UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Genetic determinants of phenotypic evolution in Saccharomyces yeast

Permalink
https://escholarship.org/uc/item/72r8n39k

Author
Roop, Jeremy Roop

Publication Date
2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Genetic determinants of phenotypic evolution in *Saccharomyces* yeast

By

Jeremy Ian Roop

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Rachel B. Brem, Co-chair
Professor John W. Taylor, Co-chair
Professor Jasper Rine
Professor Adam P. Arkin

Summer 2016

Abstract

Genetic determinants of phenotypic evolution in *Saccharomyces* yeast

By

Jeremy Ian Roop

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Rachel B Brem, Co-chair
Professor John W. Taylor, Co-chair

Years of trait mapping studies have uncovered the genetic determinants underlying phenotypic evolution over both short and long timescales in a variety of organisms. A more general understanding of the mechanisms through which evolution produces phenotypic novelty, however, still remains out of reach. Many outstanding questions remain, among them how coding sequence divergence and gene expression divergence respectively contribute to trait evolution, as well as what role rare and large effect genetic variants play in the genetic networks underlying evolving phenotypes. Using *Saccharomyces* yeast as a genetically tractable model system in which to probe these questions, we identify and dissect several examples of trait divergence within and between yeast species. We first investigate an instance of intraspecific variation among yeasts isolated from a single population, finding that large effect rare variants at hypermutable loci are drivers of many common morphological and growth phenotypes. Next, in a comparison of both recently-diverged and long-diverged species, we investigate the role that gene expression variation plays in phenotypic evolution. We find evidence in multiple gene networks for complex and coherent regulatory evolution, and in one instance we succeed in identifying the loci that effect this change, as well as several phenotypic novelties they produce. Collectively, these investigations will contribute to the growing body of literature that describe the genetic mechanisms that underlie phenotypic change and seek, eventually, to achieve a greater understanding of basic evolutionary principles.

TABLE OF CONTENTS

CHAPTER 1 **1**
AN INTRODUCTION TO THE GENETIC DETERMINATES OF EVOLUTION
Background 2
What is the role of regulatory divergence in trait evolution? 2
Does evolution favor large or small effect mutations? 3
Do rare or common alleles underlie common traits? 4
Summary and dissertation outlook 5
References 6

CHAPTER 2 **12**
RARE VARIANTS IN HYPERMUTABLE GENES UNDERLIE COMMON MORPHOLOGY AND GROWTH TRAITS IN WILD *SACCHAROMYCES PARADOXUS*
Abstract 13
Introduction 13
Results 14
Discussion 20
Materials and Methods 22
Acknowledgments 26
Figures 27
Tables 33
Supplementary Figures 34
Supplementarty Tables 50
References 52

CHAPTER 3 **59**
EVOLUTION OF A MEMBRANE PROTEIN REGULON IN *SACCHAROMYCES*
Abstract 60
Introduction 61
Results 61
Discussion 64
Materials and Methods 66
Acknowledgments 70
Figures 71

Tables 74
Supplementary Figures 75
Supplementarty Tables 79
Appendix 84
References 85

CHAPTER 4 **89**
POLYGENIC EVOLUTION OF A SUGAR SPECIALIZATION TRADE-OFF IN YEAST
Abstract 90
Introduction and Results 90
Discussion 93
Materials and Methods 93
Acknowledgments 98
Figures 99
Extended Data Figures 103
Extended Data Tables 108
Supplementary Tables 111
References 113

Chapter 1

An introduction to the genetic determinants of evolution

Background

Evolutionary biologists have long been motivated by a desire to understand the molecular determinants of phenotypic novelties and the resultant diversity of life that exists in the natural world. Driven by academic as well as more applied motivations, questions such as the identity of the mutations that differentiate humans from chimpanzees, domesticated crops from wild ancestors, and individuals susceptible to disease from those less so, have been of great interest. Recent research efforts have succeeded in providing at least partial answers to these questions. We now know, for example, many of the loci involved in the evolution of maize from its ancestor teosinte[1], the location of several regulatory regions underlying human-specific morphological traits[2,3], and the identify of several alleles strongly associated with breast cancer[4].

Given the vast organismal diversity that exists in the world, these investigations as well as other similar efforts represent only a small first step towards a comprehensive understanding of the general relationship between mutations, genes, and phenotypes. After decades of research, we are still largely ignorant as to the number of distinct loci that contribute to most variant human traits, we know even less about the genetic architecture of traits that differ between distinct species, and we have only begun to investigate the role that environment plays in modulating genetic contributions to phenotype. Undoubtedly, recent technological advances that allow for the generation of huge experimental datasets will be a great boon to researchers pursuing answers to these questions, and in numerous instances they have already yielded novel insights. A persistent challenge in these investigations, however, is that we lack a foundational understanding of the molecular processes through which traits vary and diverge [5,6]. For researchers interested in mapping the alleles underlying traits diverged between distant species, for example, a more thorough understanding of when and how evolution proceeds via coding sequence mutations as opposed to regulatory mutations would be of great assistance[7]. Similarly helpful would be an investigation of whether strong selective pressures drive adaptation via large effect mutations whereas weaker selective pressure favors adaptation via mutations of small effect, as has been suggested[6,7]. Fundamental questions such as these provide motivation for research utilizing genetically tractable organisms to advance an understanding of basic evolutionary processes, as is described in the following chapters of this dissertation. To provide additional context for these investigations, several of the most relevant outstanding questions that motivated this work are described below.

What is the role of regulatory divergence in trait evolution?

In 1975, King and Wilson first suggested that much of the phenotypic variation between humans and chimpanzees was driven not by divergence in protein coding sequence, but rather by divergence in gene-expression levels [8]. Their rationale was that the protein coding regions of these two species' genomes were not

sufficiently diverged to explain their markedly different phenotypes, and therefore the more divergent regulatory regions were likely responsible [8].
In the decades since, although evidence has been found to both support and contradict this specific claim, the more general concept that regulatory evolution acts as a significant force in phenotypic evolution has gained widespread support [9-13]. Many groups have identified examples of gene-expression changes involved in the evolution of morphological[14-16], sexual[17], and disease resistance traits and, in several cases, the changes have been shown to be adaptive[18].

This growing list of traits with a basis in gene expression divergence has led some to suggest that, in an expansion of the hypothesis proposed by King and Wilson, regulatory change may be the principal driver of phenotypic evolution, playing a more prominent role than divergence in protein coding regions [19-21]. Arguments in favor of this hypothesis draw on the observations that regulatory DNA is selectively less constrained than coding regions, and that mutations in modular regulatory units may have greater temporal and tissue specificity than those in coding regions [19,21]. And yet, the number of reported examples of trait divergence effected by regulatory changes are vastly outnumbered by those that point to coding sequence changes as the underlying drivers of evolution [7,22]. It is reasonable to assume that ascertainment bias resulting from the relative ease of identifying causative mutations in coding regions is likely responsible, at least in part, for their increased representation in the literature. Nevertheless, the comparatively few examples of regulatory evolution underling phenotypic novelties has led some to question their proposed predominant role in evolution [7,22]. The latter arguments contend that the situation is likely far more complex than a simple dichotomy between evolution driven by regulatory or coding sequence evolution, and that different genetic and biological contexts may favor certain evolutionary paths over others[7]. Thus, additional trait mapping studies that investigate the genetic determinants of a broad set of morphological and physiological traits, both in intra- as well as inter-specific comparisons, will be necessary to advance our understanding of the roles that coding and regulatory divergence play in trait evolution.

Does evolution favor large or small effect mutations?

For much of the 20th century, it was thought that phenotypic evolution was driven by a large number of small-effect mutations scattered throughout the genome that cumulatively produced novel traits [23-25]. This "micromutationist" hypothesis was founded on Fisher's "infinitesimal" model which assumed that alleles contributing small phenotypic effects were less likely to be detrimental to fitness, and thus should predominate in evolution [24,26]. With the advent of genome-wide approaches to mapping the determinants of diverged traits, however, numerous examples were found in which large-effect alleles were major drivers of evolutionary change [27-32]. Kimura's neutral theory of evolution provided an additional line of evidence supporting the importance of large-effect alleles by showing that small-effect mutations were far more likely to be lost by drift than

those with larger effect sizes [33]. Consequently, consensus surrounding the distribution of mutational effect sizes underlying trait divergence shifted away from Fisher's "infinitesimal" model and instead towards Orr's 1998 model which proposed that the distribution of mutational effect sizes was roughly exponential, with few mutations conferring large effects and a large number of mutations contributing to a smaller extent [34-36]. Experimental work has supported Orr's model, raising the possibility that it represents a reasonable standard model for describing the genetic determinants of phenotypic change [35,37].

However, as Orr himself made clear, his model was intended to deal only with a specific evolutionary scenario in which an organism was adaptively evolving towards a single fixed phenotypic optimum with no relevant standing variation present in the initial population [38,39]. Obviously this scenario represents only a small portion of all potential evolutionary paradigms, and it is unclear how the role of standing variation, neutral evolution, and the strength of natural selection will impact the exponential distribution of allele effect sizes that Orr proposed [40]. Additionally, experimental limitations have largely limited genome-scale mapping studies to investigating the genetic determinants of traits that vary within species, rather than between species. As such, the degree to which the infinitesimal or exponential models approximate the genetic architecture of divergence between species remains predominantly unknown. Interspecific evolution over long time scales may expose organisms to more diverse environments that would favor many small effect and putatively less pleiotropic mutations, but minimal experimental evidence exists with which to test this hypothesis [7,41]. Thus, more empirical work that identifies loci and individual nucleotides that are causative of phenotypic evolution under diverse evolutionary paradigms will be necessary before we develop a more comprehensive understanding of the role that classes of mutations play in evolution.

Do rare or common alleles underlie common traits?

Genome wide association studies (GWAS), first pioneered in 2005, have successfully led to the identification of thousands of variants contributing to intraspecific trait variation in both model and non model organisms [42,43]. Traits as diverse as flowering time in *Arabidopsis*[44], cholesterol levels in mice[45], and milk production in cattle[46] have all been investigated by GWAS, leading to the identification of a variety of small and large effect contributing loci. In humans, interest in disease relevant traits has motivated the sequencing of thousands of individual genomes, providing even more statistical power for association studies. As a result, tens and even hundreds of variants have been associated with traits such as diabetes[47], bipolar disorder[48], arthritis[49], inflammatory bowl disease[50] and body mass index[51]. While these examples highlight some of the major successes GWAS have had in identifying significant allelic associations, the cumulative contributions of these loci to the investigated traits, is, in most studies, relatively minimal. Most studies have managed to identify associations explaining less than 10% of the heritable variation of the trait in question, and even a recent

study utilizing an unprecedented number of human subjects to investigate the genetic basis of height managed to find associations explaining only 36% of the heritable phenotypic variation [52,53].

It has been proposed that this so called "missing heritability" results from a variety of factors. Among these are large numbers of very small-effect alleles that are difficult to detect with current sample sizes; structural variants not accounted for by current methodologies; and gene by environment interactions that modulate allelic effects [54]. Significant recent attention has also been given to an additional potential complicating factor, the possibility that traits common to multiple individuals within a population are often effected by distinct rare allelic variants, rather than the same common variants as was initially assumed [55,56]. This so called "common trait, rare variant hypothesis," if shown to apply generally to trait variation, would greatly complicate the identification of causative allelic variants and would necessitate alterations to current experimental and analytical procedures[57].

In human studies in which large sample sizes afford the power to identify rare variant alleles, autism, schizophrenia and epilepsy have all been shown to be associated with rare structural variants [58,59]. Simulations have also been used to support the rare variant hypothesis by suggesting that many common alleles previously identified as associated with disease traits may in fact result from a synthetic association of many distinct rare variant alleles that happen to be linked to a common measured polymorphism [60]. However, several recent studies that sought to directly assess the role of rare variants in human trait variation have failed to find evidence for a significant role for these rare alleles. The synthetic association hypothesis has also received substantial criticism and it seems no consensus on the role of such synthetic interactions will emerge without future experimental work [61-63]. As such, the only conclusions that can currently be made with respect to the contributions of rare and common variants is that it seems both contribute to trait differences within and between species [55]. Future studies employing larger sample sizes, or alternative methodologies such as those that seek to identify variants clustered in candidate biological pathways or at highly mutable loci, will be necessary before a more refined consensus is reached [56,64,65].

Summary and dissertation outlook

Each of the questions described above has been the subject of years of experimental and theoretical work, and it will be decades before researchers begin to arrive at satisfactory answers. It is far beyond the scope of any one investigation to comprehensively address any of these unknowns. The following chapters of this dissertation report the results of investigations that sought to incrementally advance our understanding of the relationship between mutations, genes, and evolving phenotypes. The results of these studies will add to a growing body of literature that explores this relationship in a variety of biological contexts, all of

which will eventually inform a more foundational understanding of evolutionary principles.

The investigations described in the following chapters utilize *Saccharomyces* yeast as a model organism in which biological questions can be asked in a relatively simplified experimental context. In Chapter 2, an investigation into the genetic determinants of common trait variation among *Saccharomyces paradoxus* is described, the findings of which shed light on both the role of rare variants and large effect alleles in phenotypic divergence. Chapter 3 investigates gene expression divergence between the sister species *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*, finding evidence for species specific gene expression differences likely underlying adaptive phenotypes. Chapter 4 extends the investigation of the role of regulatory divergence in trait evolution in a comparison of *Saccharomyces cerevisiae* with *Saccharomyces bayanus*, in this case pinpointing the loci responsible for the expression changes as well describing several resultant phenotypic effects.

References

1. Briggs WH, McMullen MD, Gaut BS, Doebley J. Linkage Mapping of Domestication Loci in a Large Maize Teosinte Backcross Resource. Genetics. 2007;177: 1915–1928. doi:10.1534/genetics.107.076497

2. McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature. 2011;471: 216–219. doi:10.1038/nature09774

3. Indjeian VB, Kingman GA, Jones FC, Guenther CA, Grimwood J, Schmutz J, et al. Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. Cell. 2016;164: 45–56. doi:10.1016/j.cell.2015.12.007

4. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990;250: 1684–1689.

5. Stern DL, Orgogozo V. Is genetic evolution predictable? Science. 2009;323: 746–751. doi:10.1126/science.1158997

6. Rockman MV. THE QTN PROGRAM AND THE ALLELES THAT MATTER FOR EVOLUTION: ALL THAT'S GOLD DOES NOT GLITTER. Evolution. Wiley Online Library; 2012.

7. Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution? Evolution. Wiley Online Library; 2008;62: 2155–2177.

8. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975;188: 107–116.

9. Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, et al. A Neutral Model of Transcriptome Evolution. PLoS Biol. 2004;2: e132. doi:10.1371/journal.pbio.0020132

10. Torgerson DG, Boyko AR, Hernandez RD, Indap A, Hu X, White TJ, et al. Evolutionary Processes Acting on Candidate cis-Regulatory Regions in Humans Inferred from Patterns of Polymorphism and Divergence. Begun DJ, editor. PLoS Genetics. 2009;5: e1000592. doi:10.1371/journal.pgen.1000592.t006

11. Khaitovich P, Tang K, Franz H, Kelso J, Hellmann I, Enard W, et al. Positive selection on gene expression in the human brain. Current Biology. 2006;16: R356–R358. doi:10.1016/j.cub.2006.03.082

12. estro CCN, Yokoi H, Postlethwait JH. Evolutionary developmental biology and genomics. Nat Rev Genet. 2007;8: 932–942. doi:doi:10.1038/nrg2226

13. Wray G, Hahn M, Abouheif E, Balhoff J, Pizer M, Rockman M, et al. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol. 2003;20: 1377–1419. doi:10.1093/molbev/msg140

14. Sylvain Marcellini PS. Two or Four Bristles: Functional Evolution of an Enhancer of scute in Drosophilidae. PLoS Biol. Public Library of Science; 2006;4. doi:10.1371/journal.pbio.0040386

15. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010;327: 302–305. doi:10.1126/science.1182213

16. Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB. Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population. Science. 2009;326: 1663–1667.

17. Shirangi TR, Dufour HD, Williams TM, Carroll SB. Rapid Evolution of Sex Pheromone-Producing Enzyme Expression in Drosophila. PLoS Biol. 2009;7: e1000168. doi:10.1371/journal.pbio.1000168

18. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, et al. A single p450 allele associated with insecticide resistance in Drosophila. Science. 2002;297: 2253–2256. doi:10.1126/science.1074170

19. Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007;8: 206–216. doi:10.1038/nrg2063

20. Rockman MV, Wray GA. Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol. 2002;19: 1991–2004.

21. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008;134: 25–36. doi:10.1016/j.cell.2008.06.030

22. Hoekstra HE, Coyne JA. THE LOCUS OF EVOLUTION: EVO DEVO AND THE GENETICS OF ADAPTATION. Evolution. 2007;61: 995–1016. doi:10.1111/j.1558-5646.2007.00105.x

23. Lewontin RC. The genetic basis of evolutionary change. Columbia University Press, New York; 1974.

24. Fisher RA. The Genetical Theory of Natural Selection. Oxford Univ Press; 1930.

25. Charlesworth B, Lande R, Slatkin M. A neo-Darwinian commentary on macroevolution. Evolution. 1982.

26. Fisher RA. XV.—The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the royal society of Edinburgh. 1919.

27. Abzhanov A. Bmp4 and Morphological Variation of Beaks in Darwin's Finches. Science. 2004;305: 1462–1465. doi:10.1126/science.1098095

28. Hauswirth R, Haase B, Blatter M, Brooks SA, Burger D, Drögemüller C, et al. Mutations in MITF and PAX3 Cause “Splashed White” and Other White Spotting Phenotypes in Horses. Barsh GS, editor. PLoS Genetics. 2012;8: e1002653. doi:10.1371/journal.pgen.1002653.s009

29. Kerje S. The Dominant white, Dun and Smoky Color Variants in Chicken Are Associated With Insertion/Deletion Polymorphisms in the PMEL17 Gene. Genetics. 2004;168: 1507–1518. doi:10.1534/genetics.104.027995

30. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature. 2004;428: 717–723. doi:10.1038/nature02415

31. Ghosh R, Andersen EC, Shapiro JA, Gerke JP, Kruglyak L. Natural Variation in a Chloride Channel Subunit Confers Avermectin Resistance in C. elegans. Science. 2012;335: 574–578. doi:10.1126/science.1214318

32. McGrath PT, Xu Y, Ailion M, Garrison JL, Butcher RA, Bargmann CI. Parallel evolution of domesticated Caenorhabditis species targets pheromone receptor genes. Nature. 2011;477: 321–325. doi:10.1038/nature10378

33. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge Univ Press; 1983.

34. Farrall M. Quantitative genetic variation: a post-modern view. Human Molecular Genetics. 2004;13 Spec No 1: R1–7. doi:10.1093/hmg/ddh084

35. Bell G. The oligogenic view of adaptation. Cold Spring Harbor Symposia on Quantitative Biology. 2009;74: 139–144. doi:10.1101/sqb.2009.74.003

36. Orr HA. The genetic theory of adaptation: a brief history. Nat Rev Genet. 2005;6: 119–127. doi:10.1038/nrg1523

37. 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. doi:10.1038/nature11867

38. Orr HA. Theories of adaptation: what they do and don't say. Genetica. 2005;123: 3–13.

39. Orr HA. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution. 1998;42: 934–949.

40. Olson-Manning CF, Wagner MR, Mitchell-Olds T. Adaptive evolution: evaluating empirical support for theoretical predictions. Nat Rev Genet. 2012;13: 867–877. doi:10.1038/nrg3322

41. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet. 2008;40: 346–350. doi:10.1038/ng.77

42. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308: 385–389. doi:10.1126/science.1107573

43. Li MJ, Wang P, Liu X, Lim EL, Wang Z, Yeager M, et al. GWASdb: a database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res. 2011;40: D1047–D1054. doi:10.1093/nar/gkr1182

44. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, et al. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature. Nature Publishing Group; 2010;465: 627–631. doi:10.1038/nature08800

45. Flint J, Eskin E. Genome-wide association studies in mice. Nat Rev Genet. 2012;13: 807–817. doi:10.1038/nrg3335

46. Olsen HG, Hayes BJ, Kent MP, Nome T, Svendsen M, Larsgard AG, et al.

Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. Animal Genetics. 2011;42: 466–474. doi:10.1111/j.1365-2052.2011.02179.x

47. Polychronakos C, Li Q. Understanding type 1diabetes throughgenetics: advances and prospects. Nat Rev Genet. 2011;12: 781–792. doi:10.1038/nrg3069

48. Smith EN, Koller DL, Panganiban C, Szelinger S, Zhang P, Badner JA, et al. Genome-Wide Association of Bipolar Disorder Suggests an Enrichment of Replicable Associations in Regions near Genes. Gibson G, editor. PLoS Genetics. 2011;7: e1002134. doi:10.1371/journal.pgen.1002134.s013

49. Okada Y, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014;506: 376–381. doi:10.1038/nature12873

50. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008;40: 955–962. doi:10.1038/ng.175

51. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518: 197–206. doi:10.1038/nature14177

52. Wood AR, Esko T, Yang J, Vedantam S, Pers TH. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature. 2014. doi:10.1038/ng.3097

53. Visscher PM, Brown MA, McCarthy MI, Yang J. Five Years of GWAS Discovery. The American Journal of Human Genetics. 2012;90: 7–24. doi:10.1016/j.ajhg.2011.11.029

54. 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

55. Cirulli ET, Goldstein DB. Uncovering the roles of rarevariants in common disease throughwhole-genome sequencing. Nat Rev Genet. 2010;11: 415–425. doi:10.1038/nrg2779

56. McClellan J, King M-C. Genetic heterogeneity in human disease. Cell. 2010;141: 210–217. doi:10.1016/j.cell.2010.03.032

57. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. Am J Hum Genet. 2014;95: 5–23. doi:10.1016/j.ajhg.2014.06.009

58. Stankiewicz P, Lupski JR. Structural Variation in the Human Genome and its Role in Disease. Annu Rev Med. 2010;61: 437–455. doi:10.1146/annurev-med-100708-204735

59. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014;506: 185–190. doi:10.1038/nature12975

60. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB, editors. PLoS Biol. 2010;8: e1000294. doi:10.1371/journal.pbio.1000294.t001

61. Wray NR, Purcell SM, Visscher PM. Synthetic Associations Created by Rare Variants Do Not Explain Most GWAS Results. Flint J, editor. PLoS Biol. 2011;9: e1000579. doi:10.1371/journal.pbio.1000579.s001

62. Goldstein DB. The Importance of Synthetic Associations Will Only Be Resolved Empirically. PLoS Biol. 2011;9: e1001008. doi:10.1371/journal.pbio.1001008

63. Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. PLoS Biol. 2011. doi:10.1371/journal.pbio.1000580.g001

64. Michils G, Tejpar S, Fryns J-P, Legius E, Van Cutsem E, Cassiman J-J, et al. Pathogenic mutations and rare variants of the APC gene identified in 75 Belgian patients with familial adenomatous polyposis by fluorescent enzymatic mutation detection (EMD). Eur J Hum Genet. 2002;10: 505–510. doi:10.1038/sj.ejhg.5200825

65. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, et al. Searching for missing heritability: Designing rare variant association studies. Proceedings of the National Academy of Sciences. 2014;111: E455–E464. doi:10.1073/pnas.1322563111

Chapter 2

Rare variants in hypermutable genes underlie common morphology and growth traits in wild *Saccharomyces paradoxus*

The contents of this chapter are based on the following publication:

Roop, JI, Brem RB. Rare Variants in Hypermutable Genes Underlie Common Morphology and Growth Traits in Wild *Saccharomyces paradoxus. Genetics*, October 2013

Abstract

Understanding the molecular basis of common traits is a primary challenge of modern genetics. One model holds that rare mutations in many genetic backgrounds may often phenocopy one another, together explaining the prevalence of the resulting trait in the population. For the vast majority of phenotypes, the role of rare variants and the evolutionary forces that underlie them are unknown. In this work, we use a population of *S. paradoxus* yeast as a model system for the study of common trait variation. We observed an unusual, flocculation and invasive-growth phenotype in a third of *S. paradoxus* strains, which were otherwise unrelated. In crosses with each strain in turn, these morphologies segregated as a recessive Mendelian phenotype mapping either to *IRA1* or *IRA2*, yeast homologs of the hypermutable human neurofibromatosis gene *NF1*. The causal *IRA1* and *IRA2* haplotypes were of distinct evolutionary origin and, in addition to their morphological effects, associated with hundreds of stress-resistance and growth traits, both beneficial and disadvantageous, across *S. paradoxus*. Single-gene molecular genetic analyses confirmed variant *IRA1* and *IRA2* haplotypes as causal for these growth characteristics, many of which were independent of morphology. Our data make clear that common growth and morphology traits in yeast result from a suite of variants in master regulators, which function as a mutation-driven switch between phenotypic states.

Introduction

A primary goal of modern genetics is to understand the molecular basis of traits that segregate at high frequency in populations. Toward this end, hundreds of studies have sought to map causal genes using tests for allele-sharing among affected but unrelated individuals. Against a backdrop of recent successes in fruit fly and plant populations [1-8], association mapping in many systems has yielded loci that explain only a small part of the variation in a given trait across individuals [9]. The latter challenges have motivated the proposal that the bulk of common phenotypic variation may be attributable to rare, highly penetrant mutations [10-12], including recurrent variation at hypermutable loci [13-15]. As yet, the genetic architecture of most common traits remains unknown, and experimental systems in which to

investigate the principles of common trait variation have been at a premium in the literature.

Saccharomyces yeasts have long been a workhorse of the molecular-genetic research community, with resources now becoming available for analyses of population-level variation [16]. Among the most well-studied phenotypes in the classic yeast literature are cell-clumping and filamentation-like behaviors in certain laboratory strains, which serve as models for fungal pathogenesis and have been subject to elegant molecular dissection [17-20]. These morphologies can arise spontaneously in laboratory and brewing yeast owing to variants in a number of hypermutable genes, including effectors and regulators of budding and adhesion [21,22]. In wild populations, the prevalence of morphological trait variation, and its consequences for growth and fitness, have not been well characterized. Wild yeast have, however, been the focus of phenomic profiling efforts [16,23] which have revealed strain differences in growth phenotypes in hundreds of conditions. Though some genome-scale mapping analyses have been reported [23-25], for the vast majority of growth attributes varying among yeasts, the molecular basis remains to be identified.

We set out to use yeast morphologies as a testbed for the study of common trait variation, and the role of hypermutable loci, in a system of genetically tractable, wild-collected *S. paradoxus* strain backgrounds. Our goal was to establish whether and how differences in hypermutable morphology genes segregate in wild yeast populations, and to what extent they underlie morphology and growth behaviors. The pursuit of these questions led us to the discovery of a suite of hundreds of common traits in wild yeast, and to the mapping of their genetic determinants.

Results

A polyphyletic S. paradoxus clade displays a flocculent and invasive phenotype

We used strains derived from wild isolates of *S. paradoxus* in a screen for flocculation, the formation of macroscopic cell aggregates during growth in liquid medium. Of the 24 strains in the well-defined European population of this yeast [16], eight were flocculent (Figure 1A). These strains formed flocs across a range of cell densities and environmental treatments (Figure S1), in contrast to reports of condition-specific flocculation in *S. cerevisiae* [26-29]. We also surveyed European *S. paradoxus* strains for invasive growth, the ability of yeast colonies to adhere to and penetrate a solid substrate [30-32]. Most strains invaded a solid rich-medium agar substrate to some extent (Figure 1B), again in contrast to the requirement for nutrient limitation often seen in *S. cerevisiae* [33,34]. *S. paradoxus* strains with the most dramatic invasive phenotype were also those flocculating in liquid media, such that the two traits were tightly correlated across the population (Spearman's rank correlation = 0.81, p = 2.14e-6).

To begin to investigate the evolutionary history of these morphology traits, we inferred the phylogeny of European *S. paradoxus* using genome-scale polymorphism data [16]. Surprisingly, flocculent/highly invasive strains were scattered across the phylogeny, and strains collected from neighboring locations, even those from within the same English county park, often had dissimilar morphologies (Figure 1C). Such a pattern was consistent with either of two interpretations. On the one hand, the flocculent/invasive phenotype could have been independently acquired in multiple strains, following descent from a non-flocculent, non-invasive ancestor. Alternatively, the phenotype could have been lost in multiple strains following descent from a flocculent, invasive common ancestor. Under either model, morphological variation in *S. paradoxus* would be a product of independent mutational events in distinct lineages.

Flocculation and invasive growth are linked Mendelian traits

To characterize further the molecular and evolutionary basis of flocculation and invasive growth in *S. paradoxus*, we sought to evaluate the genetic complexity of these traits. For this purpose, we carried out crosses between each flocculent, homothallic European strain and the non-flocculent European strain Z1. Seven European flocculent strains successfully formed hybrid diploids when mated as single cells with Z1. Each such hybrid strain was non-flocculent but showed a degree of invasive growth intermediate between that of its parents (see Figure 3 and Figure S6); thus, flocculation acted as a recessive phenotype and invasive growth as incompletely dominant. For each homothallic hybrid diploid, we induced sporulation, dissected the resulting tetrads, and allowed each recombinant spore to form an isogenic colony of homozygous diploid cells. Over 70% of the progeny formed colonies in each cross, except for the mating between Z1 and the flocculent European strain DBVPG4650, in which 50% of the spores dissected from each tetrad failed to yield viable colonies. For each of the remaining crosses, we cultured progeny strains in liquid culture and observed, in five cases, a 1:1 ratio of flocculent to non-flocculent morphologies (Table 1 and Figure S2). Among the progeny of each such cross, flocculation and invasive growth were coincident (Figure S2 and data not shown), indicating that for a given parent strain, a single variant locus was causal for both traits. Segregation patterns among the progeny of the flocculent, invasive European strain CBS432 suggested a polygenic model and were not investigated further.

We reasoned that if flocculation and invasive growth were a monogenic trait in a given European strain, some or all of the affected strains could share the same causal locus. To test this, we carried out complementation analyses as follows. In a cross between haploid cells of two flocculent, invasive European strains harboring variants at the same causal locus, all progeny are expected to exhibit both morphological phenotypes. In a cross between two haploids harboring causal variants at distinct loci, 25% of progeny will inherit neither causal allele and will exhibit a non-flocculent, non-invasive phenotype. From crosses of pairs of the five flocculent European strains, we inferred two linkage groups: the UK strains Q31.4

and W7 and the Siberian strain KPN3829 showed evidence for a single causative locus, distinct from that shared by a second Siberian strain, KPN3828, and the Italian strain CBS5829 (Table 1 and Figure S3). The evidence for unlinked causal loci in the two Siberian strains was particularly striking given the >99.7% identity of these two genomes, and suggested that the determinants of flocculation and invasive growth among *S. paradoxus* strains could be highly mutable over short evolutionary timescales.

FLO9 and FLO11 are terminal effectors of flocculation and invasive growth

We expected that the variants underlying flocculation and invasive growth in *S. paradoxus* were likely to lie in genes of the adhesion or cell polarity regulatory networks, which in *S. cerevisiae* comprise hundreds of components [19,31]. To streamline a candidate-based search among these genes, we elected to identify terminal effectors that mediated flocculation and invasive growth in *S. paradoxus*. We first tested for the involvement of adhesion proteins of the flocculin (FLO) family by treating liquid cultures of several flocculent European strains with mannose, a known FLO protein inhibitor. We observed a dose-dependent inhibition of flocculation by mannose, suggesting that FLO proteins did indeed play a role in the flocculent phenotype (Figure S4). We next tested directly which members of the FLO family were required for flocculation and invasive growth by knocking out, in turn, each of the *S. paradoxus* orthologs of five known *S. cerevisiae FLO* genes in the flocculent European strain W7. Deletion of *FLO9* abolished flocculation in liquid culture but had no effect on invasive growth on solid media; by contrast, a *FLO11* mutant was fully flocculent but almost entirely non-invasive (Figure 2). Deletion of *FLO1*, *FLO5*, and *FLO10* had no morphological effect (data not shown). We repeated these deletion experiments in the flocculent, invasive European strain CBS5829, which fell into a different linkage group from W7 (Table 1) and thus harbored a distinct causal variant for the morphologies. We also analyzed *FLO* gene deletions in a third flocculent, invasive European strain, KPN3829. The results in each case mirrored those from the W7 strain: *FLO9* was required for flocculation and *FLO11* for invasive growth (Figure S5). We conclude that in the context of multiple distinct genetic variants that drive morphological traits in *S. paradoxus*, *FLO9* is an effector of flocculation and *FLO11* mediates invasive growth.

Variants at IRA1 and IRA2 underlie flocculation and invasive growth

In pursuing the genetic basis of variation in flocculation and invasive growth in *S. paradoxus*, we reasoned that the causal polymorphisms were likely to lie in master regulators upstream of the *FLO* genes. In *S. cerevisiae*, the MAPK and PKA signaling pathways act as primary regulators of *FLO* gene expression, and as central control points for the sensing of and response to environmental conditions [35]. To assess the role of regulators of the MAPK and PKA cascades in control of morphological phenotypes, we carried out candidate-based linkage analyses in crosses between flocculent European strains and the non-flocculent European strain Z1. In a first test of candidate genes in the flocculent European strain W7, only inheritance at *IRA1*, a

negative regulator of Ras1/2 and a homolog of the hypermutable human neurofibromatosis gene *NF1* [13,36], was correlated with flocculation and invasive growth phenotypes among progeny in a cross with Z1 (Figure S2). We likewise detected co-inheritance between *IRA1* and morphology traits in the other strains of this linkage group, and no evidence for linkage at other tested candidate loci (data not shown). As expected, inheritance at *IRA1* was not linked to flocculation and invasive growth in KPN3828 or CBS5829, which had formed a distinct linkage group in complementation analysis (Table 1). Instead, in these two strains our candidate-based linkage analysis detected co-inheritance between both morphological traits and *IRA2*, an inhibitor of Ras1/2 paralogous to *IRA1* (Figure S2 and not shown).

To establish variants in *IRA1* and *IRA2* as causal for morphological phenotypes, we used the reciprocal hemizygote approach [37], in which a gene of interest is deleted in each homolog in turn of a hybrid diploid to create a pair of hemizygotes that differ only at the manipulated locus. We constructed reciprocal hemizygotes at *IRA* genes in hybrid diploids formed by mating each flocculent, invasive European strain to Z1. Phenotyping of these diploids confirmed the *IRA* loci as causal for morphologies in each case: the hemizygote harboring the *IRA1* or *IRA2* allele from the flocculent, invasive parent strain exhibited these morphologies, and the hemizygote bearing the *IRA1* or *IRA2* allele from the Z1 parent was non-flocculent and minimally invasive (Figure 3 and Figure S6). This held true not only for the strains we had analyzed by linkage mapping, but also for the flocculent European strain DBVPG4650, which was refractory to linkage analyses: reciprocal hemizygote experiments revealed *IRA1* to be the causal locus in this strain (Figure S6). We conclude that *IRA1* underlies flocculent and invasive growth traits in strains W7, KPN3829, Q31.4 and DBVPG4650, and *IRA2* underlies these traits in strains KPN3828 and CBS5829, validating our linkage and complementation analyses and highlighting *IRA* genes as a nexus of evolutionary change in this yeast population.

IRA1 and IRA2 variants are loss of function alleles of distinct evolutionary origin

Both Ira1 and Ira2 promote the GTP hydrolysis activity of Ras and facilitate the conversion of Ras1/2 from a GTP-bound (active) to a GDP bound (inactive) state [38]. We hypothesized that the *IRA1* and *IRA2* alleles underlying morphological traits in *S. paradoxus* were likely loss-of-function variants, since such mutations in laboratory strains of *S. cerevisiae* deregulate Ras1/2, leading to hyperactivation of the PKA signaling cascade which in turn promotes flocculation and invasive growth [22,35,39]. In order to test this hypothesis, we disrupted the variant *IRA* coding region in each strain in which we had found these variants to confer flocculation and invasive growth morphologies. Phenotypes of the resulting *IRA* null strains differed only marginally from those of their wild-type parents (Figure S7), making clear that the *IRA* gene variants present in the wild *S. paradoxus* population are complete, or nearly complete, loss-of-function alleles. Additionally, disruption of either *IRA1* or *IRA2* in the non-flocculent Z1 strain resulted in a flocculent and invasive phenotype, confirming that a loss-of-function mutation in either gene is sufficient to confer these traits in *S. paradoxus* (Figure S8).

To investigate the evolutionary history of the *IRA1* and *IRA2* alleles in European *S. paradoxus*, we used Sanger sequencing to fill gaps and confirm polymorphisms in publicly available DNA sequence data [16], and we inferred gene trees for each *IRA* locus. The results revealed no consistent relationship between variant alleles in a given *IRA* gene; instead, the strains whose *IRA* gene alleles we had identified as causal for morphological phenotypes formed polyphyletic clades, reflecting the likely origin of each allele through distinct mutational events (Figure 4). Gene trees inferred from promoter regions of either gene gave similar results (Figure S9). We observed no derived alleles common to the flocculent/invasive strains and absent from the remainder of the population, further arguing against a shared evolutionary origin of the morphologies (Figure S10). Taken together, our mutational and sequence analyses supported a model in which independently derived, loss-of-function mutations in *IRA1* and *IRA2* have led to the convergence of six *S. paradoxus* strains on flocculent and invasive growth phenotypes.

Hundreds of fitness traits are associated with IRA1 and IRA2 variation

Given that the target of *IRA1* and *IRA2*, yeast Ras, is central to metabolic and stress response signaling processes, we hypothesized that variant alleles of these genes could have wide-ranging pleiotropic effects. To investigate this, we used publicly available fitness measurements of strains derived from European *S. paradoxus* isolates in 199 growth conditions [23]. For each growth condition, our analyses considered lag time prior to exponential growth, doubling rate during exponential growth, and final cell density at the end of the culture, for a total of 592 environment-parameter combinations. In each case, we tested for a significant difference in fitness between European strains whose flocculent and invasive morphologies we had established to be the result of variation at *IRA1* or *IRA2*, and non-flocculent, non-invasive European strains. This analysis identified 371 cases in which a growth trait was significantly associated with mutations in *IRA1* or *IRA2* (false discovery rate < 0.05; Figure S11 and Dataset S1). Associations were apparent in a variety of conditions, including alternative carbon and nitrogen sources and toxin treatments. Flocculent and invasive strains bearing variants in *IRA1* and *IRA2* had advantages with respect to some growth parameters and conditions, and defects in others (Figure S11). Thus, *IRA1* and *IRA2* variants were associated with hundreds of differences in growth traits, implicating these loci as global correlates of phenotype across *S. paradoxus* strains.

To validate the role of *IRA1* and *IRA2* mutations in growth phenotypes across *S. paradoxus*, we used the phenotypic profiles in Figure S11 to identify traits likely to be affected by *IRA* gene variants, and we evaluated these predictions in reciprocal hemizygote assays. We grew reciprocal hemizygote strain pairs interrogating the *IRA* genes, and the respective parent strains, in eight environmental conditions and measured growth parameters in each case. The results revealed extensive differences between the hemizygotes of strain pairs (Figure 5, Figure S12, and Dataset S2), confirming sequence variation at *IRA1* and *IRA2* as causal for the

majority of growth traits tested. In most cases, the strains of a reciprocal hemizygote pair mirrored differences between the respective parents, and the effect of an *IRA* gene variant in one hemizygote pair usually phenocopied the effects in other pairs (Figure 5 and Figure S12). The most marked benefits of *IRA1* and *IRA2* mutations were to growth rates (Figure 5), and the most uniformly deleterious effects were on cell densities reached by cultures after nutrient exhaustion (Figure S12). These results establish variants in *IRA1* and *IRA2* as toggles between distinct phenotypic states in *S. paradoxus*, underlying growth benefits and disadvantages in a range of conditions and genetic backgrounds.

Growth traits in strains bearing IRA gene mutations can be independent of morphology

Given that cell clumping can affect growth and viability [22], we asked whether variation in growth traits across *S. paradoxus* could be attributed to mechanisms independent of morphology. For this purpose, in each of three flocculent European strains bearing variants in *IRA1* and *IRA2*, we eliminated flocculation by mutating the effector *FLO9*, and we compared growth in these non-flocculent strains to isogenic controls across a panel of conditions. The results identified growth traits fully or partly independent of flocculation (Figure 6, Figure S13, and Dataset S3). For example, in all strains tested, *IRA* variant alleles compromised the maximal cell density attainable by cultures at stationary phase in the presence of 2% arabinose (Figure S12), and this fitness defect was unaffected by *FLO9* mutation (Figure 6 and Figure S13). Likewise, in cultures using melibiose as a carbon source, *IRA* variant alleles conferred an advantage in several growth parameters (Figure 5 and Figure S12) which persisted, though in some cases with reduced magnitude, in non-flocculent *FLO9* mutant backgrounds (Figure 6 and Figure S13). Interestingly, in a few conditions the growth effects of *FLO9* mutation differed across strains, likely reflecting the action of modifier loci segregating in the population (Figure S13). We conclude that flocculation contributes to, but cannot fully explain, the growth profile of strains bearing variant alleles at *IRA1* and *IRA2*, highlighting the pleiotropic effects of these polymorphisms on yeast growth and stress resistance.

Variant IRA1 and IRA2 alleles are present in wild yeast

In light of previous reports of rapid evolution of *IRA1* and *IRA2* in laboratory *S. cerevisiae* [22], we aimed to trace the history of these loci during the processing of *S. paradoxus* strains from wild collection to laboratory derivatives. For each of the six wild-derived strains in which we had identified *IRA* gene variants, we obtained the originally isolated wild progenitor diploid and cultured it with minimal passaging. Five of these cultures (progenitors of CBS5829, DBVPG4650, KPN3828, KPN3829 and Q31.4) displayed a flocculent and invasive growth phenotype, confirming the prevalence of these phenotypes in wild yeast populations (Figure S14). The sixth strain, the progenitor of W7, was non-flocculent and minimally invasive (Figure S14). Progeny from sporulation of this strain were non-flocculent and non-invasive (data not shown), ruling out heterozygosity of a loss-of-function allele in the wild

diploid and establishing that the mutation had been acquired after introduction into the laboratory in this case.

To investigate the genetic basis of morphological traits in wild, flocculent progenitor strains, we mated a haploid from each of three such strains to the non-flocculent strain Z1, with the remaining two, KPN3828 and Q31.4, refractory to repeated mating attempts. In each of the three hybrids successfully generated between wild progenitors and Z1, sporulation revealed a pattern of Mendelian inheritance of the flocculation phenotype (Figure S15). Reciprocal hemizygote experiments implicated variation at *IRA1* in the progenitor strain of KPN3829 and *IRA2* in the progenitor strain of CBS5829, consonant with our analyses of their laboratory derivatives (Figure S6 and Figure S16). Interestingly, in reciprocal-hemizygote analysis of the mating between Z1 and the progenitor of DBVPG4650, we validated *IRA2* as causal for flocculation, in contrast to our discovery of *IRA1* as causal in the monosporic laboratory derivative of this strain (Figure S6 and Figure S16), which could reflect heterozygosity at both *IRA1* and *IRA2* in the wild DBVPG4650 diploid. Taken together, these data make clear that flocculation and invasive growth segregate among wild *S. paradoxus* and can be attributed to variation at *IRA1* and *IRA2*, underscoring the relevance to yeast evolution of our observations in laboratory derivative strains.

Discussion

The search for the molecular basis of common trait variation dominates the modern study of genetics. In *Saccharomyces* yeasts, although elegant linkage studies have mapped phenotypic differences between pairs of genetically distinct strains [37,40-45], the determinants of common phenotypes, and the evolutionary forces that underlie them, are less well understood [23,46,47]. In this work we have characterized a panel of flocculent and invasive, but otherwise unrelated, European *S. paradoxus* strains. Our results showed flocculation and invasive growth to be linked Mendelian traits, caused by loss-of-function alleles of either *IRA1* or *IRA2* which are of distinct evolutionary origin. Because each independent allele of *IRA1* and *IRA2* largely phenocopies the others, *in toto* they represent a suite of rare variants which underlie common traits in the *S. paradoxus* population.

Evolvability of IRA1 and IRA2

The independent loss-of-function alleles of *IRA1* and *IRA2* we have uncovered in *S. paradoxus* echo previous reports of mutations in these genes in wild and laboratory *S. cerevisiae*, including some acquired during the course of experimental culture [22,45,48,49]. The recurrent focus on *IRA* gene variants in the yeast literature *a priori* could reflect ascertainment bias, given the dramatic phenotypes of these alleles. However, in unbiased sequence analyses, we detected elevated nucleotide diversity at *IRA1* and *IRA2* in most *S. paradoxus* populations (Table S2), consistent either with hypermutability or relaxed selection at these loci. Because the striking fitness consequences of *IRA* gene mutations render the latter model unlikely, the

most compelling interpretation of results in the field is as evidence for hypermutability at *IRA1* and *IRA2*. These findings parallel the pattern of widespread *de novo* variation in the human homolog, *NF1*, which underlies susceptibility to neurofibromatosis [13,36,50], raising the possibility that the mechanism for hypermutability at these loci may be shared between yeast and human. Our work leaves open the question of which nucleotide changes, among the dozens segregating in the population, underlie growth and morphology behaviors. Sequence analyses identified a nonsense mutation in *IRA1* in both the progenitor and the laboratory derivative of Q31.4 (Figure S10), representing a plausible causal variant in this background. Likewise, in several other cases, *IRA* genes of wild flocculent progenitor strains harbored one or more private non-synonymous or regulatory mutations, which may underlie the observed growth and morphology phenotypes (Figure S10).

If *IRA1* and *IRA2* are prone to mutation, are variants at these loci evolutionarily relevant? Our work makes clear that *IRA* gene polymorphisms confer fitness benefits and disadvantages in a range of environmental conditions. And as a complement to landmark studies of the advantages of morphological traits in *S. cerevisiae* [51,52], our results show that growth traits in *S. paradoxus* strains harboring mutations at the *IRA* loci can be independent of cell aggregation. The strong and pleiotropic effects of *IRA* gene mutations set them apart from rare, recessive growth defects [25,53] and weakly deleterious variants [16,54,55] thought to be maintained in yeast populations as a consequence of their peculiar demography. Instead, *IRA1* and *IRA2* exhibit the features of contingency loci [56], whose hypermutability serves as a bet-hedging strategy. Under this model, constantly arising mutations at the *IRA* genes would provide short-term benefits in some environments, and be eliminated or compensated for when conditions change [22]. A specialist role for mutants at the *IRA* genes is further suggested by the loss of *IRA1* function arising under laboratory selection for tolerance to low glucose [48]. In contrast to the classic understanding of diversifying selection [57], the distinct wild alleles we have catalogued at the *IRA* genes all yield very similar, convergent phenotypes. Thus, in addition to the phenotypic switches driven by stochastic noise in biochemical processes and by protein aggregation [58-60], our work establishes rapid evolution of master regulators as a mechanism of phenotype switching in yeast.

Mechanisms of IRA1 and IRA2 variant phenotypes

In investigating morphologies of *S. paradoxus*, we identified *FLO11* as an effector of invasive growth, dovetailing with the roles of this gene in multicellularity behaviors of *S. cerevisiae* [61-63]. By contrast, our dissection of flocculation in *S. paradoxus* implicated *FLO9* as the major effector, a gene whose functional role in wild *S. cerevisiae* is unknown [64]. The emerging picture from our work and that of others [21,51,65,66] is one in which the activity of flocculin genes, and the traits they underlie, are highly variable on short evolutionary timescales. Our findings make clear that morphology traits driven by the *FLO* genes have fitness benefits in certain

growth conditions, plausibly mediated by the increased efficiency of sugar uptake observed in clumping yeast cultures [52] and protection of cells in the interior of flocs from soluble chemical stressors [51].

Additional fitness effects of *IRA* gene variants are likely to be mediated by other downstream targets of the cAMP/PKA pathway. We expect that poor growth and low cell density in stationary phase, which we observe in *S. paradoxus IRA1* and *IRA2* variants in many stress conditions, result from the repression of protective stress responses seen in hyperactive PKA mutants [67-69] as well as the disadvantages of flocculation in late phases of growth [51]. Some growth advantages that we have mapped to *IRA* gene variants, including resistance to cobalt chloride, a known hypoxia mimic [70], are likely tied to the increased respiratory capacity of hyperactive Ras mutants [71]. Our findings of rapid growth by *IRA1* and *IRA2* variants in some conditions after transfer from nutrient-poor stationary phase may be a consequence of the metabolic program activated by unregulated Ras1 and Ras2, which in wild-type cells is associated with starved cultures upon addition of glucose [72,73]. It is also tempting to speculate that the release of checks on cell growth may also underlie some of the advantages we observe in *IRA* gene variants, at least on the relatively short timescales we analyze here.

Mapping rare variants that underlie common traits

Whether rare variants underlie common traits is one of the most controversial questions in the current genetics literature [11,12,74], with a few landmark studies implicating hypermutable loci as the determinants of common phenotypes [13-15]. Our work establishes the hypermutable *IRA* genes as drivers of a broad swath of yeast phenotypes. Such a central role for mutational hotspots in yeast trait variation would be consistent with the evidence for allelic series at other loci in recent mapping studies [43]. In the ongoing search for such loci in populations, our results underscore the power of a candidate-based approach, drawing on knowledge of gene networks and population-genomic data to pinpoint rare variants with sizeable fitness consequences. With the increasing availability of sequence compendia and functional-genetic resources, this strategy holds promise for application to common traits in many organisms.

Materials and Methods

Yeast strains

Strains used in this study are listed in Table S1. With the exception of wild progenitor strains, all were homothallic diploid strains obtained from the *Saccharomyces* Genome Resequencing Project (SGRP) collection [16] or engineered from these SGRP strains. All data was generated with these laboratory derived strains except for the results presented in Supplementary Figures 10, 15, and 16; the

latter used homothallic progenitor strains were obtained directly from either V. Koufopanou or G. Liti.

Hybrid strains were generated by sporulation and tetrad dissection of two strains immediately followed by single cell mating of progeny from each, and confirmed by PCR amplification of regions containing strain-specific indels. Homozygous deletion strains were constructed by transformation with a KanMX cassette amplified from the pUG6 plasmid [75], followed by sporulation, tetrad dissection, and selection. Reciprocal hemizygotes were constructed by transformation of the KanMX cassette into a given hybrid background, followed by selection and PCR at the appropriate locus to determine which allele of the target gene had been deleted. Deletion of *FLO10* and *FLO11* in all backgrounds, as well as *IRA1* in the W7 X Z1 background, was accomplished by replacing the entire ORF with the KanMX cassette. Engineered mutation of *IRA1* and *IRA2* in all other backgrounds, and *FLO9* in all backgrounds, was accomplished by replacing 2000-3000 bp of the 5' end of the respective gene, including the start codon, with the KanMX cassette.

Invasive growth

Cultures of each strain were grown to stationary phase, followed by normalization to an optical density of 2.0. 4µl of a 1:100 dilution of this culture was then spotted onto a YPD plate (2% dextrose, 2% bacto-peptone, 1% yeast extract, 2% agar) and allowed to grow at room temperature for five days. Plates were photographed, washed under a stream of distilled water such that all non-adherent cells were removed, and photographed again. Custom Python scripts were used to quantify invasive growth from digital images by calculating the proportion of the original colony that remained after washing.

DNA sequence analysis

For the species tree in Figure 1, whole-genome alignments of laboratory derivatives of *S. paradoxus* strains were downloaded from the Saccharomyces Genome Resequencing Project [16] and accessed using the alicat.pl script. Sites with an error probability > 0.0001 as inferred by [16] were excluded from analysis. For species phylogenies, 19,695 polymorphic sites across *S. paradoxus* from these published data were used to infer a tree using FastTree [76], which was visualized using FigTree. These publicly available sequences were used, in Table S2, as input into custom Python scripts to calculate genetic identity between strains KPN3828 and KPN3829, and to calculate nucleotide diversity [77] within each *S. paradoxus* population, as

$$\pi = \sum_{ij} x_i x_j \pi_{ij}$$

where x_i and x_j are the respective frequencies of the *i*th and *j*th sequences in the population and π_{ij} is the number of nucleotide differences per nucleotide site between the *i*th and *j*th sequences.

In Figure 4 and Figure S9, inference of gene trees of the coding regions of *IRA1* and *IRA2* in laboratory derivatives of *S. paradoxus* strains, and trees of the regions 800bp directly upstream of the start codon of each of these loci used a combination of published and amended sequence data as follows. We used additional Sanger sequencing to fill gaps and confirm polymorphisms in the *IRA1* orthologs of laboratory derivatives of KPN3829 and W7, and the *IRA2* orthologs of laboratory derivatives of KPN3828 and CBS5829, from [16]. Reads were assembled with the PhredPhrap script [78]. For other strains analyzed in gene trees in Figure 4, additional high-coverage genomic sequences used to confirm polymorphisms in *IRA1* and *IRA2* were downloaded from ftp://ftp.sanger.ac.uk/pub/users/dmc/yeast/SGRP2. The complete set of gene sequences was used as input into FastTree as above.

For analysis of original isolates of Q31.4, KPN3829, KPN3828, DBVPG4650, and CBS5829 in Figure S10, we generated complete sequence data of *IRA1* and *IRA2* coding regions, and the regions 800 bp upstream of each of these two genes, by Sanger sequencing, and we visualized the results with Jalview [79].

Linkage analysis

For each strain, marker SNPs were identified in *IRA1* or *IRA2* and strains were genotyped by Sanger sequencing of a region containing the marker. Genotypes at a given locus in a given cross were used in linkage analysis as follows. Two progeny were obtained from each of four tetrads, comprising two flocculent pairs and two non-flocculent pairs. Among the flocculent pairs, the probability p_{floc} of observing by chance *a* pairs sharing the flocculent parent's allele, *b* pairs sharing the non-flocculent parent's allele, and *c* pairs with distinct alleles was calculated as $p_{floc} = n!/(a!b!c!)[(1/6)^a(2/3)^c(1/6)^b]$, and the chance probability $p_{non\text{-}floc}$ of observing a given pattern of allele-sharing among non-flocculent pairs was calculated analogously. The final probability of the observed allele-sharing under the binomial null was taken as the product of p_{floc} and $p_{non\text{-}floc}$.

Quantification of flocculation

Quantification of flocculation in the presence of mannose was conducted for a given strain as in [64] with several modifications as follows. A pre-culture was grown overnight in liquid medium and then pipetted vigorously to disrupt flocs. 0.6mL of this culture was then added to a well containing 0.6mL YPD and, separately, to a well containing 0.6mL of 0.5M EDTA, in a 24-well plate. The plate was shaken at room temperature for 15 minutes to settle flocs at the bottom of the well, after which 200µL of culture media was pipetted off the top of each well. Optical density at 600 nm (OD600) of this 200µL aliquot of media was measured, and flocculation

ability of a given strain was calculated as 1-(OD600 in YPD/OD600 in EDTA). The final measure of the degree of flocculation for each strain in a given condition was calculated as the mean of three such assays from the same overnight culture.

Association analysis of growth fitness data

Fitness data for *S. paradoxus* strains was downloaded from [23]. These values included fitness measurements for 5 of the strains we had determined to harbor variant *IRA* gene alleles and 15 non-flocculent strains. For each of the 592 environment-parameter measurements in the data set, fitness of the strains bearing variant *IRA* genes was grouped together and compared to fitness of the non-flocculent strains using a Wilcoxon rank-sum test. At a given p-value threshold t, we estimated the expected number of false positives n_f as the product of p_0 and the number of tests, with the false discovery rate (FDR) then calculated as n_f divided by the number of true tests with p-value < p_0; we set p_0 to attain an FDR of 0.05.

Fitness profiling

A pre-culture of each strain was grown for 24 hours at room temperature in complete synthetic media (CSM; 0.67% yeast nitrogen base without amino acids, 0.079% complete supplement mixture) supplemented with 2% glucose. Each strain was back-diluted to an OD of 0.02 in a 96-well plate containing 150μL of a given treatment in CSM. Plates were vigorously shaken for 10 seconds and placed in a VersaMax MicroPlate reader (Molecular Devices). OD600 was measured every ten minutes for 72 hours of growth at a temperature of 27°C without shaking. Each experiment that characterized a flocculent European strain included on the plate a replicate culture of the non-flocculent European strain Z1, which was used to normalize growth measurements of the respective flocculent strain in Figure 5 and Figure S12.

From a given time-course of OD600 measurements, custom Python scripts were used to quantify lag time, growth rate, and density change parameters as described in [23] with minor modifications as follows: the growth rate parameter was calculated in a sliding window as the slope of the line of best fit between 10 consecutive growth rate measurements corresponding to 100 minutes of growth. The top slope was discarded and a mean was taken of the second to the fourth highest slopes. Population doubling time was calculated as the ratio of ln(2) to this mean as in [23]. Each fitness parameter value was calculated as the average of at least two replicate wells on the same plate inoculated from the same pre-culture well. For a given parameter and condition in Figure 5, Figure S12, and Dataset S2, significance of the effect of *IRA* variants was assessed in parental strains by comparing the set of measurements of flocculent European strains to the set of all replicates of Z1 using the Wilcoxon rank-sum test, and for a given parameter, p-values were corrected for multiple testing across conditions by the Bonferroni method. Separately, for reciprocal hemizygotes, we estimated analogously the significance of differences between the set of measurements across all strains with

variant alleles and the set of strains with wild-type alleles. In Figure S14 and Dataset S3, for a given parameter and condition, the set of measurements across all strains with wild-type *FLO9* was compared to the set of measurements in strains mutant for *FLO9*, using the Wilcoxon rank-sum test, and for a given parameter, *p*-values were corrected for multiple testing across conditions by the Bonferroni method. In each of these analyses, results were comparable using a paired Wilcoxon test (data not shown).

Acknowledgements

We thank Jasper Rine and Sarah Bissonette for assistance with automated growth assays; Gianni Liti for generously providing sequence data before publication; Vassiliki Koufopanou, Hana Lee, Perihan Saygin and Jonas Warringer for helpful discussions; Gianni Liti and Vassiliki Koufopanou for strains; and Chris Ellison, Leonid Kruglyak, Hilary Martin, and Joshua Schraiber for insightful comments on the manuscript.

Figures

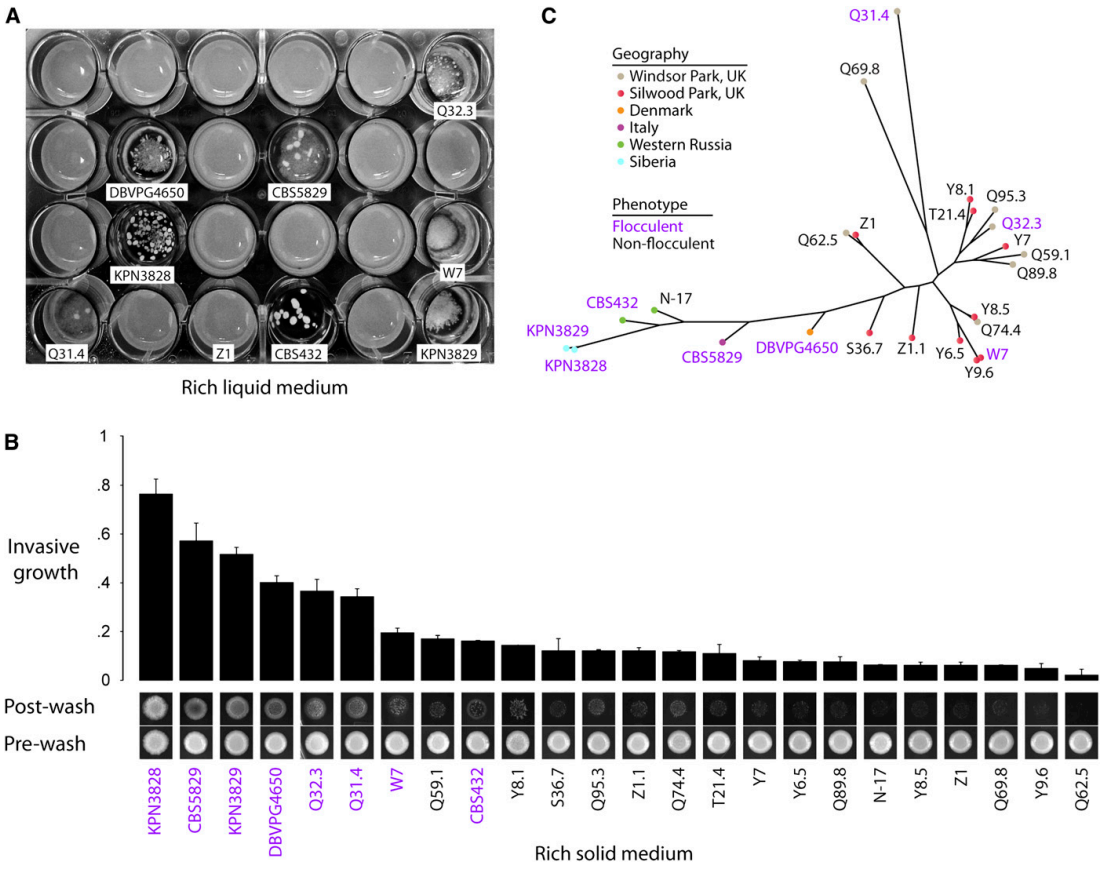


Figure 1. Flocculation and invasive growth vary across European *S. paradoxus*. (A) Photograph of cultures of European *S. paradoxus* strains after overnight growth in rich liquid yeast peptone dextrose (YPD) medium containing 2% glucose. The eight flocculent strains and the non-flocculent strain Z1 are labeled with identifiers as in [16]. (B) Photographs at bottom are of invasive growth assays: a colony of each strain was grown on YPD solid medium for five days and then photographed before ("pre-wash") and after ("post-wash") removal of non-adherent cells by a water wash. At top, each bar height reports the ratio of cell density in colonies after and before a water wash as a mean of two replicate colonies; error bars represent one standard deviation. (C) Maximum-likelihood genome tree of European *S. paradoxus*. Branch lengths are proportional to the number of segregating sites that differentiate each pair of strains. In (B) and (C), identifiers of flocculent strains are in purple.

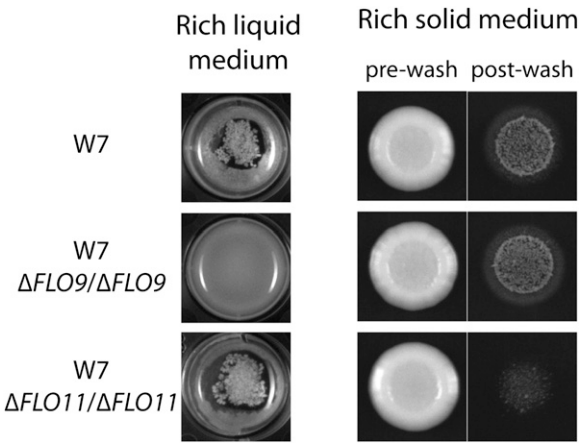


Figure 2. *FLO9* and *FLO11* are terminal effectors of flocculation and invasive growth. Each row reports morphologies of one homozygous derivative of the flocculent, invasive European strain W7. Δ, engineered loss-of-function allele. Left photographs show overnight cultures in rich liquid medium, and right photographs show the results of invasive growth assays of colonies grown on rich solid medium, both as in Figure 1.

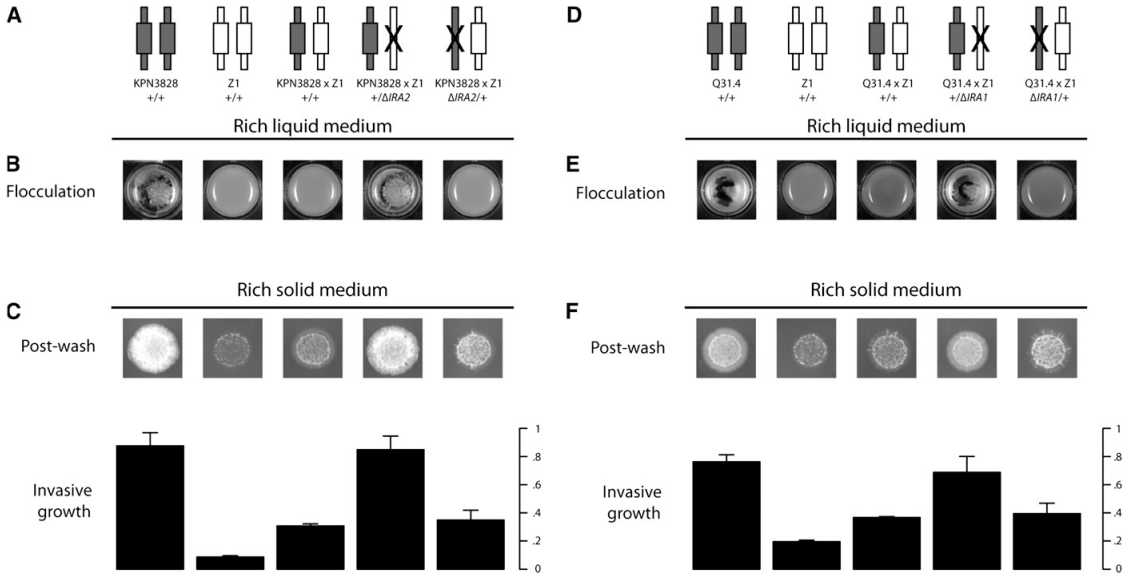


Figure 3. Variation at *IRA1* and *IRA2* underlie flocculation and invasive growth. Each panel reports results of reciprocal hemizygote analysis of genetic variation at an *IRA* gene, between one flocculent, invasive European strain and the non-flocculent European strain Z1. Each column represents one strain, with each element in the bottom panels showing results from the strain indicated at top. (A), (D) Each cartoon represents one diploid strain, with the haploid genome inherited from a flocculent parent or Z1 represented as a dark or light symbol, respectively, and a black X indicating an engineered loss-of-function of an *IRA* gene as indicated in labels. Δ, engineered loss-of-function allele. The fourth and fifth strains in each experiment are isogenic to one another at all loci except the *IRA* gene indicated, such that trait variation between them can be attributed to genetic differences at the manipulated *IRA* locus. (B), (E) Overnight cultures in rich liquid medium as in Figure 1. (C), (F) Invasive growth assays of colonies grown on solid medium as in Figure 1. Bar heights report mean invasive growth measurements of two replicate colonies and error bars represent one standard deviation.

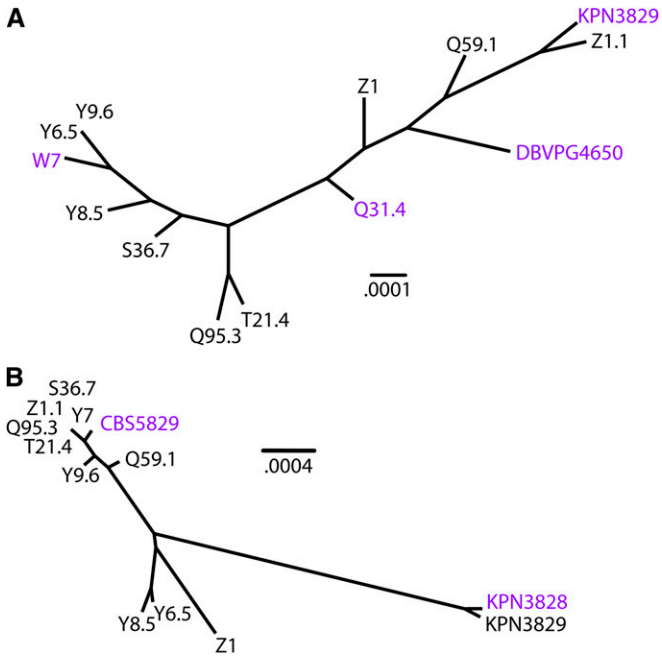


Figure 4. *IRA1* and *IRA2* haplotypes from flocculent/invasive strains are not closely related. Each panel shows a maximum-likelihood phylogeny inferred from coding sequences of *IRA1* (A) or *IRA2* (B). Identifiers of strains in which *IRA1* or *IRA2* underlies flocculent and invasive growth traits are colored purple. Scale bars indicate frequencies of base pair substitutions per site.

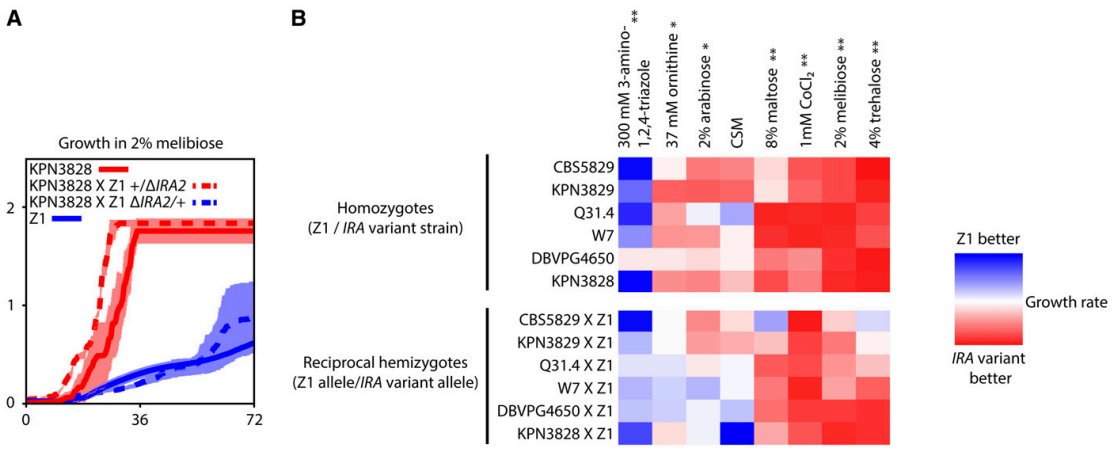


Figure 5. Variation at *IRA1* and *IRA2* underlies growth rate differences in multiple conditions. (A) Shown is an example growth time-course in liquid medium containing 2% melibiose as the sole carbon source, with time after inoculation on the *x* axis and cell density, measured as the optical density (OD) at 600 nm, on the *y* axis. Each curve reports the mean of at least two measurements of one homozygous European diploid (the flocculent strain KPN3828 or the non-flocculent strain Z1) or reciprocal hemizygote diploid constructed in the KPN3828 X Z1 hybrid background (+/Δ*IRA2*, bearing only the KPN3828 allele of *IRA2*, or Δ*IRA2*/+, bearing only the Z1 allele of *IRA2*; see Figure 3A for schematic). Error bars indicate one standard deviation. (B) Each cell reports the results of a growth experiment as in (A). Color in each cell represents a ratio of the growth rates, during log-phase growth, of two strains measured in liquid culture, with each row showing data from one strain pair and each column showing data from one media condition. In the top panel, each cell reports the ratio of growth of the homozygous non-flocculent, non-invasive European strain Z1 to growth of the indicated homozygous European strain harboring a variant *IRA1* or *IRA2* allele. In the bottom panel, each row gives results from a pair of hybrid reciprocal hemizygote strains interrogating *IRA1* or *IRA2* in the indicated strain and Z1: each cell reports the growth of the hemizygote bearing the Z1 allele of the respective *IRA* gene, relative to the growth of the hemizygote bearing the variant allele at the *IRA* gene. Media labels at top marked by an asterisk are those in which homozygote European strains bearing variant *IRA* genes differed significantly from Z1 (Wilcoxon rank-sum test, $p < 0.05$ after Bonferroni correction). Media marked by ** are those inducing two significant growth effects: homozygote European strains bearing variant *IRA* genes differed from Z1, and hemizygote strains bearing variant *IRA* genes differed from hemizygotes bearing the Z1 allele at the *IRA* genes (Wilcoxon rank-sum test, $p < 0.05$ after Bonferroni correction). CSM, complete synthetic medium. Δ, engineered loss-of-function allele. Raw measurements and significance estimates are given in Supplementary Dataset S2.

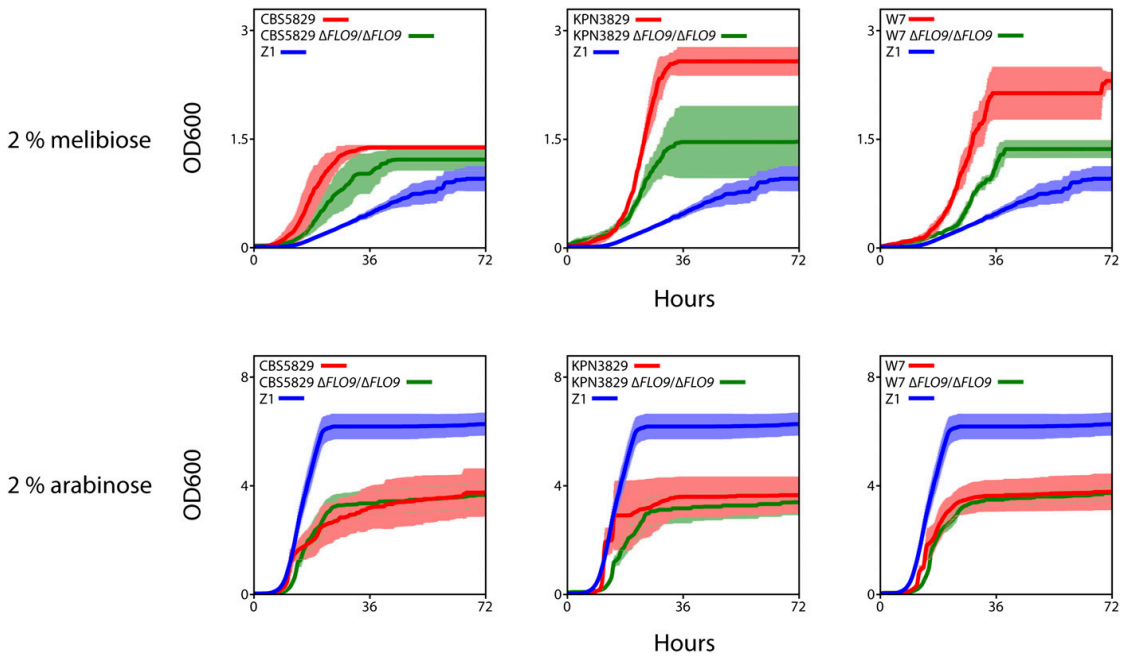


Figure 6. Fitness of strains bearing variant *IRA* gene alleles is partly or completely independent of flocculation. Each column of panels shows the results of growth experiments with one flocculent homozygote European strain, its engineered non-flocculent derivative, and the non-flocculent European strain Z1. Each row of panels shows results from cultures grown in one environmental condition. In a given panel, each trace reports the time-course of growth of one strain, with solid lines reporting the mean of three replicate cultures and error bars indicating one standard deviation. Δ, engineered loss-of-function allele. OD600, optical density of cultures at 600 nm. Raw measurements and significance estimates are given in Supplementary Dataset S3.

Table 1. Two distinct Mendelian loci underlie flocculation in the *S. paradoxus* population

Strain[a]	Proportion flocculent recombinants from cross[b]	Linkage group[c]
W7	.5 (8/16)	1
Q31.4	.5 (8/16)	1
KPN3829	.5 (6/12)	1
KPN3828	.5 (6/12)	2
CBS5829	.5 (8/16)	2
CBS432	.3 (6/20)	ND

[a] Flocculent, highly invasive European strains mated to the non-flocculent, non-invasive European strain Z1.
[b] Results of phenotype scoring of segregants from a cross between the indicated flocculent strain and Z1. For all strains except CBS432, the invasive growth trait segregated with the flocculation trait; photographs of segregants from representative crosses are shown in Supplementary Figure 2.
[c] Results of complementation crosses among flocculent, invasive European strains. ND, not done. Photographs of segregants are shown in Supplementary Figure 3.

Supplementary Figures

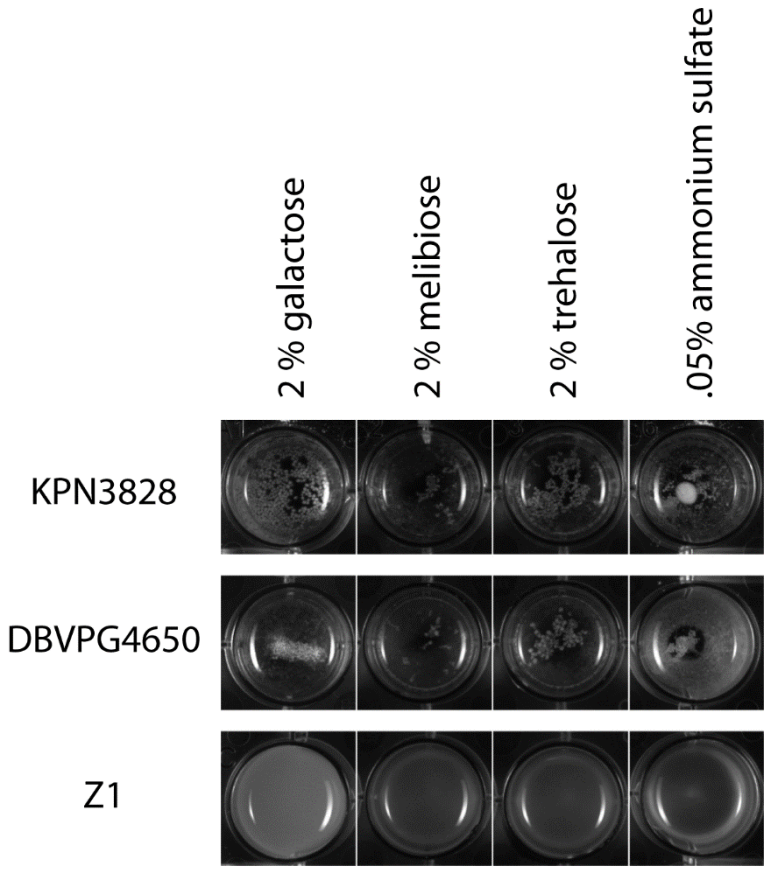


Figure S1. A subset of European *S. paradoxus* strains flocculate in a variety of conditions. Each photograph shows a liquid culture of a homozygous European *S. paradoxus* strain after overnight growth, with each row showing one strain and each column showing one treatment condition. Galactose, melibiose, and trehalose indicate media with the respective sugar as the sole carbon source. Ammonium sulfate indicates synthetic complete medium with 2% glucose and ammonium sulfate as the sole nitrogen source. The top two rows show representative flocculent strains and the bottom row shows the non-flocculent strain Z1.

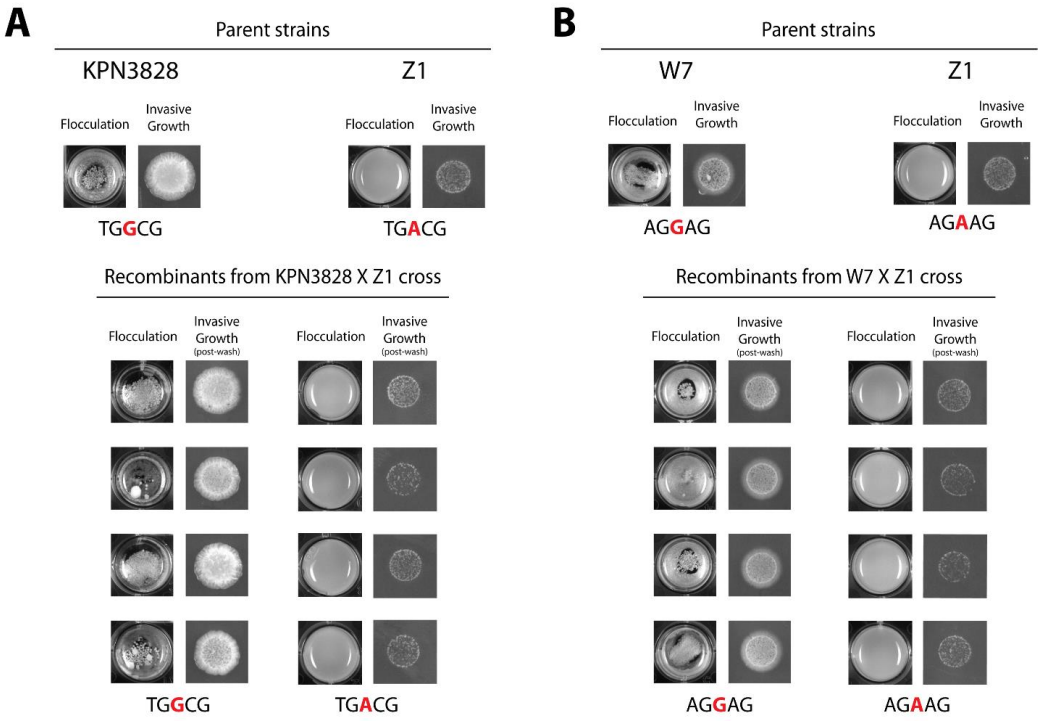


Figure S2. Flocculation and invasive growth are Mendelian traits linked to genetic variation at *IRA1* or *IRA2*. Each panel shows the results of linkage analysis of morphological traits and inheritance at an *IRA* gene, in a cross between a flocculent, invasive European strain and the non-flocculent, non-invasive European strain Z1. Each pair of photographs shows flocculation and invasive growth in one strain, assayed as in Figure 1 of main text. In each panel, the top row shows homozygote European strains used as parents in the respective cross, and each pair of photographs at the bottom shows one recombinant segregant from the cross, with segregants inheriting the variant allele of the *IRA* gene at left, and segregants inheriting the Z1 allele of the *IRA* gene at right. (A) The flocculent, invasive European homozygote KPN3828 crossed to Z1. Black and red text indicates genotype at positions 4437-4441 of *IRA2*. (B) The flocculent, invasive European homozygote W7 crossed to Z1. Black and red text indicates genotype at positions 5015-5019 of *IRA1*. In both (A) and (B), linkage between inheritance at the *IRA* gene and flocculation is significant at $p = 0.0007$.

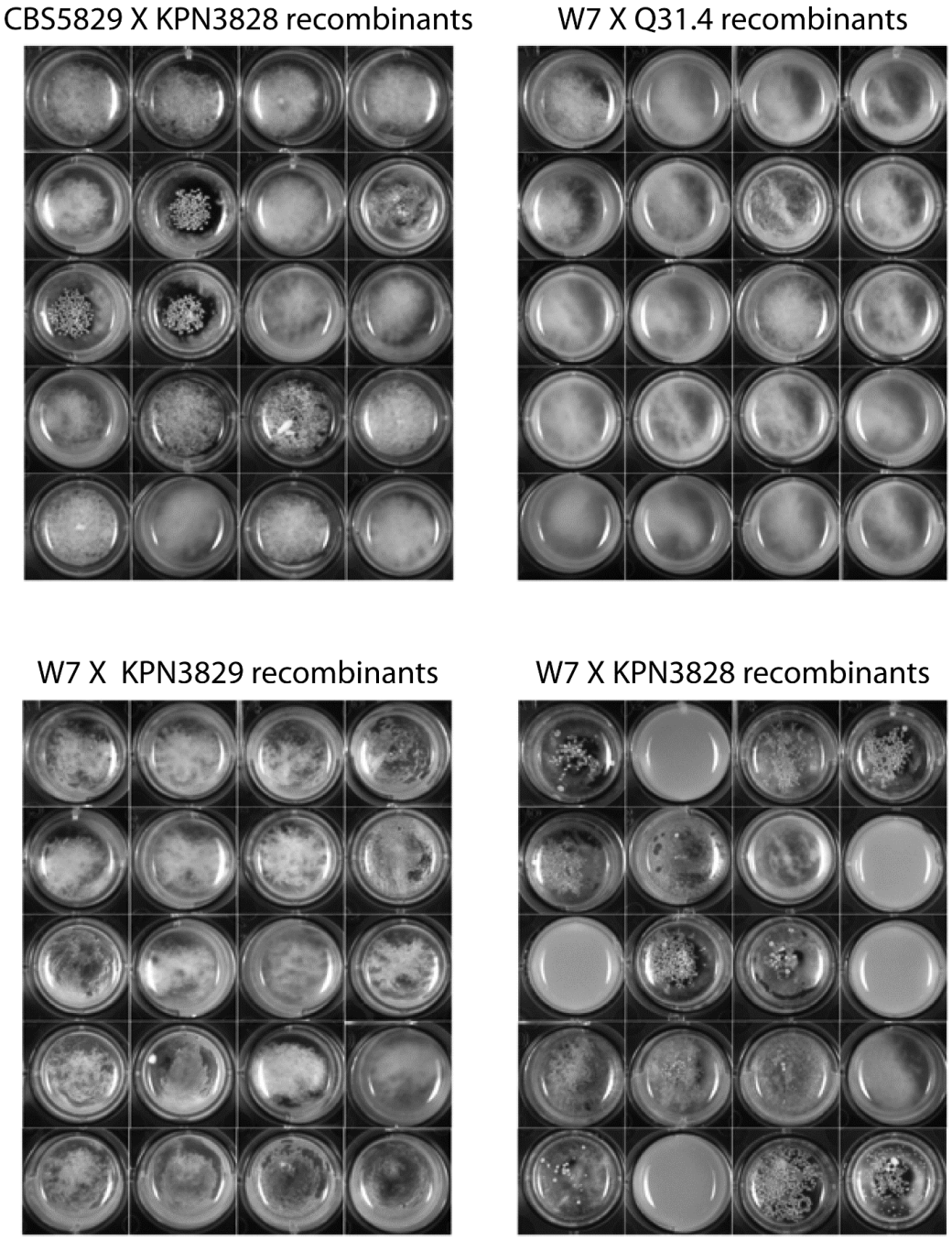


Figure S3. Two distinct loci underlie flocculation across strains. Each panel shows liquid cultures of recombinant progeny from one complementation cross between two European flocculent parent strains. Each row shows cultures of YPD medium inoculated from each of the four spores of one tetrad and grown overnight as in Figure 1A of the main text. All progeny from each cross exhibit flocculation except the cross of W7 with KPN3828 (lower right), indicating that distinct loci underlie the flocculation trait in the latter and for each other pair of European parent strains tested, the two strains share the same causal locus.

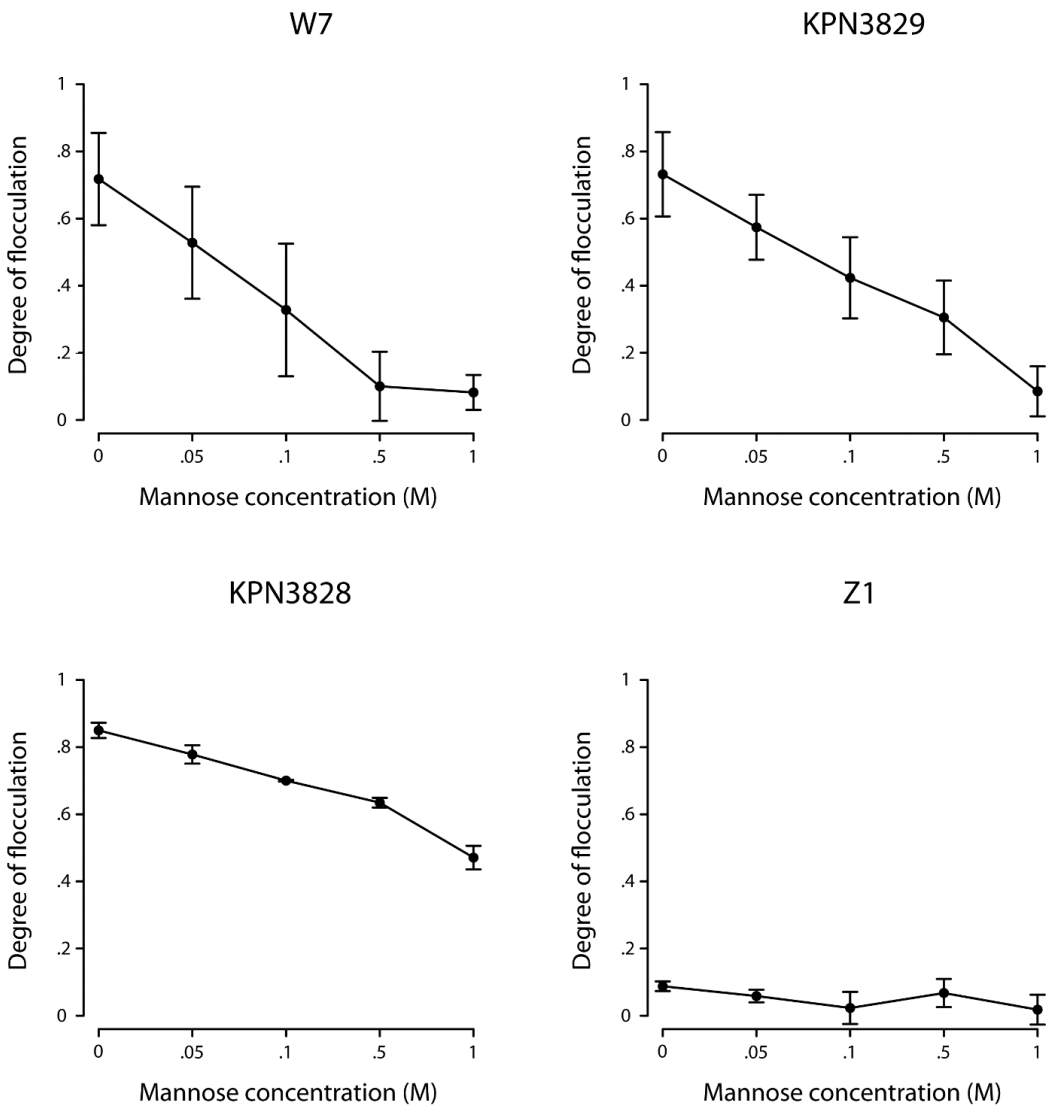


Figure S4. Flocculation is inhibited by mannose. Each panel shows the response of flocculation to increasing concentrations of mannose in one European *S. paradoxus* strain. In each panel, the *x* axis reports mannose concentration and the *y* axis reports flocculation, as measured by cell density at the top of a liquid culture allowed to settle, normalized by the analogous quantity from a culture treated with EDTA. Each data point represents the mean of three replicate cultures, and error bars report one standard deviation. Note that mannose had no effect on spatial inhomogeneity of cultures of the non-flocculent European strain Z1 (lower right), while the other three flocculent strains (see Figure 1A of the main text), are sensitive to mannose

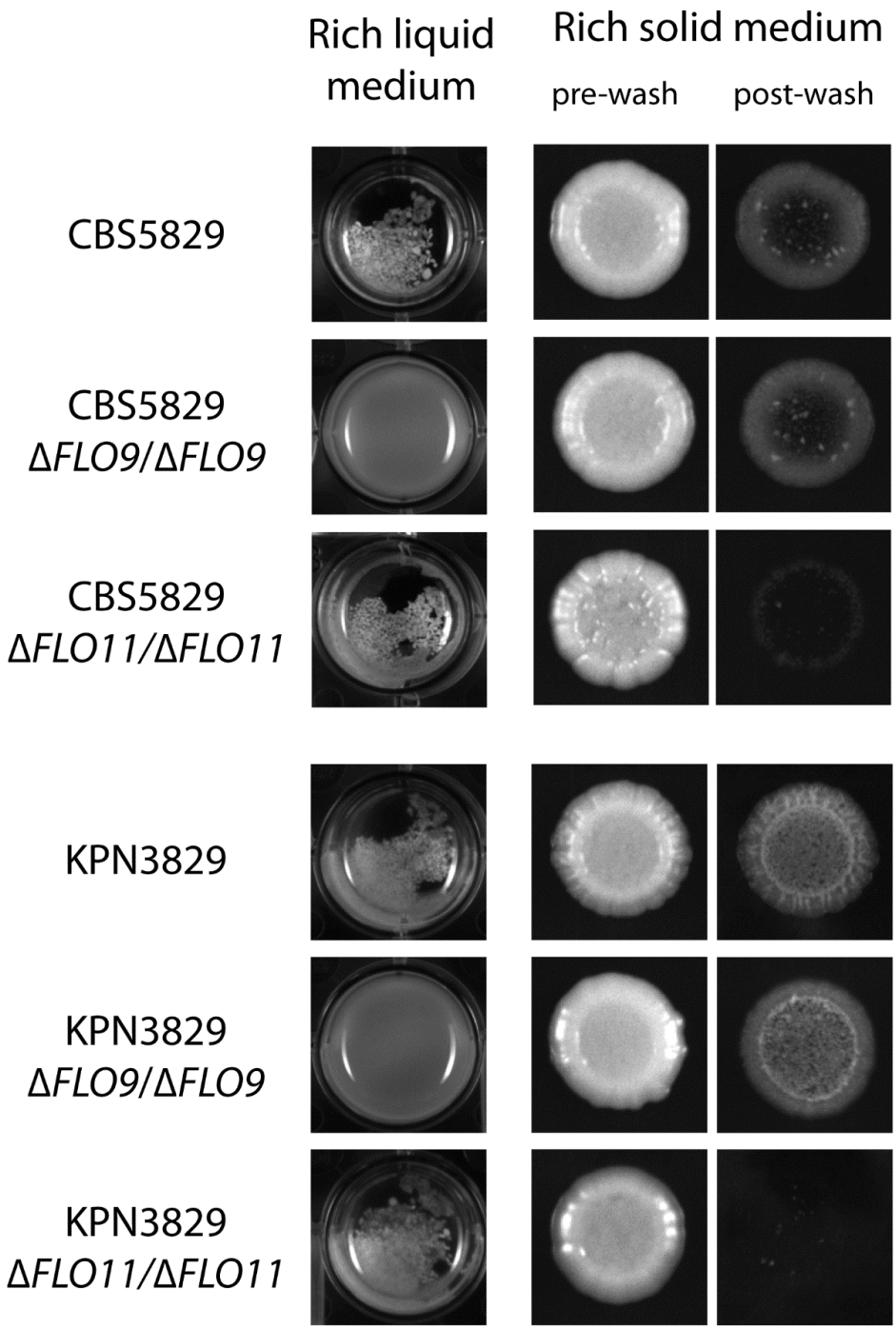


Figure S5. *FLO9* and *FLO11* are terminal effectors of flocculation and invasive growth in multiple *S. paradoxus* strains. Data are as in Figure 2 of the main text except that flocculent European strains CBS5829 and KPN3829 were analyzed.

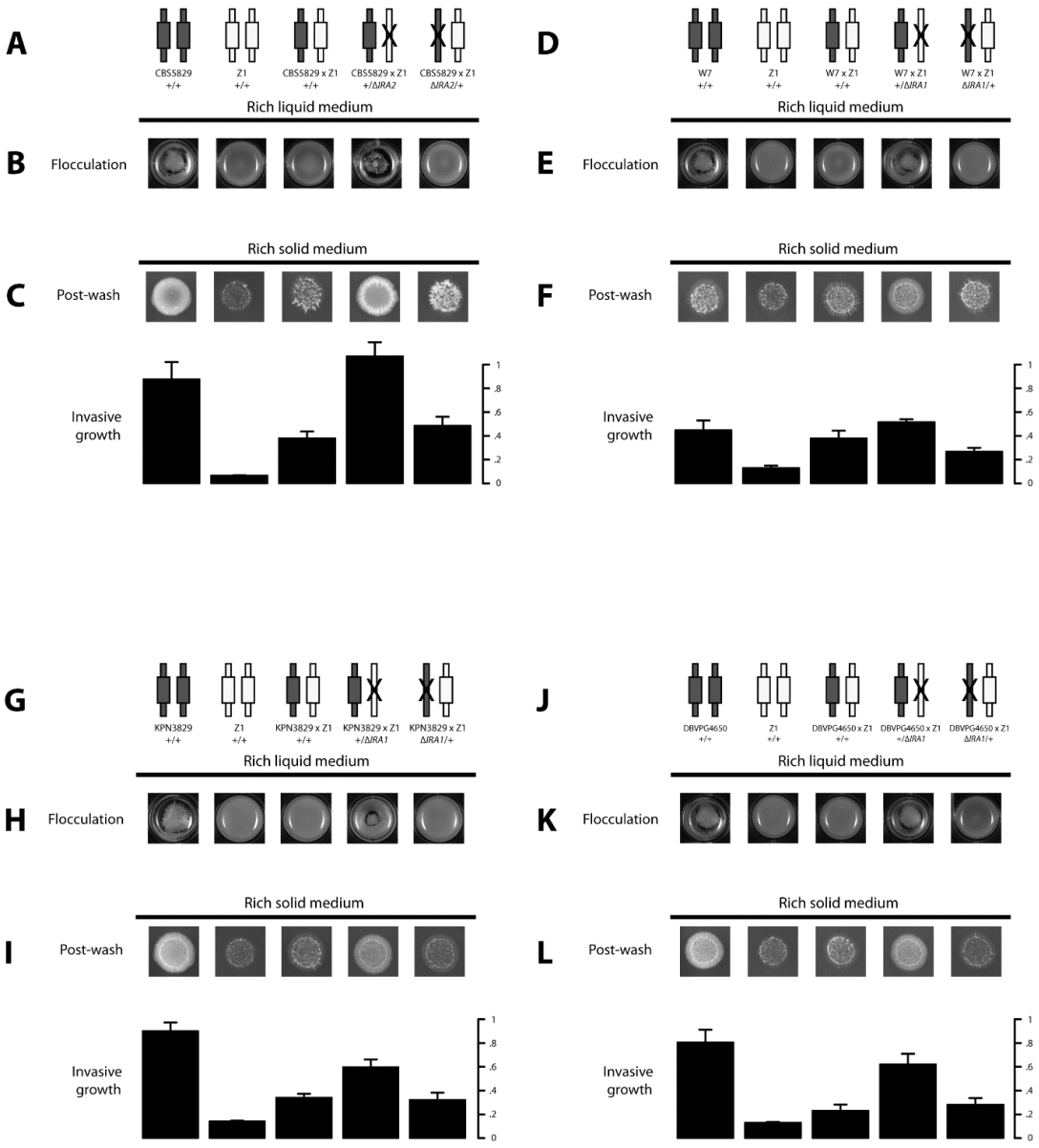


Figure S6. Variation at *IRA1* and *IRA2* underlie flocculation and invasive growth. Data are as in Figure 3 of the main text, except that flocculent European strains CBS5829 (A-C), W7 (D-F), KPN3829 (G-I), and DVBPG4650 (J-L) were analyzed.

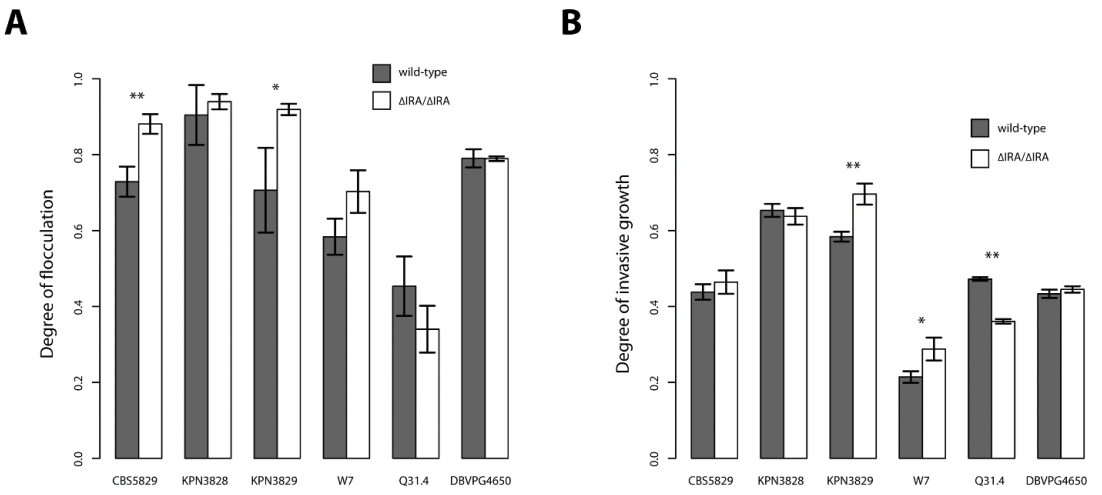


Figure S7. Variant *IRA1* and *IRA2* alleles act as partial or complete losses of function. Each panel reports measurements of one morphology in European *S. paradoxus* strains and their derivatives harboring engineered mutations in either *IRA1* or *IRA2*. In a given panel, each pair of bars reports data for two strains: a European strain and its derivative bearing an engineered mutation at the *IRA* gene found to underlie its morphologies (see Figure 3 and Figure S6). (A) Flocculation, measured as in Supplementary Figure 4. (B) Invasive growth, measured as in Figure 1B of the main text. Quantitative differences between wild type and *IRA* deletion strain phenotypes in both panels were assessed by Welch's *t* test (*, $p < 0.05$, **, $p < 0.001$). Δ, engineered loss-of-function allele.

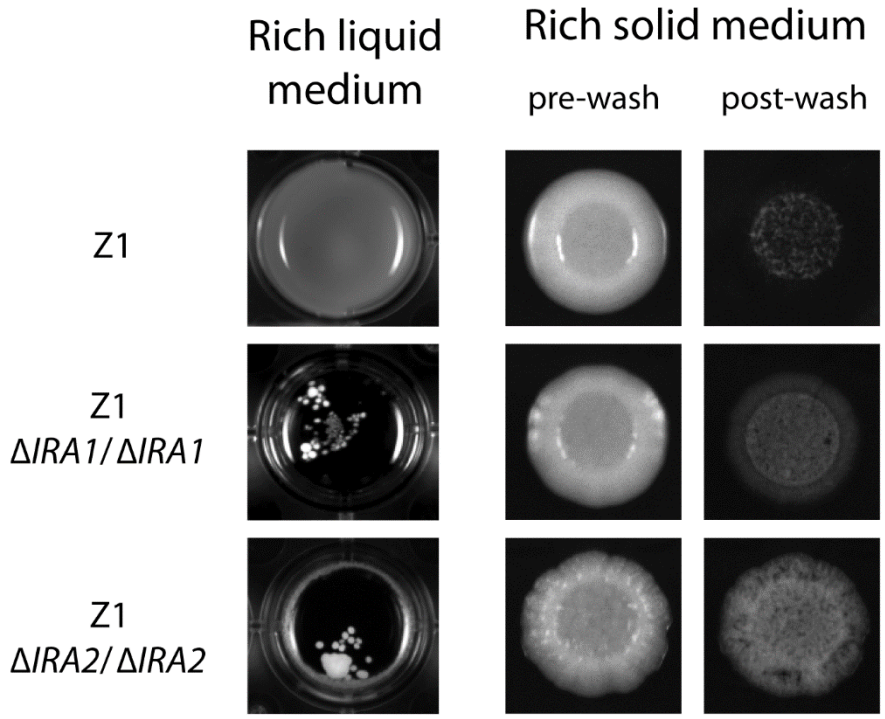


Figure S8. Engineered mutation of either *IRA1* or *IRA2* is sufficient for flocculation and invasive growth. Each row reports morphologies of one homozygous derivative of the non-flocculent, non-invasive European strain Z1. Left photographs show overnight cultures in rich liquid medium, and right photographs show the results of invasive growth assays of colonies grown on rich solid medium, both as in Figure 1 of the main text. Δ, engineered loss-of-function allele.

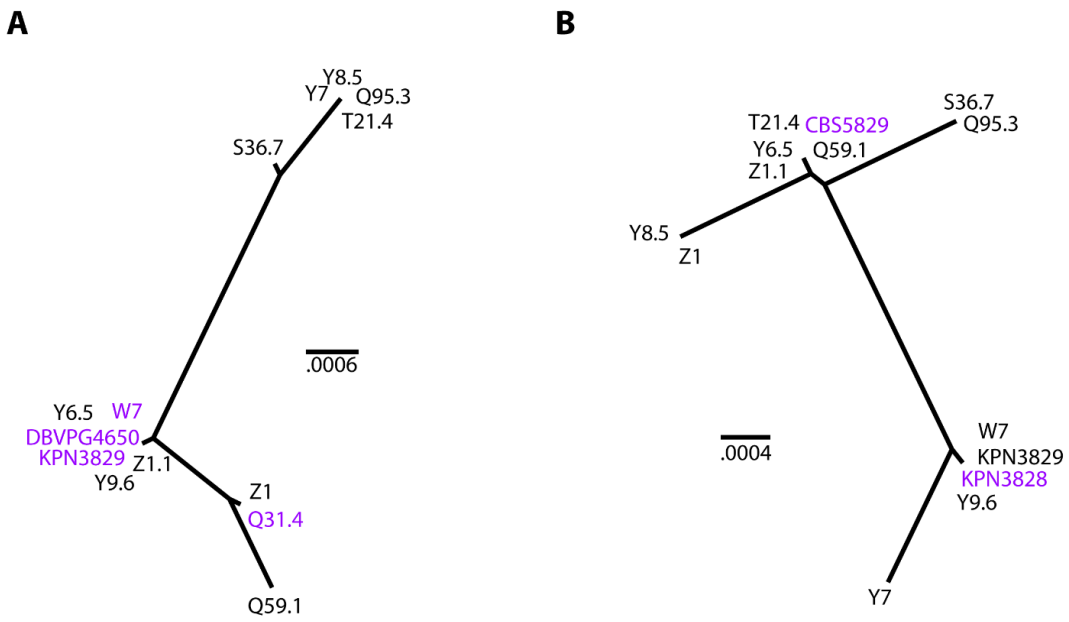


Figure S9. Phylogeny of promoter regions of *IRA1* and *IRA2*. Maximum likelihood phylogeny inferred from 800bp of nucleotide sequence upstream of the *IRA1* (A) or *IRA2* (B) coding region. Identifiers of strains in which *IRA1* or *IRA2* underlie flocculent and invasive growth traits are colored purple. Scale bars indicate frequencies of base pair substitutions per site.

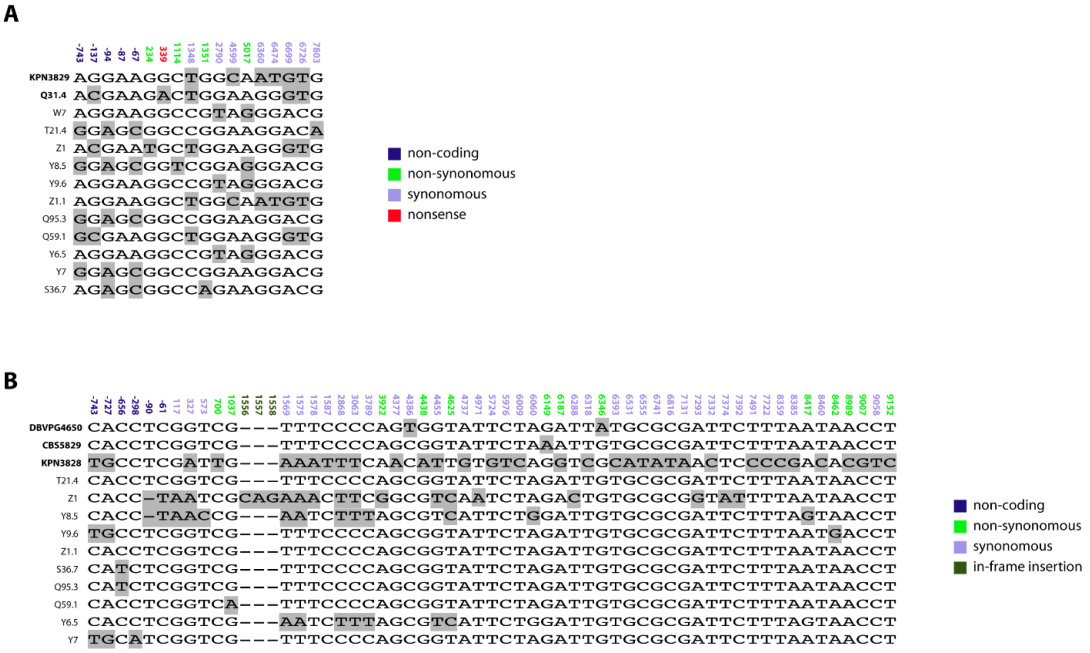


Figure S10. *IRA1* and *IRA2* genes from flocculent/invasive strains do not share derived polymorphisms. Shown are polymorphic sites in the *IRA1* (A) or *IRA2* (B) coding sequence and 800bp of promoter region for the indicated *S. paradoxus* strains. Minor alleles at each site in the alignment are highlighted in grey. Positions of each site relative to the start of the coding sequence are indicated above the alignment and color-coded to indicate the effect of each minor allele on the protein sequence. Original isolates of the five flocculent/invasive strains were used for sequencing, with names indicated in bold.

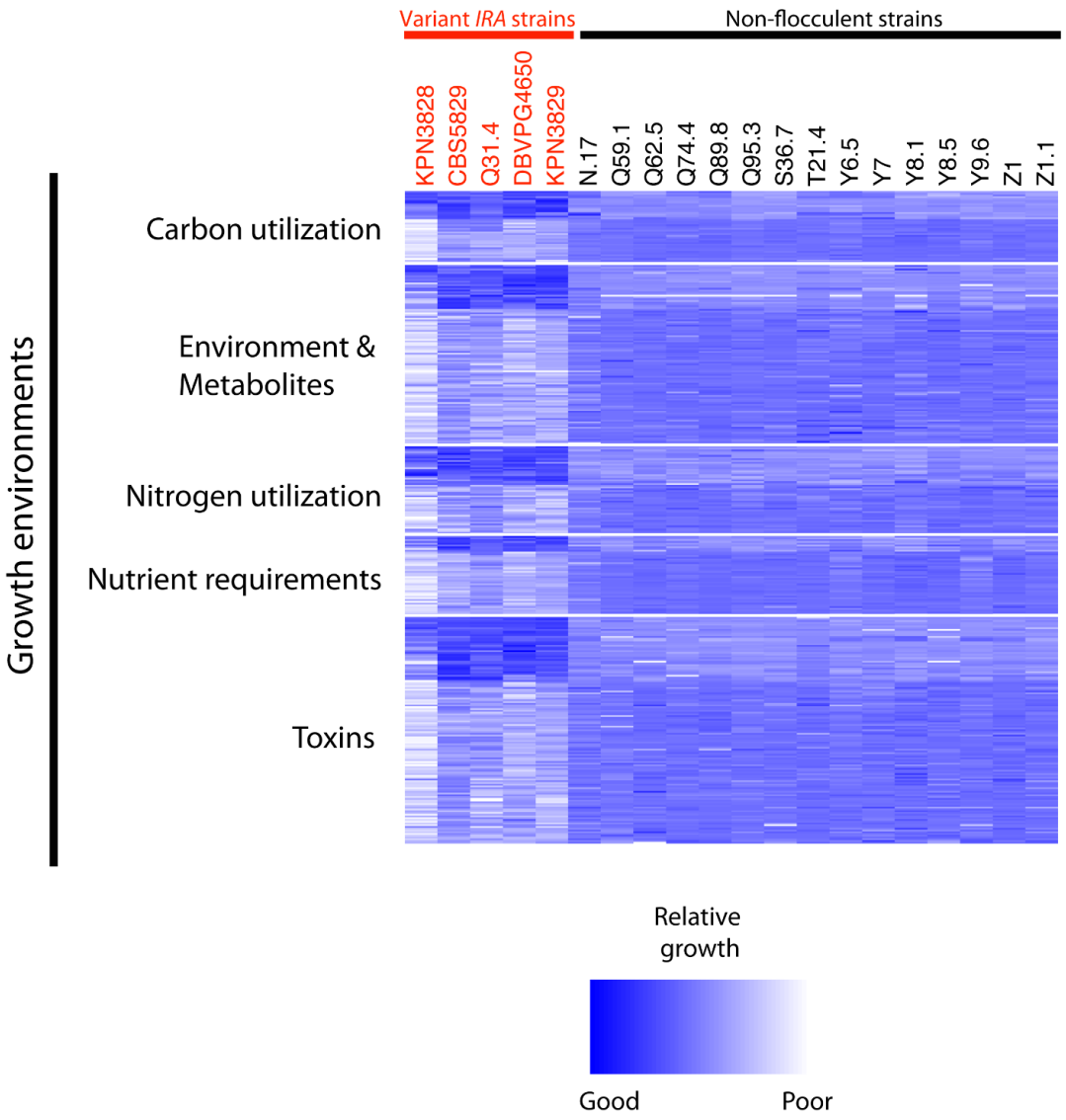


Figure S11. Hundreds of growth traits across environmental treatments associate with variants in *IRA1* and *IRA2*. Each column reports fitness measurements from a liquid culture of one strain, and each row reports measurements in one condition. For each row, measurements of lag time before resumption of log-phase growth after dilution into fresh medium, growth rate in log phase, or final culture density in stationary phase were taken from [23] and normalized against the median across strains. Shown is each environment-parameter combination for which the difference between flocculent and invasive strains bearing variants in *IRA1* or *IRA2,* and the remainder of strains in the data set, reached a significance level corresponding to a false discovery rate less than 5%. Raw measurements for all environment-parameter combinations and significance estimates are provided in Dataset S1.

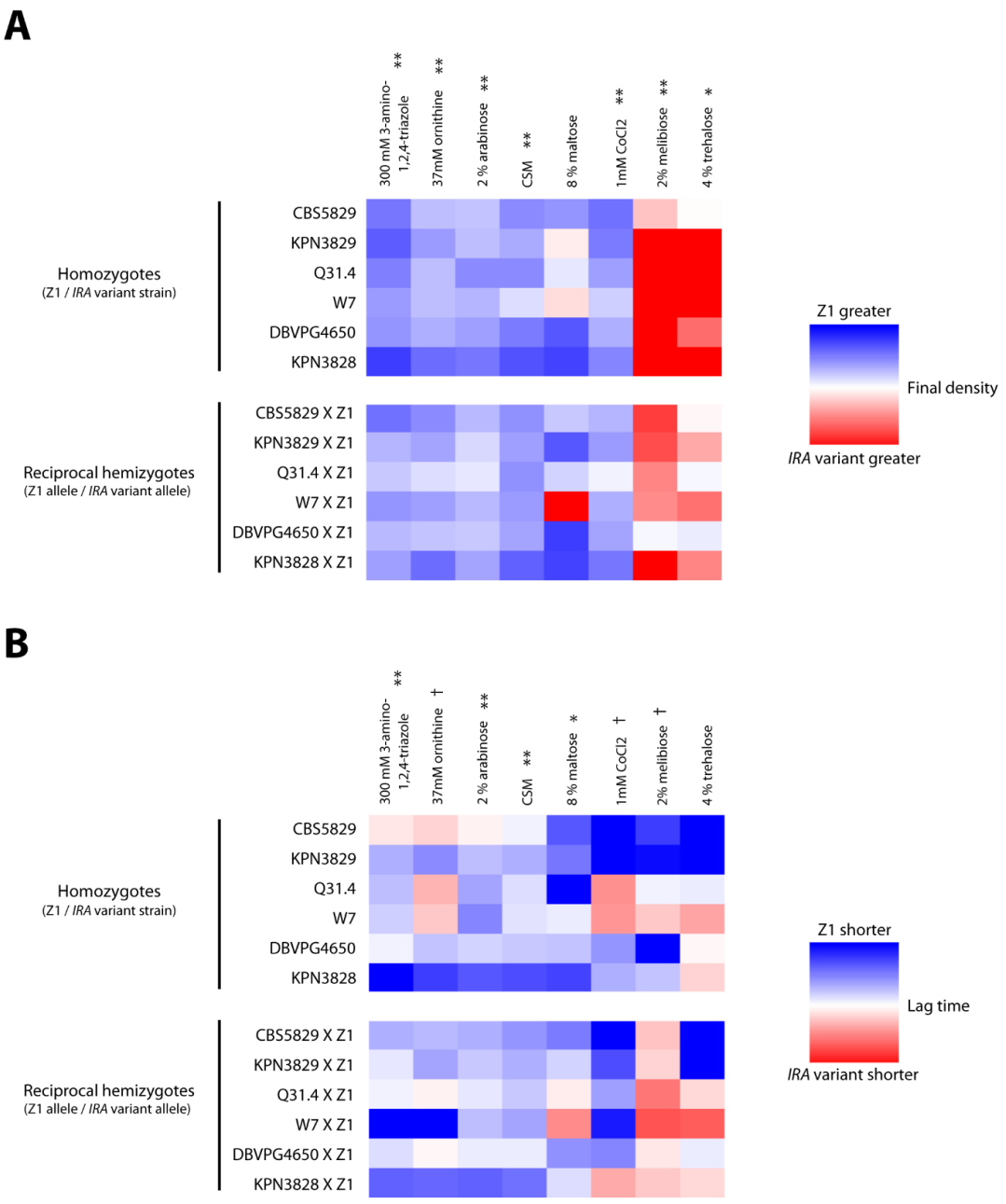


Figure S12. Variation at *IRA1* and *IRA2* underlies differences between strains in final culture density and lag in multiple conditions. Data are as in Figure 5B of the main text, except that final culture density in stationary phase (A) or lag time before resumption of log-phase growth after dilution into fresh medium (B) were analyzed. Media marked by † are those in which hemizygote hybrid strains bearing the variant *IRA* genes differed significantly from hemizygote hybrid strains bearing the Z1 allele at the *IRA* genes (Wilcoxon rank-sum test, $p < 0.05$ after Bonferroni correction), but homozygote European strains bearing variant *IRA* genes did not differ significantly from Z1.

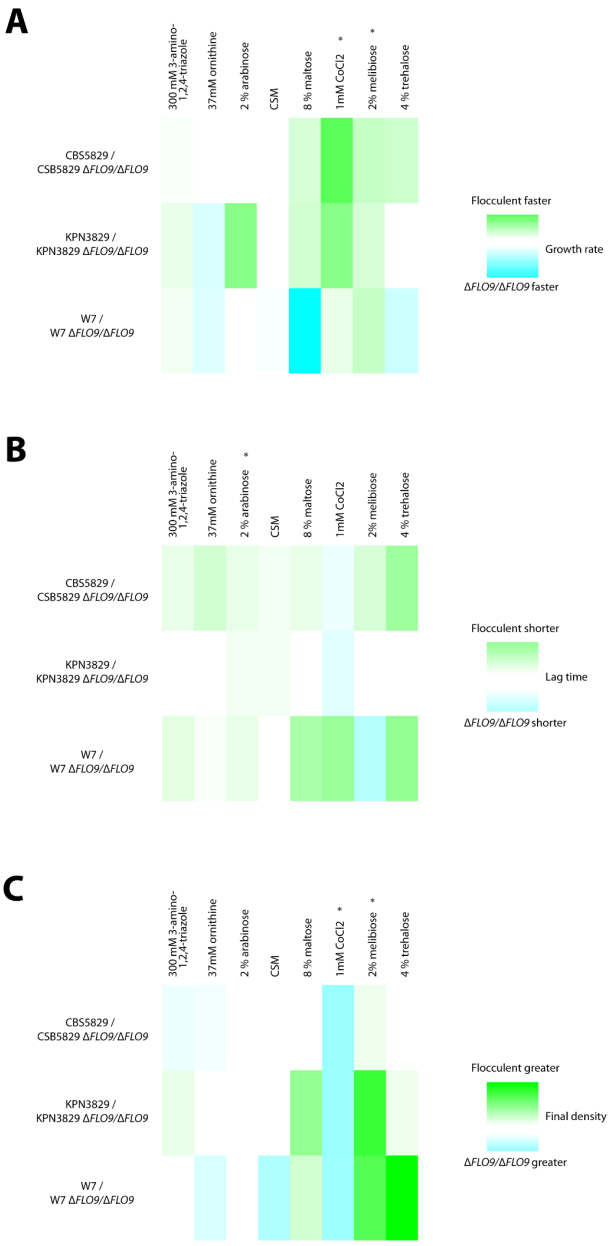


Figure S13. Effect of flocculation on culture growth. Each panel reports one growth parameter measured in liquid cultures of homozygous flocculent European strains and their engineered non-flocculent derivatives. In a given panel, each row reports results from one European strain and each row reports measurements in one medium. Color in each cell reports the indicated growth parameter in the indicated condition as a ratio of measurements from two strains: the indicated flocculent European strain and an isogenic non-flocculent strain homozygous for a null allele of *FLO9*. (A) Growth rate; (B) lag time; (C) cell density in stationary phase. In a given panel, media marked by * are those in which growth of flocculent European strains differed significantly from non-flocculent *FLO9* mutants (Wilcoxon rank-sum test, $p < 0.05$ after Bonferroni correction). Δ, engineered loss-of-function allele. Raw measurements and significance estimates are given in Supplementary Dataset S3.

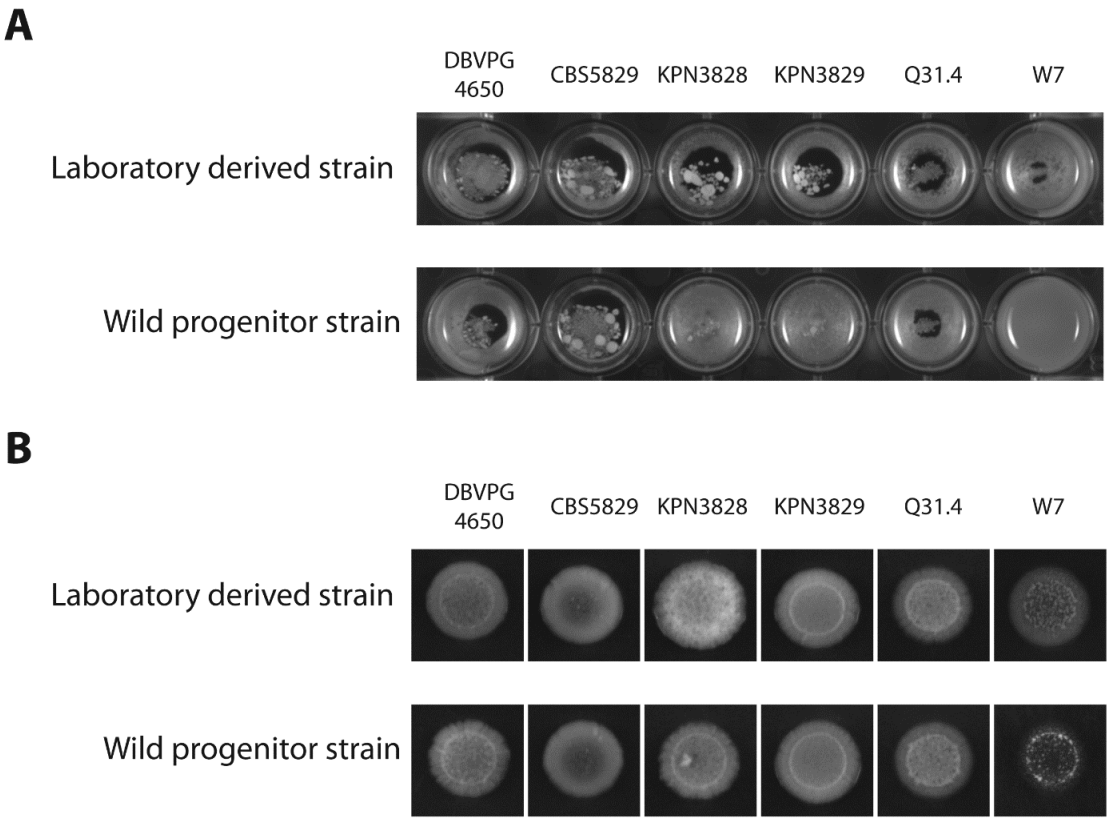


Figure S14. Most wild European *S. paradoxus* strains exhibit the same morphologies as their laboratory derivatives. Each panel shows morphologies in wild European *S. paradoxus* strains (top row) and their monosporic laboratory derivatives (bottom row). (A) Photographs of cultures after overnight growth in liquid rich medium as in Figure 1A of the main text. (B) Results of invasive growth assays of colonies grown on solid medium as in Figure 1B of the main text.

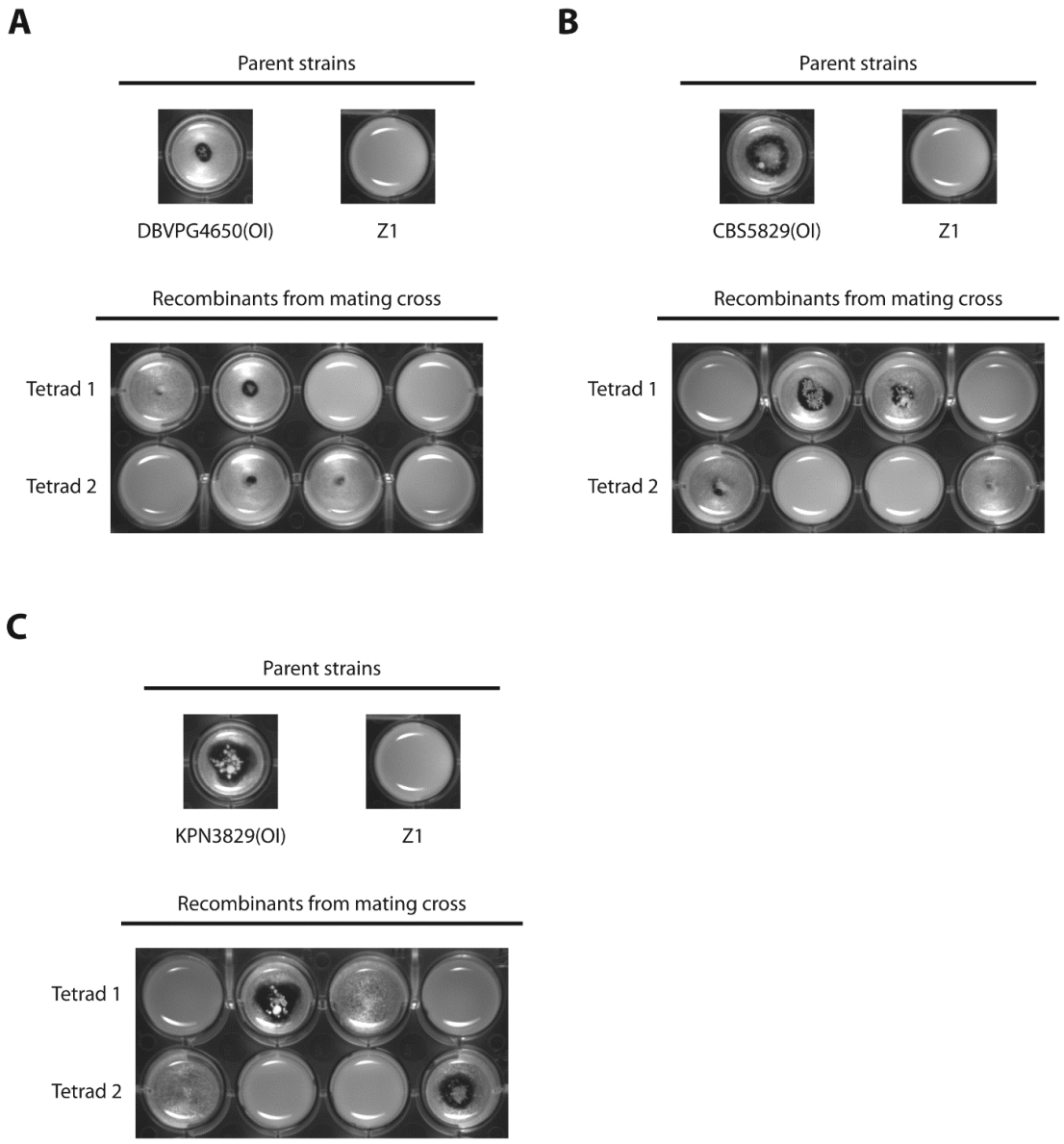


Figure S15. Flocculation is a Mendelian trait in wild progenitor strains. Each panel shows the growth phenotype in rich liquid media of a wild progenitor strain (original isolate, OI) and the Z1 non-flocculent strain (top), and recombinant progeny from a mating between these two parental strains (bottom). Recombinant progeny represent all four spores from two tetrad dissections, as indicated. (A) The flocculent wild progenitor of the DBVPG4650 strain crossed to Z1. (B) The flocculent wild progenitor of the CBS5829 strain crossed to Z1. (C) The flocculent wild progenitor of the KPN3829 strain crossed to Z1.

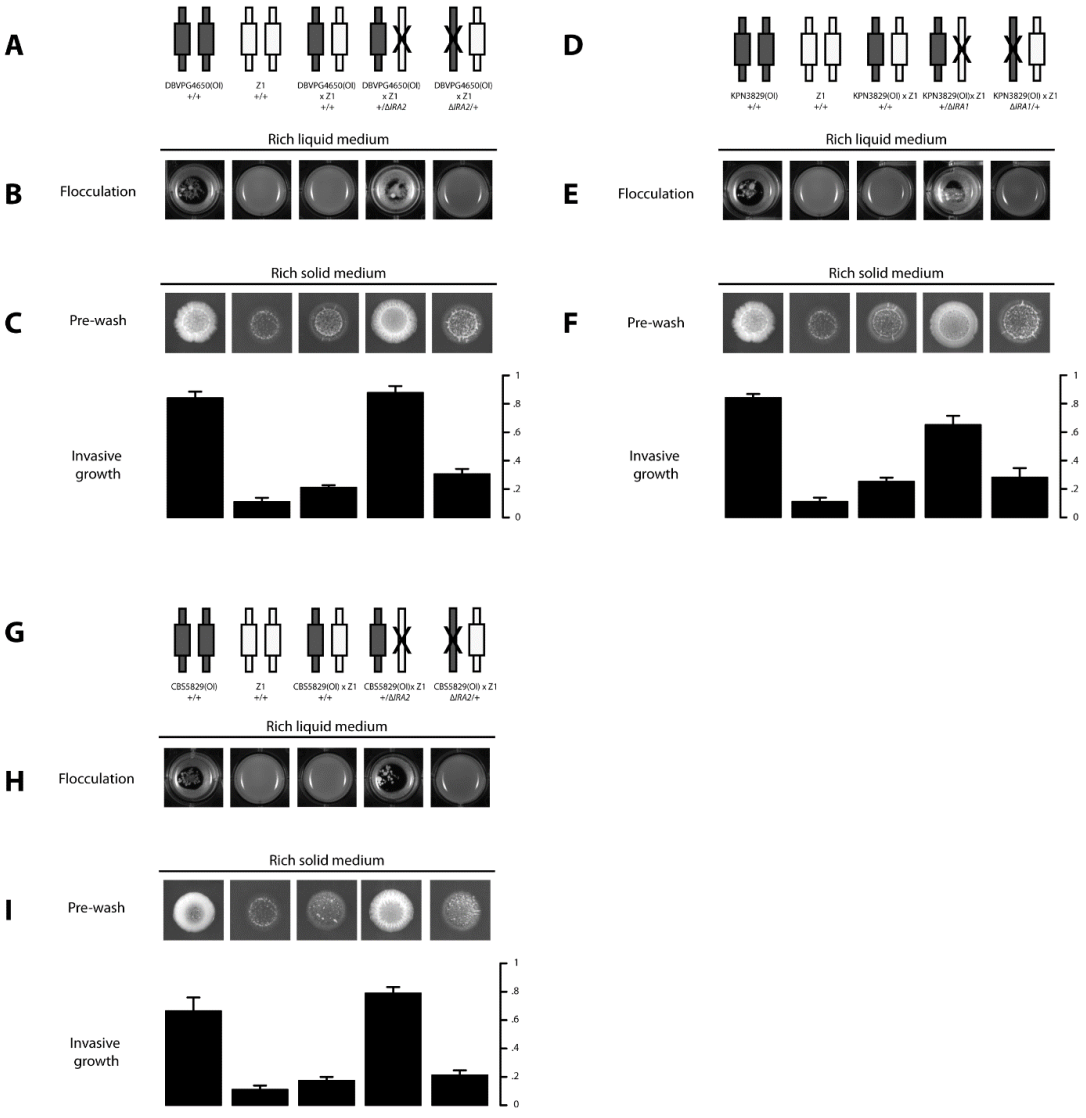


Figure S16. Variation at *IRA1* and *IRA2* underlie flocculation and invasive growth in wild progenitor strains. Data are as in Figure 3 of the main text and Figure S6, except that the wild progenitors (original isolates, OI) of strains DBVPG4650 (A-C), KPN3829 (D-F) and CBS5829 (G-I), were analyzed.

Supplementary Tables

Strain	Parent(s)	Genotype	Source
Q31.4			NCYC
Q32.3			NCYC
Q59.1			NCYC
Q62.5			NCYC
Q69.8			NCYC
Q74.4			NCYC
Q89.8			NCYC
Q95.3			NCYC
S36.7			NCYC
T21.4			NCYC
W7			NCYC
Y6.5			NCYC
Y7			NCYC
Y8.1			NCYC
Y8.5			NCYC
Y9.6			NCYC
Z1			NCYC
Z1.1			NCYC
N-17			NCYC
CBS432			NCYC
CBS5829			NCYC
DBVPG4650			NCYC
KPN3828			NCYC
KPN3829			NCYC
YJR66	Q31.4 X Z1		this work
YJR44	W7 X Z1		this work
YJR119	CBS432 X Z1		this work
YJR78	CBS5829 X Z1		this work
YJR45	KPN3828 XZ1		this work
YJR39	KPN3829 X Z1		this work
YJR65	DBVPG4650 X Z1		this work
YJR46	W7 X KPN828		this work
YJR129	KPN3829 X W7		this work
YJR130	CBS5829 X KPN3828		this work
YJR91	W7	*ΔFLO9::kanMX/ΔFLO9::kanMX*	this work
YJR123	KPN3829	*ΔFLO9::kanMX/ΔFLO9::kanMX*	this work
YJR125	CBS5829	*ΔFLO9::kanMX/ΔFLO9::kanMX*	this work
YJR60	W7	*ΔFLO11::kanMX/ΔFLO11::kanMX*	this work
YJR59	KPN3829	*ΔFLO11::kanMX/ΔFLO11::kanMX*	this work
YJR127	CBS5829	*ΔFLO11::kanMX/ΔFLO11::kanMX*	this work
YJR67	W7	*Δflo1::kanMX/Δflo1::kanMX*	this work
YJR68	W7	*Δflo5::kanMX/Δflo5::kanMX*	this work
YJR69	W7	*Δflo10::kanMX/Δflo10::kanMX*	this work
YJR97	YJR66	*ΔIRA1::kanMX/IRA1* (Q31.4)	this work
YJR98	YJR66	*ΔIRA1::kanMX/IRA1* (Z1)	this work
YJR58	YJR44	*ΔIRA1::kanMX/IRA1* (W7)	this work
YJR57	YJR44	*ΔIRA1::kanMX/IRA1* (Z1)	this work
YJR120	YJR39	*ΔIRA1::kanMX/IRA1* (KPN3829)	this work
YJR121	YJR39	*ΔIRA1::kanMX/IRA1* (Z1)	this work
YJR117	YJR65	*ΔIRA1::kanMX/IRA1* (DBVPG4650)	this work
YJR118	YJR65	*ΔIRA1::kanMX/IRA1* (Z1)	this work
YJR105	YJR78	*ΔIRA2::kanMX/IRA2* (CBS5829)	this work
YJR106	YJR78	*ΔIRA2::kanMX/IRA2* (Z1)	this work
YJR134	YJR45	*ΔIRA2::kanMX/IRA2* (KPN3828)	this work

YJR135	YJR45	*ΔIRA2::kanMX/IRA2* (Z1)	this work
YJR100	Z1	*ΔIRA2::kanMX/ΔIRA2::kanMX*	this work
YJR101	Z1	*ΔIRA1::kanMX/ΔIRA1::kanMX*	this work
Q31.4(OI)			V. Koufopanou
W7(OI)			V. Koufopanou
KPN3828(OI)			G. Liti
CBS5829(OI)			G. Liti
DBVPG4650(OI)			G. Liti
KPN3829(OI)			G. Liti

OI, Original isolate.

Table S1. Strains used in this work.

References

1. Magwire MM, Fabian DK, Schweyen H, Cao C, Longdon B, Bayer F, et al. Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in Drosophila melanogaster. PLoS Genetics. 2012;8: e1003057. doi:10.1371/journal.pgen.1003057

2. Weber AL, Khan GF, Magwire MM, Tabor CL, Mackay TFC, Anholt RRH. Genome-wide association analysis of oxidative stress resistance in Drosophila melanogaster. PLoS ONE. 2012;7: e34745. doi:10.1371/journal.pone.0034745

3. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The Drosophila melanogaster Genetic Reference Panel. Nature. Nature Publishing Group; 2012;482: 173–178. doi:10.1038/nature10811

4. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, et al. Linkage and association mapping of Arabidopsis thaliana flowering time in nature. PLoS Genetics. Public Library of Science; 2010;6: e1000940. doi:10.1371/journal.pgen.1000940.s013

5. Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in Arabidopsis thaliana. PLoS Biol. 2011;9: e1001125. doi:10.1371/journal.pbio.1001125

6. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, et al. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature. Nature Publishing Group; 2010;465: 627–631. doi:10.1038/nature08800

7. Filiault DL, Maloof JN. A genome-wide association study identifies variants underlying the Arabidopsis thaliana shade avoidance response. PLoS Genetics. 2012;8: e1002589. doi:10.1371/journal.pgen.1002589

8. Dunn B, Paulish T, Stanbery A, Piotrowski J, Koniges G, Kroll E, et al. Recurrent Rearrangement during Adaptive Evolution in an Interspecific Yeast Hybrid Suggests a Model for Rapid Introgression. PLoS Genetics. 2013;9: e1003366. doi:10.1371/journal.pgen.1003366

9. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9: 356–369. doi:10.1038/nrg2344

10. Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nat Rev Genet. 2012;13: 565–575. doi:10.1038/nrg3241

11. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB, editors. PLoS Biol. 2010;8: e1000294. doi:10.1371/journal.pbio.1000294.t001

12. McClellan J, King M-C. Genetic heterogeneity in human disease. Cell. 2010;141: 210–217. doi:10.1016/j.cell.2010.03.032

13. Shen MH, Harper PS, Upadhyaya M. Molecular genetics of neurofibromatosis type 1 (NF1). Journal of medical genetics. BMJ Publishing Group Ltd; 1996;33: 2–17.

14. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell. 2012;151: 1431–1442. doi:10.1016/j.cell.2012.11.019

15. Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science. 2010;327: 302–305. doi:10.1126/science.1182213

16. 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

17. Soares EV. Flocculation in Saccharomyces cerevisiae: a review. J Appl Microbiol. 2011;110: 1–18. doi:10.1111/j.1365-2672.2010.04897.x

18. Carstens E, Lambrechts M. Flocculation, pseudohyphal development and invasive growth in commercial wine yeast strains. S Afr J Enol Vitic. 1998.

19. Jin R, Dobry CJ, McCown PJ, Kumar A. Large-scale analysis of yeast filamentous growth by systematic gene disruption and overexpression. Mol Biol Cell. 2008;19: 284–296. doi:10.1091/mbc.E07-05-0519

20. Ryan O, Shapiro RS, Kurat CF, Mayhew D, Baryshnikova A, Chin B, et al. Global Gene Deletion Analysis Exploring Yeast Filamentous Growth. Science. 2012;337: 1353–1356. doi:10.1126/science.1224339

21. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37: 986–990. doi:10.1038/ng1618

22. Halme A, Bumgarner S, Styles C, Fink GR. Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. Cell. 2004;116: 405–415.

23. Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, et al. Trait

variation in yeast is defined by population history. PLoS Genetics. 2011;7: e1002111. doi:10.1371/journal.pgen.1002111

24. Cubillos FA, Billi E, Zörgö E, Parts L, Fargier P, Omholt S, et al. Assessing the complex architecture of polygenic traits in diverged yeast populations. Mol Ecol. 2011;20: 1401–1413. doi:10.1111/j.1365-294X.2011.05005.x

25. Zörgö E, Gjuvsland A, Cubillos FA, Louis EJ, Liti G, Blomberg A, et al. Life history shapes trait heredity by accumulation of loss-of-function alleles in yeast. Mol Biol Evol. 2012;29: 1781–1789. doi:10.1093/molbev/mss019

26. Smit G, Straver MH, Lugtenberg BJ, Kijne JW. Flocculence of Saccharomyces cerevisiae cells is induced by nutrient limitation, with cell surface hydrophobicity as a major determinant. Appl Environ Microbiol. 1992;58: 3709–3714.

27. Sampermans S, Mortier J, Soares EV. Flocculation onset in Saccharomyces cerevisiae: the role of nutrients. J Appl Microbiol. 2005;98: 525–531. doi:10.1111/j.1365-2672.2004.02486.x

28. Bayly JC, Douglas LM, Pretorius IS, Bauer FF, Dranginis AM. Characteristics of Flo11-dependent flocculation in Saccharomyces cerevisiae. FEMS Yeast Res. 2005;5: 1151–1156. doi:10.1016/j.femsyr.2005.05.004

29. Dengis PB, Nélissen LR, Rouxhet PG. Mechanisms of yeast flocculation: comparison of top- and bottom-fermenting strains. Appl Environ Microbiol. 1995;61: 718–728.

30. Guo B, Styles CA, Feng Q, Fink GR. A Saccharomyces gene family involved in invasive growth, cell-cell adhesion, and mating. Proc Natl Acad Sci USA. 2000;97: 12158–12163. doi:10.1073/pnas.220420397

31. Palecek SP, Parikh AS, Kron SJ. Genetic analysis reveals that FLO11 upregulation and cell polarization independently regulate invasive growth in Saccharomyces cerevisiae. Genetics. 2000;156: 1005–1023.

32. Bester MC, Pretorius IS, Bauer FF. The regulation of Saccharomyces cerevisiae FLO gene expression and Ca2+-dependent flocculation by Flo8p and Mss11p. Curr Genet. 2006;49: 375–383. doi:10.1007/s00294-006-0068-z

33. Vinod PK, Sengupta N, Bhat PJ, Venkatesh KV. Integration of global signaling pathways, cAMP-PKA, MAPK and TOR in the regulation of FLO11. PLoS ONE. 2008;3: e1663. doi:10.1371/journal.pone.0001663

34. Zaman S, Lippman SI, Zhao X, Broach JR. How Saccharomyces responds to nutrients. Annu Rev Genet. 2008;42: 27–81. doi:10.1146/annurev.genet.41.110306.130206

35. Palecek SP, Parikh AS, Kron SJ. Sensing, signalling and integrating physical processes during Saccharomyces cerevisiae invasive and filamentous growth. Microbiology (Reading, Engl). 2002;148: 893–907.

36. Ballester R, Marchuk D, Boguski M, Saulino A, Letcher R, Wigler M, et al. The NF1 locus encodes a protein functionally related to mammalian GAP and yeast IRA proteins. Cell. 1990;63: 851–859.

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

38. Tanaka K, Nakafuku M, Satoh T, Marshall MS, Gibbs JB, Matsumoto K, et al. S. cerevisiae genes IRA1 and IRA2 encode proteins that may be functionally equivalent to mammalian ras GTPase activating protein. Cell. 1990;60: 803–807.

39. Tanaka K, Matsumoto K, Toh-E A. IRA1, an inhibitory regulator of the RAS-cyclic AMP pathway in Saccharomyces cerevisiae. Mol Cell Biol. 1989;9: 757–768.

40. Gerke J, Lorenz K, Cohen B. Genetic Interactions Between Transcription Factors Cause Natural Variation in Yeast. Science. 2009;323: 498–501. doi:10.1126/science.1166426

41. Lee HN, Magwene PM, Brem RB. Natural variation in CDC28 underlies morphological phenotypes in an environmental yeast isolate. Genetics. 2011;188: 723–730. doi:10.1534/genetics.111.128819

42. 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. doi:10.1534/genetics.109.108068

43. 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 Genetics. 2012;8: e1002570. doi:10.1371/journal.pgen.1002570

44. 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. doi:10.1038/nature11867

45. Parts L, Cubillos FA, Warringer J, Jain K, Salinas F, Bumpstead SJ, et al. Revealing the genetic structure of a trait by sequencing a population under selection. Genome Res. 2011;21: 1131–1138. doi:10.1101/gr.116731.110

46. Hittinger CT, Goncalves P, Sampaio JP, Dover J, Johnston M, Rokas A. Remarkably ancient balanced polymorphisms in a multi-locus gene network.

Nature. 2010;464: 54–U61. doi:10.1038/nature08791

47. Will JL, Kim HS, Clarke J, Painter JC, Fay JC, Gasch AP. Incipient balancing selection through adaptive loss of aquaporins in natural Saccharomyces cerevisiae populations. PLoS Genetics. 2010;6: e1000893. doi:10.1371/journal.pgen.1000893

48. Kao KC, Sherlock G. Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. Nat Genet. 2008;40: 1499–1504. doi:10.1038/ng.280

49. Smith EN, Kruglyak L. Gene–Environment Interaction in Yeast Gene Expression. Mackay T, editor. PLoS Biol. 2008;6: e83. doi:10.1371/journal.pbio.0060083.st004

50. Ars E, Serra E, García J, Kruyer H, Gaona A, Lázaro C, et al. Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Human Molecular Genetics. 2000;9: 237–247.

51. Smukalla S, Caldara M, Pochet N, Beauvais A, Guadagnini S, Yan C, et al. FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. Cell. 2008;135: 726–737. doi:10.1016/j.cell.2008.09.037

52. Koschwanez JH, Foster KR, Murray AW. Sucrose utilization in budding yeast as a model for the origin of undifferentiated multicellularity. PLoS Biol. 2011;9: e1001122. doi:10.1371/journal.pbio.1001122

53. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang S-P, et al. A catalog of neutral and deleterious polymorphism in yeast. PLoS Genetics. 2008;4: e1000183. doi:10.1371/journal.pgen.1000183

54. Li YF, Costello JC, Holloway AK, Hahn MW. “Reverse ecology” and the power of population genomics. Evolution. 2008;62: 2984–2994. doi:10.1111/j.1558-5646.2008.00486.x

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

56. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. Current Biology. 1994;4: 24–33.

57. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genetics. 2012;8: e1002764. doi:10.1371/journal.pgen.1002764

58. Acar M, Becskei A, van Oudenaarden A. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435: 228–232. doi:10.1038/nature03524

59. Li L, Kowal AS. Environmental regulation of prions in yeast. PLoS Pathog. 2012;8: e1002973. doi:10.1371/journal.ppat.1002973

60. Halfmann R, Jarosz DF, Jones SK, Chang A, Lancaster AK, Lindquist S. Prions are a common mechanism for phenotypic inheritance in wild yeasts. Nature. 2012;482: 363–368. doi:10.1038/nature10875

61. Lo WS, Dranginis AM. FLO11, a yeast gene related to the STA genes, encodes a novel cell surface flocculin. Journal of Bacteriology. 1996;178: 7144–7151.

62. Lambrechts MG, Bauer FF, Marmur J, Pretorius IS. Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal differentiation in yeast. Proc Natl Acad Sci USA. 1996;93: 8419–8424.

63. Reynolds TB, Fink GR. Bakers' yeast, a model for fungal biofilm formation. Science. 2001;291: 878–881. doi:10.1126/science.291.5505.878

64. Van Mulders SE, Christianen E, Saerens SMG, Daenen L, Verbelen PJ, Willaert R, et al. Phenotypic diversity of Flo protein family-mediated adhesion in Saccharomyces cerevisiae. FEMS Yeast Res. 2009;9: 178–190. doi:10.1111/j.1567-1364.2008.00462.x

65. Douglas LM, Li L, Yang Y, Dranginis AM. Expression and characterization of the flocculin Flo11/Muc1, a Saccharomyces cerevisiae mannoprotein with homotypic properties of adhesion. Eukaryotic Cell. 2007;6: 2214–2221. doi:10.1128/EC.00284-06

66. Govender P, Kroppenstedt S, Bauer FF. Novel wine-mediated FLO11 flocculation phenotype of commercial Saccharomyces cerevisiae wine yeast strains with modified FLO gene expression. FEMS Microbiol Lett. 2011;317: 117–126. doi:10.1111/j.1574-6968.2011.02219.x

67. Smith A, Ward MP, Garrett S. Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. EMBO J. Nature Publishing Group; 1998;17: 3556–3564.

68. Markwardt DD, Garrett JM, Eberhardy S, Heideman W. Activation of the Ras/cyclic AMP pathway in the yeast Saccharomyces cerevisiae does not prevent G1 arrest in response to nitrogen starvation. Journal of Bacteriology. 1995;177: 6761–6765.

69. Budovskaya YV, Stephan JS, Reggiori F, Klionsky DJ, Herman PK. The Ras/cAMP-dependent protein kinase signaling pathway regulates an early

step of the autophagy process in Saccharomyces cerevisiae. J Biol Chem. ASBMB; 2004;279: 20663–20671.

70. Kwast KE, Burke PV, Staahl BT, Poyton RO. Oxygen sensing in yeast: evidence for the involvement of the respiratory chain in regulating the transcription of a subset of hypoxic genes. Proc Natl Acad Sci USA. 1999;96: 5446–5451.

71. Dejean L, Beauvoit B, Alonso A-P, Bunoust O, Guérin B, Rigoulet M. cAMP-induced modulation of the growth yield of Saccharomyces cerevisiae during respiratory and respiro-fermentative metabolism. Biochim Biophys Acta. 2002;1554: 159–169.

72. Wang Y, Pierce M, Schneper L, Güldal CG, Zhang X, Tavazoie S, et al. Ras and Gpa2 mediate one branch of a redundant glucose signaling pathway in yeast. PLoS Biol. 2004;2: E128. doi:10.1371/journal.pbio.0020128

73. Santangelo GM. Glucose signaling in Saccharomyces cerevisiae. Microbiol Mol Biol Rev. Am Soc Microbiol; 2006;70: 253–282. doi:10.1128/MMBR.70.1.253–282.2006

74. Wray NR, Purcell SM, Visscher PM. Synthetic Associations Created by Rare Variants Do Not Explain Most GWAS Results. Flint J, editor. PLoS Biol. 2011;9: e1000579. doi:10.1371/journal.pbio.1000579.s001

75. Güldener U, Heck S, Fiedler T, Beinhauer J, Hegemann JH. A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res. Oxford Univ Press; 1996;24: 2519–2524.

76. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5: e9490. doi:10.1371/journal.pone.0009490

77. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA. National Acad Sciences; 1979;76: 5269–5273.

78. Ewing B, Green P. Base-calling of automated sequencer traces usingPhred. II. error probabilities. Genome Res. Cold Spring Harbor Lab; 1998;8: 186–194. doi:10.1101/gr.8.3.186

79. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25: 1189–1191. doi:10.1093/bioinformatics/btp033

Chapter 3

Evolution of a membrane protein regulon in *Saccharomyces*

The contents of this chapter are based on the following publication:

Martin, HC*, Roop, JI*, Schraiber JG*, Hsu, TY, Brem, RB. Evolution of a membrane protein regulon in *Saccharomyces. Mol Biol Evol,* January 2012
*equal contribution by these authors

Abstract

Expression variation is widespread between species. The ability to distinguish regulatory change driven by natural selection from the consequences of neutral drift remains a major challenge in comparative genomics. In this work, we used observations of mRNA expression and promoter sequence to analyze signatures of selection on groups of functionally related genes in omycete yeasts. In a survey of gene regulons with expression divergence between *S. cerevisiae* and *S. paradoxus*, we found that most were subject to variation in *trans*-regulatory factors that provided no evidence against a neutral model. However, we identified one regulon of membrane protein genes controlled by unlinked *cis*- and *trans*-acting determinants with coherent effects on gene expression, consistent with a history of directional, non-neutral evolution. For this membrane protein group, *S. paradoxus* alleles at regulatory loci were associated with elevated expression and altered stress responsiveness relative to other yeasts. In a phylogenetic comparison of promoter sequences of the membrane protein genes between species, the *S. paradoxus* lineage was distinguished by a short branch length, indicative of strong selective constraint. Likewise, sequence variants within the *S. paradoxus* population, but not across strains of other yeasts, were skewed toward low frequencies in promoters of genes in the membrane protein regulon, again reflecting strong purifying selection. Our results support a model in which a distinct expression program for the membrane protein genes in *S. paradoxus* has been preferentially maintained by negative selection as the result of an increased importance to organismal fitness. These findings illustrate the power of integrating expression- and sequence-based tests of natural selection in the study of evolutionary forces that underlie regulatory change.

Introduction

An outstanding question in comparative genomics is the evolutionary importance of gene expression differences between genetically distinct individuals. Comparative expression studies in many taxa have made clear that a large fraction of the transcriptome varies in expression level between species [1-3], but the vast majority of this divergence is expected to be the product of neutral genetic drift. Due in part to the complexities of detecting selection from observations of regulatory sequence and gene expression [4-9], the prevalence and the mechanisms of natural selection on regulatory change remain incompletely understood.

In the search for an evolutionary logic underlying species changes in gene expression, many analyses have focused on trends across groups of functionally related genes. One powerful approach has been to trace gains or losses of a given *cis*-regulatory motif in the promoter sequences of genes in a pathway [10-20], where each *cis*-regulatory change has arisen via an independent genetic event. Such analyses have highlighted a complex, polygenic mechanism for the evolution of a given regulatory program, involving a suite of subtly acting variants in unlinked genes that function together. As a complementary approach for the study of polygenic regulatory evolution, we recently developed an expression-based strategy [1-3,7] to identify cases in which a species has accumulated *cis*-regulatory variants that predominantly up-regulate, or predominantly down-regulate, genes of common function. Given the neutral expectation of equivalent numbers of variants acting in one direction versus another [4-9,21]an imbalance in the signs of *cis*-regulatory effects represents a key line of evidence for a change in selective pressure between species on the pathway: positive selection or a relaxation of purifying selection in a given lineage could give rise to concerted regulatory change across genes of related function[7,10-20].

To date, a primary challenge of this and related expression-based tests for polygenic regulatory evolution [7,8,22] has been the ability to interpret the results in the context of classical sequence-based signatures of natural selection. In this work, we set out to develop an analysis pipeline to study cases of pathway-level regulatory evolution, harnessing both expression and DNA sequence data. To provide a tractable and data-rich model system for our analysis, we chose the well-characterized Saccharomycete yeasts.

Results

We first sought to screen groups of functionally related genes for coherent patterns of regulatory change between *S. cerevisiae* and *S. paradoxus*. For this purpose, we used measurements of gene expression in each species grown in rich medium [23]. In each of a set of gene groups defined on the basis of co-regulation in *S. cerevisiae*[10], we assessed the tendency for one species to express the genes of the pathway at predominantly higher, or predominantly lower, levels relative to the other species. The results revealed 11 gene groups with evidence for coherent regulatory variation between *S. cerevisiae* and *S. paradoxus*, at a level where ~2 groups were expected under a null model of independent evolution across genes (Table 1). Top-scoring gene groups in this analysis included heat-shock and stress-response genes, translation genes, cell cycle factors, and protein processing genes, indicating that a range of stress-response and housekeeping functions have been subject to directional expression change between the two yeasts.

Because coherent differential expression between species in a group of functionally related genes does not provide evidence for natural selection per se, we next analyzed the mechanisms of expression change between *S. cerevisiae* and *S. paradoxus* in regulons. For this purpose, we used expression measurements from a

stable hybrid diploid formed by mating the two species. Combining observations of allele-specific expression in the hybrid with analysis of the species when grown independently allows regulatory divergence to be partitioned into *cis*- and *trans*-acting contributions for each gene[23]. In each top-scoring gene group from our screen for coherent expression change, we tested a model in which one species' *cis*-regulatory alleles drove expression in the same direction relative to the allele from the other species[7], and we also carried out an analogous test using *trans*-regulatory effects. Across the groups, statistical significance measures were more striking for the latter test method (Table 1); thus, in the divergence between this pair of Saccharomycetes, detectable cases of coherent regulatory variation most often followed a model in which genetic change at *trans*-acting factors affects expression of a set of downstream targets in the same direction.

To investigate the role of non-neutral evolutionary forces underlying expression variation between *S. cerevisiae* and *S. paradoxus*, we focused on a set of genes mediating membrane protein trafficking and function and membrane lipid composition (Figure 1 and Supplementary Table 1). This gene group showed strong evidence for coherent regulatory change of *cis*- and *trans*-acting factors (reaching a significance level in each case where <0.1 group would be expected under the null; Table 1). For both *cis*- and *trans*-acting variants affecting expression of this gene group, the *S. paradoxus* allele at the respective locus conferred high expression relative to that of *S. cerevisiae* (Figure 1). The presence of many unlinked *cis*-regulatory variants between species acting in the same direction is unlikely under a neutral model[7], and indicative of a change in selective pressure on the regulation of the membrane protein set. The additional pattern of reinforcement between *cis*- and *trans*-mediated changes impinging on this regulon provides evidence against a model in which compensatory variants in a given species have arisen, in regulators and their targets, to preserve a constant degree of DNA-binding activity.

We next aimed to trace the evolutionary history of expression change in the membrane protein regulon across Saccharomycetes. We developed an analysis scheme harnessing allele-specific expression data sets from multiple inter-species yeast hybrids grown in rich medium. We first used RNA-seq to assess allele-specific expression for each gene in turn in a diploid hybrid formed from the mating of *S. kudriavzevii* and *S. cerevisiae*, and we also tabulated the analogous measurements from a hybrid of the latter and *S. bayanus*[7]. We then integrated these data with allele-specific expression measurements from the hybrid between *S. cerevisiae* and *S. paradoxus*[23], and, for each species comparison involving *S. paradoxus*, we tested for an imbalance in the direction of *cis*-regulatory changes in the membrane protein regulon. The results mirrored our analysis of *S. cerevisiae* (Supplementary Table 1): *cis*-regulatory variants at the membrane protein genes were associated with high expression in *S. paradoxus* relative to *S. bayanus* ($p = 0.007$) and to *S. kudriavzevii* ($p = 0.01$). Conclusions were unchanged when we analyzed the qualitative signs of *cis*-regulatory variation between species, rather than the quantitative effects (Supplementary Table 1). We conclude that the *cis*-regulatory alleles harbored by *S. paradoxus* at the genes of the membrane protein regulon confer elevated expression

relative to three other Saccharomycete species, providing strong evidence for a history of distinct selective pressure at these loci in *S. paradoxus*.

Given the regulatory divergence between *S. paradoxus* and other yeasts in the membrane protein group, we hypothesized that sequence-based signatures of selection would exhibit distinct characteristics across the genes of the regulon in the former species. To test this, we first sought to estimate the strength of selection on promoters of the membrane protein regulon, using genome sequences of isolates from the European population of *S. paradoxus*[24], the wine/European population of *S. cerevisiae*[24], and the Portuguese population of *S. kudriavzevii*[25]. For each population, we tabulated allele frequencies at single-nucleotide polymorphisms in all gene promoters, and we used the resulting folded site frequency spectra as input into the Poisson Random Field method for estimation of population-scaled selection coefficients γ[26]. As expected [27,28], for the whole-genome set of promoters in each species, likelihood ratio testing strongly rejected a model of neutrality in favor of an inference of purifying selection (*S. paradoxus*: $\gamma = -2.64$, $p = 7.87e\text{-}122$; *S. cerevisiae*: $\gamma = -1.1$, $p = 3.2e\text{-}7$; *S. kudriavzevii*: $\gamma = -0.72$, $p = 4.1e\text{-}17$). In *S. paradoxus*, the membrane protein gene promoters harbored an excess of low-frequency alleles compared to the genomic promoter set, reflecting the stronger action of purifying selection in culling variants from the population (Figure 2). Likelihood ratio testing confirmed the more negative selection coefficient for the membrane protein gene promoters as a better fit to the data than the parameter value inferred from the whole-genome promoter set (Figure 2). As an independent test, we conducted a resampling-based analysis of selection coefficients (see Methods); the results confirmed the difference between the membrane protein group and the rest of the genome for *S. paradoxus* ($p = 0.03$). By contrast, in *S. cerevisiae* and *S. kudriavzevii*, selection coefficients inferred from promoters of the membrane protein regulon were indistinguishable from those of the rest of the genome (likelihood ratio test $p = 0.4$ and 0.94 and resampling $p = 0.2$ and 0.54, respectively). Analyses of raw allele frequencies rather than selection coefficients (see Methods) also confirmed the difference between the membrane protein regulon and the rest of the genome in *S. paradoxus* (resampling $p = 0.03$) but not *S. cerevisiae* or *S. kudriavzevii* ($p = 0.66$ and 0.43, respectively). Thus, in *S. paradoxus* alone, allele frequencies were indicative of tight constraint on the membrane protein group, reflecting a particular importance of regulation at these loci in the niche occupied by this species.

To substantiate this conclusion by an independent method, we next applied a phylogenetic strategy, using promoter sequences from five Saccharomycete type strains[29]. We inferred evolutionary rates for each branch of the *Saccharomyces* phylogeny (Supplementary Figure 2) with PAML[30], for promoters of genes in the membrane protein regulon, and we compared the rates along the terminal branches to those inferred from promoters of randomly sampled gene groups. In *S. paradoxus*, but none of the other Saccharomycetes, the genes of the membrane protein regulon exhibited shorter branch lengths than the genomic null, corresponding to a slower rate of evolution (Supplementary Table 2). Additionally,

branch-specific rates for the membrane protein regulon were elevated in the distantly related species *S. bayanus*, suggesting a separate trend for relaxed purifying selection in the latter (Supplementary Table 2). We conclude that the rate of evolutionary change has been constrained at the promoters of the membrane protein genes in *S. paradoxus*, echoing results from population genetic analyses (Figure 2) and lending further support to the inference that regulation of these genes has been particularly important in the life history of *S. paradoxus*.

The membrane protein regulon includes a number of genes known to mediate response to cell wall and salt stress in *S. cerevisiae* (Supplementary Table 1). We expected that expression of the regulon was likely to be responsive to stress, and given the change in expression between species in rich-medium conditions (Figure 1 and Supplementary Table 1), we hypothesized that aspects of stress regulation of the membrane protein genes would have diverged between *S. paradoxus* and other yeasts. To test these hypotheses, we used gene expression measurements from *S. cerevisiae* and *S. paradoxus* grown in nutrient starvation conditions and during exposure to toxic agents[31]. As predicted, the membrane protein regulon stood out relative to the rest of the genome for its expression regulation across stress conditions (Figure 3 and Supplementary Figure 1): both *S. cerevisiae* and *S. paradoxus* induced the regulon under nitrogen starvation and growth on glycerol, and repressed it in response to the DNA-damaging agent MMS. Critically, however, expression response to environmental treatments was markedly different between the species (Figure 3). *S. paradoxus* repressed membrane gene expression to a greater extent in MMS than did *S. cerevisiae* and induced the membrane genes less in other conditions than did *S. cerevisiae*. Standard sequence search analyses had little power to detect motif changes between species in promoters of the membrane protein regulon (data not shown), as expected if the complexity of regulatory inputs at these loci served to obscure the signal of gain or loss of any given motif. In light of our observation of elevated expression levels of the membrane protein genes in rich medium in *S. paradoxus* (Figure 1), we conclude that the two species ultimately achieve similar regulatory programs for this gene group in stress conditions, starting from distinct set points in the rich-medium basal state: since the membrane protein genes are expressed in rich medium at a higher level in *S. paradoxus* than in *S. cerevisiae*, it follows that *S. paradoxus* requires more dramatic repression upon a switch to MMS treatment and less avid induction during nitrogen limitation and growth on glycerol. Such patterns of divergent stress response further highlight the evolutionary dynamics of regulation of the membrane protein gene group between yeast species.

Discussion

Distinguishing between natural selection and neutral drift as forces underlying regulatory variation remains a major challenge in evolutionary biology. Analysis of directional expression change in gene groups of common function can be a powerful tool toward this end. To date, however, methods have been at a premium for the incorporation of expression-based tests for selection with those based on DNA

sequence. In this work, we have used Saccharomycete yeasts as a testbed for these complementary paradigms in the dissection of pathway-level regulatory change.

We have shown that, in a comparison of *S. cerevisiae* and *S. paradoxus*, most instances of directional regulatory change in groups of functionally related genes can be explained by variation in *trans*-acting factors. For a given such pathway, an appealing mechanism invokes a variant in a single upstream factor, or a small number of such loci, driving directional changes in expression of downstream targets. Under this model, the relatively short waiting time required for a species to accumulate mutations in a small number of *trans*-acting factors would be consistent with the prevalence of *trans*-acting variation we observe across pathways. Importantly, any case of simple, Mendelian *trans*-acting regulatory change provides no a priori means to reject a neutral model [32,33]. We hypothesize that many of the *trans*-regulatory changes in the pathways we study may be present in yeast genomes as a consequence of drift.

As a model system for the study of non-neutral regulatory evolution, we focused on a group of genes involved in membrane protein function and trafficking and membrane lipid composition. *Cis*- and *trans*-regulatory variants harbored by *S. paradoxus* at these loci were associated with upregulation in rich-medium conditions and altered stress responsiveness relative to other yeasts. Our observation of multiple independent *cis*-regulatory variants driving expression of the membrane protein genes in the same direction is unlikely under neutrality [7] and provides evidence for a change in selective pressure in *S. paradoxus*. Additionally, the fact that alleles of *cis*- and *trans*-acting factors in *S. paradoxus* affect this gene set in the same direction argues against a model in which target genes have accumulated locally acting variants to compensate for changes in a soluble regulatory factor. Instances of such compensation have been common during the divergence of yeast species [23,34-36], consistent with the hypothesis that directional selection rarely underlies these patterns [37]. By contrast, an inference of a change in selective pressure in a given species becomes strongest when *cis*- and *trans*-acting variants impinging on a suite of genes drive expression in a consistent direction, as has been reported in a handful of previous studies [8,14] and in the present observations of the membrane protein group.

A priori, polygenic, directional expression change across the membrane protein regulon could be the result of either adaptation or relaxed purifying selection in *S. paradoxus*. Our findings provide strong evidence against the latter hypothesis. Inference using membrane protein promoters from *S. paradoxus* revealed short branch lengths on the *Saccharomyces* phylogenetic tree, and a strongly negative selection coefficient within species, relative to genomic controls. These findings dovetail with previous reports of shifts in the strength of purifying selection between gene groups [28] and between populations [38] in yeast. How might we reconcile the distinct expression program of the membrane protein genes in *S. paradoxus* with the evidence for tight selective constraint? Our results are consistent with either of two possible interpretations. The *S. paradoxus* expression

program could represent an ancestral state from which all other species have diverged; under this model, ancestral alleles have been subject to tight constraint in *S. paradoxus* as a consequence of particular importance to organismal fitness relative to that in the niches of other yeasts. Alternatively, the *S. paradoxus* expression program could represent a derived state that arose through a series of selective sweeps and since then has been preferentially maintained by negative selection. As we show in the Appendix, the latter model could manifest as a reduced number of substitutions along the *S. paradoxus* lineage if the period of adaptive evolution were sufficiently ancient, a plausible scenario for the membrane protein group given that *S. paradoxus* diverged from its closest relative, *S. cerevisiae*, ~7 million years ago[29]. Under either model, it is tempting to speculate that the distinct regulatory program we have uncovered in *S. paradoxus* affects acute tolerance to environmental insults in the wild, as Saccharomycetes respond to many stress treatments by upregulating the membrane protein gene group (Supplementary Figure 1). The elevated constitutive expression of this regulon in *S. paradoxus* is in keeping with other switches between constitutive and stress-responsive expression observed in yeasts[39], whose fitness advantages remain an area of active research.

Decades of work in genetic mapping have revealed suites of unlinked, often weakly-acting loci to be the rule rather than the exception in explaining trait variation within species [40]. The notion of genetic complexity has also become increasingly relevant in comparisons between species, as genomic methods enable studies of polygenic evolution over long timescales [10-20,35]. Understanding the evolutionary pressures at play in polygenic pathway evolution will require both observational molecular approaches and formal tests of natural selection, which together will continue to accelerate the dissection of the genetic basis of evolutionary novelties.

Materials and methods

Selection of regulons

We downloaded regulons from [10] and, from each, eliminated genes annotated as dubious ORFs (www.yeastgenome.org). We then filtered groups as follows. Any groups with fewer than 10 genes were eliminated from further analysis. Of the remainder, we compared the composition of each pair of gene groups and, if any two groups overlapped by more than 30%, we eliminated the smaller group from consideration. The final data set comprised 104 regulons.

Expression-based tests for directional regulatory evolution between S. paradoxus and S. cerevisiae

We downloaded estimates of the *cis*- and *trans*-acting contributions to gene expression divergence between *S. cerevisiae* and *S. paradoxus* from[23]. These data derived from two measurements for each gene: one of the $\log_2$ of the ratio of

expression of the *S. cerevisiae* allele to the expression of the *S. paradoxus* allele in an interspecific hybrid, reflecting the *cis*-acting contribution to expression variation between the species ($R_{Sc/Sp;cis}$), and the other of the total difference in expression between the parent species when grown independently in culture, $R_{Sc/Sp;total}$ = $\log_2(expression_{Scer}/expression_{Spar})$. As detailed in[23], the *trans*-acting contribution to variation for a given gene, $R_{Sc/Sp;trans,}$ is estimated as $R_{Sc/Sp;total}$ - $R_{Sc/Sp;cis}$; thus, from the $R_{Sc/Sp;cis}$ and $R_{Sc/Sp;trans}$ values publicly available from[23], we calculated $R_{Sc/Sp;total}$ for each gene. To first survey regulons for directional expression change between species without distinguishing between *cis*- and *trans*-acting mechanisms, we used the sum of expression effects across a gene group as a statistic to assess the coherence of regulatory divergence. Specifically, for each regulon we summed $R_{Sc/Sp;total}$ across all genes in the regulon; to evaluate significance, we conducted a two-sided test against a null distribution of 5000 groups comprised of randomly chosen genes from the genome, in which each such null group was of the same size as the regulon of interest. In Table 1, the experiment-wise false positive count for this test was calculated as the product of the number of groups tested (104) and the *p*-value threshold used (0.0204). For each regulon, we then conducted analogous tests for imbalance in *cis*- and *trans*-acting contributions to expression between the species using $R_{Sc/Sp;cis}$ and $R_{Sc/Sp;trans}$, respectively, except that a one-sided *p*-value was calculated in each case (upper-tailed if $\Sigma R_{Sc/Sp;total} > 0$ and lower-tailed otherwise). The experiment-wise significance for the membrane protein group for the latter tests was calculated as the product of the number of groups tested (11) and the *p*-value for the membrane protein regulon (0.0056 and 0.001, respectively). In Supplementary Table 1, we also converted each $R_{Sc/Sp;cis}$ value to a sign statistic (+1 if *S. cerevisiae* was associated with higher expression and -1 otherwise) and repeated resampling tests as above, using a one-sided test in each case.

Expression-based test for directional regulatory evolution between S. paradoxus and S. bayanus

We downloaded measurements of $R_{Sb/Sc,}$ the *cis*-acting contribution to regulatory variation between *S. bayanus* and *S. cerevisiae,* for each gene, from[7]. From this data set, we eliminated from consideration all genes that were not represented in the data set comparing *S. paradoxus* and *S. cerevisiae*[23], as well as those annotated as dubious ORFs. We then calculated $R_{Sb/Sp}$, the sum of $R_{Sb/Sc}$ and $R_{Sc/Sp}$, for each gene, and used this quantity in a one-sided resampling test for directional *cis*-regulatory evolution as above.

Expression-based test for directional regulatory evolution between S. paradoxus and S. kudriavzevii

We mated *S. cerevisiae* BY4716 (Open Biosystems) and *S. kudriavzevii* strain JRY9187 [29] to generate a diploid hybrid. To generate measurements of $R_{Sk/Sc,}$ the log of the fold-change in allele-specific expression levels between *S. cerevisiae* and *S. kudriavzevii,* for each gene, we conducted yeast culture, RNA isolation, and mapping as previously described [7] except that one sequencing lane was used from one

biological replicate, with the total RNA-seq data set comprising 7.04 million mapped reads. We eliminated from consideration all genes that were not represented in the data set comparing *S. paradoxus* and *S. cerevisiae*[23], as well as those annotated as dubious ORFs. We then formulated $R_{Sk/Sp}$, the sum of $R_{Sk/Sc}$ and $R_{Sc/Sp}$, for each gene, and used this quantity in a one-sided resampling test for directional *cis*-regulatory evolution as above.

Inter-specific sequence analyses

Reference sequences for *S. cerevisiae*, *S. paradoxus, S. bayanus*, *S. mikatae* and *S. kudriavzevii* were downloaded from www.*Saccharomyces*sensustricto.org[29]. For each species, we extracted the region 1000 bp upstream of each gene for all genes not annotated as dubious ORFs, and we eliminated any gene that was not annotated in each of the five species in[29]. For each such set of promoter sequences from a given gene, a five-species alignment was generated using FSA[41]. Alignments in which >500 sites were gaps or ambiguous base calls were eliminated from analysis, yielding a final data set of promoters from 4295 genes. For branch length inference, the aligned promoters from the genes of the membrane protein regulon were concatenated and used as input to the baseml module of PAML [30] along with a fixed unrooted Saccharomycete tree topology[29]. A general time reversible substitution model was used without a molecular clock, and sites in the alignment that contained gaps or ambiguous data were not used in the analysis. To evaluate statistical significance of branch lengths for the membrane protein regulon in Supplementary Table 2, the concatenation of promoter alignments and branch length inference was repeated on each of 10,000 randomly sampled gene groups of the same size as the membrane protein group, and a one-sided empirical *p*-value for each branch was then calculated as the proportion of such null groups with branch length greater than or equal to the true value inferred from the real membrane protein genes. A complementary analysis applying branch length inference to aligned sequences for each promoter separately yielded identical results (data not shown). For Supplementary Figure 2, sequences of 5167 open reading frames were downloaded from[29], aligned, and concatenated, and branch lengths were inferred as above; the tree image was generated using MEGA[42].

Promoter data from Saccharomyces population genomic sequences and estimation of selection coefficients

Chromosome alignments for strains of the European clade of *S. paradoxus* and the wine/European clade of *S. cerevisiae* populations were downloaded from [24] and accessed using the alicat.pl script. Sites with an error probability of >0.0001 were eliminated from analyses, as were genes whose promoter alignments contained >5 segregating sites. Contig alignments for strains from the Portuguese clade of *S. kudriavzevii* were downloaded from[25]. For each species, we extracted the region 1000 bp upstream of each gene for all genes not annotated as dubious ORFs, or in the case of *S. kudriavzevii*, the longest possible upstream region up to 1000 bp for which sequence data were available. The final data sets comprised promoters for

5284, 5268, and 2748 genes in *S. paradoxus*, *S. cerevisiae*, and *S. kudriavzevii*, respectively, harboring 2670, 1908, and 7356 single-nucleotide polymorphisms, respectively.

To establish a data set for a given species in which the number of strain genomes without missing data was the same for each polymorphic site[38], we set a cutoff strain count *c* equal to 15, 6, and 10 for *S. paradoxus*, *S. cerevisiae*, and *S. kudriavzevii* respectively. If a given variant site had fewer than *c* strains with non-missing allele calls, we eliminated it from consideration, and if a site had more than *c* strains with non-missing allele calls, we randomly subsampled alleles from *c* of these strains for inclusion in the analysis. Given the resulting allele set for each promoter for a given species, we tabulated allele frequencies at all sites of single-nucleotide polymorphism in the complete set of promoters for a given gene group of interest. We implemented the Poisson Random Field (PRF) method [26] in R (www.r-project.org), using the nlm function to find maximum likelihood estimates of selection coefficients γ. For a given population, γ inferred from the whole-genome set of promoters was evaluated against the null hypothesis of a value of zero; γ from the membrane protein gene promoters was evaluated against the whole-genome estimate. In each case, likelihood ratio test *p*-values were calculated assuming a chi-square distribution of the test statistic with one degree of freedom. As a complementary approach, we evaluated the γ inferred from the promoters of the membrane protein regulon *via* resampling. For this purpose, we generated 10,000 null gene sets by randomly selecting genes from the genome, and using the promoters of each such set, we estimated a maximum likelihood selection coefficient as above; a one-sided empirical *p*-value was then calculated as the proportion of null groups with γ as negative or more so than the true value inferred from the real membrane protein genes. We also tested for a distinction between raw allele frequencies in promoters of the membrane protein regulon and those of the rest of the genome, by computing the mean allele frequency across all segregating sites in the former and comparing against analogous means of 10,000 null gene sets as above. Conclusions from all analyses were unaffected by changes to the cutoff used in filtering promoters based on the number of segregating sites (data not shown).

Expression-based tests for divergence in stress responsiveness between S. cerevisiae and S. paradoxus

We downloaded microarray measurements of expression divergence between *Saccharomyces* species when grown in independent culture in stress conditions from [31] and excluded dubious ORFs from consideration. We analyzed the transcriptional response to each of five conditions: growth on glycerol as a carbon source, nitrogen starvation, heat shock, and treatment with methyl methanesulfonate (MMS) and H_2O_2. For each gene and each stress, at each of six timepoints, we calculated the expression response as the $\log_2$ fold-change relative to the rich medium YPD, as an average across all probes on the microarray affiliated with the gene. For each species, in Supplementary Figure 1 we compared the

distributions of response values between the membrane protein regulon and the rest of the genome using a two-sided Wilcoxon test. In Figure 3, we calculated the difference between species in the stress response for each gene as $\log_2$(expression in $\text{stress}_{\text{Scer}}$/expression in YPD_{Scer}) - $\log_2$(expression in $\text{stress}_{\text{Spar}}$/expression in YPD_{Spar}), and we compared these inter-species differences in responsiveness between the membrane protein regulon and the rest of the genome using a one-sided Wilcoxon test.

Acknowledgements

We thank Chris Hittinger for providing the *S. kudriavzevii* strain; Angela Kaczmarczyk, Ivy McDaniel, and Oh Kyu Yoon for strain handling and culture; Adam Boyko for providing PRFREQ software; Jeremiah Degenhardt, Dan Pollard, Eva Stukenbrock, and Oliver Zill for discussions; and Chris Ellison, Nicole King, Daniel Ortiz-Barrientos, Hua Tang, and John Taylor for helpful comments on the manuscript. This work was supported by a Burroughs-Wellcome Career Award at the Scientific Interface, an Ellison New Scholar Award in Aging, and NIH R01-GM087432 to R.B.B. and NIH R01-GM40282 to M. Slatkin

Figures

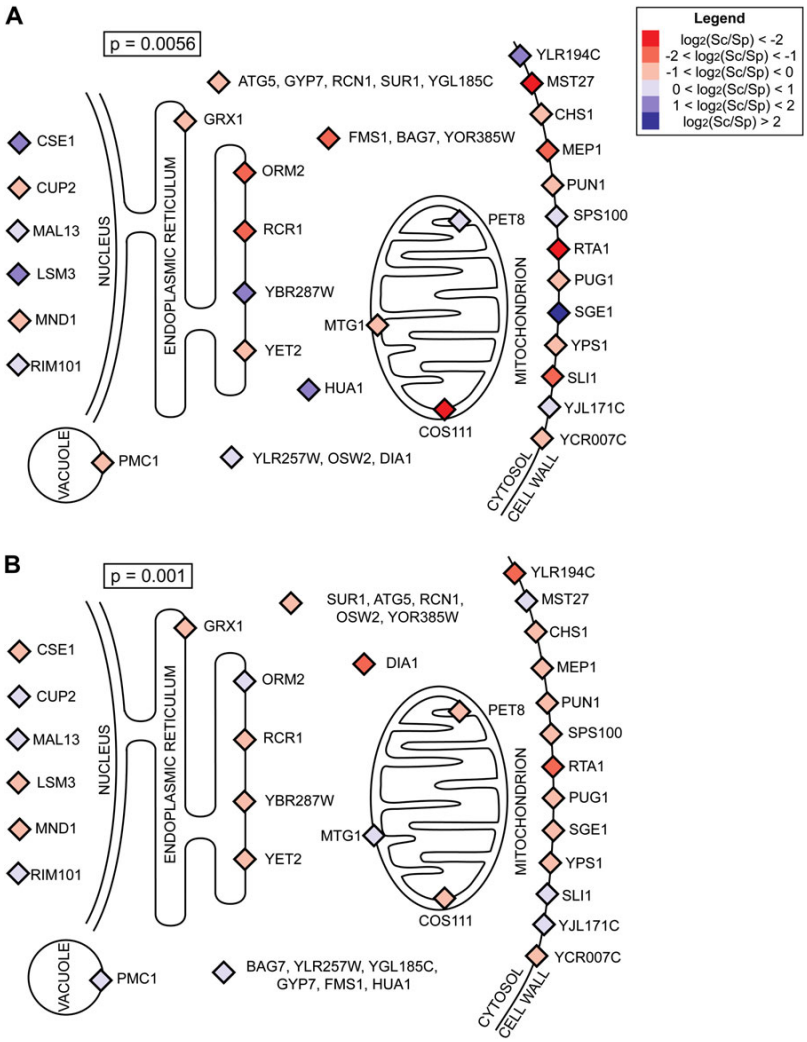


Figure 1. Directional regulatory change between Saccharomycetes in the membrane protein regulon. Each colored element represents regulatory changes between *S. cerevisiae* and *S. paradoxus* in rich medium [23] at one membrane protein gene, overlaid on a cartoon localization of the encoded protein. p, proportion of randomly sampled gene groups showing regulatory changes whose directions were at least as coherent as those observed in the membrane regulon (Table 1). (A) Signs of *cis*-regulatory effects. Orange, the allele from *S. paradoxus* was the most highly expressed in an interspecific hybrid; blue, the *S. cerevisiae* allele was most highly expressed. (B) Signs of *trans*-regulatory effects, which derive from expression measurements of interspecific hybrids and of parental strains grown independently[23]. Orange, the regulatory allele from *S. paradoxus* was associated with higher expression; blue, the *S. cerevisiae* allele was associated with higher expression.

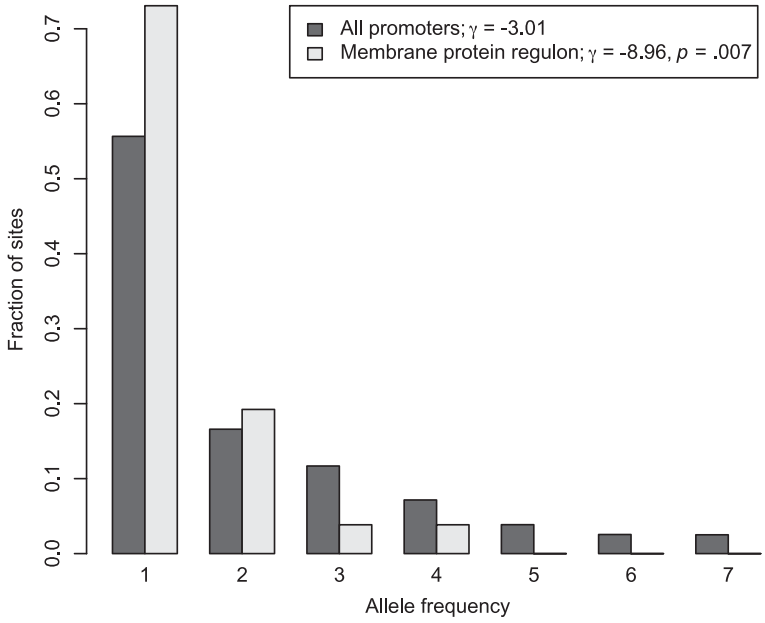


Figure 2. Increased strength of purifying selection in the membrane protein regulon in *S. paradoxus*. Each shade of grey represents analysis of single-nucleotide polymorphisms in one set of promoter sequences from European strains of *S. paradoxus*[24]. Each set of bars represents a histogram of frequencies of minor alleles. All promoters, analysis of promoter regions from all genes in the genome. Membrane protein regulon, analysis of promoter regions from the membrane protein gene group (Figure 1 and Supplementary Table 1). γ, population-scaled selection coefficient inferred from the allele frequency spectrum from the indicated set of promoters. *p*, significance of the distinction between the genomic selection coefficient and the coefficient inferred from the membrane protein group, according to likelihood ratio testing. Results are based on a data set of resampled alleles to ensure fully informative sequence data at each site (Materials and Methods).

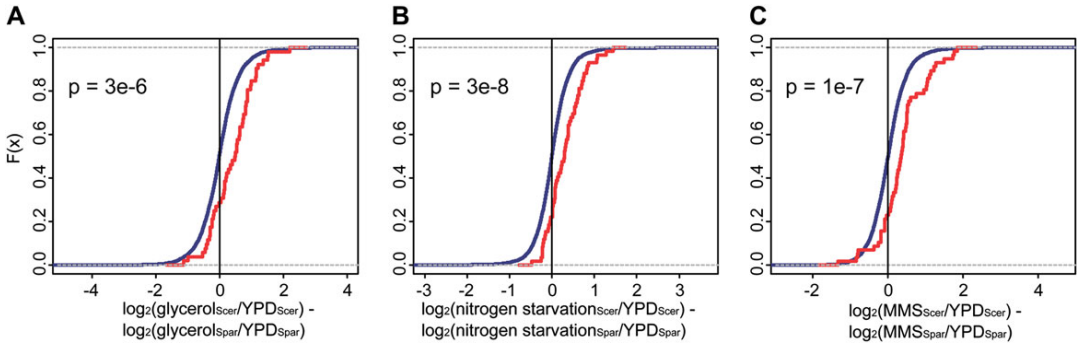


Figure 3. Inter-species differences between Saccharomycetes in stress responsiveness of the membrane protein regulon. Each trace represents a cumulative distribution function (F(x)) of the difference between *S. cerevisiae* and *S. paradoxus* in induction of gene expression after stress exposure, relative to expression in the rich medium YPD[31]. In each panel, the x-axis reports $\log_2$(expression in $\text{stress}_{\text{Scer}}$/expression in YPD_{Scer}) - $\log_2$(expression in $\text{stress}_{\text{Spar}}$/expression in YPD_{Spar}). The blue trace reports the distribution across all genes, and the red trace reports that across the membrane protein regulon (Figure 1 and Supplementary Table 1). (A) Switch to glycerol as the carbon source, 20-minute timepoint. (B) Nitrogen starvation, 45-minute timepoint. (C) Treatment with methyl methanesulfonate, 1.5-hour timepoint. p, one-sided *p*-value from a Wilcoxon test comparing expression changes in the membrane regulon to changes across the genome.

Tables

name[a]	# genes	$\Sigma R_{Sc/Sp;total}$[b]	p total[c]	p *cis*[d]	p *trans*[e]	Annotation
Cluster_adata-CalciumSpecific*	47	-25.8	<0.0002	0.0056	0.001	Membrane proteins
Node 7	183	85.8	<0.0002	0.0506	<0.0002	Heat shock proteins, stress
Cluster_PUF3	193	-46.3	<0.0002	0.2372	<0.0002	Mitochondrion
Node 73	392	-140.1	<0.0002	0.4644	<0.0002	Ribosome, translation
Cluster_ManualHAP1targets	36	-28.2	<0.0002	0.5044	<0.0002	Respiration
Cluster_HAP5	24	-12.6	0.0072	0.0114	0.108	Respiration
Cluster_Vacuole	56	17.1	0.0114	0.019	0.0434	Vacuole
Cluster_PHD1	27	-12.2	0.0114	0.0618	0.0232	Stress, signaling
Cluster_SWI4	48	-16.1	0.0136	0.071	0.0252	Cell cycle
Node 6	10	7.0	0.0178	0.1296	0.003	Protein modification and breakdown
Cluster_ARG80	18	-9.1	0.0204	0.1622	0.01	Amino acid biosynthesis

Table 1. Tests for directional regulatory change between *S. cerevisiae* and *S. paradoxus* expression in co-regulated gene groups.

[a] Gene groups defined by[10]. Results from analysis of the membrane protein regulon are denoted with an asterisk.

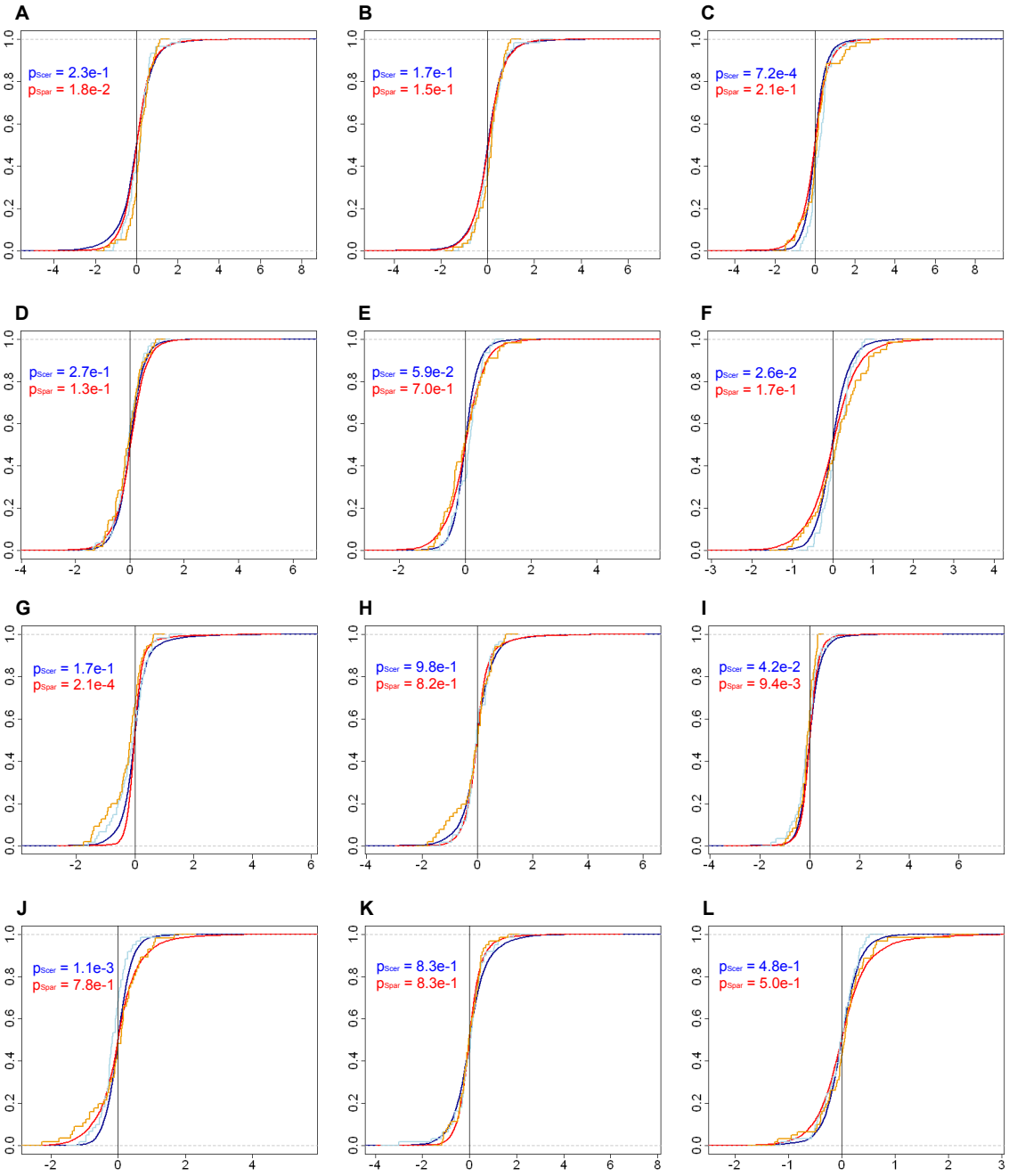
[b] Sum, across genes in the indicated regulon, of the $\log_2$ ratio of expression in *S. cerevisiae* to expression in *S. paradoxus*, when each species was grown in independent culture ($R_{Sc/Sp;\ total}$).

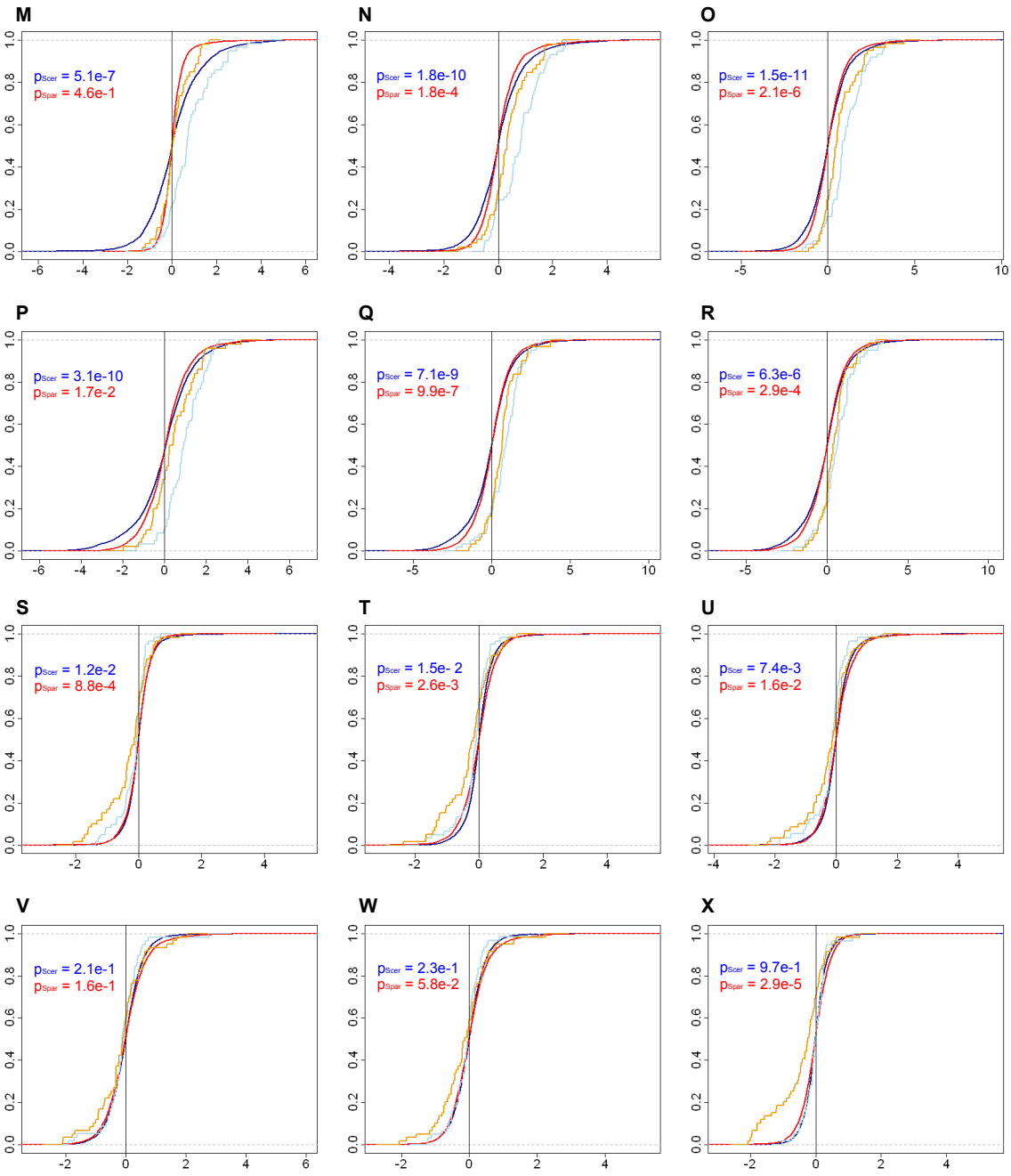
[c] Significance of a two-sided resampling test relative to the genomic null for an extreme value of the sum, across genes in the indicated regulon, of the $\log_2$ ratio of expression in *S. cerevisiae* to expression in *S. paradoxus* when each species was grown in independent culture ($R_{Sc/Sp;\ total}$). Shown are all groups meeting a significance level where 2.1 groups would be expected under the null.

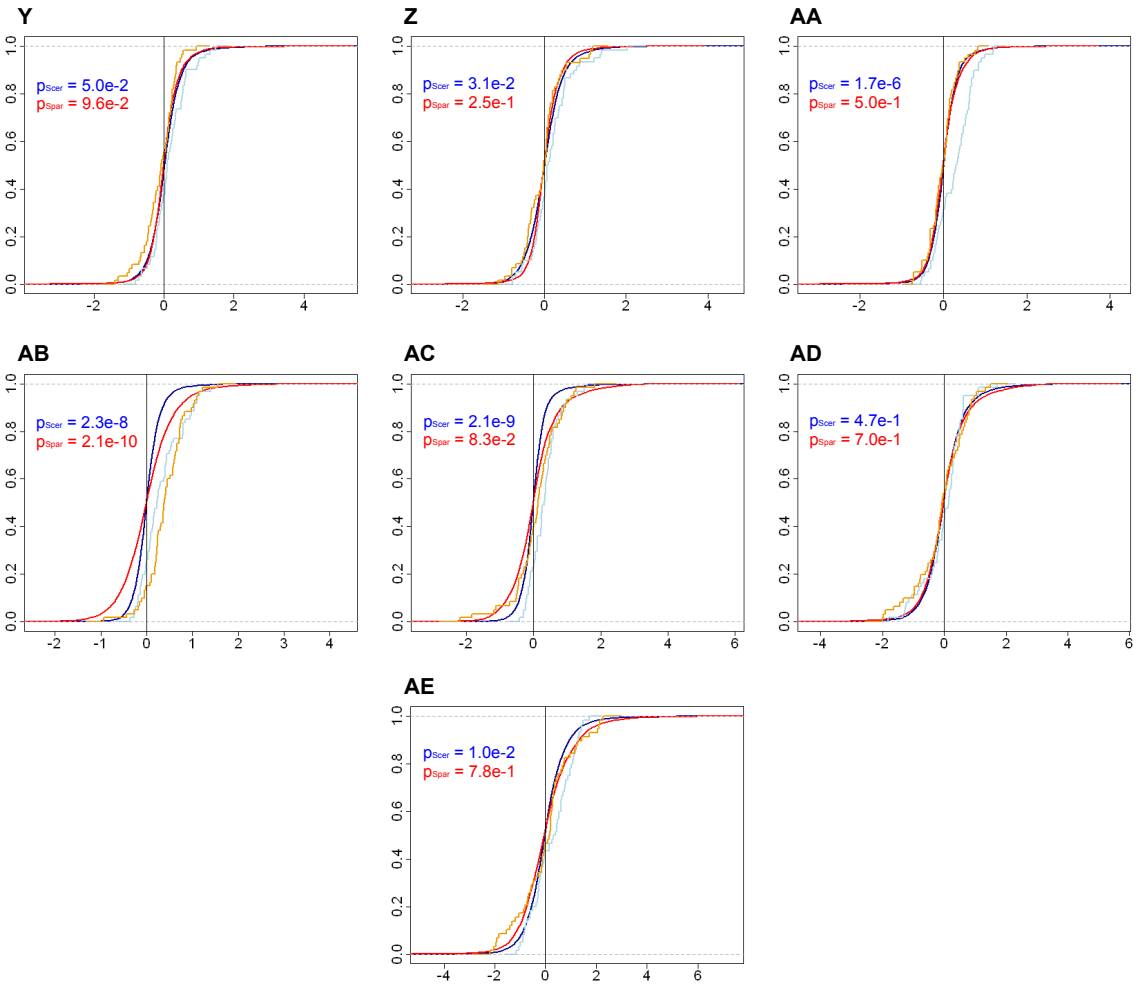
[d] Significance of a one-sided resampling test relative to the genomic null for an extreme value of the sum, across genes in the indicated regulon, of the $\log_2$ ratio of the expression of the *S. cerevisiae* allele to that of the *S. paradoxus* allele in an interspecific hybrid.

[e] Significance of a one-sided resampling test relative to the genomic null for an extreme value of the sum, across genes in the indicated regulon, of the $\log_2$ ratio of the effects of *trans*-acting regulatory variation between *S. cerevisiae* and *S. paradoxus*, where the latter is derived from expression measurements of the interspecific hybrid and of parent strains grown independently[23].

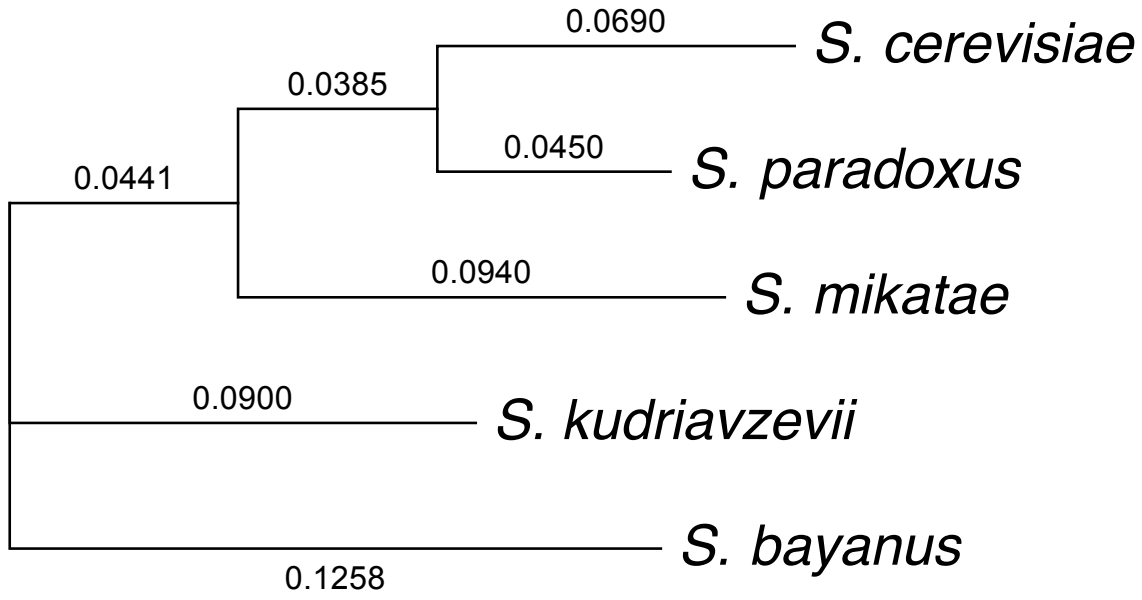
Supplementary figures




M
pScer = 5.1e-7
pSpar = 4.6e-1
N
pScer = 1.8e-10
pSpar = 1.8e-4
O
pScer = 1.5e-11
pSpar = 2.1e-6
P
pScer = 3.1e-10
pSpar = 1.7e-2
Q
pScer = 7.1e-9
pSpar = 9.9e-7
R
pScer = 6.3e-6
pSpar = 2.9e-4
S
pScer = 1.2e-2
pSpar = 8.8e-4
T
pScer = 1.5e- 2
pSpar = 2.6e-3
U
pScer = 7.4e-3
pSpar = 1.6e-2
V
pScer = 2.1e-1
pSpar = 1.6e-1
W
pScer = 2.3e-1
pSpar = 5.8e-2
X
pScer = 9.7e-1
pSpar = 2.9e-5



Supplementary Figure 1. The membrane protein regulon is regulated in response to stress in *S. paradoxus* and *S. cerevisiae*. Each trace represents a cumulative distribution function of expression change in one species and gene group after exposure to a given stress, relative to expression in the rich medium YPD[31]. In each panel, the x-axis reports $\log_2$(expression in stress/expression in YPD); dark blue, all genes in *S. cerevisiae*; light blue, genes of the membrane protein regulon in *S. cerevisiae*; red, all genes in *S. paradoxus*; orange, genes of the membrane protein regulon in *S. paradoxus*; p, two-sided *p*-value from Wilcoxon test comparing expression changes in stress between the indicated groups of genes for each species. Expression levels were profiled 10 minutes, 20 minutes, 30 minutes, 45 minutes, 1 hour and 1.5 hours after transfer from 30°C to 37°C (A-F), from YPD to media containing H_2O_2 (G-L), glycerol (M-R) and methyl methanesulfate (S-X), and after 10 minutes, 20 minutes, 45 minutes, 2 hours, 4 hours, 8 hours and 1 day (Y-AE) of nitrogen starvation.



Supplementary Figure 2. ***Saccharomyces sensu stricto*** **genome tree.** Numerical values indicate branch lengths in substitutions/site.

Supplementary Tables

A

systematic name	gene name	$R_{Sc/Sp;cis}$[a]	$R_{Sk/Sp;cis}$[b]	$R_{Sb/Sp;cis}$[c]	annotation[d]
YGR213C	RTA1	-4.40	-3.06	-2.04	Member of the fungal lipid-translocating exporter (LTE) family of proteins
YOR134W	BAG7	-2.00	-2.68	-0.96	Rho GTPase activating protein
YPL149W	ATG5	-0.76	-2.57	-1.23	Involved in autophagy and the Cvt pathway
YBR203W	COS111	-2.30	-2.29	-1.93	Protein required for resistance to the antifungal drug ciclopirox olamine
YNR059W	MNT4	-0.01	-1.98	N/A	Putative alpha-1,3-mannosyltransferase
YBR005W	RCR1	-1.44	-1.36	-1.35	Protein of the ER membrane involved in cell wall chitin deposition
YMR020W	FMS1	-1.22	-1.35	0.71	Polyamine oxidase, converts spermine to spermidine
YGR212W	SLI1	-1.18	-1.21	0.29	N-acetyltransferase, confers resistance to the sphingolipid biosynthesis inhibitor myriocin
YGR121C	MEP1	-1.24	-1.08	-0.07	Ammonium permease
YLR350W	ORM2	-1.04	-0.76	-0.28	Controls membrane biogenesis by coordinating lipid homeostasis with protein quality control
YDR319C	YDR319C	0.22	-0.53	N/A	Ortholog of the FIT family of proteins involved in triglyceride droplet biosynthesis
YML118W	NGL3	-1.43	-0.33	0.57	Putative endonuclease
YLR120C	YPS1	-0.65	-0.28	N/A	Aspartic protease, member of the yapsin family of proteases involved in cell wall growth and maintenance
YGL166W	CUP2	-0.53	-0.25	-0.50	Copper-binding transcription factor
YGL185C	YGL185C	-0.76	-0.24	-0.61	Putative protein with sequence similarity to hydroxyacid dehydrogenases
YDL010W	GRX6	-0.12	-0.22	-0.54	*Cis*-golgi localized monothiol glutaredoxin that binds an iron-sulfur cluster; involved in the oxidative stress response

YPL057C	SUR1	-0.78	-0.20	-0.12	Probable catalytic subunit of a mannosylinositol phosphorylceramide (MIPC) synthase; involved in sphingolipid biosynthesis
YMR097C	MTG1	-0.09	-0.12	-0.13	Putative GTPase peripheral to the mitochondrial inner membrane
YLR257W	YLR257W	0.06	-0.11	0.71	Putative protein of unknown function
YDL234C	GYP7	-0.17	-0.10	-0.30	GTPase-activating protein for yeast Rab family members; involved in vesicle mediated protein trafficking
YKL159C	RCN1	-0.73	-0.01	-1.13	Protein involved in calcineurin regulation during calcium signaling
YHL027W	RIM101	0.35	0.09	N/A	Transcriptional repressor involved in response to pH and in cell wall construction
YLR414C	PUN1	-0.62	0.23	-1.34	Plasma membrane protein with a role in cell wall integrity
YNL192W	CHS1	-0.02	0.26	0.29	Catalyzes the transfer of N-acetylglucosamine (GlcNAc) to chitin
YMR316W	DIA1	0.04	0.56	-1.27	Involved in invasive and pseudohyphal growth
YGL006W	PMC1	-0.24	0.63	N/A	Vacuolar Ca2+ ATPase involved in depleting cytosol of Ca2+ ions
YNL003C	PET8	0.78	0.68	0.97	S-adenosylmethionine transporter of the mitochondrial inner membrane
YMR096W	SNZ1	-1.44	0.78	N/A	Protein involved in vitamin B6 biosynthesis
YBR287W	YBR287W	1.41	0.96	0.89	Protein of unknown function
YGR268C	HUA1	1.63	1.23	N/A	Cytoplasmic protein containing a zinc finger domain with sequence similarity to that of Type I J-proteins
YJL171C	YJL171C	0.06	1.35	0.46	GPI-anchored cell wall protein of unknown function; induced in response to cell wall damaging agents
YNR010W	CSE2	1.89	1.89	2.24	Subunit of the RNA polymerase II mediator complex
YLR194C	YLR194C	1.87	2.14	1.44	Structural constituent of the cell wall attached to the plasma membrane by a GPI-anchor; upregulated in response to cell wall stress
YNR064C	YNR064C	-0.95	N/A	-3.61	Epoxide hydrolase, member of the alpha/beta hydrolase fold family

YOR385W	YOR385W	-1.62	N/A	-1.65	Putative protein of unknown function
YGL183C	MND1	-0.12	N/A	0.15	Protein required for recombination and meiotic nuclear division
YLR054C	OSW2	0.15	N/A	1.05	Protein of unknown function proposed to be involved in the assembly of the spore wall
YGR288W	MAL13	0.86	N/A	N/A	MAL-activator protein, part of complex locus MAL1
YER185W	PUG1	-0.81	N/A	N/A	Plasma membrane protein with roles in the uptake of protoprophyrin IX and the efflux of heme
YHR139C	SPS100	0.12	N/A	N/A	Protein required for spore wall maturation
YMR040W	YET2	-0.22	N/A	N/A	Protein of unknown function that may interact with ribosomes
YKL218C	SRY1	3.01	N/A	N/A	3-hydroxyaspartate dehydratase, deaminates L-threo-3-hydroxyaspartate to form oxaloacetate and ammonia
YPR198W	SGE1	3.26	N/A	N/A	Plasma membrane multidrug transporter of the major facilitator superfamily
YGL051W	MST27	-3.00	N/A	N/A	Putative integral membrane protein, involved in vesicle formation
YCR007C	YCR007C	-0.66	N/A	N/A	Putative integral membrane protein
YER184C	YER184C	0.00	N/A	N/A	Putative zinc cluster protein
YLR438C-A	LSM3	1.12	N/A	N/A	Lsm (Like Sm) protein; involved in mRNA decay

B

	S. cerevisiae-S. paradoxus	*S. kudriavzevii-S. paradoxus*	*S. bayanus-S. paradoxus*
number of genes	47	33	30
$\Sigma R_{Sx/Sp;cis}$	-13.75	-9.94	-9.31
$p(\Sigma R_{Sx/Sp;cis})$[d]	0.006	0.014	0.007
$p(\Sigma \text{sign}[R_{Sx/Sp;cis}])$[e]	0.043	0.029	0.070

Supplementary Table 1. *Cis*-regulatory changes between *Saccharomycete* species in the genes of the membrane protein regulon.

[a] The $\log_2$ of the ratio of expression of the *S. cerevisiae* allele and that of the *S. paradoxus* allele in the interspecific hybrid.

[b] The estimated effect of *cis*-regulatory variation between *S. kudriavzevii* and *S. paradoxus*, based on expression measurements in interspecific hybrids between each species and *S. cerevisiae*. N/A, missing data.
[c] The estimated effect of *cis*-regulatory variation between *S. bayanus* and *S. paradoxus*, based on expression measurements in interspecific hybrids between each species and *S. cerevisiae*. N/A, missing data.
[d] Significance of a one-sided resampling test relative to the genomic null for an extreme value of the sum of the estimated effects of *cis*-regulatory variation between the indicated species and *S. paradoxus*.
[e] Significance of a one-sided resampling test relative to the genomic null for an extreme value of the sum of the signs of estimated effects of *cis*-regulatory variation between the indicated species and *S. paradoxus*.

species	branch length, membrane group[a]	branch length, genome[b]	p[c]
S. paradoxus	0.077	0.082 (.0026)	0.020
S. cerevisiae	0.129	0.131 (.0034)	0.236
S. mikatae	0.194	0.192 (.0038)	0.753
S. kudriavzevii	0.176	0.180 (.0038)	0.196
S. bayanus	0.276	0.257 (.0054)	0.999

Supplementary Table 2. Tests for extreme branch lengths inferred from promoters of the membrane protein group.

[a] Evolutionary rate, in substitutions per site, for the terminal branch in the *Saccharomyces* phylogeny leading to the indicated species, inferred from the concatenated promoters of the membrane protein regulon and normalized with respect to total tree length.
[b] Evolutionary rate, in substitutions per site, for the terminal branch in the *Saccharomyces* phylogeny leading to the indicated species normalized with respect to total tree length, as an average of inferences from the concatenated promoters of each of 10,000 randomly chosen gene groups. The standard deviation is given in parentheses.
[c] Significance of a one-sided resampling test relative to the genomic null for short length of the terminal branch to the indicated species.

Appendix

Consider two species with effective population sizes $N_e^{(1)}$ and $N_e^{(2)}$ which diverged t generations ago, and assume that they have the same generation time. Further, assume that until t_0 generations after their divergence, one of the species (say, species 1) underwent positive selection with selection coefficient s_{11}, after which it underwent negative selection with selection coefficient s_{12}, while species 2 has been under constant negative selection with coefficient s_2.

In general, the expected number of substitutions along a given branch after time t is $u_i p(s_i)$. Here u_i is the mutation rate in species i; $p(s_i)$ is the probability of fixation of a new mutant and depends on the selection coefficient and the effective population size. Then the expected difference in the number of substitutions D along the lineages to the two species is

$$D_1 - D_2 = u_1(p(s_{11})t_0 + p(s_{12})(t-t_0)) - u_2 p(s_2)t = u_1 t_0(p(s_{11}) - p(s_{12})) + t(u_1 p(s_{12}) - u_2 p(s_2))$$

The first term in the final result is positive by the assumption that $s_{11} > s_{12}$; as such, the quantity $D_1 - D_2$ can be negative so long as $s_{12} < s_2$ and t is sufficiently big compared to t_0.

Thus, a lineage undergoing a short ancient period of positive selection followed by a long period of tight constraint can exhibit a reduced rate of fixed changes, relative to a lineage under constant and more modest negative selection.

References

1. Hodgins-Davis A, Townsend JP. Evolving gene expression: from G to E to G×E. Trends in Ecology & Evolution. 2009;24: 649–658. doi:10.1016/j.tree.2009.06.011

2. Wilson MD, Odom DT. Evolution of transcriptional control in mammals. Current Opinion in Genetics & Development. 2009;19: 579–585. doi:10.1016/j.gde.2009.10.003

3. Dowell RD. Transcription factor binding variationin the evolution of gene regulation. Trends Genet. Elsevier Ltd; 2010;26: 468–475. doi:10.1016/j.tig.2010.08.005

4. Kohn MH, Fang S, Wu CI. Inference of positive and negative selection on the 5′ regulatory regions of Drosophila genes. Mol Biol Evol. 2004. doi:10.1093/molbev/msh026

5. Blekhman R, Oshlack A, Chabot AE, Smyth GK, Gilad Y. Gene regulation in primates evolves under tissue-specific selection pressures. PLoS Genetics. 2008;4: e1000271. doi:10.1371/journal.pgen.1000271

6. Torgerson DG, Boyko AR, Hernandez RD, Indap A, Hu X, White TJ, et al. Evolutionary Processes Acting on Candidate cis-Regulatory Regions in Humans Inferred from Patterns of Polymorphism and Divergence. Begun DJ, editor. PLoS Genetics. 2009;5: e1000592. doi:10.1371/journal.pgen.1000592.t006

7. Bullard JH, Mostovoy Y, Dudoit S, Brem RB. Polygenic and directional regulatory evolution across pathways in Saccharomyces. Proc Natl Acad Sci USA. 2010;107: 5058–5063. doi:10.1073/pnas.0912959107

8. Fraser HB, Moses AM, Schadt EE. Evidence for widespread adaptive evolution of gene expression in budding yeast. Proceedings of the National Academy of Sciences. 2010;107: 2977–2982. doi:10.1073/pnas.0912245107

9. He BZ, Holloway AK, Maerkl SJ, Kreitman M. Does Positive Selection Drive Transcription Factor Binding Site Turnover? A Test with Drosophila Cis-Regulatory Modules. Petrov DA, editor. PLoS Genetics. 2011;7: e1002053. doi:10.1371/journal.pgen.1002053.t001

10. Gasch A, Moses A, Chiang D, Fraser H. Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol. 2004.

11. Ihmels J, Bergmann S, Gerami-Nejad M, Yanai I, McClellan M, Berman J, et al.

Rewiring of the yeast transcriptional network through the evolution of motif usage. Science. 2005;309: 938–940. doi:10.1126/science.1113833

12. Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, Kreitman M. Functional Evolution of a cis-Regulatory Module. Levine M, editor. PLoS Biol. 2005;3: e93. doi:10.1371/journal.pbio.0030093.st002

13. Tanay A, Regev A, Shamir R. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. Proc Natl Acad Sci USA. 2005;102: 7203–7208. doi:10.1073/pnas.0502521102

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

15. Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Seringhaus MR, et al. Divergence of transcription factor binding sites across related yeast species. Science. 2007;317: 815–819. doi:10.1126/science.1140748

16. Hinman VF, Nguyen A, Davidson EH. Caught in the evolutionary act: Precise cis-regulatory basis of difference in the organization of gene networks of sea stars and sea urchins. Developmental Biology. 2007;312: 584–595. doi:10.1016/j.ydbio.2007.09.006

17. Hogues H, Lavoie H, Sellam A, Mangos M, Roemer T, Purisima E, et al. Transcription Factor Substitution during the Evolution of Fungal Ribosome Regulation. Mol Cell. 2008;29: 552–562. doi:10.1016/j.molcel.2008.02.006

18. Tuch BB, Galgoczy DJ, Hernday AD, Li H, Johnson AD. The evolution of combinatorial gene regulation in fungi. PLoS Biol. Public Library of Science; 2008;6: e38. doi:10.1371/journal.pbio

19. Perez JC, Shin D, Zwir I, Latifi T, Hadley TJ, Groisman EA. Evolution of a bacterial regulon controlling virulence and Mg(2+) homeostasis. PLoS Genetics. 2009;5: e1000428. doi:10.1371/journal.pgen.1000428

20. Piasecki BP, Burghoorn J, Swoboda P. Regulatory Factor X (RFX)-mediated transcriptional rewiring of ciliary genes in animals. Proc Natl Acad Sci USA. 2010;107: 12969–12974. doi:10.1073/pnas.0914241107

21. Orr HA. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. Genetics. 1998;149: 2099–2104.

22. Fraser HB, Babak T, Tsang J, Zhou Y, Zhang B, Mehrabian M, et al. Systematic detection of polygenic cis-regulatory evolution. PLoS Genetics. 2011;7: e1002023. doi:10.1371/journal.pgen.1002023

23. Tirosh I, Reikhav S, Levy AA, Barkai N. A yeast hybrid provides insight into the evolution of gene expression regulation. Science. American Association for the Advancement of Science; 2009;324: 659.

24. 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

25. Hittinger CT, Goncalves P, Sampaio JP, Dover J, Johnston M, Rokas A. Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature. 2010;464: 54–U61. doi:10.1038/nature08791

26. Bustamante CD, Wakeley J, Sawyer S, Hartl DL. Directional selection and the site-frequency spectrum. Genetics. 2001.

27. Arthur RK, Ruvinsky I. Evidence That Purifying Selection Acts on Promoter Sequences. Genetics. 2011;189: 1121–1126. doi:10.1534/genetics.111.133637

28. Vishnoi A, Sethupathy P, Simola D, Plotkin JB, Hannenhalli S. Genome-wide survey of natural selection on functional, structural, and network properties of polymorphic sites in Saccharomyces paradoxus. Mol Biol Evol. 2011. doi:10.1093/molbev/msr085

29. Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, et al. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. G3 (Bethesda). 2011;1: 11–25. doi:10.1534/g3.111.000273

30. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 2007;24: 1586–1591. doi:10.1093/molbev/msm088

31. Tirosh I, Weinberger A, Carmi M, Barkai N. A genetic signature of interspecies variations in gene expression. Nat Genet. 2006;38: 830–834. doi:10.1038/ng1819

32. Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. Nat Genet. 2005;37: 544–548. doi:10.1038/ng1554

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

34. Kuo D, Licon K, Bandyopadhyay S, Chuang R, Luo C, Catalana J, et al. Coevolution within a transcriptional network by compensatory trans and cis mutations. Genome Res. 2010;20: 1672–1678. doi:10.1101/gr.111765.110

35. Lavoie H, Hogues H, Mallick J, Sellam A, Nantel A, Whiteway M. Evolutionary tinkering with conserved components of a transcriptional regulatory network. PLoS Biol. 2010;8: e1000329. doi:10.1371/journal.pbio.1000329

36. Baker CR, Tuch BB, Johnson AD. Extensive DNA-binding specificity divergence of a conserved transcription regulator. Proc Natl Acad Sci USA. National Acad Sciences; 2011;108: 7493–7498. doi:10.1073/pnas.1019177108/-/DCSupplemental

37. Takahasi KR, Matsuo T. Two types of cis-trans compensation in the evolution of transcriptional regulation. 2011. doi:10.1073/pnas.1105814108/-/DCSupplemental

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

39. Tirosh I, Wong KH, Barkai N, Struhl K. Extensive divergence of yeast stress responses through transitions between induced and constitutive activation. Proc Natl Acad Sci USA. 2011;108: 16693–16698. doi:10.1073/pnas.1113718108

40. Weiss KM. Tilting at quixotic trait loci (QTL): an evolutionary perspective on genetic causation. Genetics. 2008;179: 1741–1756. doi:10.1534/genetics.108.094128

41. Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, et al. Fast statistical alignment. PLoS Comput Biol. Public Library of Science; 2009;5: e1000392.

42. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol. 2011;28: 2731–2739. doi:10.1093/molbev/msr121

Chapter 4

Polygenic evolution of a sugar specialization trade-off in yeast

The contents of this chapter are based on the following publication:

Roop JI, Chang KC, Brem RB. Polygenic evolution of a sugar specialization trade-off in yeast. Nature. 2016;530

Abstract

The evolution of novel traits can involve many mutations scattered throughout the genome[1,2]. Detecting and validating such a suite of alleles, particularly if they arose long ago, remains a key challenge in evolutionary genetics[1-3]. Here we dissect an evolutionary tradeoff of unprecedented genetic complexity between long-diverged species. When cultured in 1% glucose medium supplemented with galactose, *Saccharomyces cerevisiae*, but not *S. bayanus* or other *Saccharomyces* species, delayed commitment to galactose metabolism until glucose was exhausted. Promoters of seven galactose (*GAL*) metabolic genes from *S. cerevisiae*, when introduced together into *S. bayanus*, largely recapitulated the delay phenotype in 1% glucose-galactose medium, and most had partial effects when tested in isolation. Variation in *GAL* coding regions also contributed to the delay when tested individually in 1% glucose-galactose medium. When combined, *S. cerevisiae GAL* coding regions gave rise to profound growth defects in the *S. bayanus* background. In medium containing 2.5% glucose supplemented with galactose, wild-type *S. cerevisiae* repressed *GAL* gene expression and had a robust growth advantage relative to *S. bayanus*; transgenesis of *S. cerevisiae GAL* promoter alleles or *GAL* coding regions was sufficient for partial reconstruction of these phenotypes. *S. cerevisiae GAL* genes thus encode a regulatory program of slow induction and avid repression, and a fitness detriment during the glucose-galactose transition but a benefit when glucose is in excess. Together, these results make clear that genetic mapping of complex phenotypes is within reach, even in deeply diverged species.

Introduction and Results

A central goal of evolutionary genetics is to understand how organisms acquire phenotypic novelties. Such traits, if they are have evolved over long timescales, can have a genetic basis quite distinct from those arisen more recently[4]. In landmark cases, single genes underlying species differences have been pinpointed and validated[5], but the polygenic architecture of ancient traits has remained a mystery.

In hybrids formed by mating *Saccharomyces cerevisiae* with other *Saccharomyces* species[6], we noted a pattern of coherent, *cis*-regulatory variation in the seven genes of the galactose metabolic pathway. During growth in medium with glucose as the sole carbon source, the *S. cerevisiae* allele at each *GAL* gene conferred low expression relative to other *Saccharomyces*, except for the repressor *GAL80*, at which the *S. cerevisiae* allele drove expression up (Figure 1b). Likewise, purebred *S.*

cerevisiae expressed *GAL* effectors at low levels in glucose, and *GAL80* at high levels, relative to other species (Figure 1b and [7]). *S. paradoxus,* the sister species to *S. cerevisiae*, had an intermediate expression phenotype (Figure 1b). Thus, the *S. cerevisiae GAL* program is one of heightened glucose repression relative to other species, as a product of *cis*-regulatory changes at the five loci that encode the seven *GAL* genes. Because such a pattern is unlikely under neutrality[7], these data raised the possibility that selective pressure on the *GAL* pathway had changed along the *S. cerevisiae* lineage.

In *S. cerevisiae*, pre-expression of metabolic genes in glucose medium can boost fitness upon a switch to other carbon sources[8-10]. We therefore expected that *GAL* expression divergence in glucose could have phenotypic correlates in other conditions. Culturing cells in 1% glucose medium supplemented with galactose, we observed a qualitative distinction between species (Figure 2a). In *S. cerevisiae,* growth was retarded by a diauxic lag midway through the timecourse, reflecting the expected delay in assembling galactose metabolic machinery once glucose is exhausted[8,9,11]. In more distantly related yeasts, we observed no lag in 1% glucose-galactose medium supplemented with galactose (Figure 2a-b), although *S. paradoxus* had a modest lag (Figure 2a-b) that echoed its intermediate regulatory phenotype (Figure 1b). Glucose mixtures with maltose and raffinose engendered a lag in all members of the clade (Extended Data Figure 1). *S. cerevisiae* strains from distinct populations all exhibited a lag in glucose-galactose cultures (Figure 2c). These data highlight *S. cerevisiae* as an extreme among *Saccharomyces* with respect to two attributes of galactose metabolism: reduced *GAL* gene expression during growth in pure glucose, and diauxic lag in 1% glucose-galactose medium supplemented with galactose.

To dissect further the divergence in galactose metabolic behaviors, we focused on a comparison of *S. cerevisiae* with its distant relative *S. bayanus* var. *uvarum (S. bayanus)*. In 1% glucose-galactose medium, both species initially metabolized glucose with similar rates, indicating that neither used the sugars simultaneously (Figure 2d). In *S. bayanus* cultures, galactose consumption began at a point just before the complete exhaustion of glucose. For *S. cerevisiae*, glucose exhaustion triggered the diauxic lag, during which galactose levels in its culture medium were largely unchanged. After the lag, with the eventual resumption of log-phase growth by *S. cerevisiae*, galactose levels finally dropped (Figure 2d). These results implicate the transition between glucose and galactose metabolism as a nexus of phenotypic differences between the species.

For direct tests of the phenotypic impact of divergence at the *GAL* genes, we replaced *GAL* gene sequences in one species by those of the other at the endogenous loci (Figure 3a). In a first investigation of *GAL* promoters, *S. cerevisiae* alleles of the regions upstream of *GAL1, GAL3, GAL4*, and *GAL10* were each sufficient for a partial gain in diauxic lag in *S. bayanus*, in 1% glucose-galactose medium (Figure 3b-c). Control experiments established the inverse effect of *S. bayanus GAL* promoter alleles, reducing lag in the *S. cerevisiae* background (Extended Data Figure 2).

We next aimed at a more complete reconstruction of *S. cerevisiae*-like galactose metabolic behaviors, which we inferred to be derived, in *S. bayanus* as a representative of the likely ancestral state. An *S. bayanus* strain harboring all seven *GAL* promoters from *S. cerevisiae* recapitulated 69% of the lag phenotype of the *S. cerevisiae* parent (Figure 3b-c), with *GAL* gene expression peaking at the same timepoint as that of wild-type *S. cerevisiae* and at similar amplitude (Figure 3d). Comparison to the sum of lag effects from individual promoter swaps revealed negative epistasis in the seven-promoter replacement strain (Figure 3c), and in strains harboring intermediate *S. cerevisiae* promoter combinations (Extended Data Figure 3).

Transgenesis of individual *S. cerevisiae GAL* coding regions was also sufficient for a partial lag in *S. bayanus*, in the case of *GAL1*, *GAL3*, and *GAL4* (Figure 3b-c). Swaps of *GAL* promoter-coding fusions revealed negative epistasis at *GAL1* and *GAL3*: for these genes, the sum of phenotypes from the respective promoter- and coding transgenics was far more dramatic than the effect of the promoter-coding fusion (Figure 3c). Combining all seven *S. cerevisiae GAL* coding or promoter-coding regions in *S. bayanus,* we observed an exaggerated, long-term growth delay in 1% glucose-galactose medium, distinct from the temporary lag of wild-type strains and promoter transgenics (Figure 3b). This defect reflected dysfunction of multiple modules of the *S. cerevisiae GAL* pathway in *S. bayanus*, as it could be elicited by just the two regulators Gal3 and Gal4 swapped from *S. cerevisiae*, or just *S. cerevisiae* alleles of the enzymes Gal1, Gal7, and Gal10 (Extended Data Figure 3). Coding and promoter-coding swap strains did ultimately resume active growth (Figure 3b) and metabolize galactose from mixed-sugar medium (Extended Data Figure 4), and their *GAL* expression induction was markedly delayed (Figure 3d). These strains also grew poorly in pure galactose medium (Extended Data Figure 5). Together, our data make clear that diauxic lag in 1% glucose-galactose medium can be largely recapitulated by divergent *GAL* gene promoters; *GAL* protein alleles from *S. cerevisiae* make a partial contribution to lag when tested in isolation and, when combined in *S. bayanus*, confer growth defects far exceeding those of either wild-type.

In light of the conservation of diauxic lag across *S. cerevisiae* (Figure 2c), we hypothesized that this species had maintained its divergent galactose metabolic behavior on the basis of a fitness benefit. Among the potential mechanisms for such an advantage, we focused on the possibility that as *S. cerevisiae* represses *GAL* genes in glucose-replete conditions (Figure 1), it avoids the liability of expressing unused proteins and enables rapid growth[9,12,13]. When cultured in 2.5% glucose medium also containing galactose, wild-type *S. cerevisiae* exhibited a 10% faster growth rate (Figure 4a-b), and 4-9 fold lower expression of *GAL* enzymes (Figure 4c), than *S. bayanus*. Both species metabolized glucose almost exclusively across the timecourse (Figure 4d-e). Replacement of all seven *S. bayanus GAL* promoters with *S. cerevisiae* alleles recapitulated the program of low *GAL* gene expression (Figure 4c), and conferred a growth rate halfway between those of the wild-type species

(Figure 4a-b), in 2.5% glucose medium supplemented with galactose. The *S. bayanus* strain harboring all seven *S. cerevisiae GAL* coding regions also expressed *GAL* genes at low levels (Figure 4c), which mirrored this strain's exaggerated delay in *GAL* gene induction (Figure 3b-d), and was associated with a partial growth benefit (Figure 4a-b). Promoter-coding replacement conferred no additional phenotype over and above the effects of transgenesis of either region type alone (Figure 4a-c). We conclude that *S. cerevisiae GAL* promoters, by shutting down expression of the galactose metabolic pathway, are adaptive in conditions of abundant glucose, and this program can be phenocopied by *S. cerevisiae GAL* protein alleles in *S. bayanus*. Sequence analyses revealed a high ratio of inter-specific divergence to intra-species polymorphism in *GAL* gene promoters, though not in *GAL* coding regions (Extended Data Tables 1-3), suggestive of a history of directional evolution at these loci.

Discussion

In this work, we have dissected glucose-specialist phenotypes that distinguish *S. cerevisiae* from other members of the *Saccharomyces* clade. *S. cerevisiae* is reluctant to transition from glucose to galactose metabolism, and has a growth advantage in a high-glucose environment. Additionally, the *S. cerevisiae* program confers an increase in biomass accumulation during growth in pure galactose (Extended Data Figure 4c) and could also be beneficial when glucose availability fluctuates rapidly[10]. As *S. cerevisiae* alleles of *GAL* gene promoters are largely sufficient for this family of traits, they may have served as an easily evolvable, and likely adaptive, origin of these characters. By contrast, the *S. cerevisiae GAL* proteome, which confers synthetic growth defects in modern-day *S. bayanus*, may have evolved slowly over a rugged fitness landscape, under distinct forces or at a different period. Such a model would dovetail with the *cis*-regulatory basis of a related, but genetically simple, galactose metabolism trait that evolved more recently between yeasts[12]. For any suite of divergent regulatory regions, observing *cis*-acting effects on gene expression can open a first window onto their phenotypic relevance and that of the gene products they control. With this paradigm, evolutionary biologists need not be limited by polygenicity in the mapping of genotype to phenotype, even between long-diverged species.

Materials and methods.

Yeast strains

Strains used in this study are listed in Supplementary Table 1. Abbreviations in figures and tables are as follows: *S. cer, Saccharomyces cerevisiae; S. par, Saccharomyces paradoxus; S. mik, Saccharomyces mikatae; S. bay, Saccharomyces bayanus; S. cas, Saccharomyces castellii*. Allele-swap strains constructed in haploid *S. bayanus* JRY294 or JRY296 (isogenic MATa and MATα derivatives of type strain CBS7001) used the MIRAGE method[14] with several modifications as follows. A

1.7kb region containing the *K. lactis* URA3 coding sequence and regulatory region was amplified from the pCORE-UH plasmid[15] and used for each half of the inverted repeat. The *S. cerevisiae GAL* region to be swapped in was attached to one half of the inverted repeat cassette by overlap extension PCR, after which the two halves of the final cassette were ligated together. Due to the different sizes of the two halves of the inverted repeat cassette, a second restriction digestion step as described in [14] was not necessary to remove non-desired ligation products. Transformation with the cassette, followed by confirmation of positive transformants and plating onto 5-FOA medium, resulted in excision of the inverted repeat from the target genome, leaving behind a marker-less allele swap at the locus. Sanger sequencing was used to verify the correct nucleotide sequence of each swapped allele. The *S. cerevisae* allele of each promoter and CDS was amplified from genomic DNA of YHL068 [6]. Promoter, coding, and promoter-coding fusion swap strains were engineered by replacing 600bp of intergenic region directly upstream of the CDS, the CDS, or these two regions combined, respectively, in *S. bayanus* with orthologous regions from *S. cerevisiae*. For allele swaps in the *S. cerevisiae* background in Extended Data Figure 2, 720bp of the region between the *GAL1* and *GAL10* open reading frames was amplified from *S. bayanus* strain CBS7001 and used to construct a MIRAGE cassette as above, and transformed into *S. cerevisiae* strain JRY313 (isogenic MATa derivative of BY4743) and selected as above. For each transgenic, two or more independent transformants were used as replicates for growth profiling and sugar concentration measurements. Combining unlinked allele swaps into a single genome was accomplished by single-cell mating of single-locus swaps, followed by sporulation, tetrad dissection, and diagnostic PCR to identify segregant colonies with the allele combinations of interest. *S. cerevisiae* alleles of *GAL1, GAL7* and *GAL10* were combined in the *S. bayanus* background by successive allele swap transformations.

Growth curves and quantification

All growth experiments were conducted at 26°C in YP media (2% bacto-peptone, 1% yeast extract) supplemented with various carbon sources as follows. Experiments measuring diauxic lag in 1% glucose medium supplemented with galactose utilized medium containing 1% glucose and 1% galactose. Experiments measuring growth profiles in other non-galactose carbon sources utilized media containing 1% glucose and 1% of the secondary carbon source as indicated. Experiments measuring maximum growth rates in high-glucose media containing galactose utilized medium containing 2.5% glucose and 10% galactose. Experiments measuring growth profiles in pure galactose medium utilized media containing 2% galactose.

Growth timecourses were carried out as follows. Strains were grown in YP containing 2% galactose (Figure 4 and Extended Data Figure 5) or 2% glucose (all other figures) for 24 hours with shaking at 200rpm. Each strain was then back-diluted into the same medium to an OD of 0.1 and grown for an additional 6 hours. These log-phase cultures were then back-diluted to an OD of 0.02 in a 96-well plate

containing 150µl of YP with the appropriate amount of a given carbon source. Plates were covered with a gas-permeable membrane, placed in a Tecan F200 plate reader and incubated with orbital shaking for the duration of each experiment. OD_{600} measurements were made every 30 minutes.

Gain in lag was calculated from growth curves as (1 - geometric mean rate) of a given strain. Geometric mean rate (GMR) was calculated as in [11], with the following differences. A window of 0.1 to 0.8 OD units was used for quantification in all figures apart from Figure 2b. In Figure 2b, GMR was calculated within a window bounded by 20% and 80% of the maximum final yield attained during the time course for each species. Maximum growth rate was calculated as in [10]except that a window of 0.01 to 0.3 OD units was used and a geometric mean of growth rates was calculated. Final growth yield was calculated as the difference between initial and final OD_{600} measurements as in [16].

Replication schemes and analysis were as follows. To enable qualitative comparisons among species in Figures 2a, 2b and 2c, on a given 96-well plate, six biological replicate cultures of each strain were assayed. On each growth plate, replicate fitness values greater than two standard deviations from the mean of fitness values for that strain were considered artifacts of technical error and discarded. Displayed data for a given strain are the results of growth measurements from one plate and are representative of at least two plates cultured on different days. Experiments in Figure 2d were as in Figure 2a-2c except that four biological replicate cultures were measured.

To enable highly powered quantitative comparisons among strains of growth in 1% glucose-galactose medium in Figure 3b-c, 12 biological replicate cultures of each transgenic strain were assayed across several plates and days, in each case alongside replicates of wild-type *S. bayanus* and *S. cerevisiae*. The growth rate of a given strain measured on a given plate was normalized to the value for wild-type *S. bayanus* on that plate, and these normalized measurements were then averaged across plates. Because we included wild-type strains on each plate, their growth measurements as displayed in Figure 3b-c are averages over 78 and 54 replicates of *S. bayanus* and *S. cerevisiae*, respectively. Artifact filtering was as above. Differences in growth among strains were assessed for statistical significance by a two-sided Wilcoxon rank-sum test and a Bonferroni correction was applied in instances of multiple tests.

To enable highly powered quantitative comparisons among strains of growth in medium containing 2.5% glucose and 10% galactose in Figure 4a-b, we assayed growth of 250, 170, 204, 136, and 192 biological replicate cultures of wild-type *S. bayanus*, wild-type *S. cerevisiae*, the combinatorial promoter transgenic strains, the combinatorial promoter-coding transgenic strains, and the combinatorial coding transgenic strains, respectively, across several plates and days. As above, growth measurements for each strain in turn assayed on a given plate was normalized to the wild-type *S. bayanus* cultured on that plate, and normalized growth rate

measurements were combined across plates. Artifact filtering and statistical testing were as above.

Experiments in Extended Data Figure 1 were as in Figure 2a-2c. Experiments for Extended Data Figure 2 were as in Figure 3 except that 12 replicate cultures were measured for each strain. Experiments for Extended Data Figure 3 were as in Figure 3b-c. Experiments for Extended Data Figure 4 were as in Figure 2a-2c.

Sugar measurements

For growth timecourse experiments in which sugar concentration was measured, an appropriate number of wells (see below) were inoculated into a 96-well plate and cultured in the Tecan F200 plate reader as above, such that media from at least two replicate wells could be harvested at each timepoint of interest and at least four replicate cultures of each strain would remain untouched for growth curve analysis. Samples were taken for sugar measurement by cutting the membrane covering the 96-well plate with a razor and extracting all 150µl of cell culture in a given well. Care was taken to only puncture the membrane above the harvested well such that adjacent wells were not affected. Cells and debris were pelleted from the sampled culture by brief centrifugation and the supernatant was extracted for quantification of glucose and galactose. Glucose was measured using the GlucCell glucose monitoring system (Chemglass Life Sciences). Galactose was measured with the Amplex Red Galactose Oxidase assay kit (Molecular Probes, Life Technologies). For galactose measurements, a Tecan Safire plate reader (Tecan) was used to quantify fluorescence and the relationship between fluorescence and galactose concentration was determined using a standard curve.

Data in Figure 2d were obtained by sampling two biological replicate cultures at each timepoint. Data in Figure 4d, 4e and Extended Data Figure 4 were obtained by sampling three biological replicate cultures at both the start and endpoints of the growth timecourse. For all experiments, three technical replicates were assayed for each biological replicate culture sampled, and mean values are reported. All data are representative of two identical experiments conducted on different days.

Quantitative PCR

Timepoint *s*amples for qPCR analysis were obtained from cultures analogously to those obtained for sugar consumption quantification detailed above. Between two and fifteen replicate wells were harvested at each timepoint and pooled in order to have sufficient biological material for RNA isolation. RNA was isolated using an RNeasy mini kit (Qiagen) and cDNA was synthesized using SuperScript III (Life Technologies). DyNAmo HS SYBR green (Thermo Scientific) was used for quantitative PCR and all quantification was done on a CFX96 machine (BioRad). Gene expression levels relative to *ACT1* were calculated using the $2^{\wedge\Delta\Delta Ct}$ method[17]. Three technical replicates per biological sample were assayed, and mean values are reported.

Sequence analyses

Custom Python scripts were used to extract coding sequences and 600bp promoter regions for type strains of *S. paradoxus*, *S. mikatae*, and *S. bayanus*[18] for each gene that had an ortholog in each of the five *Saccharomyces sensu strico* species as reported in [18] *S. cerevisiae* population sequences were downloaded from the following sources:
YJM978, UWOPS83-787, Y55, UWOPS05-217.3, 273614N, YS9, BC187, YPS128, DBVPG6765, YJM975, L1374, DBVPG1106, K11, SK1, 378604X, YJM981, UWOPS87-2421, DBVPG1373, NCYC3601, YPS606, Y12, UWOPS05-227.2, and YS2 from http://www.yeastrc.org/g2p/home.do; Sigma1278b, ZTW1, T7, and YJM789 from http://www.yeastgenome.org/; and RM11 from http://www.broadinstitute.org/annotation/genome/saccharomyces_cerevisiae. *S .cerevisiae* sequences were aligned to each of the other three species in turn using FSA [19], using the '--nucprot' option for the coding sequence alignments. For each set of species alignments, nucleotide replacement and polymorphic sites were tabulated using Polymorphorama[14,20] for coding regions and custom Python scripts for promoter regions.

Sequence analyses of the seven genes of the *GAL* pathway (*GAL1, GAL2, GAL3, GAL4, GAL7, GAL10,* and *GAL80*) using *S. paradoxus, S. mikatae*, or *S. bayanus* as an outgroup were done as follows. For a given species comparison, we first calculated the NI_{TG} statistic[15,21] for the promoter regions of the *GAL* genes using synonymous sites in the downstream gene as putative neutral sites, as

$$\frac{\sum D_{si}P_{ri}/(P_{si}+D_{si})}{\sum P_{si}D_{ri}/(P_{si}+D_{si})}$$

where *i* counts genes of the group, D_s and D_r denote the number of divergent synonymous and divergent promoter sites respectively, and P_s and P_r denote the number of polymorphic synonymous and polymorphic promoter sites respectively.

Minor alleles with a frequency of less than 0.15 were ignored[17,22,23] and all site counts were corrected for multiple hits using a Jukes-Cantor model. In computing NI_{TG} for the seven genes of the *GAL* pathway, we considered that since *GAL1* and *GAL10* are adjacent genes on chromosome II sharing a 662bp promoter region, simply counting sites within 600bp promoter regions for each gene separately would have resulted in double-counting of 532bp. To avoid this, we aligned the 662bp intergenic region of these two genes and considered it one locus in the NI_{TG} calculation. Synonymous sites for this region were taken as the sum of such sites in both the *GAL1* and *GAL10* coding sequences. To evaluate significance, we generated 10,000 randomly chosen groups of six genes each and computed NI_{TG} across the promoters of each such null group as above. The empirical significance of the true NI_{TG} value for *GAL* gene promoters was then taken as the proportion of null groups whose promoters gave an NI_{TG} less than or equal to the true value. We assessed

selection acting on non-synonymous sites in *GAL* coding sequences with a pipeline analogous to the above, except that *GAL1* and *GAL10* were considered separate loci, as the coding sequences do not overlap; thus our resampling test used 10,000 random groups of seven coding regions each. Tests for selection using the Neutrality Index statistic[18,24] [(polymorphic non-synonymous sites/polymorphic synonymous sites) / (divergent non-synonymous / divergent synonymous sites), with a similar formulation for non-synonymous sites] used the same pipeline as above except that we tabulated the average NI across genes of a group, and compared this average quantity for the *GAL* genes against resampled groups of the same size.

D_{XY} [25], the average number of pairwise differences between outgroup species and *S. cerevisiae* strains, and intrapopulation nucleotide diversity[26] (π), were calculated using custom Python scripts. Nucleotide diversity was ascertained on alignments containing only *S. cerevisiae* strains belonging to the European population, the most deeply sampled in our data set. Empirical significance of these statistics for *GAL* gene promotes and coding sequences was calculated using the resampling pipeline described above.

Code availability

Custom Python scripts used for data analysis are available upon request

Acknowledgments

This work was supported by NIH GM087432 to R.B.B. and a Hellman Graduate Fellowship from UC Berkeley to J.I.R. The authors thank Adam Arkin for his generosity with advice and resources; Jasper Rine for yeast strains; Ophelia Venturelli for technical expertise; and Avi Flamholz, Josh Schraiber and Patrick Shih for helpful discussions.

Figures

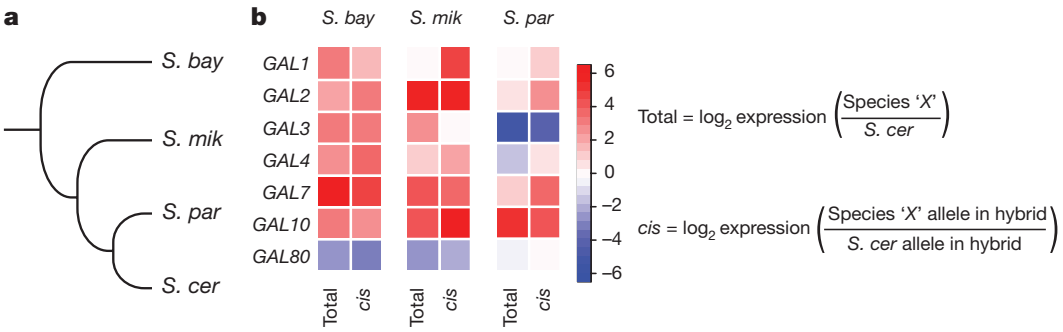


Figure 1. Polygenic *cis*-regulatory evolution among yeast species in galactose metabolic genes. a) Tree of *Saccharomyces* species studied here[18]. **b)** Ratios of expression between the indicated species and *S. cerevisiae*, of the indicated galactose metabolism gene during culture in glucose medium[6]. Total, expression measured in purebred species; *cis*, expression from the indicated species' allele in a diploid hybrid between this species and *S. cerevisiae*, reflecting effects of *cis*-regulatory divergence.

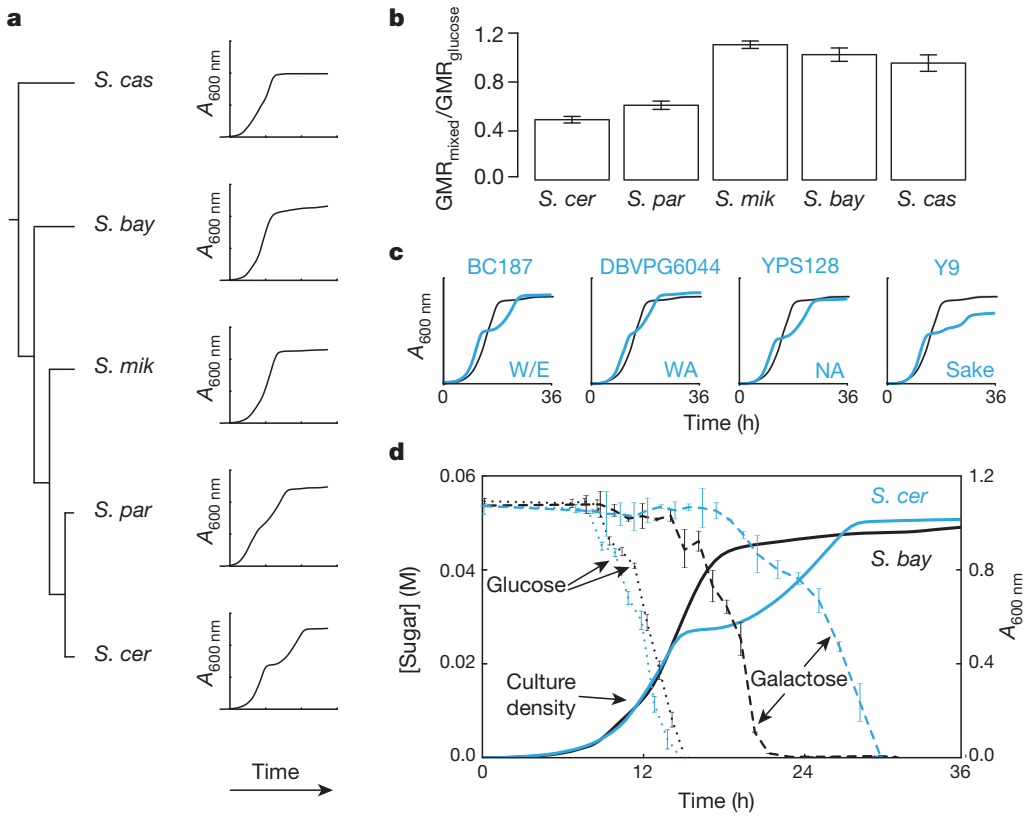


Figure 2. Diauxic lag, in 1% glucose-galactose medium, is conserved within *S. cerevisiae* and divergent among species. a) Growth of *Saccharomyces* type strains inoculated into medium containing 1% glucose and 1% galactose. **b)** Geometric means of the growth rates (GMR) from **a**, normalized to the analogous quantity in glucose medium. **c)** Growth of *S. cerevisiae* isolates (blue) from the indicated populations (W/E, Wine/European; W.A., West African; N.A., North American) and the *S. bayanus* type strain (black), inoculated into medium containing 1% glucose and 1% galactose. **d)** Growth (solid lines) of *S. cerevisiae* and *S. bayanus* inoculated into medium containing 1% glucose and 1% galactose, and medium concentrations of glucose and galactose (dotted and broken lines, respectively). Error bars report standard error of the mean.

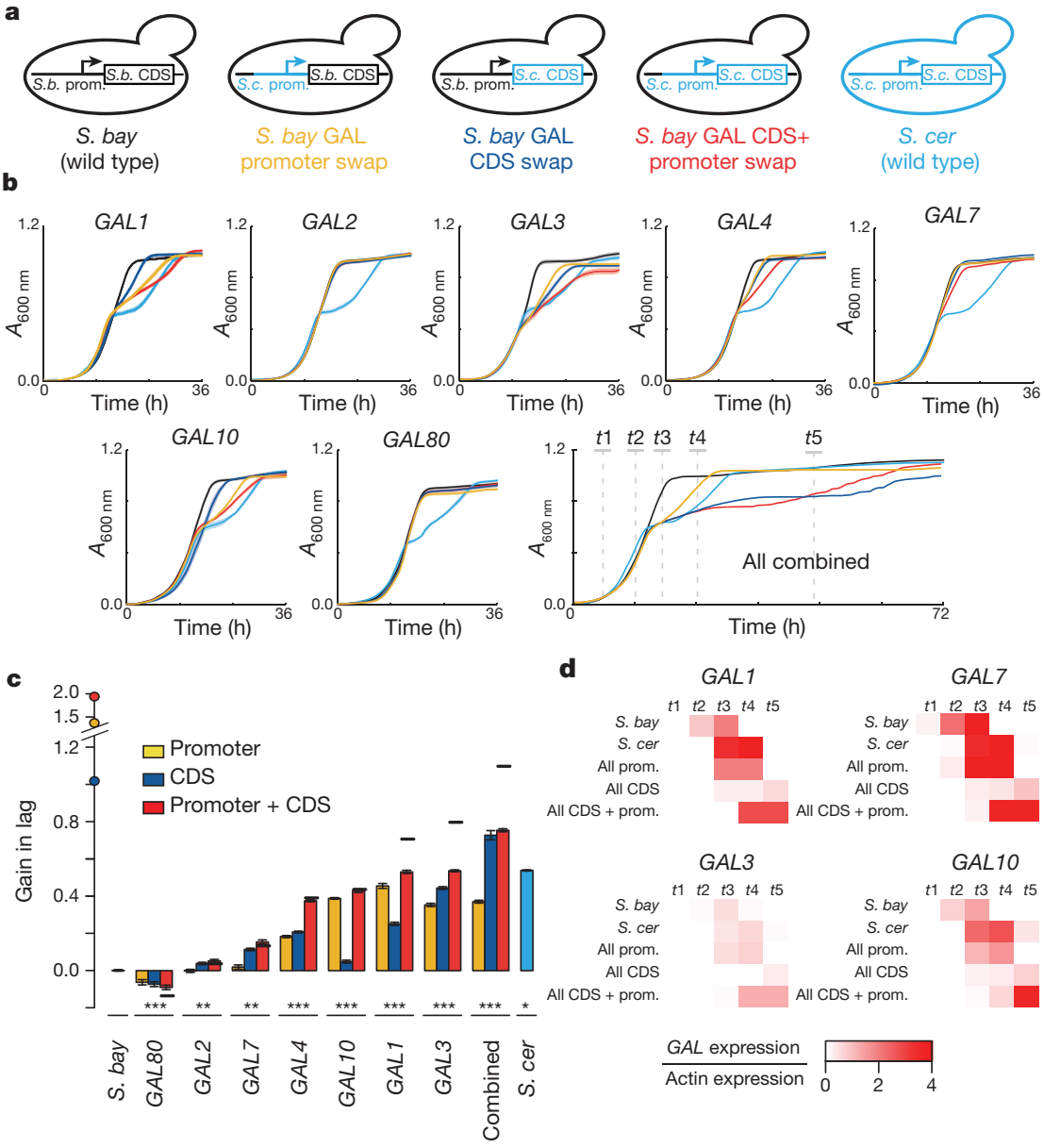


Figure 3. *S. cerevisiae* alleles of *GAL* genes confer diauxic lag in 1% glucose-galactose medium. a) Replacement of *S. cerevisiae GAL* sequences into *S. bayanus* at the endogenous loci. **b)** Growth of *S. bayanus* harboring *S. cerevisiae* alleles of a single *GAL* gene, or of all seven genes, inoculated into medium containing 1% glucose and 1% galactose. **c)** Each bar reports the ratio of the GMR from **b** to that of wild-type *S. bayanus*, subtracted from 1; negative values are GMRs faster than wild-type. Error bars report standard error of the mean. Asterisks, significant differences ($p < 0.001$, Wilcoxon rank-sum) from wild-type *S. bayanus.* Also shown are expected phenotypes of promoter-CDS transgenics for a single gene (horizontal lines) or seven-locus transgenics (circles on *y*-axis), under an additive model of contributions from the regions combined in the respective strains. **d)** *GAL* gene expression at timepoints indicated in the final panel of **b.**

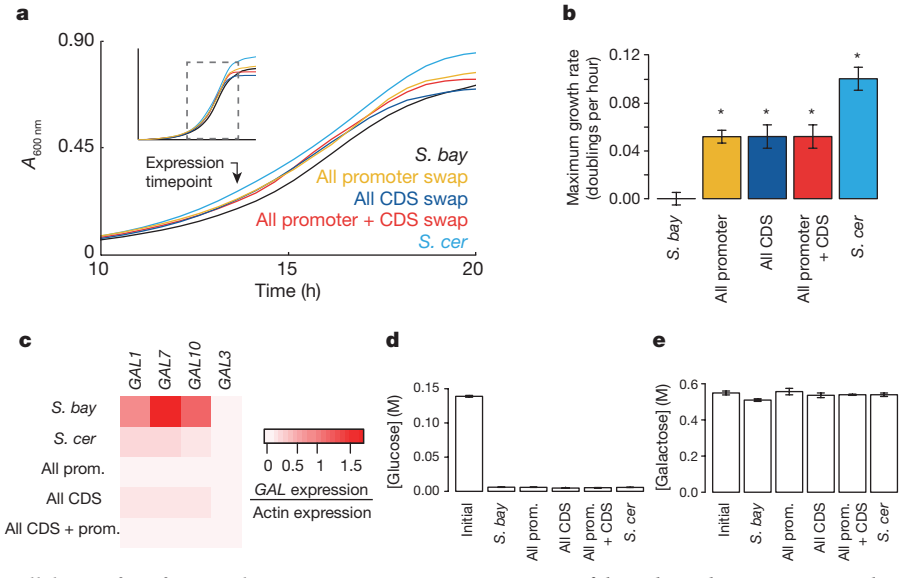
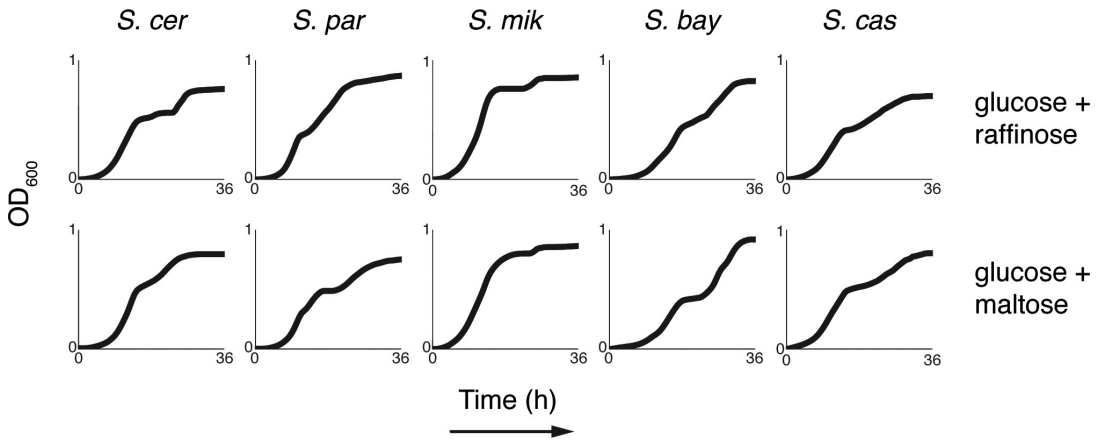
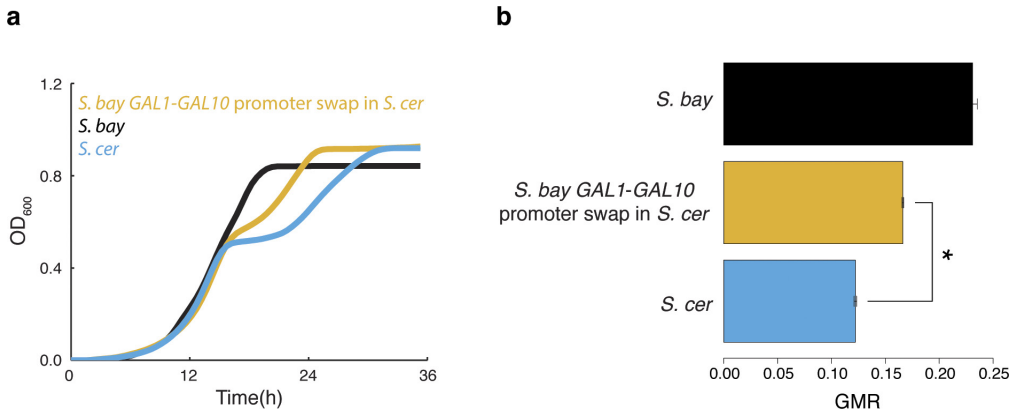


Figure 4. *S. cerevisiae GAL* alleles confer a fitness advantage in 2.5% glucose medium supplemented with galactose. a) Growth of *S. bayanus* strains harboring *S. cerevisiae* alleles of all seven *GAL* genes, and the wild-type species, inoculated into medium containing 2.5% glucose and 10% galactose. **b)** Each bar reports the difference in maximum growth rate between the indicated species and wild-type *S. bayanus* from **a**. Error bars report standard error of the mean, and asterisks indicate rates significantly different ($p < 1 \text{ X } 10^{-8}$, Wilcoxon rank-sum) from wild-type *S. bayanus*. **c)** *GAL* gene expression at the timepoint indicated by the arrow in **a**. **d-e)** In each panel, the first bar reports sugar concentration in medium before inoculation; the remaining bars report sugar concentrations in medium after growth timecourses in **a.**

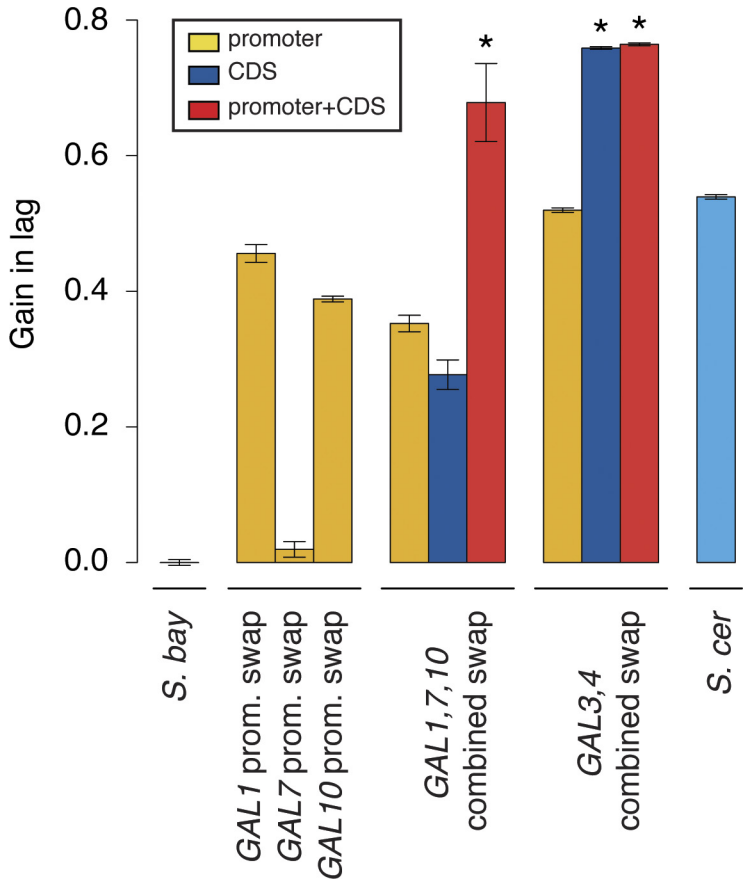
Extended Data Figures



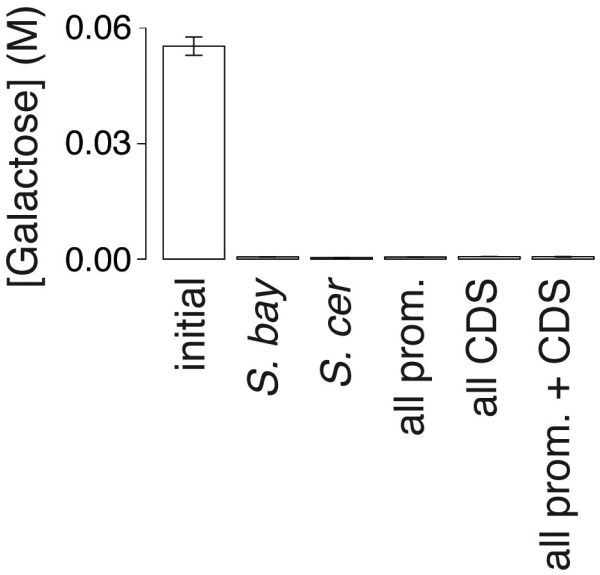
Extended Data Figure 1. **Divergence in the *S. cerevisiae* diauxic lag trait is specific to growth in glucose-galactose medium.** Each trace reports growth of the indicated yeast inoculated into medium containing 1% glucose and 1% of the indicated secondary carbon source.



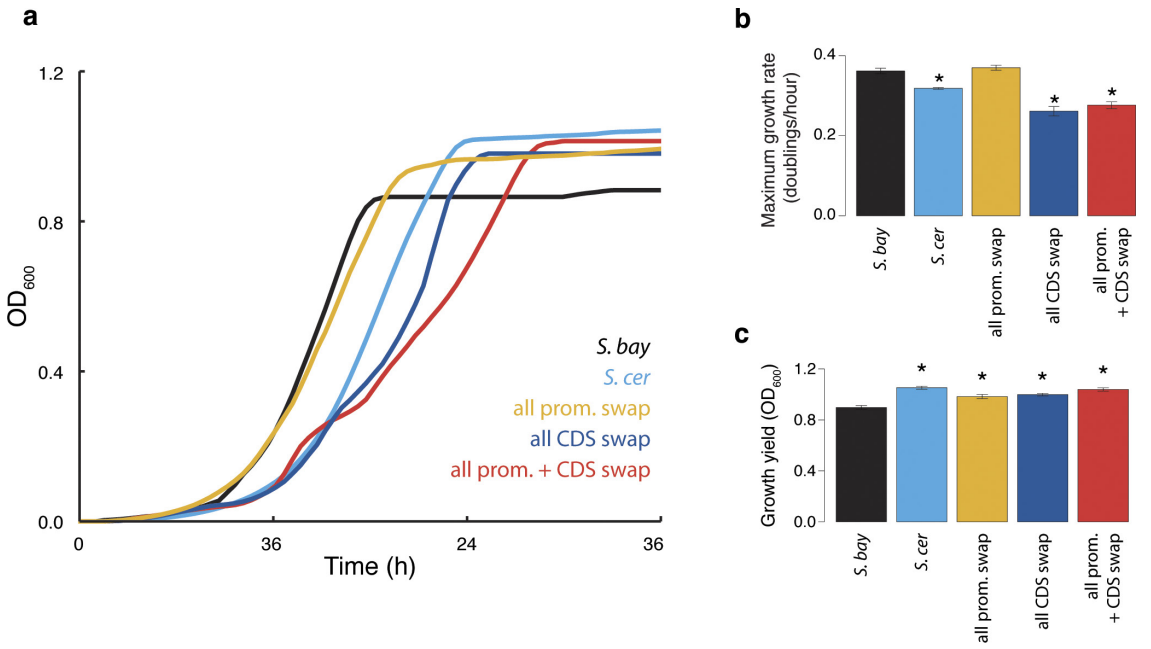
Extended Data Figure 2. The *S. bayanus* allele of the *GAL1* and *GAL10* promoters confers partial rescue of diauxic lag in *S. cerevisiae* in 1% glucose-galactose medium. **a)** Data are as in Figure 3b of the main text, except that the yellow curve reports growth of an *S. cerevisiae* strain harboring the *GAL1* and *GAL10* promoters from *S. bayanus*. **b)** Each bar reports the geometric mean growth rate (GMR) of the strains shown in **a.** Error bars report standard error of the mean, and the asterisk indicates a significantly different rate ($p < 1 \text{ X } 10^{-15}$, Wilcoxon rank-sum) between the transgenic swap strain and wild-type *S. cerevisiae*.



Extended Data Figure 3. **Intermediate combinations of *S. cerevisiae GAL* gene alleles confer an exaggerated lag defect on *S. bayanus,* in 1% glucose-galactose medium.** Each bar reports the ratio of the GMR to that of wild-type *S. bayanus*, subtracted from 1, from a timecourse of the indicated strain inoculated into medium containing 1% glucose and 1% galactose. The first through fourth bars, and the last bar, are from Figure 3c of the main text; each of the remaining bars reports results from an *S. bayanus* strain harboring *S. cerevisiae* alleles at the indicated combination of *GAL* loci. Symbols and analyses are as in Figure 3c of the main text.



Extended Data Figure 4. *S. bayanus* strains harboring *S. cerevisiae* alleles of all seven *GAL* loci, inoculated into 1% glucose-galactose medium, deplete growth media of galactose. The first bar reports galactose concentration in medium before inoculation, and the remaining bars report concentrations after growth timecourses, in the experiments in Figure 3b of the main text.



Extended Data Figure 5. In pure galactose medium, *S. cerevisiae* alleles of *GAL* loci are sufficient for increased biomass accumulation and, in the case of protein-coding regions, slow growth. a) Each trace reports growth of an *S. bayanus* strain harboring *S. cerevisiae* alleles of all seven *GAL* genes, or a wild-type control, inoculated into medium containing 2% galactose. **b)** Each bar reports maximum growth rate from the respective timecourse in **a.** Error bars report standard error of the mean, and asterisks indicate values significantly different ($p < 0.05$, Wilcoxon rank-sum) from wild-type *S. bayanus*. **c)** Each bar reports growth yield from the respective timecourse in **a.** Error bars report standard error of the mean, and asterisks indicate values significantly different ($p < 0.05$, Wilcoxon rank-sum) from wild-type *S. bayanus*.

Extended Data Tables

Extended Data Table 1a

	Outgroup		
Statistic	*S. par*	*S. mik*	*S. bay*
NI_{TG} *GAL* promoters	0.32	0.37	0.31
NI_{TG} genome promoters	0.72	0.91	0.73
p-value	0.025	0.013	0.018

Extended Data Table 1b

	Outgroup		
Statistic	*S. par*	*S. mik*	*S. bay*
NI_{TG} *GAL* non-synonymous sites	0.98	0.92	1.03
NI_{TG} genome non-synonymous sites	1.28	1.21	1.27
p-value	0.25	0.19	0.26

Extended Data Table 1. Excess of divergence relative to polymorphism in *GAL* promoter regions. Each panel reports analyses of the NI_{TG} measure[21] comparing polymorphism within *S. cerevisiae* to divergence between *S. cerevisiae* and the indicated outgroup species, taken across promoter sites (**a**) or non-synonymous coding sites (**b**), with normalization by the analogous measure from synonymous coding sites. In a given panel, the first row reports NI_{TG} across the seven *GAL* genes, the second row reports the mean NI_{TG} from 10,000 randomly drawn gene groups, and the bottom row reports empirical significance of the distinction between *GAL* genes and the genomic null.

Extended Data Table 2a

	Outgroup		
D_{XY}	*S. par*	*S. mik*	*S. bay*
GAL promoters	0.172	0.276	0.328
genome promoters	0.15	0.241	0.281
p-value	0.147	0.099	0.045
GAL CDS	0.102	0.155	0.188
genome CDS	0.096	0.153	0.190
p-value	0.337	0.429	0.573

Extended Data Table 2b

π	promoters	CDS
GAL	0.00067	0.00069
genome	0.00194	0.00113
p-value	0.181	0.513

Extended Data Table 2. Divergence between species, and polymorphism within *S. cerevisiae*, at *GAL* loci. a) The first line reports mean D_{XY} for *GAL* promoters in comparisons between *S. cerevisiae* and outgroup species. The second line reports the analogous statistics for the mean of 10,000 randomly drawn gene groups, and the third line reports empirical significance of the distinction between *GAL* genes and the genomic null. The fourth, fifth, and sixth lines are analogous to the above except that *GAL* coding regions were analyzed. **b)** The first line reports π for *GAL* promoters or coding regions within *S. cerevisiae*. The second line reports the analogous statistics for the mean of 10,000 randomly drawn gene groups, and the third line reports empirical significance of the distinction between *GAL* genes and the genomic null.

Extended Data Table 3a

	Outgroup		
Statistic	*S. par*	*S. mik*	*S. bay*
GAL promoters	0.32	0.42	0.35
genome promoters	0.94	1.13	0.92
p-value	0.003	0.009	0.014

Extended Data Table 3b

	Outgroup		
Statistic	*S. par*	*S. mik*	*S. bay*
GAL non-synonymous sites	1.43	1.24	1.44
genome non-synonymous sites	2.02	1.88	1.80
p-value	.30	.22	.36

Extended Data Table 3. Evaluation of selection acting on *S. cerevisiae* promoter and coding sequences using the neutrality index statistic. Data are as in Extended Data Table 1 except that the neutrality index[24] was used for each test.

Supplementary Tables

Species	Strain	Parent	Genotype/Details
S. cer	BY4743		MATa/@ his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0
S. cer	JRY313	BY4743	MATa haploid derivative of BY4743
S. cer	JRY594	JRY313	ΔGAL1(-668 to -1) :: sbGAL1(-720 to -1)
S. cer	YPS128		
S. cer	Y9		
S. cer	DBVPG6044		
S. cer	BC187		
S. cas	FM476		
S. mik	NCYC2888		
S. par	Z1		
S. bay	CBS7001		
S. bay	JRY9189	CBS7001	MATa
S. bay	JRY9190	CBS7001	MAT@
S. bay	JRY294	JRY9189	ΔURA3::KanMX
S. bay	JRY296	JRY9190	ΔURA3::KanMX
S. bay	JRY381	JRY296	ΔGAL1(-600 to -1) :: scGAL1(-600 to -1)
S. bay	JRY382	JRY296	ΔGAL1(-600 to -1) :: scGAL1(-600 to -1)
S. bay	JRY369	JRY294	ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP)
S. bay	JRY370	JRY294	ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP)
S. bay	JRY387	JRY296	ΔGAL2(-600 to -1) :: scGAL2(-600 to -1)
S. bay	JRY388	JRY296	ΔGAL2(-600 to -1) :: scGAL2(-600 to -1)
S. bay	JRY367	JRY296	ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP)
S. bay	JRY368	JRY296	ΔGAL2(-600 to STOP) :: scGAL2(-600 to STOP)
S. bay	JRY383	JRY294	ΔGAL3(-600 to -1) :: scGAL3(-600 to -1)
S. bay	JRY384	JRY294	ΔGAL3(-600 to -1) :: scGAL3(-600 to -1)
S. bay	JRY373	JRY294	ΔGAL3(-600 to STOP) :: scGAL3(-600 to STOP)
S. bay	JRY374	JRY294	ΔGAL3(-600 to STOP) :: scGAL3(-600 to STOP)
S. bay	JRY389	JRY296	ΔGAL4(-600 to -1) :: scGAL4(-600 to -1)
S. bay	JRY390	JRY296	ΔGAL4(-600 to -1) :: scGAL4(-600 to -1)
S. bay	JRY377	JRY294	ΔGAL4(-600 to STOP) :: scGAL4(-600 to STOP)
S. bay	JRY378	JRY294	ΔGAL4(-600 to STOP) :: scGAL4(-600 to STOP)
S. bay	JRY385	JRY294	ΔGAL7(-600 to -1) :: scGAL7(-600 to -1)
S. bay	JRY386	JRY294	ΔGAL7(-600 to -1) :: scGAL7(-600 to -1)
S. bay	JRY371	JRY294	ΔGAL7(-600 to STOP) :: scGAL7(-600 to STOP)
S. bay	JRY372	JRY294	ΔGAL7(-600 to STOP) :: scGAL7(-600 to STOP)
S. bay	JRY397	JRY294	ΔGAL10(-600 to -1) :: scGAL10(-600 to -1)
S. bay	JRY398	JRY294	ΔGAL10(-600 to -1) :: scGAL10(-600 to -1)
S. bay	JRY360	JRY296	ΔGAL10(-600 to STOP) :: scGAL10(-600 to STOP)

S. bay	JRY361	JRY296	ΔGAL10(-600 to STOP) :: scGAL10(-600 to STOP)
S. bay	JRY400	JRY296	ΔGAL80(-600 to -1) :: scGAL80(-600 to -1)
S. bay	JRY401	JRY296	ΔGAL80(-600 to -1) :: scGAL80(-600 to -1)
S. bay	JRY357	JRY296	ΔGAL80(-600 to STOP) :: scGAL80(-600 to STOP)
S. bay	JRY358	JRY296	ΔGAL80(-600 to STOP) :: scGAL80(-600 to STOP)
S. bay	JRY477	JRY296	ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP), ΔGAL10(-600 to STOP) :: scGAL10(-600 to STOP), ΔGAL7(-600 to -1) :: scGAL7(-600 to -1)
S. bay	JRY478	JRY296	ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP), ΔGAL10(-600 to STOP) :: scGAL10(-600 to STOP), ΔGAL7(-600 to -1) :: scGAL7(-600 to -1)
S. bay	JRY442	JRY296XJRY294	ΔGAL1(-600 to -1) :: scGAL1(-600 to -1), ΔGAL2(-600 to -1 :: scGAL2(-600 to -1), ΔGAL3(-600 to -1 :: scGAL3(-600 to -1), ΔGAL4(-600 to -1 :: scGAL4(-600 to -1), ΔGAL7(-600 to -1 :: scGAL7(-600 to -1), ΔGAL10(-600 to -1 :: scGAL10(-600 to -1), ΔGAL80(-600 to -1 :: scGAL80(-600 to -1)
S. bay	JRY443	JRY296XJRY294	ΔGAL1(-600 to -1 :: scGAL1(-600 to -1), ΔGAL2(-600 to -1) :: scGAL2(-600 to -1), ΔGAL3(-600 to -1 :: scGAL3(-600 to -1), ΔGAL4(-600 to -1 :: scGAL4(-600 to -1), ΔGAL7(-600 to -1 :: scGAL7(-600 to -1), ΔGAL10(-600 to -1 :: scGAL10(-600 to -1), ΔGAL80(-600 to -1 :: scGAL80(-600 to -1)
S. bay	JRY485	JRY296XJRY294	ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP), ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP), ΔGAL3(-600 to STOP) :: scGAL3(-600 to STOP), ΔGAL4(-600 to STOP) :: scGAL4(-600 to STOP), ΔGAL7(-600 to STOP) :: scGAL7(-600 to STOP), ΔGAL7(-600 to STOP) :: scGAL7(-600 to STOP), ΔGAL10(-600 to STOP) :: scGAL10(-600 to STOP), ΔGAL80(-600 to STOP) :: scGAL80(-600 to STOP)
S. bay	JRY449	JRY296XJRY294	ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP), ΔGAL1(-600 to STOP) :: scGAL1(-600 to STOP), ΔGAL3(-600 to STOP) :: scGAL3(-600 to STOP), ΔGAL4(-600 to STOP) :: scGAL4(-600 to STOP), ΔGAL7(-600 to STOP) :: scGAL7(-600 to STOP), ΔGAL7(-600 to STOP) :: scGAL7(-600 to STOP), ΔGAL10(-600 to STOP) :: scGAL10(-600 to STOP), ΔGAL80(-600 to STOP) :: scGAL80(-600 to STOP)
S. bay	JRY409	JRY294	ΔGAL1(-600 to -1) :: scGAL1(-600 to -1), ΔGAL7(-600 to -1) :: scGAL7(-600 to -1), ΔGAL10(-600 to -1) :: scGAL10(-600 to -1)
S. bay	JRY410	JRY294	ΔGAL1(-600 to -1) :: scGAL1(-600 to -1), ΔGAL7(-600 to -1) :: scGAL7(-600 to -1), ΔGAL10(-600 to -1) :: scGAL10(-600 to -1)
S. bay	JRY510	JRY294	ΔGAL7(+1 to STOP) :: scGAL7(+1 to STOP)
S. bay	JRY511	JRY294	ΔGAL7(+1 to STOP) :: scGAL7(+1 to STOP)
S. bay	JRY512	JRY294	ΔGAL3(+1 to STOP) :: scGAL3(+1 to STOP)
S. bay	JRY513	JRY294	ΔGAL3(+1 to STOP) :: scGAL3(+1 to STOP)
S. bay	JRY514	JRY296	ΔGAL80(+1 to STOP) :: scGAL80(+1 to STOP)
S. bay	JRY515	JRY296	ΔGAL80(+1 to STOP) :: scGAL80(+1 to STOP)
S. bay	JRY521	JRY296	ΔGAL10(+1 to STOP) :: scGAL10(+1 to STOP)
S. bay	JRY522	JRY296	ΔGAL10(+1 to STOP) :: scGAL10(+1 to STOP)
S. bay	JRY527	JRY294	ΔGAL2(+1 to STOP) :: scGAL2(+1 to STOP)
S. bay	JRY528	JRY294	ΔGAL2(+1 to STOP) :: scGAL2(+1 to STOP)
S. bay	JRY529	JRY296	ΔGAL4(+1 to STOP) :: scGAL4(+1 to STOP)
S. bay	JRY530	JRY296	ΔGAL4(+1 to STOP) :: scGAL4(+1 to STOP)
S. bay	JRY516	JRY294	ΔGAL1(+1 to STOP) :: scGAL1(+1 to STOP)
S. bay	JRY517	JRY294	ΔGAL1(+1 to STOP) :: scGAL1(+1 to STOP)
S. bay	JRY538	JRY294	ΔGAL1(+1 to STOP) :: scGAL1(+1 to STOP), ΔGAL7(+1 to STOP) :: scGAL7(+1 to STOP), ΔGAL10(+1 to STOP) :: scGAL10(+1 to STOP)
S. bay	JRY539	JRY294	ΔGAL1(+1 to STOP) :: scGAL1(+1 to STOP), ΔGAL7(+1 to STOP) :: scGAL7(+1 to STOP), ΔGAL10(+1 to STOP) :: scGAL10(+1 to STOP)
S. bay	JRY571	JRY296XJRY294	ΔGAL1(+1 to STOP) :: scGAL1(+1 to STOP), ΔGAL7(+1 to STOP) :: scGAL7(+1 to STOP), ΔGAL10(+1 to STOP) :: scGAL10(+1 to STOP), ΔGAL2(+1 to STOP) :: scGAL2(+1 to STOP), ΔGAL3(+1 to STOP) :: scGAL3(+1 to STOP), ΔGAL4(+1 to STOP) :: scGAL4(+1 to STOP), ΔGAL80(+1 to STOP) :: scGAL80(+1 to STOP)
S. bay	JRY572	JRY296XJRY294	ΔGAL1(+1 to STOP) :: scGAL1(+1 to STOP), ΔGAL7(+1 to STOP) :: scGAL7(+1 to STOP), ΔGAL10(+1 to STOP) :: scGAL10(+1 to STOP), ΔGAL2(+1 to STOP) :: scGAL2(+1 to STOP), ΔGAL3(+1 to STOP) :: scGAL3(+1 to STOP), ΔGAL4(+1 to STOP) :: scGAL4(+1 to STOP), ΔGAL80(+1 to STOP) :: scGAL80(+1 to STOP)
S. bay	JRY560	JRY296XJRY294	ΔGAL3(+1 to STOP) :: scGAL3(+1 to STOP), ΔGAL4(+1 to STOP) :: scGAL4(+1 to STOP)

S. bay	JRY561	JRY296XJRY294	ΔGAL3(+1 to STOP) :: scGAL3(+1 to STOP), ΔGAL4(+1 to STOP) :: scGAL4(+1 to STOP)
S. bay	JRY432	JRY296XJRY294	ΔGAL3(-600 to -1) :: scGAL3(-600 to -1), ΔGAL4(-600 to -1) :: scGAL4(-600 to -1)
S. bay	JRY461	JRY296XJRY294	ΔGAL3(-600 to -1) :: scGAL3(-600 to -1), ΔGAL4(-600 to -1) :: scGAL4(-600 to -1)
S. bay	JRY415	JRY296XJRY294	ΔGAL3(-600 to STOP) :: scGAL3(-600 to STOP), ΔGAL4(-600 to STOP) :: scGAL4(-600 to STOP)
S. bay	JRY416	JRY296XJRY294	ΔGAL3(-600 to STOP) :: scGAL3(-600 to STOP), ΔGAL4(-600 to STOP) :: scGAL4(-600 to STOP)

Supplementary Table 1. Strains used in this study.

References

1. Orr HA. The genetic theory of adaptation: a brief history. Nat Rev Genet. 2005;6: 119–127. doi:10.1038/nrg1523

2. Rockman MV. THE QTN PROGRAM AND THE ALLELES THAT MATTER FOR EVOLUTION: ALL THAT'S GOLD DOES NOT GLITTER. Evolution. Wiley Online Library; 2012.

3. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010;20: R208–15. doi:10.1016/j.cub.2009.11.055

4. SAVOLAINEN O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14: 807–820. doi:10.1038/nrg3522

5. Nadeau NJ, Jiggins CD. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends Genet. Elsevier Ltd; 2010;26: 484–492. doi:10.1016/j.tig.2010.08.004

6. Schraiber JG, Mostovoy Y, Hsu TY, Brem RB. Inferring Evolutionary Histories of Pathway Regulation from Transcriptional Profiling Data. Teichmann SA, editor. PLoS Comput Biol. 2013;9: e1003255. doi:10.1371/journal.pcbi.1003255.s012

7. Caudy AA, Guan Y, Jia Y, Hansen C, DeSevo C, Hayes AP, et al. A new system for comparative functional genomics of Saccharomyces yeasts. Genetics. 2013;195: 275–287. doi:10.1534/genetics.113.152918

8. Bullard JH, Mostovoy Y, Dudoit S, Brem RB. Polygenic and directional regulatory evolution across pathways in Saccharomyces. Proc Natl Acad Sci USA. 2010;107: 5058–5063. doi:10.1073/pnas.0912959107

9. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. 2014

Feb. doi:10.1101/002907

10. 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. doi:10.1371/journal.pbio.1002041.s017

11. 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.s025

12. De Deken RH. The Crabtree effect: a regulatory system in yeast. Microbiology (Reading, Engl). Soc General Microbiol; 1966;44: 149–156.

13. Peng W, Liu P, Xue Y, Acar M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nat Commun. Nature Publishing Group; 2015;6: 1–9. doi:10.1038/ncomms7226

14. Nair NU, Zhao H. Mutagenic inverted repeat assisted genome engineering (MIRAGE). Nucleic Acids Res. 2009;37: e9. doi:10.1093/nar/gkn943

15. Storici F, Resnick MA. The Delitto Perfetto Approach to In Vivo Site-Directed Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast. Methods in Enzymology. Elsevier; 2006. pp. 329–345. doi:10.1016/S0076-6879(05)09019-1

16. Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, et al. Trait variation in yeast is defined by population history. PLoS Genetics. 2011;7: e1002111. doi:10.1371/journal.pgen.1002111

17. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods. 2001;25: 402–408. doi:10.1006/meth.2001.1262

18. Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, et al. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. G3 (Bethesda). 2011;1: 11–25. doi:10.1534/g3.111.000273

19. Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, et al. Fast statistical alignment. PLoS Comput Biol. Public Library of Science; 2009;5: e1000392.

20. Haddrill PR, Bachtrog D, Andolfatto P. Positive and Negative Selection on Noncoding DNA in Drosophila simulans. Mol Biol Evol. 2008;25: 1825–1834. doi:10.1093/molbev/msn125

21. Stoletzki N, Eyre-Walker A. Estimation of the neutrality index. Mol Biol Evol. 2011;28: 63–70. doi:10.1093/molbev/msq249

22. Charlesworth J, Eyre-Walker A. The McDonald-Kreitman Test and Slightly Deleterious Mutations. Mol Biol Evol. 2008;25: 1007–1015. doi:10.1093/molbev/msn005

23. He BZ, Holloway AK, Maerkl SJ, Kreitman M. Does Positive Selection Drive Transcription Factor Binding Site Turnover? A Test with Drosophila Cis-Regulatory Modules. Petrov DA, editor. PLoS Genetics. 2011;7: e1002053. doi:10.1371/journal.pgen.1002053.t001

24. Rand DM, Kann LM. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol Biol Evol. 1996;13: 735–748.

25. Nei M. Molecular Evolutionary Genetics. Columbia University Press, New York; 1987.

26. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA. National Acad Sciences; 1979;76: 5269–5273.