes at LOD > 5.

This table lists genomic regions for which peak LOD > 5 in the bulk segregant analysis. 2-LOD support intervals are shown for each peak in each cross, as well as averaged support intervals that combine information from "clusters" of peaks within 20kb of each other from different crosses. A subset of genes with sacCer3 (SGD R64-1-1) annotations in the support intervals for each locus are shown.

Cross	Chromosome	Max. LOD	Max. LOD Location (bp on chromosome)	2-LOD Support Interval (start and end in bp)		Clustered locus 2-LOD S.I. (mean start & end)		Genes in (clustered) 2-LOD S.I.
YJM978 x BC187	4	197.71	465800	459600	470600	449750	468638	NTH1, YRB1, RCR2, RAD57, MAF1, SOK1, TRP1, ARS1, GAL3, SNQ2, YDR003W-A, YDR008C, YDR010C
YPS606 x Y12	4	164.7	451600	446100	458300			
YPS606 x S288C	4	92.99	461600	451300	469900			
DBVPG1106 x YJM421	4	69.24	462200	450700	467000			
I14 x BC187	4	69.12	461800	452700	472600			
I14 x YJM421	4	40.79	467400	463800	474100			
DBVPG1106 x S288C	4	32.96	457600	444700	469300			
YJM978 x Y12	4	24.49	449300	429100	467300			
I14 x BC187	13	37.28	183200	173600	193600	163867	201367	SUP5, SUR7, GAL80, AIM32, RSE1, GSF2, PRM6, PRP39, RRN11, CAT2, and 17 more annotations…
DBVPG1106 x YJM421	13	7.03	176200	164800	197400			
YJM978 x Y12	13	6.76	185100	153200	213100			
YJM978 x Y12	13	15.68	609800	594700	621600	578233	629467	CEP3, ALD3, ALD2, EAR1, HOT1, DDR48, PAI3, SIP18, ECM5, and 20 more annotations…
YPS606 x Y12	13	7.85	611200	566600	629300			
I14 x BC187	13	5.83	608200	573400	637500			

DBVPG1106 x S288C	14	31.85	460200	451800	469400	450500	472850	TOP2, TCB2, SNN1, MKT1, END3, and 4 more annotations…
YPS606 x S288C	14	12.35	464000	449200	476300			
YPS606 x S288C	4	13.95	1155900	1147200	1168900	1144850	1176950	HXT7, HXT6, HXT3, SVF1, and 17 more annotations…
DBVPG1106 x YJM421	4	8.66	1153700	1142500	1185000			
YPS606 x S288C	16	9.43	81200	71800	94000	74300	92400	GYP5, GAL4, RBD2, HUT1, SRP68, and 5 more annotations…
I14 x YJM421	16	7.46	81900	76800	90800			
YPS606 x S288C	15	7.55	746100	730800	760700	730800	760700	MGM1, STE4, SAS5, SPR2, and 3 more annotations…
YPS606 x S288C	15	7.55	746700	730800	760700			
I14 x YJM421	4	7.41	365400	342500	373800	339500	474650	STP4, SIT4, NPC2, MRP10, FAD1, MTF2, and 20 more annotations…
I14 x YJM421	4	5.97	384100	374800	425700			
YPS606 x Y12	2	7.04	568200	537700	601000	537700	601000	ICS2, AMN1, IFA38, CDC28, and 16 more annotations…
YPS606 x Y12	2	7.04	568600	537700	601000			
YPS606 x Y12	8	6.87	287400	249100	324400	251400	318450	MSR1, HXT4, AHT1, HXT1, HXT5, and 24 more annotations…
I14 x BC187	8	5.79	287600	253700	312500			
YPS606 x Y12	13	15.64	357400	339800	363700	N/A	N/A	IMP2, MIH1, MSN2, CCS1 and 14 more...
DBVPG1106 x S288C	12	11.61	1053400	1042100	1065000	N/A	N/A	FMP27, PDP3, NBP1, GAB1 and 7 more...
YJM978 x Y12	7	11	781500	766200	812800	N/A	N/A	CBF2, VPS62, BTN2, SKN1 and 30 more...
YPS606 x S288C	8	9.68	112500	96100	125900	N/A	N/A	SHU1, MRP4, LAG1, HSE1 and 21 more...
YJM978 x Y12	2	8.76	451100	408500	474400	N/A	N/A	TEC1, MIS1, RPL19A, AAC3 and 35 more...
YPS606 x S288C	10	8.74	668000	649000	688900	N/A	N/A	IBA57, RPS5, ENT3, VPS70 and 16 more...
YPS606 x S288C	16	8.17	878100	858000	898000	N/A	N/A	GPH1, SGV1, ORC4, TIF3 and 28 more...
YPS606 x S288C	15	7.89	375500	314700	393600	N/A	N/A	RPB11, SIN3, PFA4, IZH2 and 45 more...

I14 x YJM421	4	7.57	297900	284300	312100	N/A	N/A	SNU23, RPN6, PMT1, OPI6 and 16 more...
DBVPG1106 x YJM421	1	7.23	143900	129800	162300	N/A	N/A	CYS3, SWC3, MDM10, SPO7 and 20 more...
DBVPG1106 x S288C	2	7.17	416700	395000	469500	N/A	N/A	RPG1, SEC18, SPT7, UBC4 and 40 more...
YPS606 x S288C	13	6.66	108100	97600	122000	N/A	N/A	TUB1, ATP18, TDA9, DUS1 and 9 more...
I14 x BC187	15	6.38	796200	755100	822400	N/A	N/A	RCN2, MCT1, ODC2, SNR35 and 48 more...
YPS606 x Y12	15	5.92	210100	148100	248300	N/A	N/A	HAL9, MPD2, DUF1, MHF1 and 52 more...
DBVPG1106 x S288C	9	5.85	117500	91200	139800	N/A	N/A	OM45, VHS2, SNR68, FLX1 and 19 more...
I14 x BC187	7	5.5	321200	262800	344500	N/A	N/A	CEG1, RSM23, CWC23, SOH1 and 48 more...
DBVPG1106 x S288C	12	5.36	653200	633200	682100	N/A	N/A	RCK2, YEF3, SSP120, SYM1 and 28 more...
YPS606 x S288C	16	5.26	577300	533100	605700	N/A	N/A	TAF3, RET3, RQC2, CHL1 and 39 more...
YJM978 x BC187	8	5.25	31500	5100	82100	N/A	N/A	COS8, ARN2, PAU13, ECM34 and 45 more...
YPS606 x S288C	14	5.2	149400	122300	189700	N/A	N/A	TOF1, SEC2, BNI1, ALP1 and 28 more...
YJM978 x BC187	13	5.14	568600	498400	649000	N/A	N/A	SNR24, ASC1, SPC24, SHH3 and 98 more...

Figure S3.1. 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 [36]. Strains highlighted in red were used in this study, while strains in black were not.

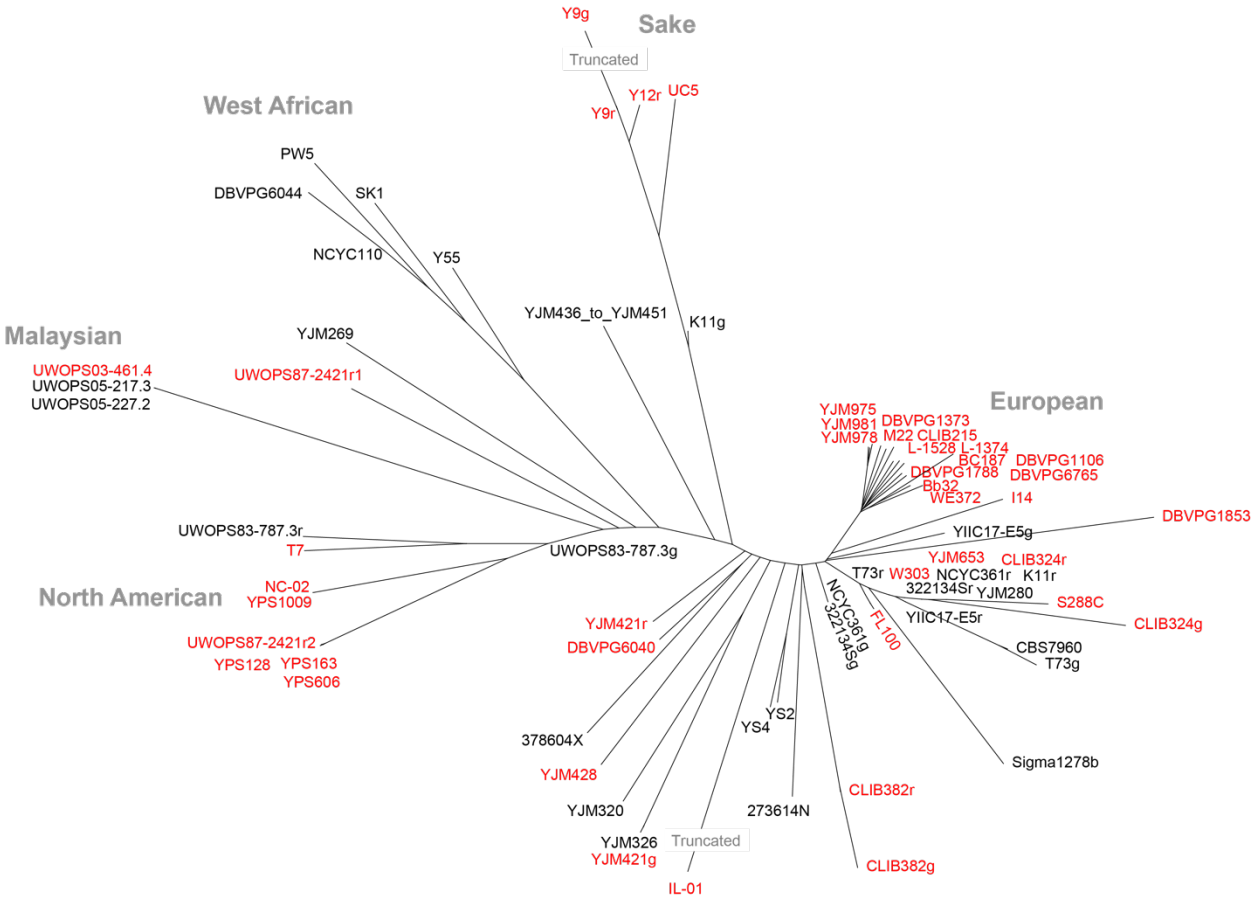


Figure S3.2. Quality control for query strains based on the reference strain.

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.

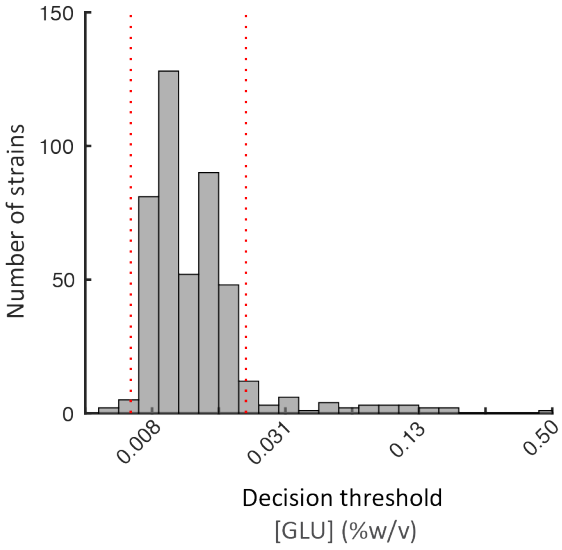


Figure S3.3. Determining a cut-off for query outliers.

(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 derived from a normal distribution of standard deviation 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.

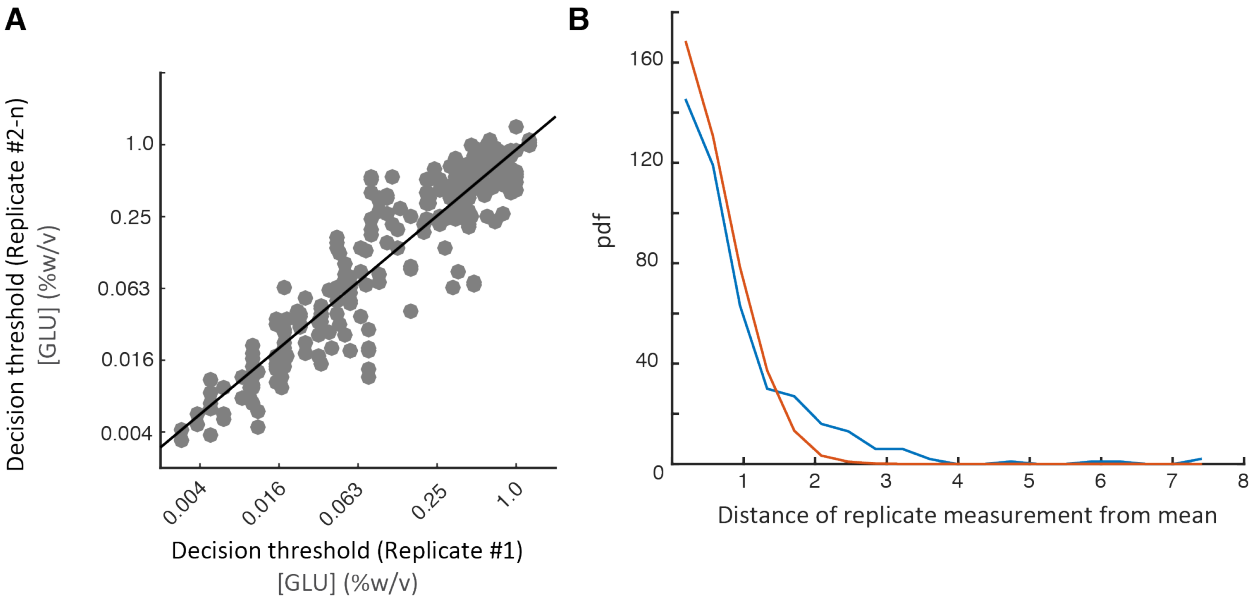


Figure S3.4. Steady-state expression of *GAL1pr-YFP* from a panel of natural isolates in mixtures of glucose and galactose.

Representative YFP induction profiles of the diploid natural isolates assayed in Figure 3.1. Cells were grown for 8 hours, a time previously determined to be sufficient for expression to reach steady-state [21], 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 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.

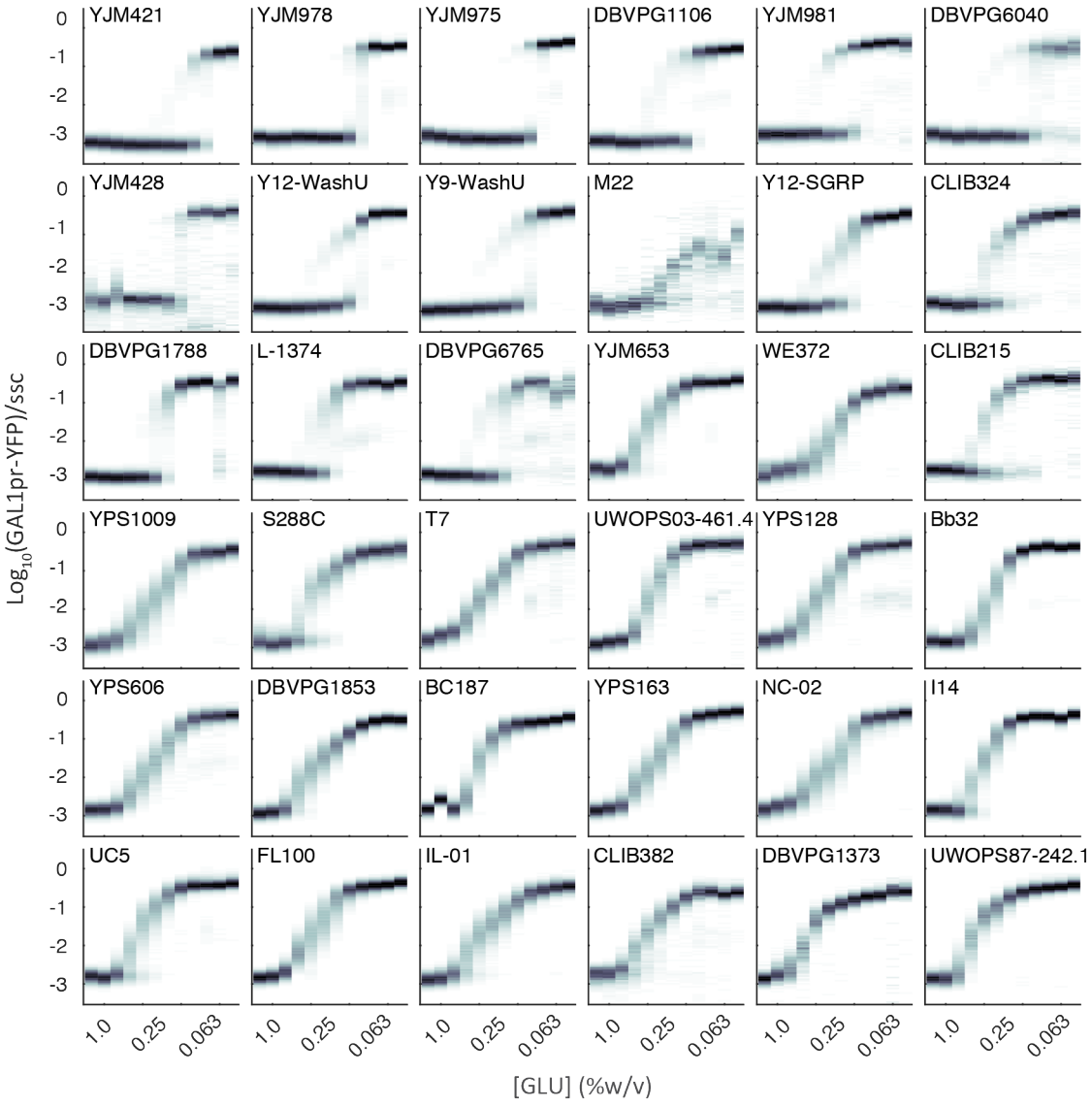


Figure S3.5. Growth rate in 0.5% glucose or 0.5% galactose is not strongly correlated with decision threshold.

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 Figure 3.1). 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.

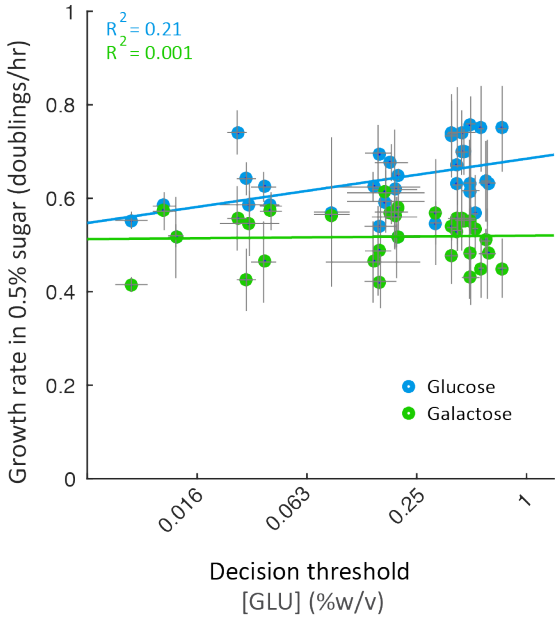


Figure S3.6. Correlation between genetic distance and phenotypic distance for decision threshold and traits from literature.

Genetic distance [36] 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.

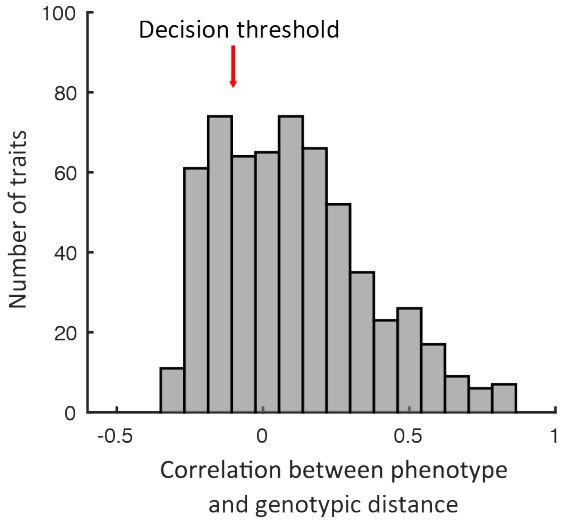


Figure S3.7. Relationship of decision threshold with phylogeny and ecological niche.
Phylogenetic tree was constructed based on the Cromie et al. distance matrix (Materials and methods) with the bar plot indicating decision threshold (from Figure 3.1). Color of bars indicate the ecological niche of strain.

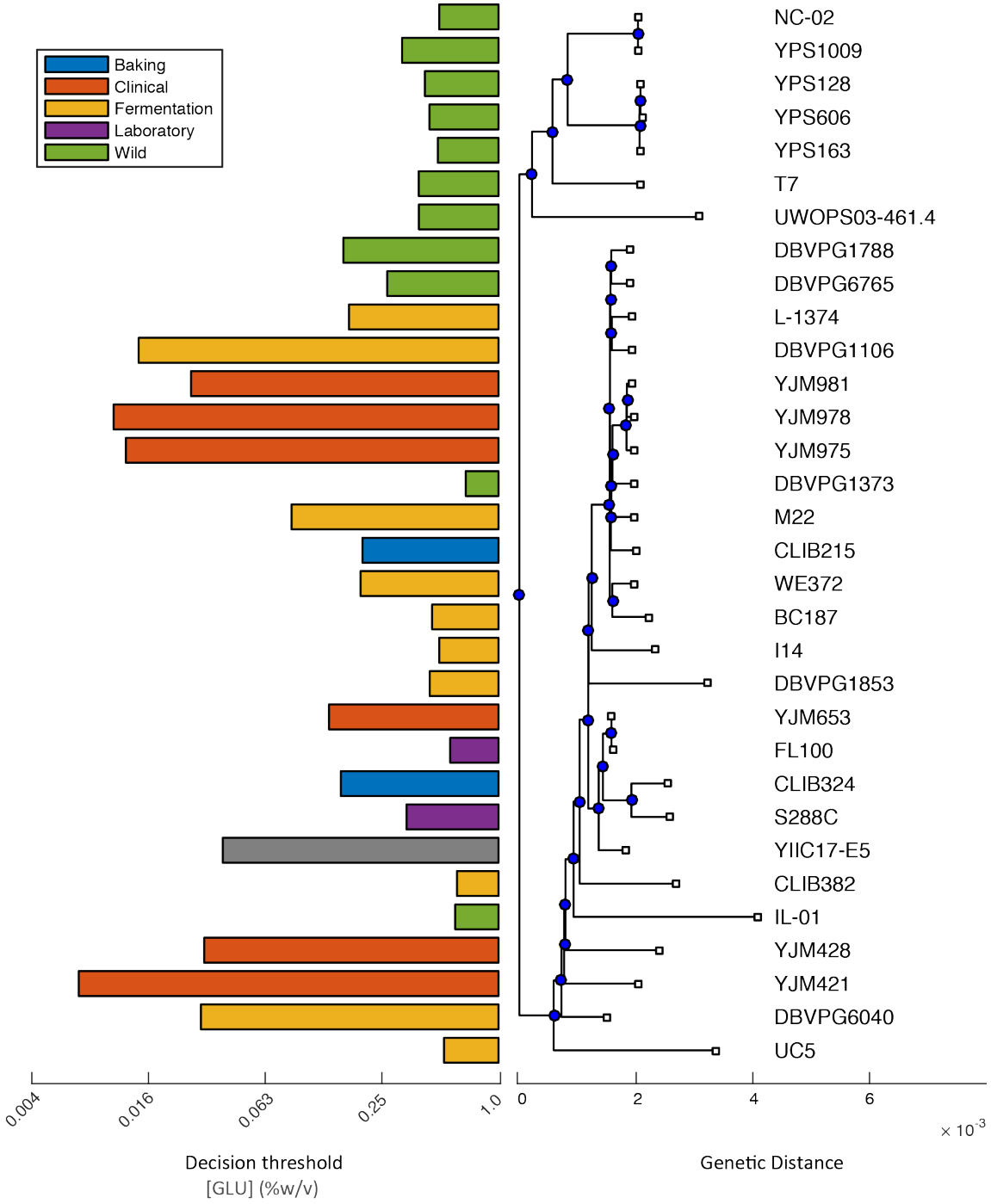


Figure S3.8. Significance and effect size of detected loci.

(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.

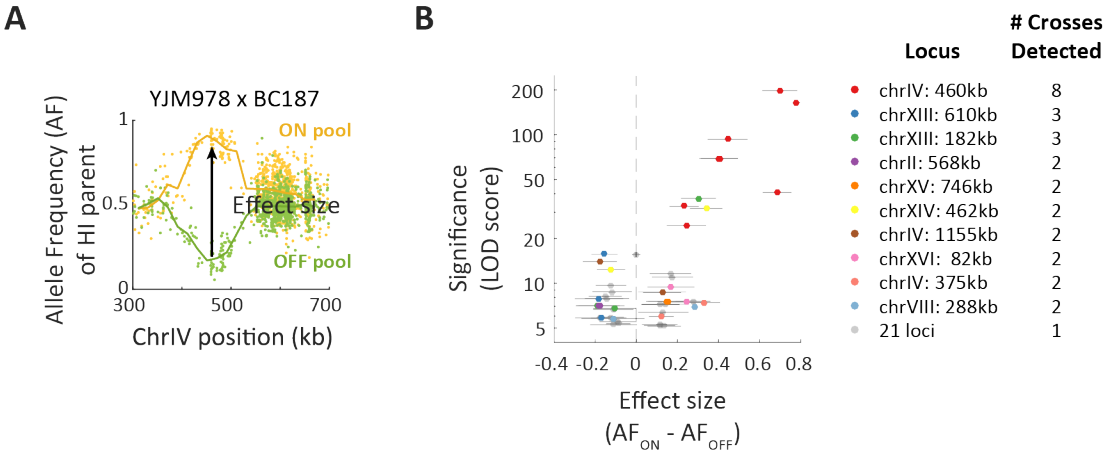


Figure S3.9. Representative YFP induction profiles of *GAL3* allele replacements.

Homologous *GAL3* allele replacement strains were assayed in a gradient of glucose in a background of 0.25% galactose (Figure 3.3A-C). The alleles were assayed in three backgrounds (A) YJM978, (B) BC187, and (C) S288C. (D) The parental strain is shown for comparison.

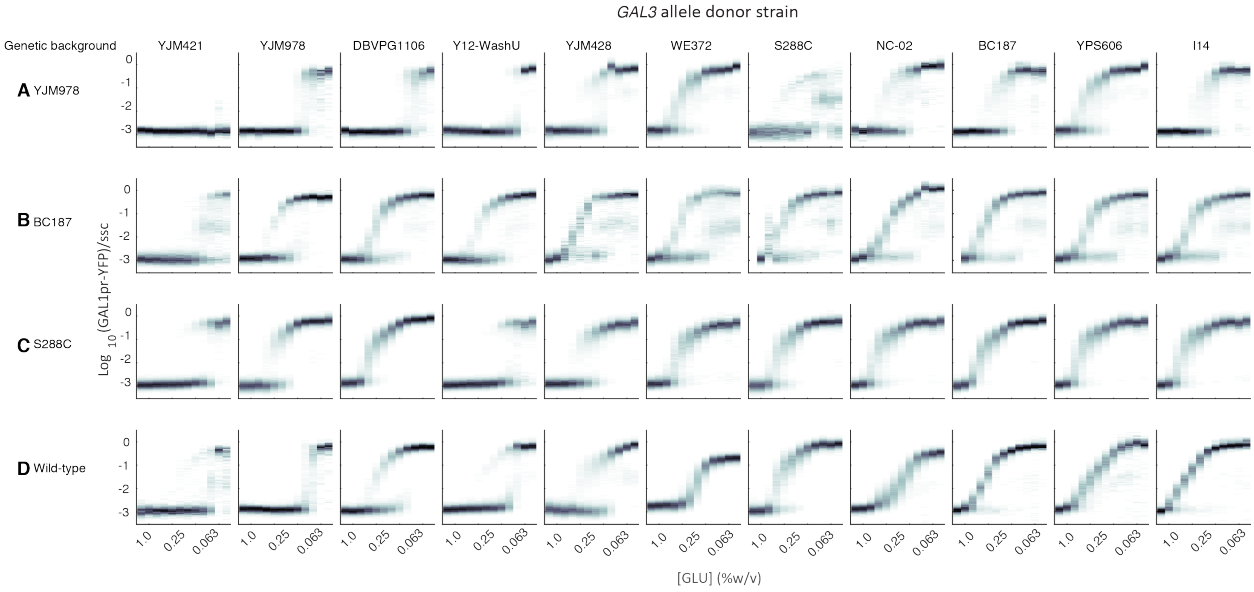


Figure S3.10. Hemizygous hybrids YFP density plots.

Homologous *GAL3* allele replacement strains were assayed in a gradient of glucose in a background of 0.25% galactose (Figure 3.3D). Three different alleles (A) YJM978, (B) BC187, and (C) S288C were assayed in seven genetic backgrounds.

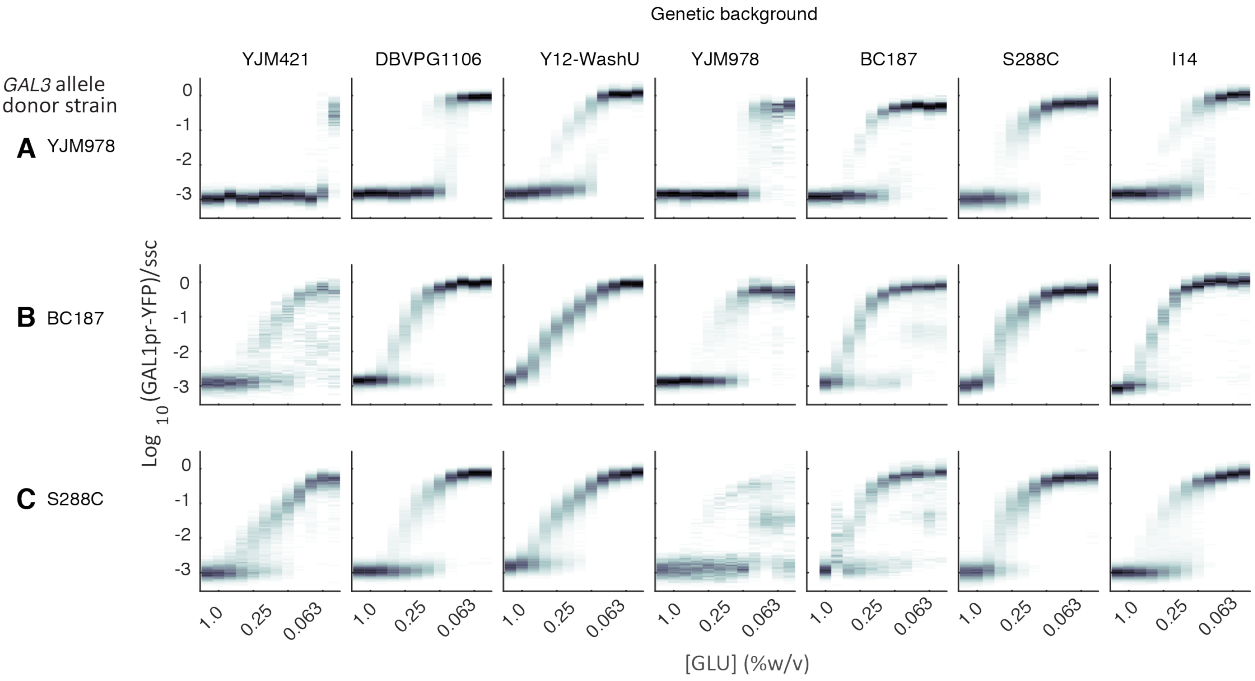


Figure S3.11. Phenotypic variation of hybrid (and hybrid conversion) segregants.

(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).

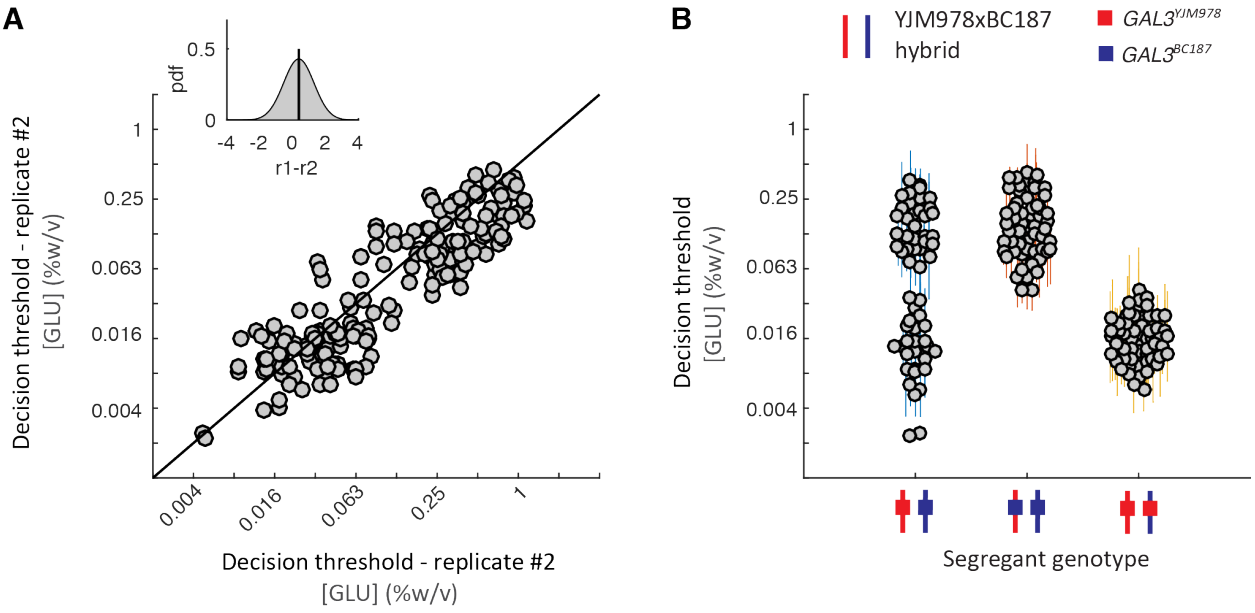


Figure S3.12. Scatter plot of decision threshold versus *GAL1*-YFP steady state expression [35].
Scatter plot of steady state *GAL1* expression levels versus decision threshold of a subset of strains from Figure 3.1. We previously showed that *GAL1* expression levels before the diauxic lag are inversely correlated with the diauxic lag length [35]. We extend that show that the decision threshold is correlated to these *GAL1* expression levels.

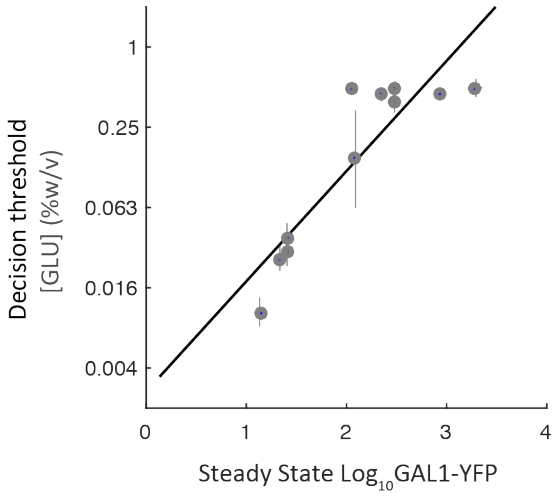
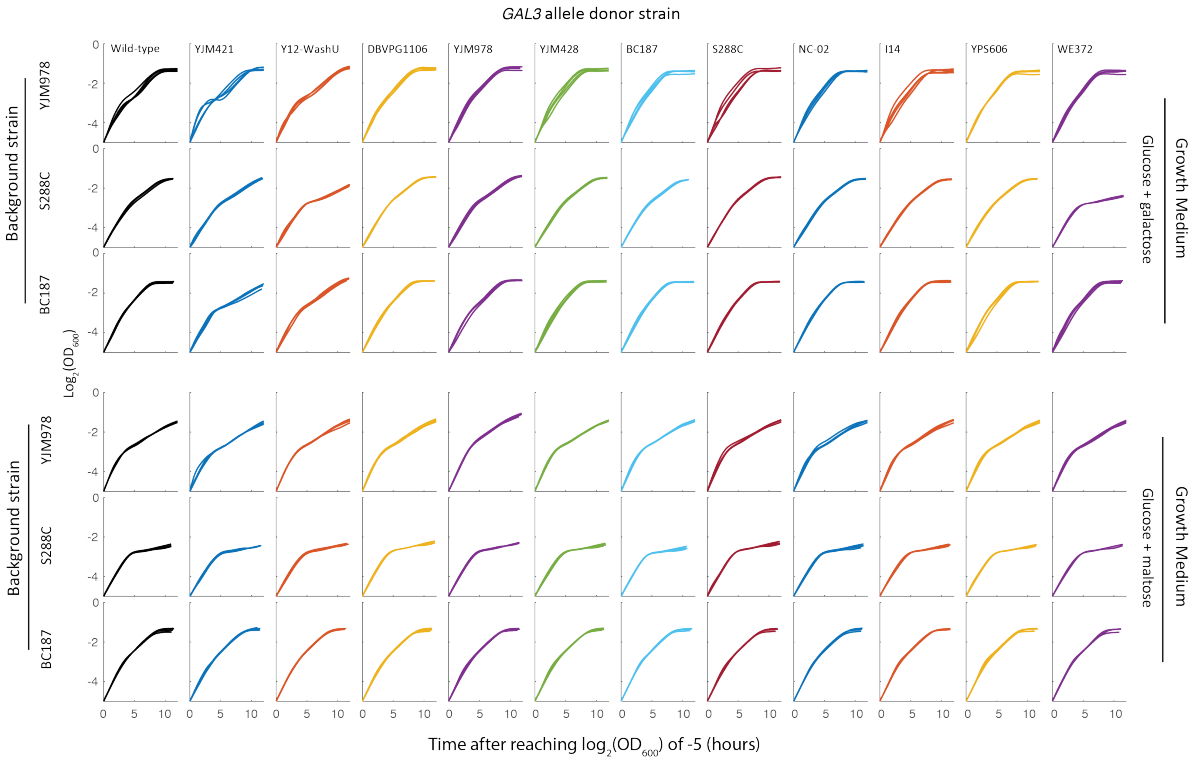


Figure S3.13. Growth curves of *GAL3* allele replacement strains.

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 III.

Supporting Information for Chapter 4

Natural genetic variation can independently tune the induced fraction and induction level of a bimodal signaling response

Figure S4.1. 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 4.1G.

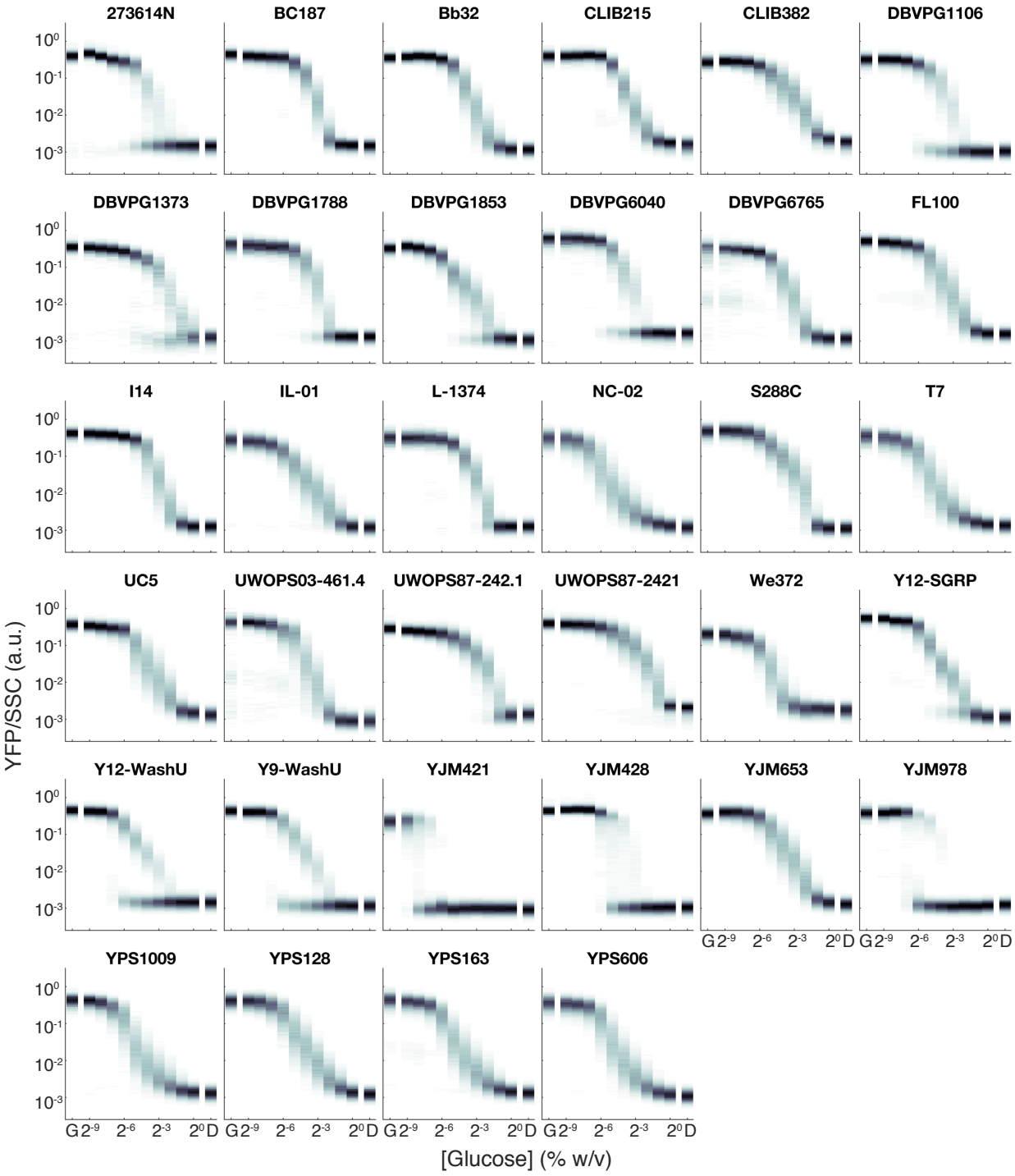


Figure S4.1 (Continued). GAL response phenotypes for 34 natural isolates

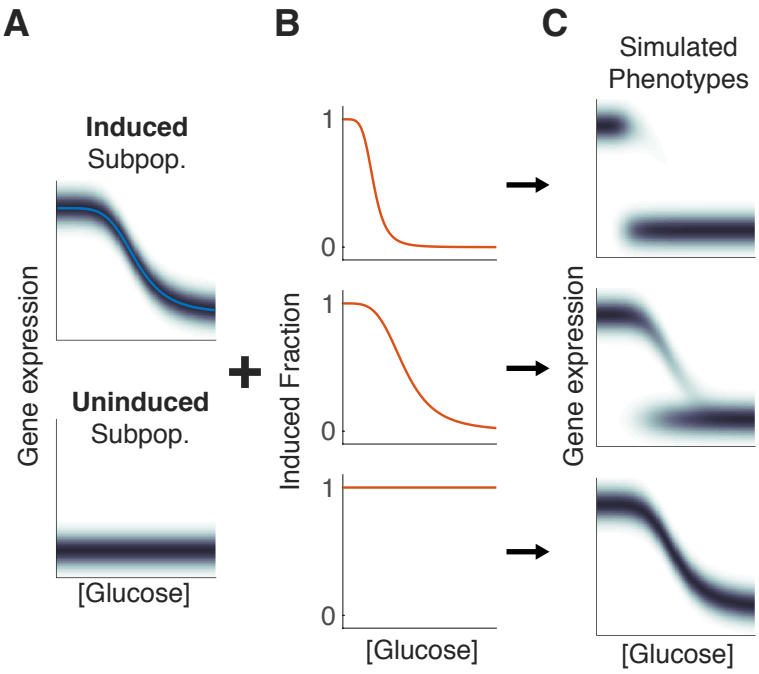


Figure S4.2. Bimodal phenotypes simulated using a subpopulation decomposition framework

(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.

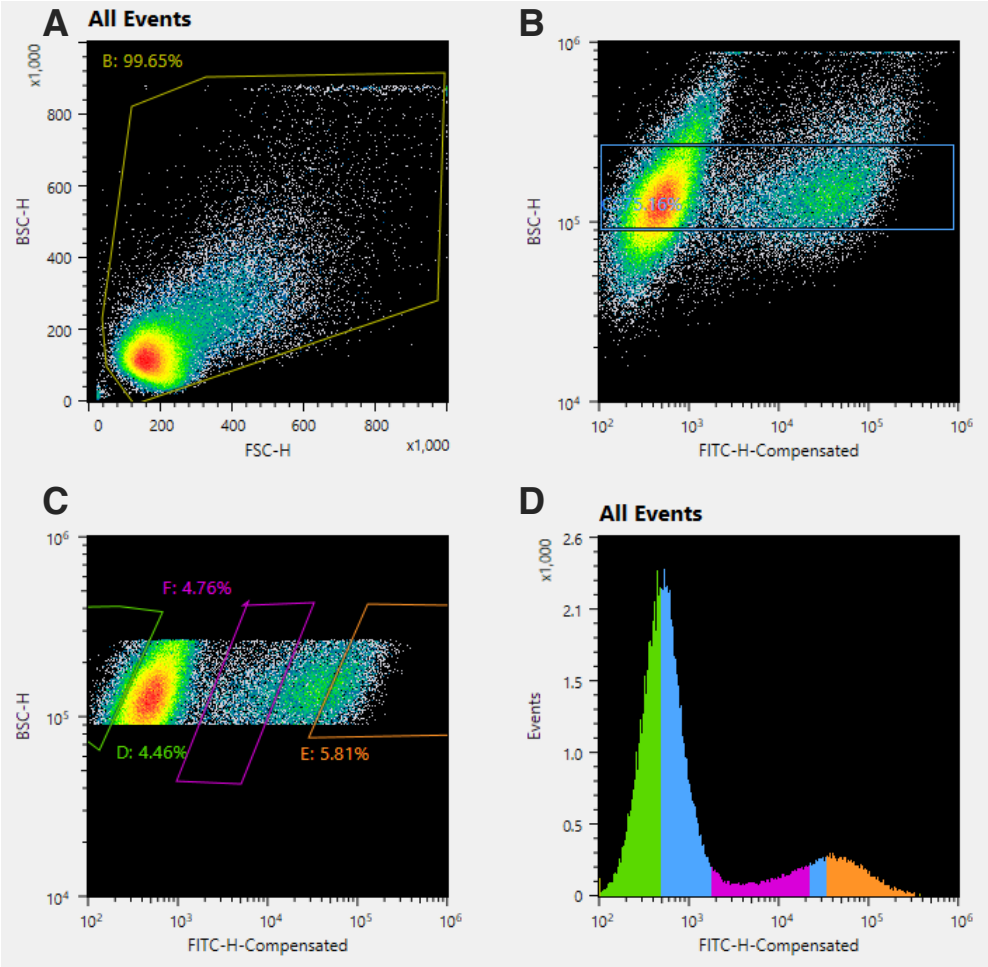


Figure S4.3. Sorting strategy for bulk-segregant analysis

(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.

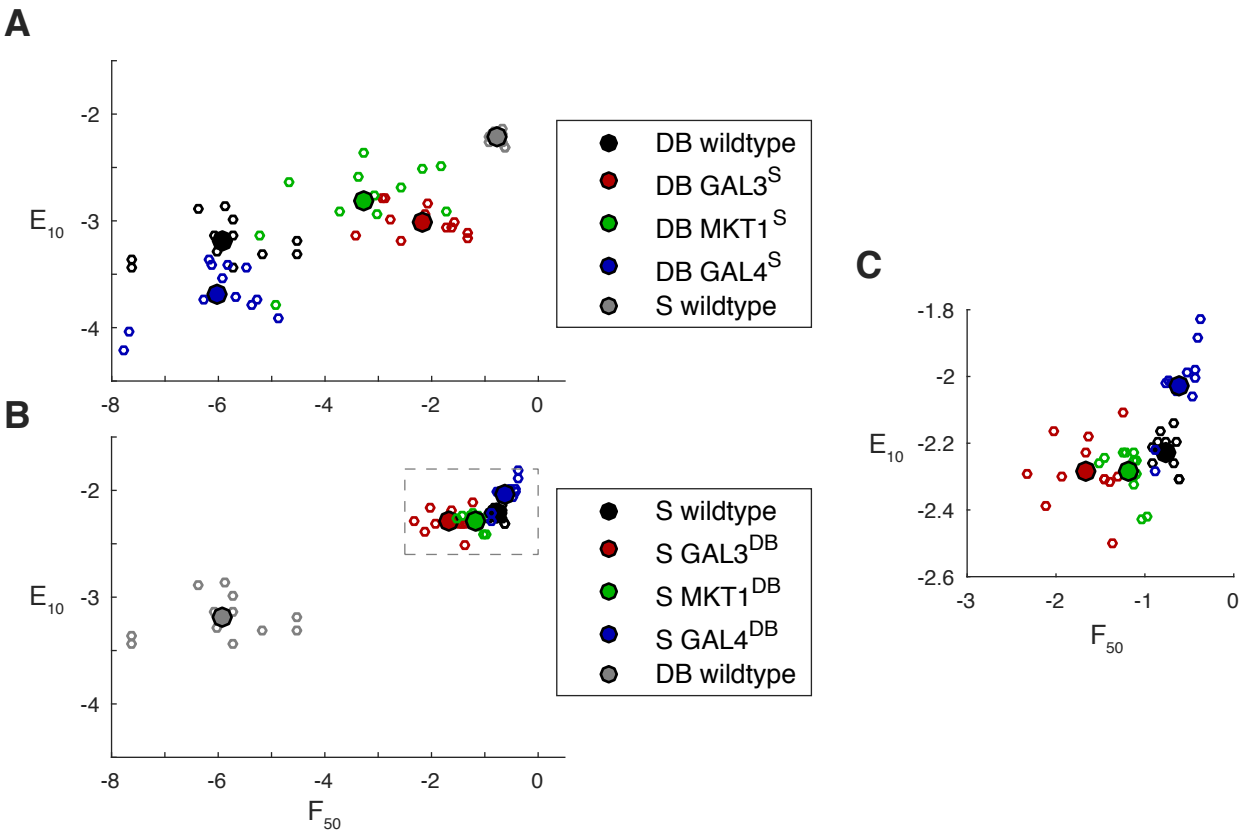


Figure S4.4. Effect of allele replacement of *GAL3*, *MKT1*, or *GAL4* in DBVPG1106 and S288C backgrounds.

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.