THE EVOLUTIONARY CONSTRAINTS AND DYNAMICS OF SUGAR METABOLISM IN YEASTS

By

Meihua Kuang

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Cellular and Molecular Biology)

at the
UNIVERSITY OF WISCONSIN–MADISON
2017

Date of final oral examination: 08/07/2017

The dissertation is approved by the following members of the Final Oral Committee:

Chris Todd Hittinger, Assistant Professor, Department of Genetics
Audrey P. Gasch, Professor, Department of Genetics
Nicole Perna, Professor, Department of Genetics
Judith Kimble, Professor, Department of Biochemistry
David C. Schwartz, Professor, Department of Genetics

ACKNOWLEDGMENTS

I sincerely acknowledge my thesis advisor, Dr. Chris Todd Hittinger, for being such a supportive mentor throughout my graduate study. I would not have accomplished any of the studies in this dissertation without his valuable training. I would also like to sincerely thank Amanda Hulfachor for helping me to learn English and to adapt to a totally new culture here. I would like to thank Dr. William G. Alexander, and Dr. Lily Wong for their help in developing my skills in experiments and troubleshooting. I would like to thank every lab member in the Hittinger lab. I really appreciate all the help I ever received from everyone, and the fun discussions and science I had.

I would give my special thank you to my parents. They always gave me their best. I would not have been in graduate school without the strict parenting from my father. I am really lucky to have all these supports from my parents. I would also like to specifically thank my husband, Bing Wang, for always being available and very supportive. I would like to thank my best friend, Dan Wu, for all the enjoyable discussions in archeology, evolution, history, physics, and mathematics. I would like to thank the teaching team in Biocore 383 and all the students in the class for the special teaching and learning experiences they provided. I have learned a lot from all my friends in Madison. This was a precious experience in my life. Furthermore, I would like to thank my past scientific mentors, Dr. Chunhua Fu, Dr. Jianfeng Liu, Dr. Ying Yang, Yaohua Pan, Guoyuan Deng, and Maoru Xu.

Lastly, I would like to thank the beautiful and fascinating nature for offering me all these wonders, and the excellent scientific community here in UW-Madison that is together built by all the outstanding scientists.

TABLE OF CONTENTS

Page

LIST OF TABLES iv

LIST OF FIGURES v

ABSTRACT viii

1 Introduction 1

1.1 Challenges for studying complex trait evolution 1
1.2 The evolutionary dynamics of duplicate gene divergence 4
1.3 From genetic divergence to GRN output modifications 6
1.4 Constraints on GRN evolution 8
1.5 Recurring patterns imply how natural laws constrain evolution 9
1.6 Yeast galactose metabolism as a model to study evolutionary constraints 12

2 Ongoing diversification of ancient paralogs regulating metabolism[1,2] 14

2.1 Abstract 14
2.2 Introduction 15
2.3 Results 17
2.3.1 *GAL* gene content and sequence differences 17
2.3.2 Less partitioned galactokinase and co-induction functions 18
2.3.3 *S. uvarum* has two co-repressors with partially overlapping functions . . . 23
2.3.4 Strains lacking the co-repressors arrest their growth 25
2.3.5 Overactive galactose catabolism precedes widespread metabolic and regulatory defects 26
2.3.6 Specific sugars can exacerbate metabolic overload 40
2.3.7 The less active *S. cerevisiae GAL* network is less susceptible to metabolic overload when derepressed 42
2.4 Discussion 45
2.4.1 Biodiversity offers a panoramic window to molecular biology 45
2.4.2 The non-equivalence of sugars in contributing to metabolic overload 46
2.4.3 Network architectures reflect metabolic constraints 46
2.4.4 Ongoing functional diversification of paralogs and their gene networks . . 47
2.5 Materials and Methods 51
2.6 Acknowledgments 55

Page

3 Repeated cis-regulatory tuning of a metabolic bottleneck gene during evolution[1] . . 57

3.1 Abstract . . . 57
3.2 Introduction . . . 58
3.3 Results . . . 60
3.4 Discussion . . . 70
3.5 Acknowledgements . . . 73
3.6 Supplementary notes . . . 73
3.7 Materials and methods . . . 74

4 Conclusions . . . 77

4.1 The evolutionary dynamics of duplicate genes often underlie GRN evolution . . . 79
4.2 General constraints in the evolution of sugar metabolism . . . 80
4.3 Evolution through cis-regulatory elements or protein coding regions . . . 81
4.4 To understand the mechanism is to predict . . . 82
4.5 Evolutionary studies can fuel the study of metabolism and molecular mechanisms . 83

APPENDICES

Appendix A: Supplementary Materials for Chapter 2 . . . 100
Appendix B: Supplementary Materials for Chapter 3 . . . 104

LIST OF TABLES

Table Page

Appendix
Table

A.1 Amino acid identity and *GAL* gene composition between *S. uvarum* and *S. cerevisiae* *GAL* network. 100

A.2 *S. uvarum* and *S. cerevisiae* have qualitatively similar *gal1* null phenotypes. 101

A.3 Metabolite ratio between *S. uvarum gal80*Δ *gal80b*Δ and wild-type in SC +2% galactose over time. 102

A.4 Normalized metabolite concentrations in *S. uvarum* wild-type and *gal80*Δ *gal80b*Δ in galactose over time. 103

B.1 A summary of 20 genetic features in 17 species of the family Saccharomycetaceae. . . 113

B.2 Signatures of relaxed selection on the *GAL* genes in the *Lachancea* species that grew slowly. 114

B.3 A summary of *Lachancea GAL/MEL* gene features and growth in galactose or melibiose. 115

LIST OF FIGURES

Figure Page

2.1 The *S. uvarum GAL* network. 18

2.2 *SuvaGAL1* and *SuvaGAL3* are not as subfunctionalized as *ScerGAL1* and *ScerGAL3*. . 19

2.3 *SuvaGAL80* and *SuvaGAL80B* encode co-repressors with partially overlapping functions. 22

2.4 In SC + 2% galactose, *S. uvarum gal80*Δ and *gal80b*Δ had *GAL1* expression levels similar to the wild-type at mid-log phase. 24

2.5 The TGA phenotype of *S. uvarum gal80*Δ *gal80b*Δ can be rescued by *S. cerevisiae GAL80* or by re-introducing *SuvaGAL80*. 27

2.6 The galactose-dependent temporary growth arrest phenotype of *S.uvarum gal80*Δ *gal80b*Δ. 28

2.7 Galactose-specific global differential expression of *S. uvarum gal80*Δ *gal80b*Δ. . . . 29

2.8 High performance liquid chromatography measurements of key metabolites in SC + 2% galactose in *S. uvarum gal80*Δ *gal80b*Δ and wild-type during the TGA phase and after the growth resumed. 31

2.9 *GAL1* expression was higher at the early stages of growth in SC + 2% galactose in the *S. uvarum gal80*Δ *gal80b*Δ background but gradually decreased. 32

2.10 Fructose, mannose, or glucose alone did not lead to a TGA phenotype or other growth defects. 33

2.11 The regulation of *PGM1* by galactose was inferred as the ancestral state. 34

2.12 Galactose-1-phosphate accumulation of *S. cerevisiae gal7*Δ and *gal10*Δ. 36

2.13 Overly rapid galactose catabolism leads to metabolic overload and bottlenecks. 37

Figure Page

2.14 The addition of sugars downstream of the trehalose cycle exacerbated metabolic overload. 41

2.15 The less active *S. cerevisiae GAL1* gene is partially responsible for a subtle Temporary Growth Arrest. 43

2.16 Ongoing diversification of the functions of the *GAL1-GAL3* and *GAL80-GAL80B* duplicate gene pairs in *Saccharomyces*. 48

3.1 Predicted Gal4-binding sites upstream of *PGM1/2* strongly correlate with galactose growth in the family Saccharomycetaceae. 61

3.2 Gal4-mediated regulation of phosphoglucomutase is necessary and sufficient to support vigorous growth on galactose for multiple yeast species. 63

3.3 Predicted Gal4-binding sites upstream of *PGM1* predicted galactose growth differences among closely related species in the genus *Lachancea*. 66

3.4 The significance of Gal4-mediated regulation of *PGM1* depends on galactose concentrations. 69

Appendix
Figure

B.1 The *PGM* association was specific to galactose and not related to the growth in glucose. 104

B.2 Mutation of Gal4-binding sites upstream of *PGM1* did not cause any growth defects on glucose in *S. uvarum* and *L. kluyveri*. 105

B.3 Temporary Growth Arrest (a phenotype caused by metabolic overload when both Gal80 co-repressors are deleted) was completely rescued by the deletion of two predicted Gal4-binding sites upstream of *SuvaPGM1*. 106

B.4 The correlation between growth on galactose and the number of predicted Gal4-binding sites upstream of *PGM1/2* is maintained in different base media and temperatures. 107

B.5 The effect of Gal4 site mutation in the promoter of *PGM1/2* in *S. kudriavzevii* or *K. lactis*. 109

Figure Page

B.6 *Lachancea* species showed strong association of having predicted Gal4 site upstream of *PGM1* with growth in galactose but not in glucose. 111

B.7 A clade in the *Lachancea* genus that grew slowly in high concentrations of galactose grew slightly better at low concentrations. 112

ABSTRACT

Our incomplete understanding of the relationship between genotypes and phenotypes makes it extremely difficult to understand and predict the evolution of diverse phenotypes in nature. Phenotypes evolve on the basis of genetic alterations. Gene regulatory networks (GRNs) build the link between genotypes and phenotypes. Using well-characterized GRNs as models may get us closer to understand the relationship between genotypes and phenotypes. Taking advantage of powerful yeast genetics, this dissertation uses the yeast *GAL*actose (*GAL*) utilization networks as a model and carefully dissects three levels of evolutionary mechanisms between genotypes and phenotypes. The three levels include the functional divergence of individual genes, divergence of GRN activities, and the co-evolution between individual genes and between interacting GRNs. Through these three levels of characterization, I reveal that *GAL* network activity is constrained by downstream glycolysis capacity. This is likely a general constraint to sugar metabolized through glycolysis. I further examine recurring genetic changes underlying repeated tuning of *GAL* network activities across multiple genera and reveal a general pleiotropic constraint at the bottleneck of galactose metabolism. These constraints even enable remarkable prediction of *GAL* network activities merely based on DNA sequences. Understanding the genetic and functional basis of trait divergence reveals general evolutionary constraints and allows us to predict evolutionary outcomes.

Chapter 1

Introduction

1.1 Challenges for studying complex trait evolution

The astonishing diversity of phenotypes among living organisms has long fascinated evolutionary biologists. In 1859, Charles Darwin's *On the Origin of Species* proposed the theory of natural selection and laid the foundation of evolutionary studies (Darwin, 1859). Natural selection describes the differential survival and reproduction of individuals due to their phenotypic differences. A phenotype is a collection of observable characteristics or traits of an organism, which is the result of both genetic and environmental factors. However, the understanding that phenotypes come from genetics was a long journey of scientific research in the 20th century. The nature of inheritance was not revealed until the rediscovery of the Mendelian laws of inheritance in the early 20th century. DNA was discovered as the hereditary material in 1940s. Between the 1950s and 1970s, the Central Dogma of Molecular Biology was gradually developed. It defines how genetic information flows to, or between RNAs and proteins. The mechanisms from genotypes to phenotypes are now relatively well understood for numerous simple traits that are determined by a single gene, but are mostly elusive for complex traits that are controlled by multiple genes. It is extremely challenging to construct the map from genotypes to phenotypes for complex traits, to say nothing of how changes in genotypes lead to alterations of phenotypes. Understanding the evolution of complex traits therefore still remains a major challenge in evolutionary studies. Instead of focusing on specific underlying genes, many have focused on overall genetic landscapes, which might be more readily addressed: whether there exists a major effector locus, or whether change is mainly constituted of numerous small-effector genes. As a link from genotypes to phenotypes, the

underlying gene regulatory networks (GRNs) are well-characterized for plenty of traits in model organisms. Using a well-characterized GRN in a model organism as a reference, we can conduct functional comparative studies to understand not just the genetic basis but more importantly, the functional basis of phenotypic trait divergence between related species. To functionally compare the two GRNs, there are multiple levels we can compare: 1) individual genetic components, 2) GRN properties, and 3) the co-evolution between genes or between GRNs. Dissecting each level will get us closer to understanding the relationship between genotypes and phenotypes.

On the first level, to dissect which genetic components are functionally different, one can create homologous mutants to compare the possibly differential phenotypic changes, compare homolog activities in vitro, or ideally, compare functional divergence in the same genetic background through reciprocal complementation or allele swaps. Specifically, reciprocal complementation is through reciprocal crosses while only one parental allele is functionally active in each cross; allele swap replaces a gene with its homolog from a related species through genetic engineering approaches. Chapter 2 of this dissertation takes advantage of the power of yeast genetic engineering and conducts a functional comparative study through a series of allele swaps. Since duplicate genes play specific roles in evolution and might continuously provide potential to GRN evolution, which will be further discussed later in this chapter, Chapter 2 focuses on whether the evolutionary state of a duplicate gene constrains the potential it offers to GRN evolution.

On the second level, overall GRN properties depend on the architecture and details of the GRNs, which will be further described later in this chapter. Chapters 2 and 3 compare GRN activities between related yeast species across a broad phylogenetic scale.

On the third level, this dissertation specifically focuses on how functional divergence of multiple genes together alters GRN activities and how the evolution of GRNs is constrained by other interacting GRNs. How gene divergence contributes to GRN divergence is not well understood, since functional divergence of multiple genes does not necessarily alter GRN activities linearly. Thus, Chapters 2 and 3 both take into account GRN architectures in understanding the co-evolution. Possible approaches to reveal constraints by other GRNs include dramatically decreasing or increasing the GRN output or studying the constraints at the intersection of the GRN we studied and

other GRNs. Chapter 2 uses the first approach, while Chapter 3 uses the latter approach to study constraints on GRN evolution.

Through studying the above three levels, we might be able to reveal the functional basis of trait divergence, and more significantly, the possible evolutionary constraints underlying trait divergence. Phenotypes evolve on the basis of multiple GRNs. GRNs evolve through the alterations of individual genes and their regulation, which leave signatures on DNA sequences. Similar genetic changes (e.g., amino acid substitutions, cis-regulatory changes) thus imply genetic constraints underlying similar phenotypes in distant lineages. Many combinations of genetic mutations can lead to similar phenotypic outcomes, but only small subsets are selected by natural selection during long-term evolution. I refer to these limitations as genetic constraints. The constraints leading to limited genetic solutions for particular phenotypes might come from any of the above three mechanistic levels: such as limited sets of amino acids conferring particular protein functions, or constraints by other genes or other GRNs on alternative changes. Both genes and GRNs interact with other genes and GRNs to perform their functions and finally lead to phenotypic outcomes in response to intrinsic or extrinsic signals. Genetic changes might perturb the functions of other genes or GRNs during evolution and therefore be constrained. The stronger the constraints are, the more repeatable and thereby more predictable evolution is. Chapter 2 systematically compare the differences at the above three levels of a homologous GRN between two related species and reveals constraints at multiple levels. Chapter 3 investigates recurring genetic changes underlying repeated tuning of the GRN activities across a broad phylogenetic scale and examines whether constraints enable prediction of GRN evolution.

In summary, this dissertation extensively dissects three mechanistic levels between genotypes and phenotypes and the generality of evolutionary constraints. Specifically, Chapters 2 and 3 study how functional divergence of multiple genes together alters the output of a GRN and in turn, how this output is constrained by other interacting GRNs. To understand the mechanism is to better predict, so lastly, Chapter 3 further investigates recurring genetic changes underlying repeated tuning of a trait at a broader phylogenetic scale to reveal general constraints and examine the predictive

power of constraints. In the following, I will discuss the background for each mechanistic level and recurrent evolution in more detail.

1.2 The evolutionary dynamics of duplicate gene divergence

In 1970, Susumu Ohno proposed the significance of gene duplication in evolution in his book *Evolution by Gene Duplication* (Ohno, 1970). Gene duplication is the creation of a second copy of a gene, or part of a gene. It can occur through small-scale tandem duplication, segmental duplication, or whole-genome duplication (WGD). Possession of only one gene limits opportunities to accumulate mutations to generate novel changes. However, if there are two copies of one gene, the selection pressure acting on either copy is relaxed. It allows increased chances to accumulate mutations without disrupting the original function because of the buffer provided by the other copy. Therefore, gene duplication creates evolutionary opportunities and might give rise to novel regulation or novel protein functions. In the last two decades, gene duplication has been a major focus of evolutionary studies. With recent advances in genome sequencing technologies, convincing evidence has shown that WGDs have occurred multiple times across the tree of life: three times in paramecium (McGrath et al., 2014), once in yeasts (Wolfe and Shields, 1997, Marcet-Houben and Gabaldon, 2015), many times in plants (Arabidopsis Genome, 2000, Blanc et al., 2000), and three times in vertebrates (Dehal and Boore, 2005), with one case specific to teleost fishes (Jaillon et al., 2004). WGD has been associated with massive changes in evolution, such as adaptive radiation. There is a temporal association between WGD events and the massive speciation in angiosperms (De Bodt et al., 2005) and teleost fishes (Postlethwait et al., 2004). The global metabolic shift from aerobic respiration to aerobic fermentation in yeasts has been proposed to have been facilitated by WGD (Conant and Wolfe, 2007, Jiang et al., 2008). Chapters 2 and 3 will both focus on yeast GRNs that are impacted by WGD to examine how WGD might have resolved any evolutionary constraints.

After the WGD, most duplicate genes were rapidly lost, followed by millions of years of slower loss, a general trend for diverse species. After WGD, only roughly 10–50% of duplicate genes survived (McGrath et al., 2014, Wolfe and Shields, 1997, Aury et al., 2006, Jaillon et al., 2004).

The evolution of duplicate genes created by WGD followed three stages (Innan and Kondrashov, 2010): 1) The initial redundancy stage: the two duplicate genes might be subject to random loss due to the complete redundancy, while minor differences will eventually accumulate between copies over time (Scannell et al., 2007); 2) The fate determination stage: both duplicate genes can be fixed in the population if their differences are visible to natural selection; 3) The maintenance stage: both copies are now under purifying selection or continuously modified by natural selection. Even after maintaining in the genome for millions of years, duplicate genes may be lost when their functions become obsolete, as a result of ecological niche shift, such as the massive loss of olfactory duplicate genes in the primate lineages (Gilad et al., 2003), or due to the evolution of alternative genetic solutions. Intriguingly, paralogs are often lost asymmetrically, with one paralog lost more frequently in independent clades. These evolutionary dynamics of duplicate genes might impact the potential of a duplicate gene contributing to GRN evolution. Chapter 2 will discuss duplicate gene loss and decreased functional redundancy between duplicate genes, and how these changes impact GRN evolution.

Although most duplicate genes are lost, likely due to their redundancy, a small subset have survived. Extensive research focusing on the mechanism of duplicate gene maintenance has proposed several hypotheses: 1) Gene dosage selection: gene duplication increases the expression level of a gene, which might be selected for by natural selection; 2) Neofunctionalization: novel functions arose in duplicate genes (Walsh, 1995, Conantand Wolfe, 2008); 3) Sub-functionalization through duplication-degeneration-complementation (Force et al., 1999) or escape from adaptive conflict (Hittinger and Carroll, 2007, Des Marais and Rausher, 2008): if the ancestral gene performs multiple functions, then the ancestral functions can be subdivided into the descendent copies. The latter process can occur neutrally, in which both copies are still required to perform the ancestral functions (Force et al., 1999), or this process can occur adaptively by allowing the ancestral functions to be optimized in two separate genes (Hittinger and Carroll, 2007, Des Marais and Rausher, 2008).

Duplicate genes do not diverge exclusively through subfunctionalization or neofunctionalization, but instead, multiple forces can co-occur and intertwine (He and Zhang, 2005, Voordeckers et

al., 2012). In fact, duplicate gene divergence might be more dynamic than previously appreciated. For instance, a large family of fungal glucosidase are descendent copies from one ancestral gene, and dynamically evolved to different enzyme activities at different time points in different clades, involving both neofunctionalization and subfunctionalization (Voordeckers et al., 2012). Another example is a group of homeodomain-containing transcription factors, the *Hox* genes. Descended from multiple ancient gene duplication events, several *Hox* duplicate genes continuously diverged, likely contributing to the morphological diversity in different lineages (Carroll, 1995, Carroll et al., 2005). For instance, *Hox3* lost its *Hox* function and gained a novel role in beetles and locusts (Falciani et al., 1996), and diverged again after another duplication in flies (Stauber et al., 2002). Fushi tarazu (Ftz) continuously diverged and developed different roles in beetles (Brown et al., 1994, Stuart et al., 1991) and in flies (Lohr et al., 2001, Damen, 2002). Ultrabithorax (Ubx) switched from a transcriptional activator to a repressor in Insecta, which probably facilitated the evolution of limbless abdomens in modern insects (Galant and Carroll, 2002). Chapter 2 will examine how the ongoing functional divergence of ancient duplicate genes contribute to divergent GRN outputs in different yeast lineages.

1.3 From genetic divergence to GRN output modifications

The architecture of GRNs include the genes involved and the regulatory interactions between each gene, and any molecules a gene product might interact with (e.g., metabolites). Regulatory interaction leads to activation or repression of the target gene in a directional, quantitative, and time-dependent manner. Each gene can be defined as a node, and the corresponding regulatory interaction can be defined as a directional edge. The architecture of a GRN can be modified through the addition or removal of the node or the edge. The molecular mechanisms of such modifications have been extensively characterized (Baker et al., 2012, McKeown et al., 2014, Sayou et al., 2014, Ihmels et al., 2005, Arnoult et al., 2013, Pougach et al., 2014, Carroll, 2008, Gasch et al., 2004, Tsong et al., 2006). The loss of an edge can occur through changes in a DNA binding domain (Sayou et al., 2014), a protein-protein interaction domain, or its cis-regulatory elements (Carroll, 2008, Ihmels et al., 2005, Gasch et al., 2004). The removal of a node can occur through gene loss

(Baker et al., 2012) or simply the loss of an edge. The addition of a new node can occur through gene duplication of an existent node (Arnoult et al., 2013, Conant and Wolfe, 2006, Pougachet al., 2014) or the emergence of DNA-binding activity (Sayou et al., 2014, Pougach et al., 2014), protein-protein interaction activity (Baker et al., 2012), or the addition of regulatory elements to an existent node (Carroll, 2008, Ihmels et al., 2005, Gasch et al., 2004, Tsong et al., 2006). Hybrid or intermediate states where both the ancestral regulatory scheme and derived scheme are present have been shown to be important for these step-wise modifications during evolutionary transition (Baker et al., 2012, Sayou et al., 2014, Pougach et al., 2014). Chapters 2 and 3 will both examine the modifications of GRN architectures in different lineages.

Functional changes of individual genes or rewiring of GRNs do not necessarily change the GRN output. Substantial alterations have occurred in the mating network in different yeast clades but few have altered the GRN output (Baker et al., 2012, Tsong et al., 2006, Baker et al., 2011). The haploid-specific genes are regulated by an activator in the *Candida albicans* clade, but switched to regulation by a repressor in the *Saccharomyces* clade (Baker et al., 2012). Matα1, a transcriptional regulator in the yeast mating pathway, has extensively altered its DNA binding specificity to bind to unrelated DNA sequences in the common ancestor of the CUG-clade, but unexpectedly, Matα1 still regulates the same core set of target genes (Baker et al., 2011). Although some genetic changes apparently do not alter phenotypes, the lack of phenotypic changes can be limited by the conditions examined and the aspects of GRN output examined.

As a unit, GRNs display characteristics different from individual genes. In some cases, the properties of a GRN are well defined and predictable (Alon, 2007). For instance, some incoherent feedforward loops will result in a specific and transient response pulse, such as the glucose signaling pathways mediated through Rtg1 or Mig2 (Kuttykrishnan et al., 2010). Some coherent feedforward loops lead to a delay in response to the input signal (Mangan et al., 2003, Alon, 2007), while some with delay after the removal of the signal (Kalir et al., 2005, Alon, 2007). How feedback or feedforward loops are wired in a GRN determine numerous characteristics, such as the response time (Mangan et al., 2006), expression noise (Chalancon et al., 2012, Macneil and Walhout, 2011), oscillation frequency (Kearns et al., 2006, Nelson et al., 2004), and bimodality

(Venturelli et al., 2012). For instance, the transition rate between turning the GRN on and off and the inducibility by different concentrations of input signals can both be determined by the differential strengths of negative feedback loop (Peng et al., 2015, Avendano et al., 2013). These properties can be important to evolutionary adaptation in the complicated and ever-changing natural environment. Since the functional divergence of individual genes does not always linearly alter the GRN output, understanding GRN architectures might shed light on specific properties being selected for. Chapters 2 and 3 will further discuss how GRN architectures are configured to modulate GRN output during evolution.

1.4 Constraints on GRN evolution

Since most GRNs are not independent of one another, occasionally there will be conflicts between GRN evolution. The evolution of glucose repression on respiration during the evolutionary transition from aerobic respiration to aerobic fermentation due to the trade-offs between glucose fermentation and respiration (Kayikci and Nielsen, 2015). This transition also de-coupled the transcription of mitochondria ribosomal genes and cytoplasmic ribosomal genes, likely because of the need to de-couple cell growth and mitochondria biogenesis (Ihmels et al., 2005). Yeast Ste12 transcription factor regulates a similar set of mating-pheromone responsive genes (**a**-specific genes) by directly binding to DNA in the *Saccharomyce* clade, but indirectly binding to DNA through a secondary regulator in *K. lactis*. An otherwise direct DNA contact caused mis-expression of target genes in *K. lactis*. This constraint is resolved through the emergence of a repressor in an overlapping GRN in *S. cerevisiae*, likely allowing repression of mis-expression (Sorrells et al., 2015). Understanding how GRNs constrain each other during evolution might reveal in what scenarios GRNs are constrained and thus evolve on more restricted paths. These constraints might enable us to better predict GRN evolution. Chapters 2 and 3 both explore the constraints on GRN evolution and the generality of constraints across a broad evolutionary time scale.

1.5 Recurring patterns imply how natural laws constrain evolution

Evolutionary trajectories are confounded by historical contingency and complicated mapping from genotypes to phenotypes, making the prediction of evolution extremely challenging. Fortunately, phylogenetically replicated evolutionary changes give us hints that evolution has probably walked on a much more restricted and thereby predictable path from time to time, which might allow us to outline what S. J. Gould called "the boundary between predictability under invariant law and the multifarious possibilities of historical contingency" (Gould, 1989). Repeated trait modifications describe convergent or parallel changes that evolved independently in different clades. When such changes evolved from similar ancestral states to the same derived states, it is termed parallel evolution. If these changes started from different ancestral states but evolved to similar states, it is termed convergent evolution. If similar genetic changes account for parallel or convergent phenotypic adaptation, it is called genetic convergence, which implies that there are restricted sets of available genetic solutions to achieve particular phenotypic adaptation (Stern, 2013, Martinand Orgogozo, 2013, Christin et al., 2010).

Due to the constraints by physics and physiology, plenty of features in phenotypic adaptation evolved repeatedly in unrelated taxa (Wake, 1991, Kelley and Pyenson, 2015). For instance, to achieve similar functions, numerous distantly related species have convergent skeletal features. Multiple lineages of vertebrates that are able to fly or glide independently evolved similar appendages through the extension of their forelimbs, such as the birds, bats, and flying squirrels. Divers evolved from terrestrial ancestors had their limbs modified into convergent hydrodynamic (i.e., flipper-like) forelimb structures (Kelley and Pyenson, 2015), such as the sea lion, elephant seal, otter, ducks, and whales. Specialized jumpers often have extended or sturdy hind legs, such as the kangaroo, jerboa, springhare, leaf hopper, grasshopper, locust, and cricket. Herbivorous animals often have grinding teeth, carnivorous predators have pointed teeth, and omnivores have both. In fact, dentition features allow archeologists to identify the diet of a fossil sample (Benton, 2010). However, both giant panda and red panda still have pointed teeth even though they have exclusive

bamboo diets, implying that transitional forms in evolution could be obscure, and obsolete features might take millions of years to degenerate.

Although some phenotypic traits have evolved in a predictable manner, which reflects the constraints by physics, the underlying genetic bases are much more difficult to predict. This is restricted by our lack of understanding of how genotypes map to phenotypes, and the mechanisms in between these two. Nonetheless, repeated genetic changes might reveal hints for prediction. Short-term experimental evolutionary studies have repeatedly revealed parallel mutations in independent evolving lines (Woods et al., 2006, Hong et al., 2011, Wong et al., 2012). Long-term evolution has also repeatedly used the same genes to cope with similar ecological challenges in some cases. Body color crypsis is a prevalent strategy that organisms evolved again and again to adapt to specific ecological niches. The coat color variation of mammals is mainly due to variation in the biosynthesis of melanin. This has repeatedly occurred through changes of *MC1R* or *agouti*, two genes in the characterized pathway of melanin biosynthesis. Alteration of these two genes has been associated with coat color variation in mice (Hoekstra et al., 2006, Nachman et al., 2003), multiple felid species (Schneider et al., 2015, Eizirik et al., 2003), and horses (Rieder et al., 2001). In addition to mammals, birds with a wide body coloration spectrum also have cases of melanin coloration changes associated with either gene (Theron et al., 2001). A different case of repeated evolution is the larva trichome patterning, which occurred through spatial expression loss of *shavenbaby* (*svb*) in different *Drosophila* species (Sucena et al., 2003). Acting as a developmental bottleneck gene and integrating all the patterning information, *svb* determines whether a cell will differentiate a trichome. The repeated spatial modification of *svb* is very likely due to pleiotropic constraints. Similar repeated spatial expression modification of *Dll* determines differential wing spot patterning in multiple wing-spotted *Drosophila* species (Arnoult et al., 2013). In addition to animals, convergent evolution is also prevalent in plants. For instance, the C_4 photosynthesis pathway in plants has evolved independently more than 45 times (Sage, 2004). A key C_4 enzyme, phosphoenolpyruvate carboxylase (PEPC) evolved from the same non-C_4 enzyme at least eight times independently in the grass family Poaceae, with extensive convergent amino acid substitution (Christin et al., 2007). Furthermore, the evolution of resistance to toxins (Zhen et al.,

2012, Ujvari et al., 2015), pesticides (Dobler et al., 2012), and antibiotics (Farhat et al., 2013, Wong et al., 2012) also repeatedly involved similar genetic changes. Major leaps in evolution arising from novel protein functions also occurred at the same sets of genes multiple times, such as lens crystallins being repeatedly recruited from heat shock proteins and enzymes (Wistow, 1993), further emphasizing the remarkable genetic constraints for some cases of adaptation.

Conservation across life kingdoms implies genetic constraints. Some genes are much more conserved than others, such as the extreme genetic conservation of ribosomal RNAs. The major metabolic pathways are well conserved from bacteria to humans, such as glycolysis and the TCA cycle. Some GRNs are well conserved without substantial changes in their genetic architectures, whereas others have diverged more extensively, such as the yeast mating pathways (Baker et al., 2012, Tsong et al., 2006).

Convergent evolution is not exclusive to single genes but can leave genome-wide scale signatures. The giant panda and red panda, separated by over 40 million years, independently evolved exclusive bamboo diets, as well as pseudo-thumbs from sesamoid bones from carnivorous ancestors. Genome sequence analyses have revealed numerous convergent amino acid substitutions in limb development genes, as well as vitamin metabolic genes in both species, likely to adapt to this specific dietary lifestyle (Hu et al., 2017). Convergent genetic mutations at a genome-wide scale are also found in echo locating mammals, such as bats and dolphins (Parker et al., 2013), multiple clades of carnivore plants (Fukushima et al., 2017), and local adaptations of two divergent conifers (Yeaman et al., 2016). These data imply that some genetic changes were selected in a more restricted and thereby more predictable manner.

To achieve the same phenotypes, even though some distant clades employ convergent genetic changes, other clades may adopt divergent genetic changes. For instance, melanic evolution involves alternative pathways. Three population of melanic mice in New Mexico and several felid species have melanism changes but no mutations are found in the coding regions of either of the two common target genes, *MC1R* or *agouti* (Hoekstra and Nachman, 2003, Eizirik et al., 2003). People of European and East Asia ancestry both have light skin colors after the ancestral migration

from Africa, but different genes are involved (Edwards et al., 2010). Why nature sometimes has selected the same sets of mutations but sometimes different sets is not well understood, but dissecting the mechanistic details underlying convergent or parallel phenotypic trait divergence might help us better understand the mechanisms. Chapter 3 will dissect the mechanisms underlying repeated tuning of galactose utilization and discuss reasons of alternative routes to convergent evolution.

1.6 Yeast galactose metabolism as a model to study evolutionary constraints

Fitness can be defined as the number of viable and fertile offspring produced in a given time. Fitness is a measure to quantify the action of natural selection. As unicellular microorganisms, yeasts can divide mitotically, and each yeast cell can generate one progeny in every cell division. The conditional fitness of yeasts can therefore be directly quantified as the changes in the number of cells in a given time at a given condition. Since the number of cells at a given size is proportional to absorbance, the conditional fitness can be scored based on absorbance measurements (Moore et al., 1988). Whether specific genetic modifications have an impact on conditional fitness can therefore be directly measured through absorbance.

Most metabolic pathways are well conserved from bacteria to humans, and therefore conclusions drawn from unicellular yeast organisms might be general. Exceptions might include those specific to ecological impacts or situations specific to single-celled organisms. Nonetheless, the patterns learned from genetic architectures or regulatory schemes might be general.

Galactose is metabolized through the Leloir pathway that is conserved from bacteria to humans. The *GAL*actose network is well characterized in the model organism *Saccharomyces cerevisiae*. As a relatively simple GRN, the yeast *GAL* network serves as a model in genetics (Johnston, 1987, Douglas and Hawthorne, 1966), molecular biology (Johnston, 1987, Kundu et al., 2007, Brickner, 2010), systems biology (Venturelli et al., 2012, Acar et al., 2005, Peng et al., 2015, Avendano et al., 2013), and evolution (Hittinger and Carroll, 2007, Hittinger et al., 2010, Hittinger et al., 2004, Roop et al., 2016, Dalal et al., 2016, Rubio-Texeira, 2005), providing the potential to integrate knowledge from multiple disciplines. Therefore, the yeast *GAL* network allows detailed dissection and integration of evolutionary constraints at multiple mechanistic levels. The *GAL* network in

S. cerevisiae includes the enzymes in the Leloir pathway, a transporter, and regulators. It contains basic regulatory features that are general in the regulatory networks of different organisms. It controls a measureable complex trait, galactose utilization, the output of which can be directly measured. Although the genetic architectures have been modified multiple times through WGD and differential duplicate gene loss in different clades, the regulatory schemes are relatively conserved in the family Saccharomycetaceae. With a small number of GRN members and powerful genetic manipulation tools, the genetic and mechanistic basis of differential *GAL* network activities and differential galactose utilization between related species can be more directly characterized.

Chapter 2 functionally compares the *GAL* network divergence between *S. cerevisiae* and its relative, *Saccharomyces uvarum*. *S. uvarum GAL* network has a few more duplicate genes, offering the opportunities to study the significance of duplicate gene divergence to GRN evolution. Chapter 2 systematically and carefully dissects three levels of evolution:

1) the functional dynamics of duplicate genes;

2) the constraints of functional divergence on duplicate genes contributing to the output of the GRN;

3) the constraints on GRN evolution with interacting GRNs.

As a quantitative trait, galactose utilization has been changed quantitatively and repeatedly among related species. Such variation in galactose utilization represents a repeated evolution of *GAL* network activity tuning and provides a good model to study whether the same genetic changes underlie the same direction of trait divergence. Since the metabolic and molecular mechanism of galactose metabolism is well characterized, Chapter 3 is able to examine the mechanisms underlying repeated metabolic evolution, in both the molecular and ecological aspects. Chapter 3 further discusses why the same sets of changes are selected in some species but alternatives are adopted in others. These phylogenetically replicated changes reveal us hints to predict evolution.

Chapter 2

Ongoing diversification of ancient paralogs regulating metabolism[1,2]

2.1 Abstract

The evolutionary mechanisms leading to duplicate gene retention are well understood, but the long-term impacts of paralog differentiation on the regulation of metabolism remain under appreciated. Here we experimentally dissect the functions of two pairs of ancient paralogs of the *GAL*actose sugar utilization network in two yeast species. We show that the *Saccharomyces uvarum* network is more active, even as over-induction is prevented by a second co-repressor that the model yeast *Saccharomyces cerevisiae* lacks. Surprisingly, removal of this repression system leads to a strong growth arrest, likely due to overly rapid galactose catabolism and metabolic overload. Alternative sugars, such as fructose, circumvent metabolic control systems and exacerbate this phenotype. We further show that *S. cerevisiae* experiences homologous metabolic constraints that are subtler due to how the paralogs have diversified. These results show how the functional differentiation of paralogs continues to shape regulatory network architectures and metabolic strategies long after initial preservation.

[1]This work was done under supervision of Chris Todd Hittinger and in collaboration with Paul D Hutchins, Jason D Russell, and Joshua J Coon. A version of this chapter was originally published as **Kuang MC, Hutchins PD, Russell JD, Coon JJ, Hittinger CT**. 2016. Ongoing resolution of duplicate gene functions shapes the diversification of a metabolic network. *Elife*. Sep 30; **5**:e19027.

[2]Author contributions. Conception and design: MCK, CTH. Acquisition of data: MCK, CTH, PDH. Analysis and interpretation of data: MCK, CTH, PDH, JDR, JJC. Drafting or revising the article: MCK, CTH, PDH, JDR, JJC.

2.2 Introduction

Gene duplication provides raw material for evolution to act upon. Even so, most duplicate genes are inactivated and become pseudogenes before fixation. The molecular mechanisms behind paralog retention and differentiation have attracted considerable attention, and several general models have been proposed, including neofunctionalization (Ohno 1970; Zhang et al. 2002), gene dosage selection (Conant and Wolfe 2007; Sandegren and Andersson 2009; Conant et al. 2014), subfunctionalization by duplication-degeneration-complementation (Force et al. 1999), and subfunctionalization by escape from adaptive conflict (Hittinger and Carroll 2007; Des Marais and Rausher 2008). Theoretical studies have proposed that the fates of duplicate genes are rapidly determined after gene duplication events (Moore and Purugganan 2003; Innan and Kondrashov 2010). These models generally treat the preservation of duplicate genes as a race to distinguish their functions prior to the complete inactivation of one of the redundant paralogs, either through neutral (Force et al. 1999; Lynch et al. 2001) or adaptive changes (Clark 1994; Lynch et al. 2001). Regardless of the initial functional changes or dosage effects facilitating the fixation of paralogs, retention is not the end of their evolutionary paths (Gordon et al. 2009; Conant et al. 2014).

Duplicate genes continue to diverge in different lineages, providing additional evolutionary opportunities for organisms to diversify. Previously fixed copies of duplicate genes can alter their expression timing and patterns (Huminiecki and Wolfe 2004; Tmpel et al. 2006), change substrate affinities (Voordeckers et al. 2012), and switch between regulatory modules (Thompson et al. 2013). In several cases, paralogs encoding enzymes have been recruited to perform regulatory functions, such as *S. cerevisiae HXK2*, *GAL3*, and *ARG82* (Gancedo and Flores 2008; Conant et al. 2014; Gancedo et al. 2014). Previously differentiated developmental roles can even be transferred from one paralog to another during evolution (Urea et al. 2016). Perhaps more significantly, long-preserved paralogs can be lost in lineage-specific manners, a common phenomenon observed across the tree of life, including in bacteria (Gmez-Valero et al. 2007), yeasts (Scannell et al. 2007), Paramecium (Aury et al. 2006; McGrath et al. 2014), plants (De Smet et al. 2013), fish (Amores et al. 2004), and mammals (Amores et al. 1998; Blomme et al. 2006). Although pervasive, the

importance of ongoing paralog diversification to the evolution of organismal traits and phenotypes remains under appreciated.

Duplicate gene differentiation has heavily impacted the evolution of regulatory and metabolic networks (Reece-Hoyes et al. 2013; Voordeckers et al. 2015). Paralogs have contributed to the expansion of regulatory networks (Teichmann and Babu 2004), the derivation of novel networks (Conant and Wolfe 2006; Wapinski et al. 2010; Prez et al. 2014; Pougach et al. 2014), the specialization of network regulation (Lin and Li 2011), and the robustness of networks to perturbation (Papp et al. 2004; Deutscher et al. 2006). The WGD has even been proposed to have facilitated the evolution of anaerobic glucose fermentation strategy called Crabtree-Warburg Effect in the lineage of yeasts that includes *Saccharomyces* (Conant and Wolfe 2007; Jiang et al. 2008). Gene regulation and metabolism are heavily intertwined biological processes, but there are few eukaryotic models that allow for an integrated study of the ongoing differentiation of paralogous genes with regulatory and metabolic diversification (Yamada and Bork 2009; Conant et al. 2014).

The *Saccharomyces cerevisiae GAL*actose sugar utilization network is one of the most extensively studied eukaryotic regulatory and metabolic networks, and its homologous networks are evolutionarily dynamic in yeasts. In *S. cerevisiae*, it includes the three enzymes of the Leloir pathway (Gal1, Gal7, and Gal10) that catabolize galactose, the galactose transporter Gal2, and three regulators. In the absence of galactose, the transcription factor Gal4 is inhibited by the co-repressor Gal80. When galactose is present, Gal80 is sequestered by the co-inducer Gal3, allowing Gal4 to activate the expression of the *GAL* network (Johnston 1987; Bhat and Murthy 2001; Egriboz et al. 2013). Numerous studies have shown that the *GAL* networks of various yeast lineages vary in gene content (Hittinger et al. 2004, 2010; Wolfe et al. 2015) and gene activity (Peng et al. 2015; Roop et al. 2016). Despite these findings, the impacts of variable network architectures on the evolution of gene regulation and metabolism are not well understood.

As a model for how duplicate gene divergence creates variable network architectures, we functionally characterized the *GAL* network of *Saccharomyces* uvarum (formerly known as *Saccharomyces bayanus var. uvarum*) and compared it to *S. cerevisiae*. Here we show that two *GAL* network paralog pairs in *S. uvarum* have diverged to different degrees and states than their *S.*

cerevisiae homologs. We further show that, unlike *S. cerevisiae*, *S. uvarum* deploys a second co-repressor that prevents over-induction of the network. *S. uvarum* mutants lacking both co-repressors revealed surprising constraints on the rapid utilization of galactose; specifically, they arrested their growth, and metabolomic investigations suggested that they experienced metabolic overload. We show that homologous constraints exist in a milder form in *S. cerevisiae*, and the degree of metabolic constraint is affected by how *GAL* network paralogs have diversified between the species. These results show how, after a hundred of millions of years of preservation, two pairs of interacting duplicate genes have continued to diverge functionally in ways that broadly impact metabolism, regulatory network structures, and the future evolutionary trajectories available.

2.3 Results

2.3.1 *GAL* gene content and sequence differences

S. uvarum has orthologs encoding all regulatory and structural genes that are present in *S. cerevisiae*, but it has duplicate copies of two additional genes. The first additional duplicate gene is *GAL80B*, which is a paralog of *GAL80*; this pair of paralogs was created by the whole genome duplication (WGD) event roughly 100 million years ago (Wolfe and Shields 1997; Marcet-Houben and Gabaldn 2015). *GAL80B* has been retained in the *S. uvarum-Saccharomyces eubayanus* clade, but it was lost in the *S. cerevisiae-Saccharomyces arboricola* clade (Hittinger et al. 2010; Scannell et al. 2011; Caudy et al. 2013; Hittinger 2013; Liti et al. 2013; Baker et al. 2015). The second one is *GAL2B*, which was created by a recent tandem duplication in *S. uvarum-S. eubayanus* clade. Both *S. cerevisiae* and *S. uvarum* also contain a pair of specialized paralogs created by the WGD, *GAL1* and *GAL3*. By comparing amino acid sequences against the *S. cerevisiae GAL* network, we found that most *GAL* genes are diverged to a similar extent (see Appendix A: Table A.1), except for *GAL4*, which is primarily conserved in its DNA-binding and other functionally characterized domains. None of the *S. uvarum GAL* homologs exhibited significantly elevated rates of protein sequence evolution (from previously calculated dN/dS ratios (Byrne and Wolfe 2005)), which might have otherwise suggested extensive neofunctionalization. Thus, we focused on whether and how the key regulatory genes partitioned functions differently between the two species.

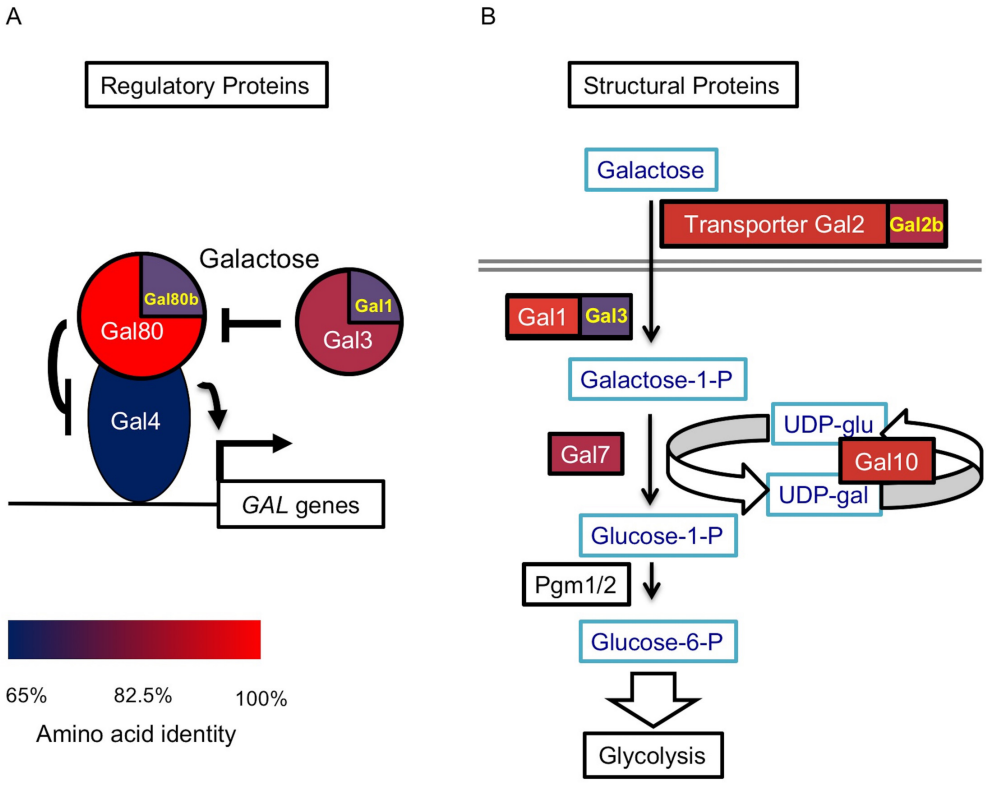


Figure 2.1: The *S. uvarum GAL* network. (A) The *GAL* regulatory network. (B) The *GAL* or Leloir metabolic pathway. The colors show the amino acid identity of each component compared to their *S. cerevisiae* homologs (full data in Appendix A: Table A.1). Proteins with two homologs in *S. uvarum* are split into two parts: Gal1/Gal3 and Gal80/Gal80b (also known as *Sbay_12.142* (Scannell et al. 2011) or *670.20* (Caudy et al. 2013)) are two pairs of paralogs from a WGD event, while Gal2/Gal2b (also known as *Sbay_10.165* (Scannell et al. 2011) or *672.62* (Caudy et al. 2013)) are paralogs from a recent tandem duplication event (Hittinger et al. 2004).

2.3.2 Less partitioned galactokinase and co-induction functions

In *S. cerevisiae*, the *GAL1* and *GAL3* paralogs are descended from an ancestral bi-functional protein that was both a co-inducer and a galactokinase (Rubio-Texeira 2005; Hittinger and Carroll 2007). They are almost completely subfunctionalized: *ScerGAL3* lost its galactokinase activity and

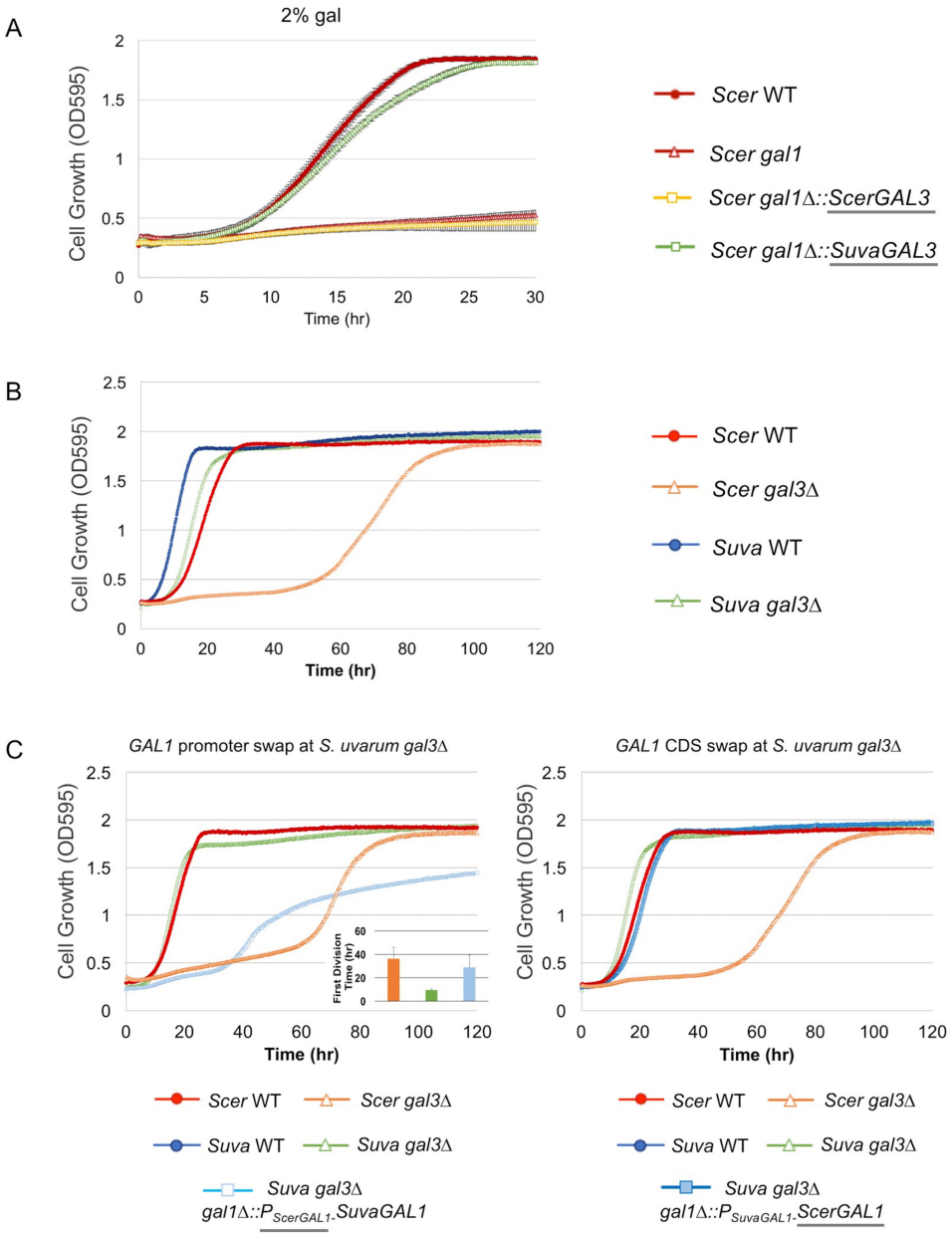


Figure 2.2: *SuvaGAL1* and *SuvaGAL3* are not as subfunctionalized as *ScerGAL1* and *ScerGAL3*.

(continued) *SuvaGAL1* and *SuvaGAL3* are not as subfunctionalized as *ScerGAL1* and *ScerGAL3*. (A) *S. uvarum GAL3* likely encodes a functional galactokinase. The error bars represent standard deviations of three biological replicates. A Wilcoxon rank sum test comparing the average times to first doubling between *S. cerevisiae gal1* and *S. cerevisiae gal1*Δ:: *SuvaGAL3* was significantly different (p = 5.2e-3, n = 6). Note that driving *ScerGAL3* from the *ScerGAL1* promoter was insufficient to support growth with galactose as the sole carbon source, but *SuvaGAL3* was sufficient. (B) Unlike *S. cerevisiae gal3*Δ, *S. uvarum gal3*Δ does not show Long-Term Adaption (LTA). Strains were cultured in SC + 2% galactose. Wilcoxon rank sum tests comparing the average times to first doubling between *S. uvarum gal3*Δ and *S. uvarum* wild-type strains were significantly different (p = 4.5e-5, n = 12). These experiments have been repeated independently at least twice with three biological replicates, but growth curves display only one representative replicate because LTA emergence is stochastic. (C) LTA was recapitulated in *S. uvarum gal3*Δ by replacing its *GAL1* promoter with the *S. cerevisiae GAL1* promoter (left panel) or, to a much lesser extent, by replacing the coding sequence (right panel). The insets show the times to the first doubling for the strains for their respective panels. The bar colors in the inset are the same as the growth curves. To highlight strain comparisons that test discrete hypotheses, three genotypes are repeated in Figure 2.2B and in both panels of Figure 2.2C: *S. uvarum gal3*Δ, *S. cerevisiae* wild-type, and *S. cerevisiae gal3*Δ. Strains were cultured in SC + 2% galactose. Wilcoxon rank sum tests comparing the average times to first doubling between strains were as follows: (1) p = 4.6e-4 and n = 9 for *S. uvarum gal3*Δ versus *S. uvarum gal3*Δ *gal1*Δ*::*$P_{SuvaGAL1}$*-ScerGAL1*, (2) p = 4.2e-5 and n = 12 for *S. uvarum gal3*Δ versus *S. uvarum gal3*Δ *gal1*Δ*::*$P_{ScerGAL1}$*-SuvaGAL1*, and (3) p = 0.21 and n = 12 for *S. uvarum gal3*Δ *gal1*Δ*::*$P_{ScerGAL1}$*-SuvaGAL1* versus *S. cerevisiae gal3*.

became a dedicated co-inducer, whereas *ScerGAL1* lost most of its co-inducer activity but maintains galactokinase activity (Platt and Reece 1998; Platt et al. 2000; Timson et al. 2002; Hittinger and Carroll 2007; Lavy et al. 2015). Unlike ScerGal3, SuvaGal3 retains a -Ser-Ala- dipeptide in its active site that is sufficient to weakly restore galactokinase activity when added back to ScerGal3 (Platt et al. 2000), so we hypothesized that *SuvaGAL3* encodes a functional galactokinase. To test this hypothesis, we precisely replaced the coding sequence of *ScerGAL1*, the gene encoding the

sole galactokinase in *S. cerevisiae* (Platt et al. 2000), with *SuvaGAL3* in *S. cerevisiae*. As expected, *SuvaGAL3* conferred robust growth in galactose when driven by the *ScerGAL1* promoter, suggesting that *SuvaGAL3* retains galactokinase activity (Figure 2.2A). Nonetheless, the *S. uvarum gal1* null mutant did not grow better in 2% galactose than it did without any carbon source, a phenotype similar to the *S. cerevisiae gal1* null mutant (Appendix A: Table A.1), indicating that the native *GAL3* promoter expression is insufficient to support robust metabolism.

To further examine the functional divergence between *SuvaGAL1* and *SuvaGAL3*, we knocked out *GAL3* in *S. uvarum*. Surprisingly, the *S. uvarum gal3* null mutant did not show the classic Long Term Adaptation (LTA) phenotype of the *S. cerevisiae gal3* null mutant (Tsuyumu and Adams 1973). Instead of a growth delay of multiple days, we observed a delay of only a few hours in *S. uvarum gal3*Δ relative to wild-type (Figure 2.2B). These results suggest that other genes in *S. uvarum* may be able to partially compensate for the deletion of *SuvaGAL3*, such as its paralog, *SuvaGAL1*. To determine whether *GAL1* differences between *S. uvarum* and *S. cerevisiae* might be responsible for the different *gal3* null phenotypes, we replaced the *SuvaGAL1* coding sequence or promoter sequence with their *ScerGAL1* counterparts in the background of *S. uvarum gal3*Δ. The *ScerGAL1* promoter swap in *S. uvarum gal3*Δ largely recapitulated LTA, while the *ScerGAL1* coding sequence swap extended the delay to a lesser extent (Figure 2.2C). Since the *GAL1-GAL10* promoter is a divergent promoter, genetic modifications (evolved or engineered) inevitably impact both genes, as well as perhaps a lncRNA previously described in *S. cerevisiae* (Cloutier et al. 2016). These results suggest that differences at the *GAL1* locus, especially within this promoter, are primarily responsible for the lack of LTA in the *S. uvarum gal3*Δ mutant. Overall, the data suggest that *SuvaGAL1* is functionally redundant with *SuvaGAL3* to a much greater extent than are *ScerGAL1* and *ScerGAL3*. Thus, it is likely that the homologs in the common ancestor of *S. uvarum* and *S. cerevisiae* were more functionally redundant than in modern *S. cerevisiae*, and considerable subfunctionalization between *ScerGAL1* and *ScerGAL3* happened after the divergence of *S. uvarum* and *S. cerevisiae*.

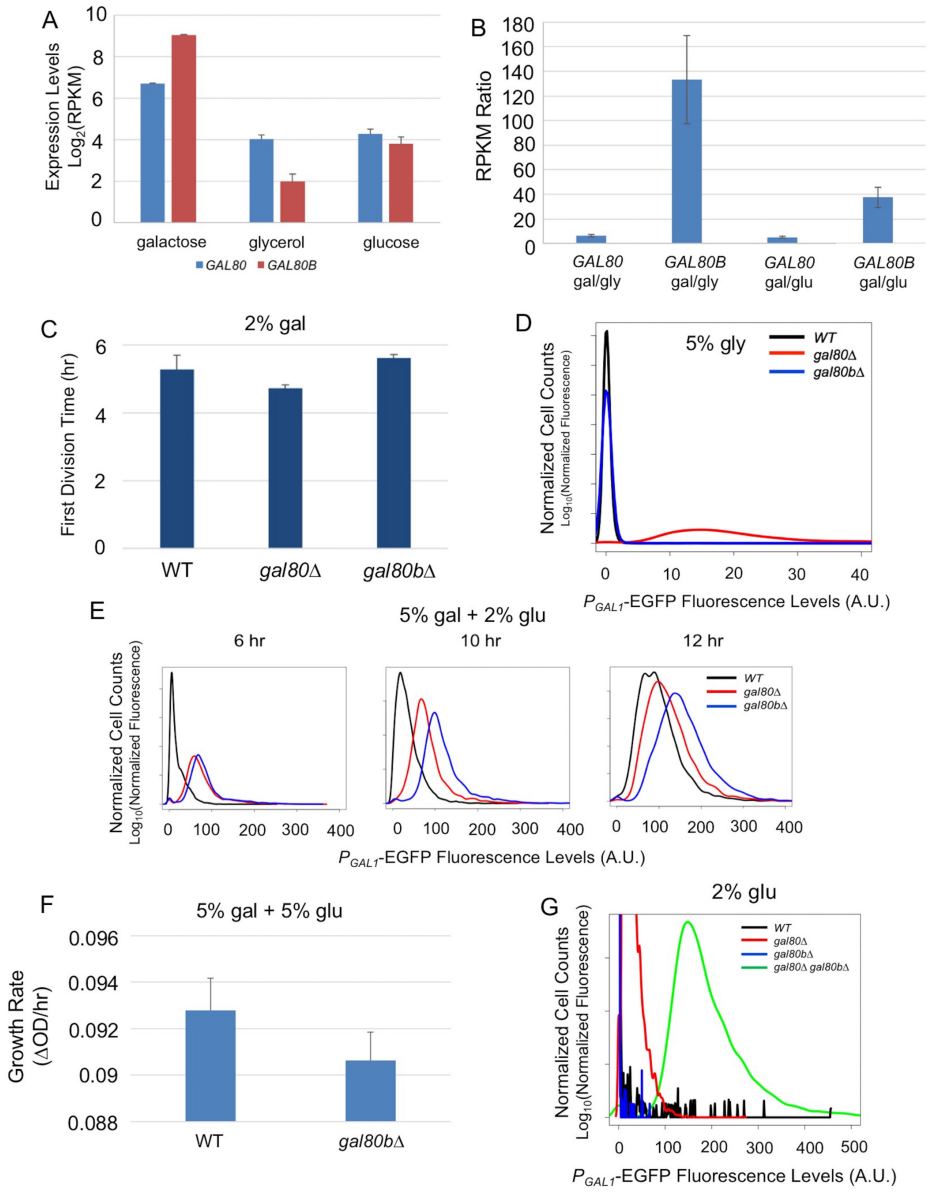


Figure 2.3: *SuvaGAL80* and *SuvaGAL80B* encode co-repressors with partially overlapping functions.

(continued) *SuvaGAL80* and *SuvaGAL80B* encode co-repressors with partially overlapping functions. (A) Expression divergence between *SuvaGAL80* and *SuvaGAL80B*. The bar graph on the left shows the mRNA levels (in log_2 of Reads Per Kilobase of transcript per Million mapped reads or RPKM) of *SuvaGAL80* and *SuvaGAL80B* in SC + 2% galactose, SC + 5% glycerol, and SC + 2% glucose. Error bars represent the standard deviations of three biological replicates. (B) Divergent galactose induction between *SuvaGAL80* and *SuvaGAL80B*. The bar graph shows the ratio of mRNA levels between galactose (gal) and glycerol (gly), or between galactose and glucose (glu) from the data in Panel A. (C) Removing *SuvaGAL80* conferred rapid initial growth in galactose. The bar graph shows the average time to first doubling of three biological replicates of each genotype in SC + 2% galactose from a representative experiment. *S. uvarum gal80*Δ grew significantly faster than wild-type (p = 1.8e-3, n = 14, Wilcoxon rank sum test), but *S. uvarum gal80b*Δ did not (p = 0.61, n = 14, Wilcoxon rank sum test). (D) Removing *SuvaGAL80* resulted in constitutive *GAL1* expression. The histogram shows the fluorescence levels of an *EGFP* reporter when driven by the *S. uvarum GAL1* promoter in SC + 5% glycerol as determined by flow cytometry. (E) Removing *SuvaGAL80B* led to the elevated *GAL1* expression in a mixture of glucose and galactose. Flow cytometry was conducted on strains cultured in SC + 5% galactose + 2% glucose. (F) Removing *SuvaGAL80B* caused a fitness defect in a mixture of glucose and galactose. The specific growth rate of *S. uvarum gal80b*Δ was significantly lower than wild-type (p = 2.7e-4, n = 18, Wilcoxon rank sum test). (G) *SuvaGAL80* and *SuvaGAL80B* were both able to partially compensate for the loss of the other in repressing conditions, but the double-knockout resulted in constitutive expression. The histogram reports flow cytometry data from strains cultured in SC + 2% glucose for 9 hr.

2.3.3 *S. uvarum* has two co-repressors with partially overlapping functions

Next, we examined the functional divergence of the other pair of paralogous regulatory genes, *SuvaGAL80* and *SuvaGAL80B*, which are homologous to the *ScerGAL80* gene that encodes the sole *GAL* gene co-repressor in *S. cerevisiae*. We first examined the expression of these two genes in the presence or absence of galactose (Figure 2.3A). RNA sequencing (RNA-Seq) showed that

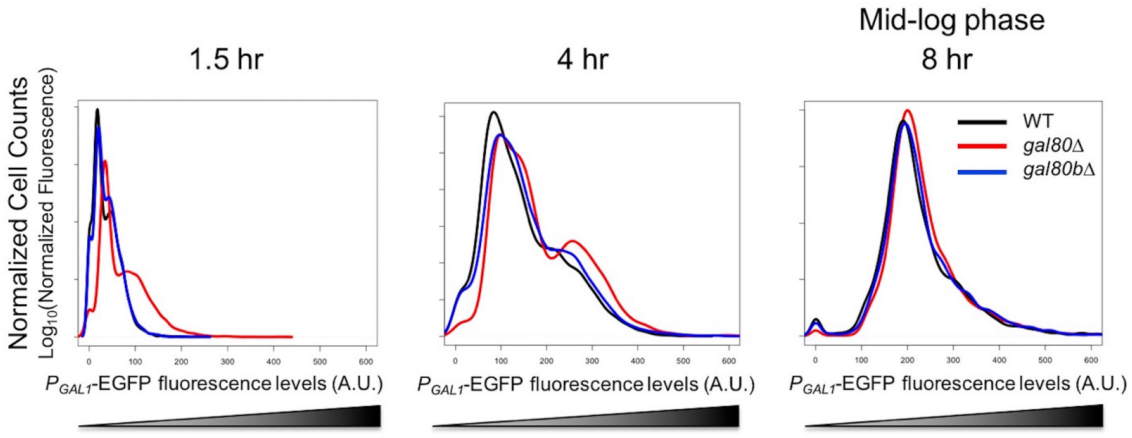


Figure 2.4: In SC + 2% galactose, *S. uvarum gal80*Δ and *gal80b*Δ had *GAL1* expression levels similar to the wild-type at mid-log phase. Flow cytometry histogram of P_{GAL1}-*EGFP* fluorescence.

SuvaGAL80 was expressed at a higher level than *SuvaGAL80B* in the absence of galactose (i.e. with glycerol or glucose as the sole carbon source). In contrast, in the presence of galactose, *SuvaGAL80B* was induced by 133-fold, much higher than the 6-fold induction observed for *SuvaGAL80* (Figure 2.3B). *S. uvarum gal80*Δ null mutants had a shorter lag time than wild-type in galactose, as seen in *S. cerevisiae gal80*Δ null mutants (Torchia et al. 1984; Segr et al. 2006; Hittinger et al. 2010), but *gal80b*Δ null mutants did not (Figure 2.3C). Deleting *SuvaGAL80* also resulted in elevated *GAL1* expression in the non-inducing condition (i.e. 5% glycerol), while deleting *SuvaGAL80B* had no detectable effect (Figure 2.3D). Therefore, we conclude that *SuvaGAL80* is the main gene responsible for repressing the *GAL* network in the absence of galactose.

Perhaps because of its dynamic expression, the deletion mutant phenotype of *S. uvarum gal80b*Δ proved condition dependent. Consistent with previous negative results (Caudy et al. 2013), no apparent phenotypic differences were observed for the *S. uvarum gal80b*Δ strain when it was grown in galactose, nor were its *GAL1* expression levels altered (Figure 2.3C and Figure 2.4). Nonetheless, in a mixture of galactose and glucose, we observed elevated *GAL1* expression in *S. uvarum*

*gal80b*Δ strains, beyond the levels observed in *S. uvarum gal80*Δ strains (Figure 2.3E). Additionally, *S. uvarum gal80b*Δ grew significantly slower than wild-type after transfer from galactose to a mixture of galactose and glucose (Figure 2.3F), suggesting that *SuvaGAL80B* plays a specific and biologically important repressive role in conditions where it is required to prevent network over-induction. We also observed strong negative epistasis when both co-repressors were removed: the co-repressor double knockout had substantially higher *GAL1* expression than either single knockout strain or the *S. uvarum* wild-type strain in the absence of galactose (Figure 2.3G). Thus, *SuvaGAL80* and *SuvaGAL80B* encode partially redundant *GAL* gene co-repressors that can each partially compensate for the loss of the other. We conclude that *SuvaGAL80B* may play a minor role in the absence of galactose, but it provides important modulation in induced conditions.

2.3.4 Strains lacking the co-repressors arrest their growth

Surprisingly, knocking out both *GAL80* and *GAL80B* in *S. uvarum* resulted in a strong Temporary Growth Arrest (TGA) phenotype in galactose (Figure 2.6A). This result stands in sharp contrast to the observation that *S. cerevisiae gal80* null mutant strains from multiple genetic backgrounds (the lab strains S288c, W303, and R21, as well as the vineyard strain RM11-1a examined here) grew faster in galactose, a phenotype shared with *Saccharomyces kudriavzevii gal80*Δ null mutants and attributed to constitutive *GAL* expression (Torchia et al. 1984; Segr et al. 2006; Hittinger et al. 2010). This growth arrest was not a genetic engineering artifact; reintroducing *SuvaGAL80* completely rescued the growth arrest, and knocking out these two genes with different markers produced the same mutant phenotype (Figure 2.5). More importantly, introducing *ScerGAL80* completely rescued the growth arrest (Figure 2.5), suggesting that the TGA phenotype was not due to novel molecular functions specific to *SuvaGAL80* or *SuvaGAL80B*. Instead, the dramatically varied phenotypes imply that these two species have different regulatory or metabolic wiring for galactose metabolism.

To test whether the TGA phenotype was associated with *S. uvarum*-specific *GAL* network members, we performed RNA-Seq on *S. uvarum gal80*Δ *gal80b*Δ in 2% glucose or 5% glycerol, conditions where the complete *GAL* network is expected to be constitutively expressed (Torchia et al.

1984; Segr et al. 2006; Hittinger et al. 2010). We identified genes as *GAL* network members if and only if they were: 1) significantly up-regulated in *S. uvarum gal80*Δ *gal80b*Δ compared to the wild-type at FDR = 0.05 (35 genes); 2) up-regulated by at least two-fold (19 genes); 3) up-regulated in both glucose and glycerol (nine genes); and 4) predicted to contain Gal4 consensus binding sites ($CGGN_{11}CCG$) upstream of their coding sequences. Using these stringent criteria, we found eight potential *GAL* network members in *S. uvarum*, seven of which were shared with *S. cerevisiae* based on previous chromatin immune precipitation and gene expression data (*GAL1*, *GAL2*, *GAL2B*, *GAL7*, *GAL10*, *MEL1*, and *GCY1*) (Torchia et al. 1984; Ren et al. 2000) (Figure 2.7A). *GAL3*, a well-established Gal4 target in *S. cerevisiae*, was considered differentially expressed using less stringent criteria, but orthologs of two other known targets were not (*MTH1* and *PCL10*). The sole novel *GAL* network member in *S. uvarum* was the *PGM1* gene, which was up-regulated 26-fold in 5% glycerol in *S. uvarum gal80*Δ *gal80b*Δ relative to wild-type. In *S. cerevisiae*, *PGM1* encodes the minor isoform of phosphoglucomutase, which, along with Pgm2, connects the Leloir pathway to glycolysis (Figure 2.1). Notwithstanding the *PGM1* gene, we conclude that the *S. uvarum* and *S. cerevisiae GAL* networks have similar compositions, and the handful of differences do not seem to readily explain the remarkably strong and unexpected TGA phenotype seen in *S. uvarum* strains lacking their co-repressors.

2.3.5 Overactive galactose catabolism precedes widespread metabolic and regulatory defects

In contrast to the constitutive expression of a fairly small network of direct Gal4 targets seen during growth in glucose and glycerol, *S. uvarum gal80*Δ *gal80b*Δ double mutants experienced global changes in gene expression that were specific to growth in galactose (Figure 2.7B, C). During the TGA phase, 1,006 genes were differentially expressed in *S. uvarum gal80*Δ *gal80b*Δ relative to wild-type (620 genes up-regulated and 386 genes down-regulated by at least two-fold with FDR = 0.05 (Figure 2.13-source data 1: https://doi.org/10.7554/eLife.19027.017)). After the mutant resumed growth in galactose, the vast majority (78%, 783 of 1,006 genes) of these genes returned to expression levels indistinguishable from wild-type, and Gene Ontology (GO)

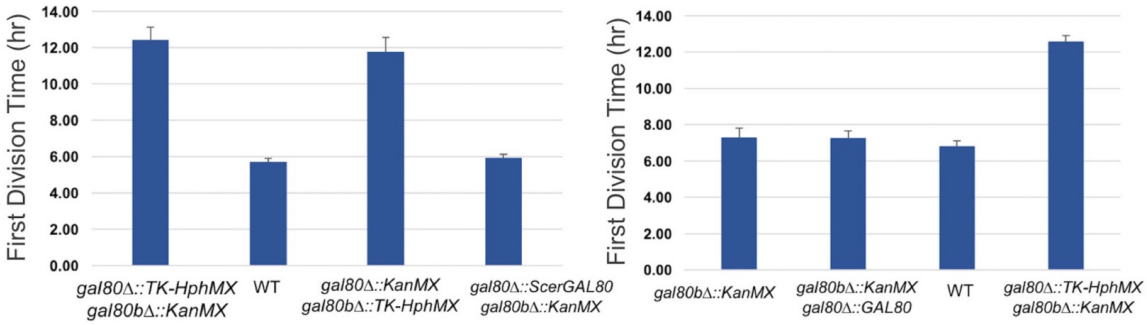


Figure 2.5: The TGA phenotype of *S. uvarum gal80*Δ *gal80b*Δ can be rescued by *S. cerevisiae GAL80* or by re-introducing *SuvaGAL80*. The bar graphs show the average times to first doubling time of three biological replicates. Strains were cultured in SC + 2% galactose.

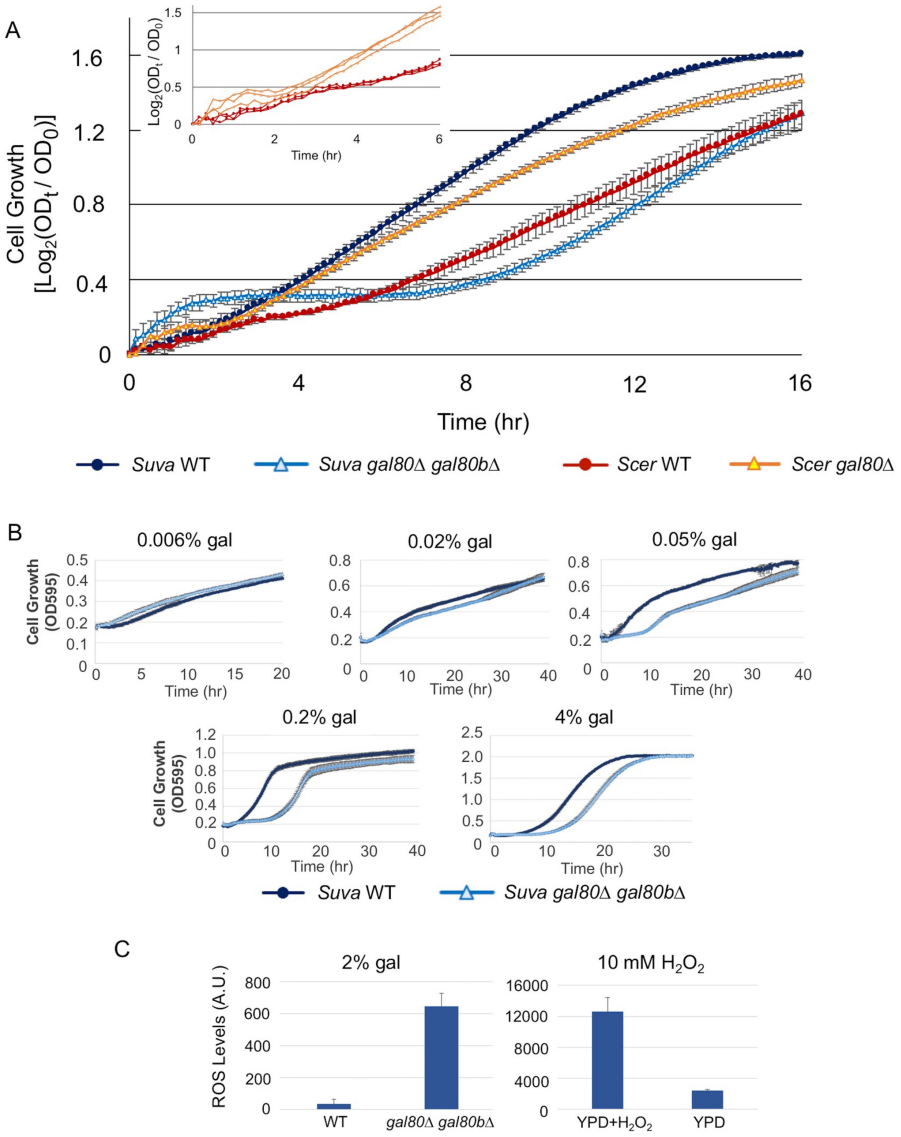


Figure 2.6: The galactose-dependent temporary growth arrest phenotype of *S.uvarum gal80Δ gal80bΔ*.

(continued) (A) The Temporary Growth Arrest (TGA) phenotype in SC + 2% galactose. The averages of the log_2 of the ratios between absorbances at each time point (OD_t) and initial absorbances (OD_0) for three biological replicates are shown. The error bars represent standard deviations. The inset shows the first six hours for three biological replicates each of *S. cerevisiae* wild-type and *gal80*Δ (in the background of *S. cerevisiae* RM11-1a, a vineyard strain). (B) The degree of the TGA phenotype was concentration dependent. A representative experiment with three biological replicates is shown; the experiment has been repeated three times. (C) Excessive reactive oxygen species (ROS) were accumulated in *S. uvarum gal80*Δ *gal80b*Δ during the TGA phase. ROS levels are reported as relative fluorescence and were measured 6.5 hr after inoculation into SC + 2% galactose (p = 8.6e-6, n = 11, Wilcoxon rank sum test). The bar graph on the right shows a positive control using *S. uvarum* wild-type in YPD and YPD + 10 mM H_2O_2.

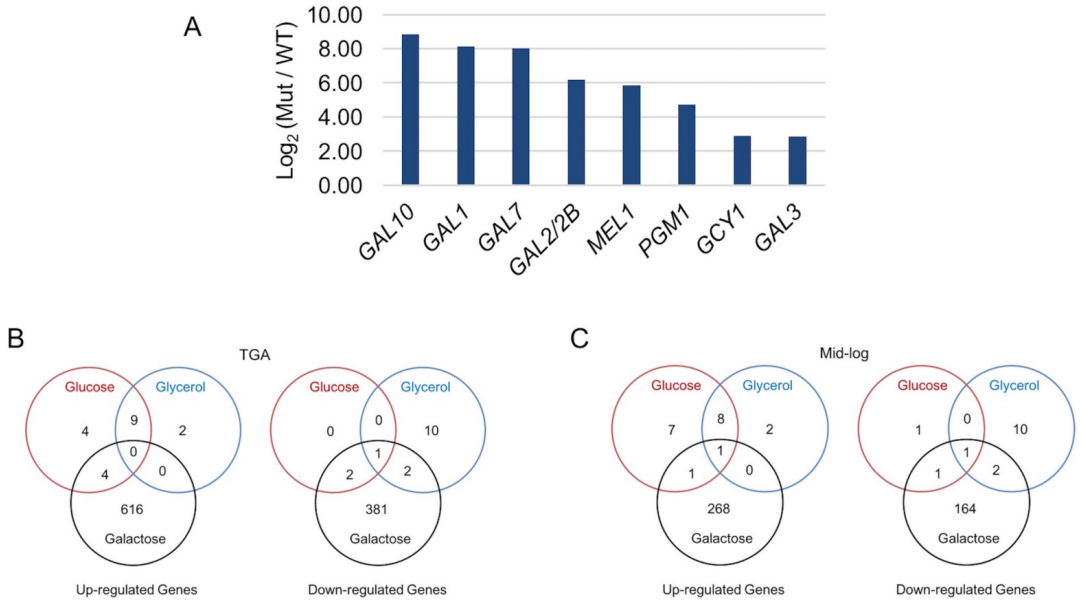


Figure 2.7: Galactose-specific global differential expression of *S. uvarum gal80*Δ *gal80b*Δ.

(continued) (A) *S. uvarum GAL* network comprises similar targets as the *S. cerevisiae GAL* network. The bar graph shows the $\log_2$ of the RPKM ratio between *S. uvarum gal80*Δ *gal80b*Δ and wild-type in SC + 5% glycerol. Note that *GAL80* and *GAL80B* are not in the list because they were knocked out in the double mutant, but both genes contain putative Gal4 binding sites in their promoters. *GAL3* was considered differentially expressed by edgeR (p-value = 2.95e-8 in the condition of glycerol at 11.2-fold and p = 2.83e-8 in glucose at 7.3-fold, both at FDR<1.4e-5), although it was not by EBSeq (posterior probability of being equally expressed was 0.13 in glycerol and 0.28 in glucose). The two other experimentally verified *S. cerevisiae* Gal4 target genes (*MTH1* and *PCL10*) were not considered up-regulated by either edgeR (p = 0.5 at 0.9-fold for *MTH1* and p = 0.8 at 1-fold for *PCL10* in glycerol, p = 0.1 at 1.5-fold for *MTH1* and p = 0.1 at 1.5-fold for *PCL10* in glucose, all at FDR = 1) or EBSeq (posterior probabilities of being equally expressed for *MTH1* and *PCL10* were 0.7 and 0.8 in glycerol, respectively, and 1.0 and 1.0 in glucose, respectively), despite having conserved putative Gal4 binding sites in their upstream sequences in *S. uvarum*. The three genes described as down-regulated in *S. uvarum gal80b*Δ strains by Caudy et al. (Caudy et al. 2013) were not differentially expressed in *S. uvarum gal80*Δ *gal80b*Δ strains in our growth conditions. (B-C) Venn diagrams of differential expression of *S. uvarum gal80*Δ *gal80b*Δ harvested at (B) the TGA phase, and (C) mid-log phase in SC + 2% glucose, + 5% glycerol, or + 2% galactose. Note that most gene expression changes were galactose-specific, suggesting that they were caused by metabolic defects, rather than direct regulation by Gal4. Note that, relative to wild-type, there were still hundreds of differentially expressed genes at the mid-log phase, but most (78%, 783 of 1,006) genes that were differentially expressed during the TGA phase had returned to normal expression.

term analysis showed that most of the biological processes affected during the TGA phase returned to normal (Supplementary File 1: https://doi.org/10.7554/eLife.19027.023). The TGA phase gene expression profile was not consistent with a global environmental stress response (e.g. nuclear ribosome biogenesis and rRNA processing were up-regulated) but instead suggested a complex and incoherent integration of the regulatory signals that govern metabolism

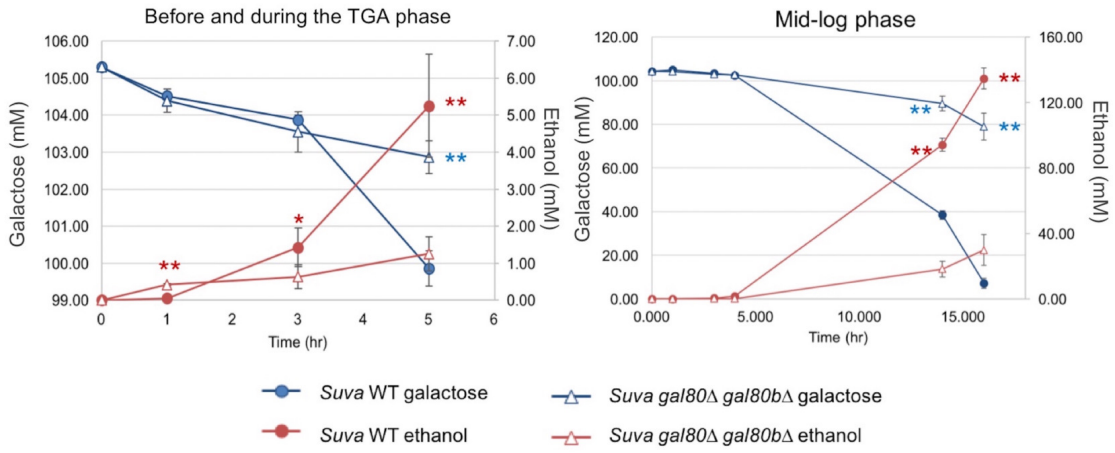


Figure 2.8: High performance liquid chromatography measurements of key metabolites in SC + 2% galactose in *S. uvarum gal80*Δ *gal80b*Δ and wild-type during the TGA phase and after the growth resumed. Statistically significant data points are marked by asterisks (*: $p<0.05$, **: $p<0.01$, one-tailed Students *t*-test). Red corresponds to ethanol, and blue corresponds to galactose; ethanol was produced by galactose catabolism, but ethanol production provided a more sensitive readout than galactose consumption in early-stage cultures. Note that *S. uvarum gal80*Δ *gal80b*Δ produced significantly more ethanol by the 1-hour time point (before the TGA phase), but the *S. uvarum* wild-type strain produced significantly more ethanol at the 3-hour and later TGA time points.

(Figure 2.13-source data 1: https://doi.org/10.7554/eLife.19027.017 and Supplementary File 1: https://doi.org/10.7554/eLife.19027.023).

Several lines of evidence suggested that this mis-regulation might be caused by overly rapid galactose catabolism immediately prior to the TGA phase. First, the optical density of the co-repressor double mutant initially increased faster than the wild-type in galactose and only plateaued after about 1.5 hours (Figure 2.6A). Second, during this early growth in galactose, the co-repressor double mutant produced more ethanol than the wild-type (Figure 2.8). Third, *GAL1* was also

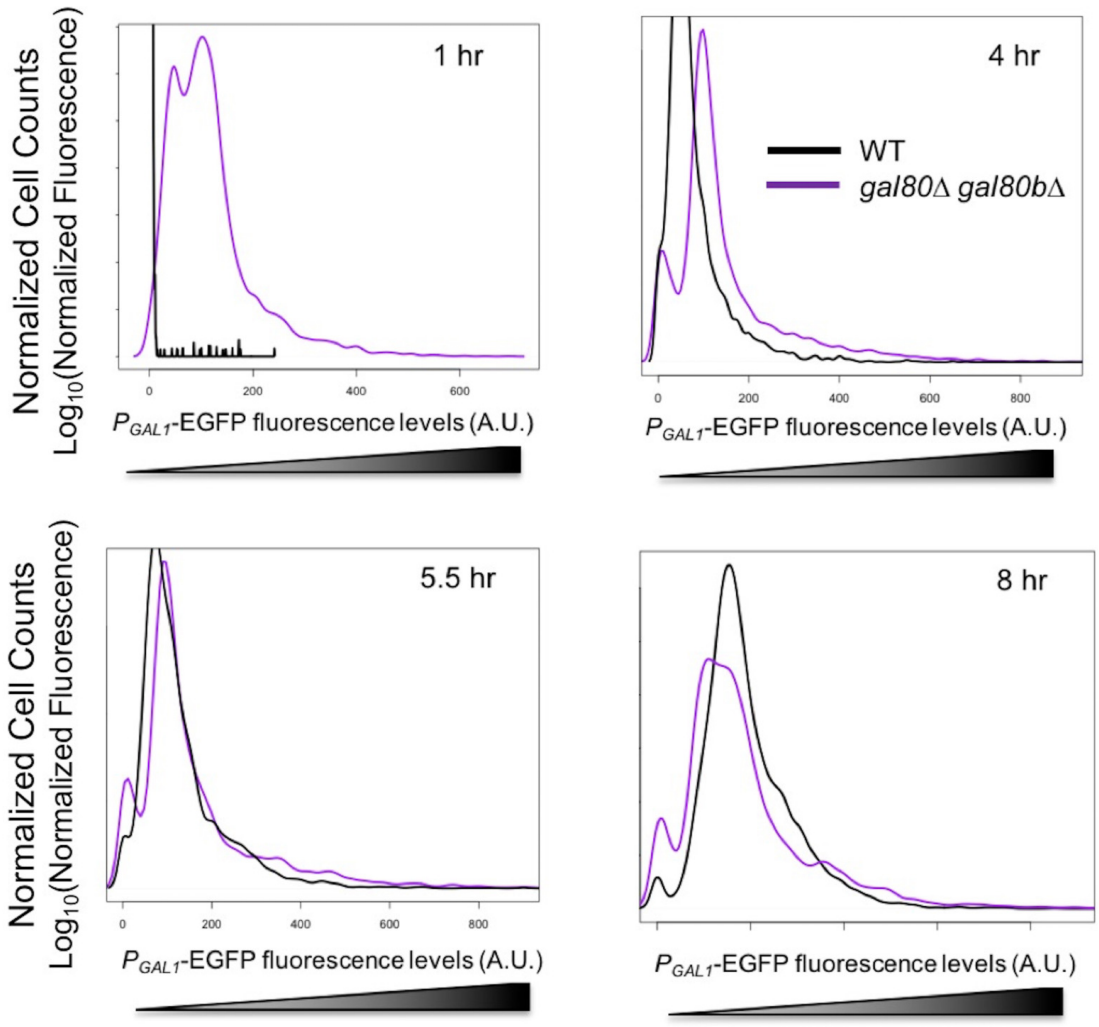


Figure 2.9: *GAL1* expression was higher at the early stages of growth in SC + 2% galactose in the *S. uvarum gal80*Δ *gal80b*Δ background but gradually decreased. Fluorescence levels were obtained by flow cytometry, normalized to forward scatter, and plotted as histograms. 4-hour was during the TGA phase, whereas 8-hour was approaching the end of the TGA phase.

strongly over-expressed in the mutant early during growth in galactose, but *GAL1* expression gradually converged with the wild-type strain as the cells transitioned into the TGA phase (Figure 2.9). Finally, the severity of the TGA phenotype depended strongly on galactose concentration (Fig-

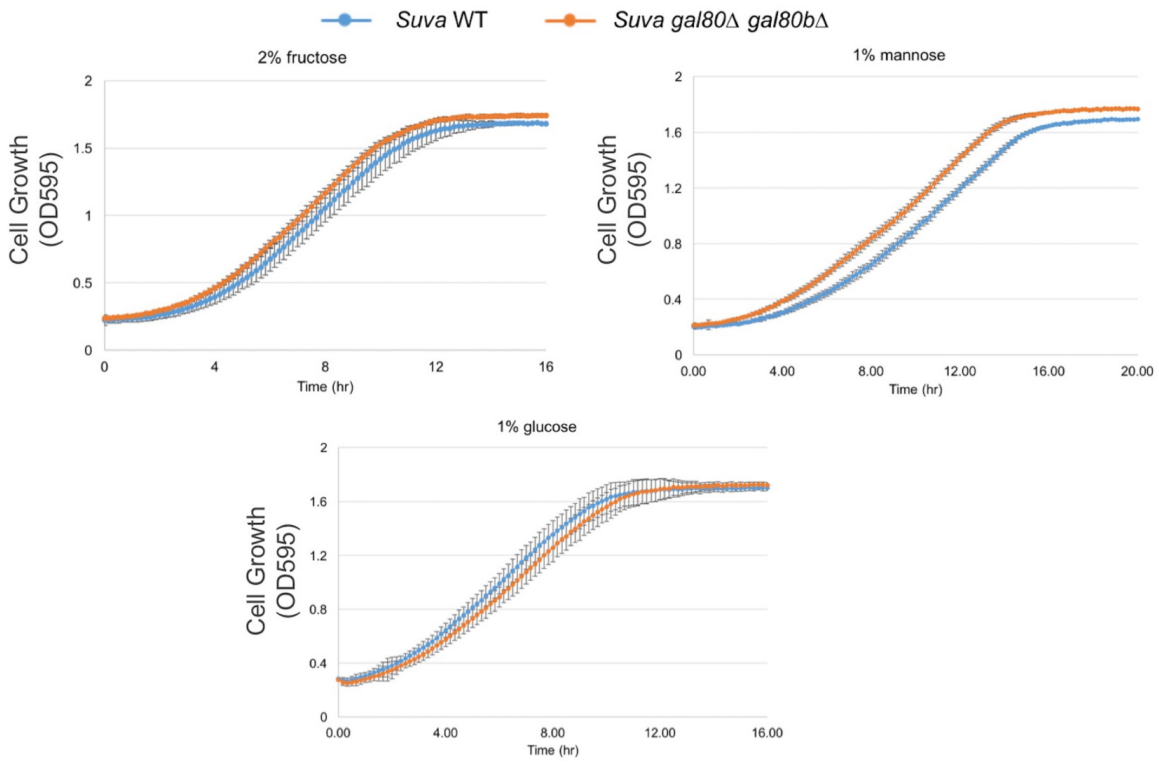


Figure 2.10: Fructose, mannose, or glucose alone did not lead to a TGA phenotype or other growth defects. All experiments were performed in SC media with the carbon sources indicated.

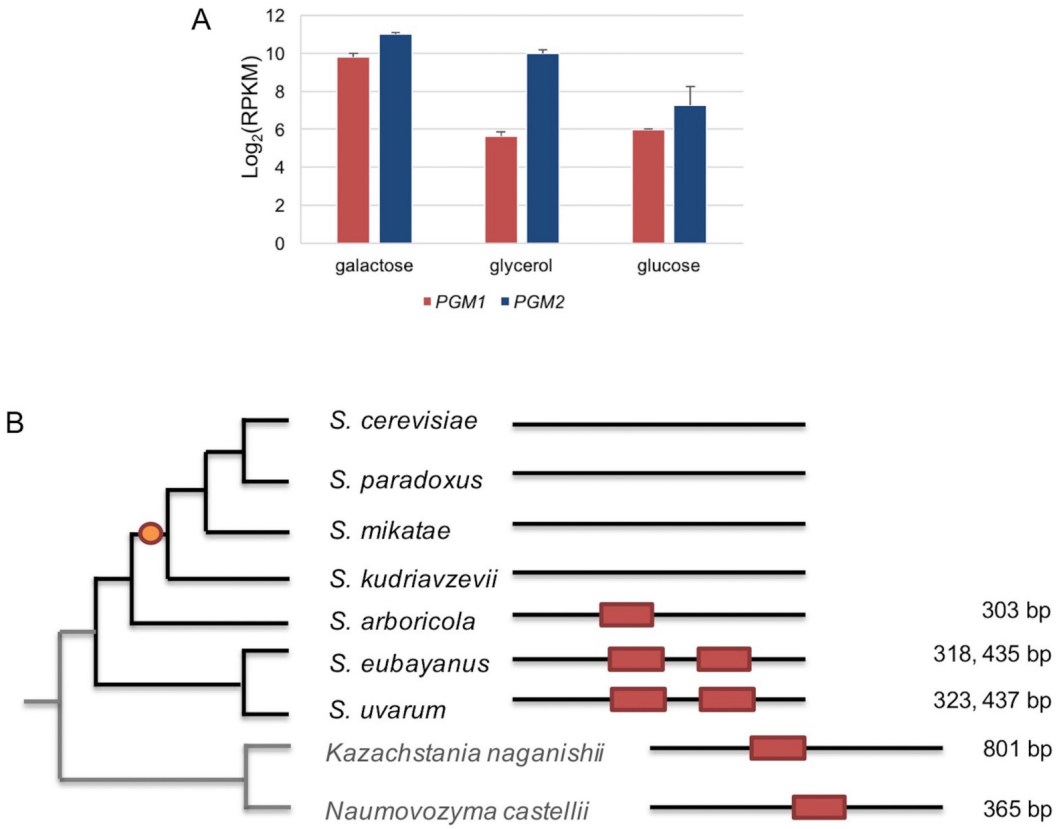


Figure 2.11: The regulation of *PGM1* by galactose was inferred as the ancestral state.

(continued) The regulation of *PGM1* by galactose was inferred as the ancestral state. A) mRNA levels of *S. uvarum PGM1* and *PGM2* during mid-log phase in SC + 2% galactose, SC + 5% glycerol, and SC + 2% glucose. Note that *PGM2*, which encodes the major isoform of phosphoglucomutase, has long been known to be transcriptionally induced by 3-4-fold in galactose, but it lacks a Gal4 binding site and does not appear to be a direct target in *S. cerevisiae* (Oh and Hopper 1990; Rubio-Texeira 2005). These features are broadly shared with *S. uvarum PGM2*, which is transcriptionally induced 2-fold by galactose relative to glycerol but is not transcriptionally upregulated in the *gal80*Δ *gal80b*Δ mutant; nor does it have a consensus Gal4 site. B) Conservation of putative Gal4 binding sites upstream of *PGM1* in *S. uvarum*, *S. eubayanus*, *S. arboricola*, and two outgroup species. The orange dot indicates the inferred loss of direct regulation of *PGM1* by Gal4 based on the presence or absence of putative Gal4 binding sites ($CGGN_{11}CCG$). The distances upstream from the start codon are shown at the right. The putative sites are shown as red boxes at the relative position of the upstream sequences of *PGM1*. Note that, in *Kazachstania nagansihii*, the upstream intergenic region of *PGM1* ortholog is 1958 bp, an unusually long intergenic region for yeasts, and contains a divergent promoter that also drives expression of the *PMU1* ortholog.

ure 2.6B), and growth defects were not seen in other carbon sources (Figure 2.10).

To further characterize how overly rapid galactose catabolism might lead to the TGA phenotype, we performed metabolomic analyses using mass spectrometry on co-repressor double mutant and wild-type strains cultured in 2% galactose. Prior to the TGA phase, the co-repressor double mutant accumulated galactose-1-phosphate, a known toxic intermediate of galactose metabolism (Petry and Reichardt 1998; de Jongh et al. 2008), but this two-fold accumulation (relative to wild-type) seemed unlikely to be sufficient to explain the TGA phenotype. The level of galactose-1-phosphate in *S. uvarum gal80*Δ *gal80b*Δ returned to normal during the TGA phase (Figure 2.13 and Appendix A: Table A.3-A.4) and was not nearly as strong as in *S. cerevisiae gal7*Δ or *gal10*Δ controls (seven- to 11-fold relative to *S. cerevisiae* wild-type) (Figure 2.12). Moreover, we did not

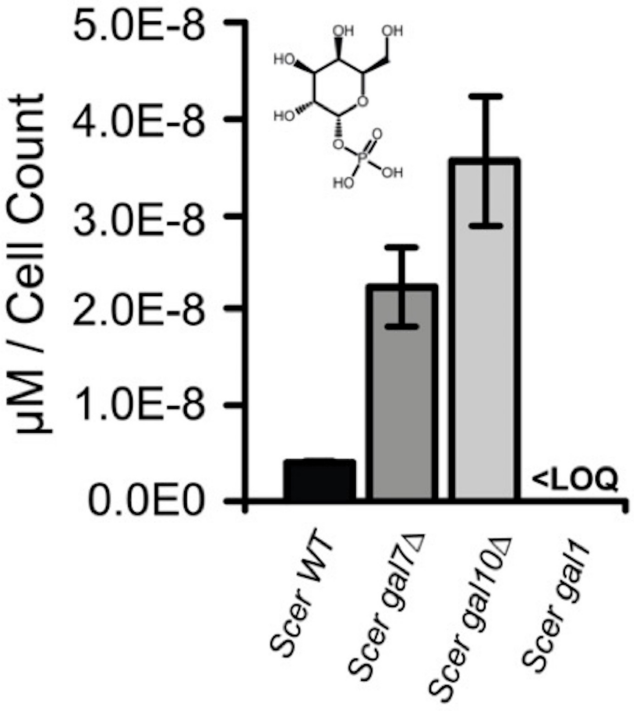


Figure 2.12: Galactose-1-phosphate accumulation of *S. cerevisiae gal7*Δ and *gal10*Δ. Galactose-1-phosphate levels were quantified by mass spectrometry. Samples were harvested after 4.5 hours of growth in 2% galactose. " LOQ" stands for "Limit of Quantification".

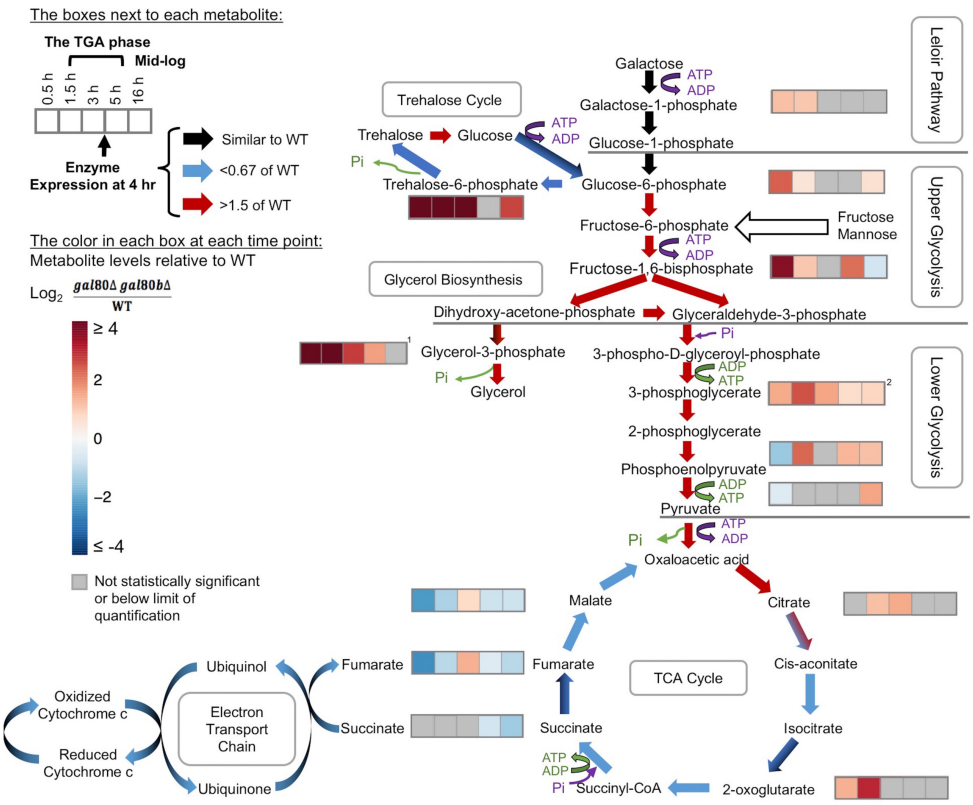


Figure 2.13: Overly rapid galactose catabolism leads to metabolic overload and bottlenecks.

(continued) Overly rapid galactose catabolism leads to metabolic overload and bottlenecks. The graph shows the metabolite levels and transcript expression for the Leloir pathway, glycolysis, trehalose cycle, glycerol biosynthesis, TCA cycle, and electron transport chain. Purple steps cost ATP or inorganic phosphate (Pi), while green steps generate ATP or Pi. Strains were cultured in SC + 2% galactose. The arrows are color-coded to represent the RNA-Seq gene expression differences of *S. uvarum gal80*Δ *gal80b*Δ relative to wild-type at 4 hr (red, up in the mutant; blue, down in the mutant; black, similar to wild-type; mixed colors (e.g. black and blue) indicate that the expression of genes involved in this step differs). The boxes next to each metabolite represent the $\log_2$ of the metabolite concentration differences relative to wild-type over time (0.5, 1.5, 3, 5, and 16 hr, respectively). The statistical significance for metabolite levels was assessed using Students *t*-tests (n = 3, $p<0.05$ with gray reported as not significant). The 1.5 hr to 5 hr time points correspond to the TGA phase, whereas the 16 hr time point corresponds to mid-log phase after recovery from the TGA phase. 1, the sum of the metabolite concentrations of glycerol-3-phosphate and glycerol-2-phosphate, the latter of which is not known to be a major metabolite in *Saccharomyces*; 2, the sum of the metabolite levels of 3-phosphoglycerate and 2-phosphoglycerate.

observe gene expression signatures consistent with the previously described responses to galactose-1-phosphate toxicity (e.g., environmental stress response, unfolded protein response) (Slepak et al. 2005; De-Souza et al. 2014).

Instead, both transcriptomic and metabolomic analyses revealed broad metabolic defects as bottlenecks developed downstream of the Leloir pathway. During the growth arrest, we observed increased expression of genes that encode glycolytic enzymes (Figure 2.13 and Figure 2.13-source data 1: https://doi.org/10.7554/eLife.19027.017). Key metabolic intermediates also accumulated in *S. uvarum gal80*Δ *gal80b*Δ strains before and during growth arrest, especially in upper glycolysis and interacting pathways (Figure 2.13 and Appendix A: Table A.3-A.4). In particular, fructose-1,6-biphosphate accumulated significantly prior to the TGA phase (12.6-fold of wild-type levels) (Figure 2.13 and Appendix A: Table A.3-A.4), a bottleneck that frequently occurs when upper glycolysis outpaces lower glycolysis (van Heerden et al. 2014). Under these conditions,

inorganic phosphate becomes a limiting factor for growth as the "investment" steps in upper glycolysis deplete the cells of ATP and phosphate to form sugar phosphates (Teusink et al. 1998; van Heerden et al. 2014). Indeed, *S. uvarum gal80*Δ gal80bΔstrains had one-fifth of the ATP as wild-type prior to the TGA phase (Appendix A: Table A.3-A.4) and had significantly up-regulated (25-fold) expression of *PHO84*, which encodes a high-affinity phosphate transporter (Figure 2.13-source data 1: https://doi.org/10.7554/eLife.19027.017).

S. cerevisiae combats metabolic overload in upper glycolysis by using two main pathways to restore phosphate pools. The trehalose cycle temporarily reroutes upper glycolysis to store sugars as trehalose (van Heerden et al. 2014), while glycerol biosynthesis offers an early exit from glycolysis (Luyten et al. 1995). Disrupting the *S. cerevisiae* trehalose cycle leads to the accumulation of fructose-1,6-biphosphate, decreased ATP levels, and ultimately growth arrest due to a metabolically unbalanced state (van Heerden et al. 2014; Gibney et al. 2015), metabolic changes similar to the *S. uvarum* TGA phenotype. Strikingly, both pathways experienced dramatic bottlenecks in *S. uvarum gal80*Δ *gal80b*Δ strains before and during the TGA phase. Specifically, *S. uvarum gal80*Δ *gal80b*Δ cells accumulated 79- to 231-fold more trehalose-6-phosphate before and during the TGA phase, while they accumulated 225-fold more glycerol-3-phosphate before the TGA phase, the latter of which lessened to some extent during the TGA phase (3- to 16-fold) (Figure 2.13 and Appendix A: Table A.3-A.4). These data are consistent with the hypothesis that the trehalose cycle and the glycerol biosynthesis pathway are unable to handle the metabolic overload when galactose is catabolized too rapidly in *S. uvarum* strains lacking the *GAL* network repression system.

The metabolic effects of the TGA phenotype also reverberated downstream, leading to the transcriptional down-regulation of the lower part of the TCA cycle and the electron transport chain (Figure 2.13). Reduced respiratory activity has been shown to increase the formation of reactive oxygen species (ROS) (Barros et al. 2004), and the co-repressor double mutant had strong signatures of mitochondrial dysfunction. GO terms related to mitochondrial structural components, mitochondrial translation, and respiration were among the most strongly down-regulated (Supplementary File 1: https://doi.org/10.7554/eLife.19027.023). Indeed, we observed significantly

higher accumulation of ROS in *S. uvarum gal80*Δ *gal80b*Δ during the TGA phase by using the general ROS indicator dichlorodihydrofluorescein diacetate (H_2DCF-DA) (Figure 2.6C). We conclude that disconnecting *S. uvarum* galactose metabolism from the negative feedback loops normally provided by the co-repressors Gal80 and Gal80b likely allows galactose to enter the Leloir pathway and glycolysis too rapidly, leading to metabolic defects far beyond the mild accumulation of galactose-1-phosphate and deep into central metabolism.

2.3.6 Specific sugars can exacerbate metabolic overload

To determine whether the TGA phenotype reflected a more general metabolic constraint imposed by the interplay between glycolysis and interacting metabolic pathways, we grew *S. uvarum gal80*Δ *gal80b*Δ in mixtures of galactose with fructose, mannose, or glucose. Fructose, mannose, and glucose are all primarily catabolized through glycolysis, but only glucose generates glycolytic intermediates that are upstream of the trehalose cycle (Figure 2.13). Thus, fructose and mannose are expected to contribute directly to metabolic overload with minimal offsetting effects from the trehalose cycle. If the interaction between glycolytic load and the trehalose cycle were important to the TGA phenotype, growing the double mutant in mixtures of galactose with fructose or mannose would exacerbate the growth arrest. In contrast, if the TGA phenotype were caused by galactose-specific metabolism, the addition of these more preferred sugars would have no effect, or perhaps mitigate the TGA phenotype. Consistent with the TGA phenotype being caused by a general overloading of upper glycolysis, both fructose and mannose strongly exacerbated the TGA phenotype in *S. uvarum gal80*Δ *gal80b*Δ, while glucose partially rescued the TGA phenotype (Figure 2.14A). Importantly, mixing fructose or mannose with galactose had much stronger defects than the identical amounts of galactose alone (Figure 2.6B and Figure 2.14A). Co-culturing wild-type *S. uvarum* strains in galactose with these sugars was not inherently toxic (Figure 2.14A), so the presence of the co-repressors allows cells to cope with this challenge. Growing *S. uvarum gal80*Δ *gal80b*Δ in fructose, mannose, or glucose alone also did not cause growth defects (Figure 2.10). Moreover, deleting *GAL1* completely rescued the TGA phenotype in the co-repressor double mutant (Figure 2.14B), while mixtures of mannose and galactose dramatically increased

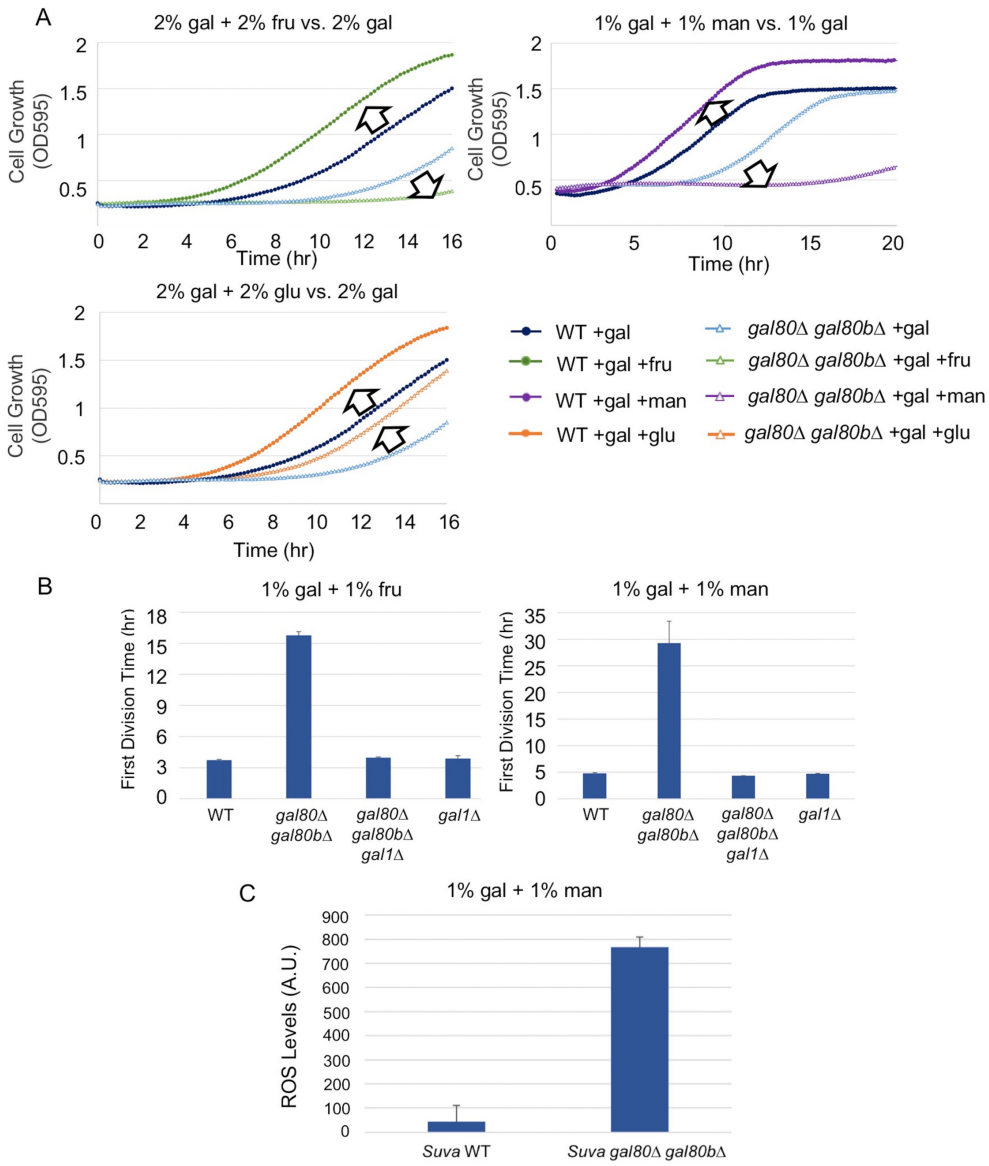


Figure 2.14: The addition of sugars downstream of the trehalose cycle exacerbated metabolic overload.

(continued) A) Fructose and mannose exacerbated the TGA phenotype in the *S. uvarum gal80*Δ *gal80b*Δ background, whereas glucose partially rescued the TGA phenotype. B) The *S. uvarum* TGA phenotype in galactose and fructose or mannose can be rescued by the deletion of *GAL1*. The average times to first doubling are shown for three biological replicates. The error bars represent standard deviations. *S. uvarum gal80*Δ *gal80b*Δ *gal1*Δ was significantly different than *S. uvarum gal80*Δ *gal80b*Δ in both SC + 1% galactose + 1% fructose (p = 4.8e-3, n = 6, Wilcoxon rank sum test) and SC + 1% galactose + 1% mannose (p = 2.9e-3, n = 6, Wilcoxon rank sum test). *S. uvarum gal80*Δ *gal80b*Δ *gal1*Δ was not significantly different from *S. uvarum gal1*Δ in SC + 1% galactose + 1% fructose (p = 0.43, n = 6, Wilcoxon rank sum test) but was marginally different from *S. uvarum gal1*Δ in SC + 1% galactose + 1% mannose (p = 0.03, n = 6, Wilcoxon rank sum test). C) Elevated accumulation of ROS in *S. uvarum gal80*Δ *gal80b*Δ in SC + 1% galactose + 1% mannose. *S. uvarum gal80*Δ *gal80b*Δ had significantly higher ROS levels than the wild-type (p = 8.6e-6, n = 11, Wilcoxon rank sum test). ROS levels are reported as relative fluorescence levels.

the levels of ROS in *S. uvarum gal80*Δ *gal80b*Δ (Figure 2.14C), implying that the phenotypic enhancement caused by this sugar mixture acts through the same mechanism observed in galactose alone. Collectively, these results suggest that overly rapid catabolism of sugars can lead to general metabolic and growth defects when the appropriate futile metabolic cycles and negative feedback regulatory loops are not able to slow down catabolism.

2.3.7 The less active *S. cerevisiae GAL* network is less susceptible to metabolic overload when derepressed

We next considered whether the differences between the *GAL* networks of *S. cerevisiae* and *S. uvarum* might explain why a similar phenotype had not been reported for *S. cerevisiae* co-repressor mutants. Recent work has convincingly shown that the *S. uvarum GAL* network is more transcriptionally active than the *S. cerevisiae GAL* network, especially in non-inducing and mixed sugar conditions (Caudy et al. 2013; Roop et al. 2016). Thus, we wondered whether *S. cerevisiae* and *S. uvarum* galactose catabolism might be under qualitatively similar constraints, even as the more poised and active state of the *S. uvarum GAL* network might render it more vulnerable to metabolic

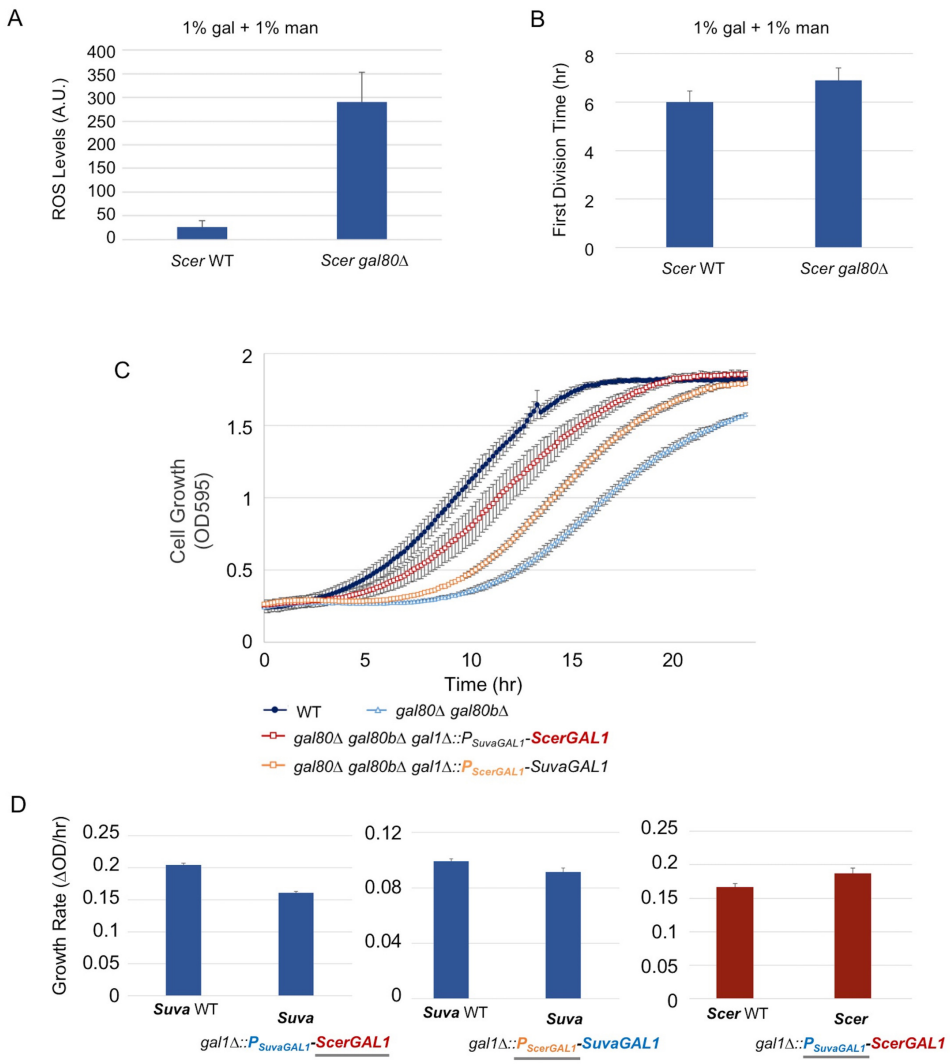


Figure 2.15: The less active *S. cerevisiae GAL1* gene is partially responsible for a subtle Temporary Growth Arrest.

(continued) A) Elevated accumulation of ROS in *S. cerevisiae gal80*Δ in SC + 1% galactose + 1% mannose. *S. cerevisiae gal80*Δ had significantly higher ROS than wild-type (p = 2.3e-6, n = 12, Wilcoxon rank sum test). ROS levels are reported as relative fluorescence levels. B) *S. cerevisiae gal80*Δ grew more slowly when galactose was mixed with mannose. The average of three biological replicates from a representative experiment is shown, and the error bars represent standard deviations (p = 0.028, n = 6, Wilcoxon rank sum test). C) Both the *ScerGAL1* coding sequence and promoter are able to partially rescue the TGA phenotype. The error bars show the standard deviation of three biological replicates. D) Both the *ScerGAL1* coding sequence and promoter reduced the growth rate of an otherwise wild-type strain of *S. uvarum* in SC + 2% galactose, while the reciprocal swap of the *GAL1* promoter in *S. cerevisiae* increased its growth rate. Wilcoxon rank sum tests comparing the specific growth rates of each subpanel were all significant: 1) p = 2.3e-6 and n = 12 for *S. uvarum gal1*Δ::$P_{SuvaGAL1}$-*ScerGAL1* versus *S. uvarum* wild-type, 2) p = 2.5e-4 and n = 9 for *S. uvarum gal1*Δ::$P_{ScerGAL1}$-*SuvaGAL1* versus *S. uvarum* wild-type, and 3) p = 8.8e-3 and n = 9 for *S. cerevisiae gal1*Δ::$P_{SuvaGAL1}$-*ScerGAL1* versus *S. cerevisiae* wild-type.

overload. First, we examined *S. cerevisiae gal80* null mutants more closely and found a similar but less-pronounced early rapid increase in optical density, followed by a brief but reproducible TGA phenotype (Figure 2.6A, inset). This observation was missed by earlier studies, which were focused on later time points, because *S. cerevisiae gal80* null mutants eventually grow much faster on galactose (Torchia et al. 1984; Segr et al. 2006; Hittinger et al. 2010). To test whether the weak TGA phenotype seen in *S. cerevisiae* was due to mechanistically similar metabolic constraints, we sought to exacerbate the phenotype of a *S. cerevisiae gal80*Δ strain in a mixture of mannose and galactose. Indeed, the co-repressor mutant produced significantly more ROS than wild-type under these conditions (Figure 2.15A) and grew slightly more slowly (Figure 2.15B).

Given the interspecific functional differences described above for *GAL1* (Figure 2.2C) and its role as the gatekeeper of the Leloir pathway, we hypothesized that the varied strengths of the TGA phenotype might be due to genetic differences in the *GAL1* locus. Thus, we precisely replaced the *S. uvarum GAL1* promoter or the *GAL1* coding sequence with their *S. cerevisiae* counterparts

in *S. uvarum gal80*Δ *gal80b*Δ. The *S. cerevisiae GAL1* promoter rescued the TGA phenotype to some extent, but the *GAL1* coding sequence swap was able to rescue the TGA phenotype to an even greater extent (Figure 2.15C). To confirm that *ScerGAL1* was less active than *SuvaGAL1* and not less toxic for other reasons, we examined the same precise allele replacements in an otherwise wild-type *S. uvarum* strain (i.e. containing functional copies of both co-repressors), as well as a precise reciprocal swap in *S. cerevisiae* replacing the *ScerGAL1* promoter with the *SuvaGAL1* promoter. Swapping the *ScerGAL1* promoter and coding sequence into *S. uvarum* both led to lower growth rates in galactose, while swapping the *SuvaGAL1* promoter into *S. cerevisiae* led to faster growth (Figure 2.15D). We conclude that the *S. uvarum GAL1* promoter and coding sequences both encode higher activity than their *S. cerevisiae* counterparts. Thus, differences in their *GAL* network activities at least partly explain the relative strengths of their TGA phenotypes and the constraints placed on their galactose metabolisms.

2.4 Discussion

2.4.1 Biodiversity offers a panoramic window to molecular biology

The deep conservation of metabolism and many molecular processes contrasts sharply with the rapid turnover in the regulatory networks that sculpt organismal and phenotypic diversity. Here we have shown how numerous genetic differences between the *S. cerevisiae* and *S. uvarum GAL* networks, especially in the functions of paralogous regulatory genes, contribute to a more poised and active state in *S. uvarum* that is coupled to more robust repression system. When genes encoding the co-repressors were deleted, *S. uvarum* displayed a strong and unexpected growth arrest in galactose, likely due to metabolic overload. Even though *S. cerevisiae* produced qualitatively similar results, decades of previous research on this iconic metabolic and regulatory network overlooked their relatively mild presentation. Just as exaggerated manifestations facilitated the discoveries of transposons in maize, RNAi in *Caenorhabditis elegans*, and telomeres in *Tretrahymena* (Blackburn et al. 2006), the striking phenotype observed in the non-traditional model organism *S. uvarum* allowed us to more fully characterize the defect caused by overly rapid galactose catabolism, while demonstrating metabolic constraints conserved across sugars and organisms.

2.4.2 The non-equivalence of sugars in contributing to metabolic overload

In contrast to glucose, fructose and mannose each had strikingly deleterious effects on cells that were already consuming galactose too rapidly. In *Saccharomyces*, these differences can be explained both by their effects on signaling pathways and by their entry points into glycolysis. Several glucose signaling pathways directly repress *GAL* gene transcription (Johnston et al. 1994) and increase the degradation rate of Gal2 protein (Horak and Wolf 2001), both of which would serve to reduce glycolytic load. In *S. cerevisiae*, fructose and mannose do not trigger glucose repression as strongly as glucose (Dynesen et al. 1998; Meijer et al. 1998). Perhaps as importantly, fructose and mannose bypass the trehalose cycle, a futile cycle recently shown to detour more than a quarter of early-stage glycolytic flux to prevent an unbalanced metabolic state and growth arrest (van Heerden et al. 2014). The challenges of the catabolism of sugars other than glucose may be widespread. For example, in humans, bypassing glucose-responsive regulatory mechanisms with fructose has been associated with diabetes (Lł et al. 2009; Kolderup and Svihus 2015) and cancer (Port et al. 2012; Jiang et al. 2016).

2.4.3 Network architectures reflect metabolic constraints

The intrinsic constraints imposed by galactose metabolism may have led to the evolution of regulatory mechanisms that protect against the risks of metabolic overload. Many of the differences between the *S. uvarum* and *S. cerevisiae GAL* networks can be explained as offering alternative protective strategies, while affording varied catabolic capabilities. For instance, the direct regulation of the *PGM1* gene by Gal4 would enhance the connection between the Leloir pathway and glycolysis in *S. uvarum* relative to *S. cerevisiae* (Fu et al. 2000; Ostergaard et al. 2000; Garcia Sanchez et al. 2010). *S. uvarum PGM1* is highly induced by galactose (Figure 2.11), but this likely ancestral regulatory connection was lost in the *S. cerevisiae-S. kudriavzevii* clade (Figure 2.11). Nearly all of the known differences between the *S. cerevisiae* and *S. uvarum GAL* networks make *S. uvarum* more active, including 1) apparent regulation of *PGM1* by Gal4; 2) the presence of genes encoding two galactose transporters (Figure 2.1); 3) the galactokinase activity of SuvaGal3

(Figure 2.2A); 4) the higher activity of both the *GAL1* coding and cis-regulatory sequences (Figure 2.15D); and 5) higher background gene expression across the network (Caudy et al. 2013; Roop et al. 2016). Indeed, the possession of a gene encoding a second co-repressor appears to be one of the few features of the *S. uvarum GAL* network that would serve to counteract its higher activity. Thus, the dramatic up-regulation of *GAL80B* during induction may offer a robust negative feedback loop that helps prevent over-induction and metabolic overload. The retention of *GAL80B* may have allowed *S. uvarum* to maintain a more active *GAL* network, while the *S. cerevisiae GAL* network evolved lower activity.

Comparison of yeast genomes beyond the *Saccharomyces* suggests that galactose metabolism may impose similar constraints across the yeast phylogeny. The genes encoding the Leloir enzymes occur in one of the few broadly conserved yeast gene clusters (Wong and Wolfe 2005; Slot and Rokas 2010; Wolfe et al. 2015; Riley et al. 2016), which has been suggested could promote enzyme co-regulation to prevent the accumulation of toxic intermediates (Price et al. 2005; Lang and Botstein 2011) or ensure that only complete networks are co-inherited (Lawrence and Roth 1996; Hittinger et al. 2010). In addition to *S. uvarum*, many yeast species that underwent the WGD retain *GAL80B* (Hittinger et al. 2004). Perhaps due to these intrinsic metabolic challenges and the limited benefits of maintaining a dedicated *GAL* network, the ability to consume galactose has been lost many times across diverse yeast lineages (Hittinger et al. 2004, 2010, 2015; Slot and Rokas 2010; Wolfe et al. 2015; Riley et al. 2016).

2.4.4 Ongoing functional diversification of paralogs and their gene networks

In contrast to more commonly studied processes of the rapid neofunctionalization and subfunctionalization of paralogs (Moore and Purugganan 2003; Innan and Kondrashov 2010), we have shown how duplicate *GAL* genes continued to diverge functionally in ways that dramatically influenced the metabolic and regulatory states of extant *Saccharomyces* species. Based on the redundancy observed between *GAL1* and *GAL3* and between *GAL80* and *GAL80B* in *S. uvarum*, we infer that the functions of these two paralog pairs overlapped more at the origin of the genus *Saccharomyces* than in *S. cerevisiae* (Figure 2.16). After the *S. uvarum-S. eubayanus* clade diverged

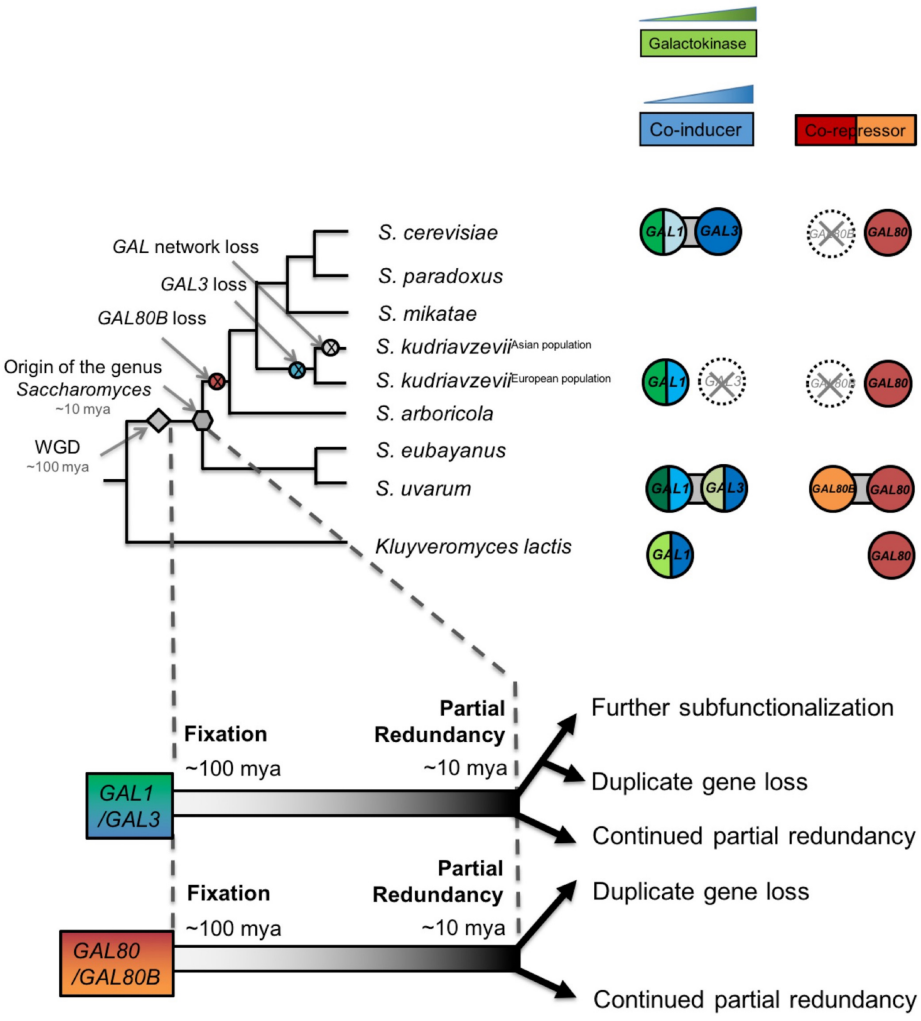


Figure 2.16: Ongoing diversification of the functions of the *GAL1-GAL3* and *GAL80-GAL80B* duplicate gene pairs in *Saccharomyces*.

(continued) Important evolutionary events are shown on the cladogram. WGD, the whole genome duplication that created the two pairs of paralogs. The inferred duplicate divergence fates are shown at the bottom of the tree. The inferred timeline is depicted by the dashed line. Roughly 100 million years ago, these two pairs of duplicate genes were fixed in the ancestral genome following a WGD event. Considerable partial redundancy was maintained in the lineage leading to the origin of the genus *Saccharomyces*. In the last 10 million years, the fates of the duplicate genes have functionally diverged along different evolutionary trajectories. The bifunctionality of the *GAL1*/*GAL3* genes is represented by green for the enzymatic function and blue for the co-induction function. The color shading represents approximate functionality for experimentally characterized genes: a darker color indicates a stronger function, whereas a lighter color indicates a weaker function. The dashed circle with a cross indicates the loss of the indicated gene. Note that the *S. kudriavzevii* Asian population lost its entire *GAL* network, while the European population retained most of its *GAL* network but lost *GAL80B* and *GAL3*. The additional co-repressor in *S. uvarum* may minimize the risk of metabolic overload due to an otherwise highly active *GAL* network.

from the *S. arboricola*-*S. cerevisiae* clade, these genes met distinct fates in different lineages (Figure 2.16). *GAL80B* was lost in the *S. cerevisiae*-*S. arboricola* clade, while it was retained in *S. uvarum* and *S. eubayanus* (Hittinger et al. 2010; Scannell et al. 2011; Caudy et al. 2013; Hittinger 2013; Liti et al. 2013; Baker et al. 2015). The fates of *GAL1* and *GAL3* were still more varied. *GAL3* was lost in a European population of *S. kudriavzevii*, resulting in an induction defect, while the entire *GAL* network was lost in an East Asian population of this species that cannot consume galactose (Hittinger et al. 2010). *GAL1* and *GAL3* were nearly completely subfunctionalized in *S. cerevisiae* (Hittinger and Carroll 2007), but we have shown here that they maintain considerable redundancy in *S. uvarum*.

For both paralog pairs, the ongoing functional diversification has been asymmetric. Deleting *GAL80B* and *GAL3* produced less striking phenotypes than deletion of their paralogs in *S. uvarum*, and some lineages have experienced inactivation or loss of these genes naturally. In the lineage leading to *S. cerevisiae*, Gal3 completely lost enzymatic activity, while a decrease in the promoter

activity of its paralog *GAL1* reduced, but did not eliminate, its ability to induce the network rapidly. Other *GAL* genes also experienced an adaptive decrease in promoter activities in the lineage leading to *S. cerevisiae* (Roop et al. 2016), which may have been enabled or necessitated by the loss of the secondary co-repressor encoded by *GAL80B*. Remarkably, the disparate resolutions of the functions of these paralogs did not happen soon after the WGD that created them. Instead, the diversification described here occurred within the last 10 million years of a 100 million year history, demonstrating that the echoes of duplication events continue to resonate through gene networks much longer than is generally appreciated (Gordon et al. 2009; Conant et al. 2014).

The ongoing functional diversification of ancient paralogs likely has an even greater impact on the evolution of plants and vertebrates, where nearly all extant species are the products of multiple rounds of WGD, and differential paralog retention is widespread (Amores et al. 1998, 2004; Aury et al. 2006; Blomme et al. 2006; Gmez-Valero et al. 2007; Scannell et al. 2007; De Smet et al. 2013; McGrath et al. 2014). Molecular and genetic dissection is much more challenging in these systems, but there are hints that the diversification of ancient paralogs continues to have functional consequences for the evolution of metabolism (Steinke et al. 2006; Conant et al. 2014) and development (Kassahn et al. 2009; Cortesi et al. 2015). Paralog diversification is often asymmetric as one paralog acquires a more specialized or auxiliary role (Force et al. 1999; Moore and Purugganan 2003; Hittinger and Carroll 2007; Des Marais and Rausher 2008; Innan and Kondrashov 2010; Conant et al. 2014). Even if this specialization is conditionally adaptive, the auxiliary paralog can become more susceptible to gene loss when conditions change. Paralog loss ends the saga of duplicate gene diversification, possibly forcing partially redundant functions back onto the remaining paralog, relieving paralog interference (Baker et al. 2013), or leading to compensatory changes elsewhere in the network. Perhaps more interestingly, paralog loss eliminates redundancy and limits the long-term potential for adaptation. The ongoing evolutionary processes affecting the *GAL* paralogs show how gene duplication facilitates phenotypic change and network diversification in ways that continue to reverberate.

2.5 Materials and Methods

Strain construction

To construct *GAL* gene knockouts, we used MX cassettes (*hphMX*, *natMX*, or *kanMX*) (Wach et al. 1994; Goldstein and McCusker 1999) to precisely replace the coding sequence from start codon to stop codon. Transformations were based on the standard lithium acetate/PEG method optimized for *S. uvarum* (room temperature incubation, followed by a 37°C heat shock) (Gietz et al. 1995; Caudy et al. 2013). To perform allele swaps, the coding sequence or promoter was first replaced by a selectable and counter-selectable *TK-hphMX* cassette, which does not require prior introduction of an auxotrophy (Alexander et al. 2014). The coding sequence or promoter of the desired replacement sequence was amplified by PCR primers with overhangs homologous to the targeted genomic flanking region. In some cases, extended homology (100-300 bp) was then introduced through PCR sewing. For each *GAL1* promoter swap, we swapped the entire upstream intergenic region. Note that the *S. cerevisiae* and *S. uvarum GAL1* promoters are both divergent promoters that also regulate *GAL10* and may also impact a lncRNA previously described in *S. cerevisiae* (Cloutier et al. 2016). Successful replacement strains were isolated by selecting for the loss of thymidine kinase activity by resistance to 5-fluorodeoxyuridine (FUdR), as well as the loss of resistance to hygromycin by replica plating (Alexander et al. 2014). GFP reporters were constructed in three parts: the *hphMX* cassette was placed upstream as the selection marker, the *S. uvarum GAL1* promoter was used to drive the expression of the reporter, and the reporter was a *yEGFP* (yeast Enhanced Green Fluorescence Protein) construct with a *S. cerevisiae CYC1* terminator that was amplified from FM1282 (Hittinger and Carroll 2007; Hittinger et al. 2010). GFP reporters were introduced to replace a *S. uvarum gto1*, an inactive pseudogene (chr7: 767, 328-766, 478) orthologous to *S. cerevisiae GTO1* (Scannell et al. 2011). The modified loci of all transformants were verified by Sanger sequencing. *S. cerevisiae* is NCBITaxon:4932, *S. uvarum* is NCBITaxon:230603, and the strains used in this study are listed in Supplementary File 2.

Media and growth assays

Strains were first streaked on YPD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose, 18 g/L agar) plates from frozen glycerol stocks. Next, a single colony of each strain was cultured in synthetic complete (SC) medium plus 0.2% glucose (1.72 g/L yeast nitrogen base without amino acids, 5 g/L ammonium sulfate, 2 g/L complete dropout mix, 2 g/L glucose) for 2-3 days, a condition that does not induce and only minimally represses the *GAL* network. There were at least two biological replicates for each genotype, generally from independent transformants. These precultures were washed with water and inoculated into the desired growth media in a 96 well plate. No explicit power analyses were performed to determine sample sizes or the number of replicates. Instead, each experiment was independently performed at least twice on separate days; details can be found in each legend. Biological replicates were defined as independent isogenic colonies on agar plates, which were used for subsequent precultures and growth assays; technical replicates were defined as independent growth assays from the same preculture. The absorbance of each well was read by an unshaken BMG FLUOstar Omega plate reader every 10 minutes at 595 nm. The number of cell divisions for each time point was calculated as $\log_2[(OD_{strain}\text{-}OD_{media})/(OD_{start}\text{-}OD_{media})$, an equation that normalized each optical density time point to its starting optical density and the optical density of the medium. The times to first doubling were calculated as the times for the optical densities to double from their normalized starting points. Specific growth rates were calculated using the Growth Curve Analysis Tool (GCAT) (Bukhman et al. 2015). Replicates that failed to grow as pre-cultures or during growth assays were considered as outliers and were excluded from subsequent analyses; no other data were excluded. For *S. cerevisiae* and *S. uvarum* *gal1* mutant growth assays (Appendix A: Table A.2), strains were pre-cultured in SC plus 0.67% fructose for 2 days and inoculated at a 1:1000 ratio into supplemented minimal medium (1.72 g/L yeast nitrogen base without amino acids, 5 g/L ammonium sulfate, 85.6 mg/L uracil, 85.6 mg/L lysine, 20 g/L galactose) plus 2% galactose or no carbon source. The growth properties of these strains were determined by subtracting the optical densities of cultures in media without a carbon source from media with galactose; differences less than 0.05 were considered as "no growth." In each case, *S. cerevisiae* strains were cultured at 30°C, while *S. uvarum* strains were cultured at

24°C, except when they were cultured in the same 96 well plate. In these cases (Figure 2.2C, Figure 2.2B and Figure 2.6A), strains were grown at 26°C, and the results were summarized in one graph.

Flow cytometry

The pre-culture and growth conditions were identical to those described above for the 96-well growth assays. At the indicated time points, 1-30 μL cultures were transferred from the 96-well plate to fresh medium of the same type in a new 96-well plate to obtain a concentration of 200-500 cells/μL for flow cytometry. There were at least three biological replicates for each genotype. The flow cytometry was conducted using a Guava EasyCyte Plus flow cytometer. Each experiment was independently conducted at least twice on separate days. The data were extracted from FCS 2.0 formatted files using FlowCore (Hahne et al. 2009) (RRID:SCR_002205). The fluorescence levels were normalized by forward scatter to control for cell size. For each genotype, histograms of normalized fluorescence levels of 6,000 cells were smoothed by Kernel density estimation and plotted using the R statistical package.

RNA sequencing

Strains were pre-cultured in SC plus 0.2% glucose for 2 days and inoculated into SC plus 2% galactose, 2% glucose, or 5% glycerol. Samples were harvested at the indicated time points and frozen using a dry ice/ethanol bath. Total RNA was extracted using the standard acidic phenol protocol (Hittinger and Carroll 2007), and residual DNA was removed through DNase I treatment. Poly-A enrichment was performed with the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490, in the experiment to examine *S. uvarum* GAL network membership) or with the NEB Magnetic mRNA Isolation kit (NEB #S1550, in the experiment sampled during the TGA phase and at mid-log phase in galactose). Illumina libraries were constructed using the NEB Ultra Directional RNA Library Prep Kit for Illumina (NEB #E7420) and sequenced using the Illumina HiSeq 2500 platform. Reads were mapped onto the *S. uvarum* reference genome (CBS7001) (Scannell et al. 2011) using Bowtie version 2.2.2 with local read alignment and otherwise default settings (Langmead et al. 2009). Read counts were quantified by HTSeq version 0.6.0 (Anders et al. 2015) (RRID:SCR_005514). Differential expression was determined using EBseq version

1.1.5 with a false discovery rate (FDR) of 0.05 (Leng et al. 2013) (RRID:SCR_003526). Analysis with edgeR (RRID:SCR_012802) using the default settings was performed in parallel to examine known *S. cerevisiae* Gal4 targets that were not scored as differentially expressed in *S. uvarum* (Robinson et al. 2010). Differentially expressed genes were further analyzed by GO term analysis (Ashburner et al. 2000; Cherry et al. 2012) (Generic GO Term Mapper, RRID:SCR_005806; SGD Gene Ontology Slim Mapper, RRID:SCR_005784). The RNA-Seq data are available at NCBI's SRA under accession number SRP077015.

^{13}C-labelled yeast metabolome extract preparation

The ^{13}C yeast metabolome extract (Bennett et al. 2008) was prepared by growing Y22-3 (McIlwain et al. 2016) aerobically on YNB (-AA) + 1% ^{13}C glucose. Yeast cultures were inoculated at an OD of 0.05 into ^{13}C medium. Samples were harvested from each culture by centrifugation and frozen in liquid N_2. Frozen pellets were first extracted with 750 µL of 40:40:20 ACN/MeOH/H_2O, followed by a second extraction with 500 µL of the same extraction solvent. Extracts were pooled, centrifuged, and the supernatant was collected for later use as an internal standard for absolute metabolite quantification (Bennett et al. 2008).

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) metabolomic analyses

Lyophilized cell culture metabolites were extracted from mutant and wild-type strains with 5 mL ice-cold 7:2:1 MeOH/$CHCl_3$/H_2O, and 100 µL of the extract was mixed with 10 µL ^{13}C-labelled yeast metabolome extract. Three biological replicates were included for the *S. uvarum* strains (Figure 2.13), while two were included for the *S. cerevisiae* strains (Figure 2.12). Chromatographic separations based on a previously described method (van Dam et al. 2002; Long et al. 2012) were carried out on an Agilent 1200 series HPLC comprising a vacuum degasser, binary pump, heated column compartment, and thermostated autosampler set to maintain 6°C. Mobile phase A (MPA) was 0.5 mM NaOH, and mobile phase B (MPB) was 100 mM NaOH. 20 µL of intracellular extract or calibrant standard mixture was separated on a Dionex IonPac AS11-HC IC column (2.0 mm x 250 mm, 9.0 µm) held at 40°C using a flow rate of 0.35 mL/min. Metabolite elution was achieved by first holding at 5% MPB for 22.5 min to separate isobaric phospho sugar species. MPB was then linearly increased from 5% to 100% over 27.5 min to elute the

remaining metabolites. MPB was held at 100% for 7 min for column cleaning followed by an 8-min re-equilibration step at 5% MPB. The LC system was coupled to a Dionex ERS 500 suppressor controlled by a Dionex Reagent-Free Controller (model RFC-10) and an Agilent 6460A Triple Quadrupole MS. The MS was operated in negative mode, acquiring MRM scans for each metabolite. Quantification based off external standard calibration curves and correction with the ^{13}C-labelled yeast standard was performed with Agilent MassHunter Quantitative Analysis software (version B.06.00).

High performance liquid chromatography (HPLC)

The pre-culture conditions were identical to those described above for the growth assays. At indicated time points, 1 mL of cells were centrifuged, and 500 µL supernatant was harvested and frozen at -80°C. HPLC was conducted at the GLBRC Metabolomics Lab using an HPLC-RID system with an Aminex HPX-87H (BioRad, Inc. Hercules, CA) following previously described protocols (Moore and Johnson 1967; Ehrman and Himmel 1994). Instrument control, data collection and analyses were conducted using ChemStation B.04.03 software (Agilent Technologies, Inc., Palo Alto, CA).

Statistical analysis

All *p*-values, except for the RNA-Seq, metabolomics (two-sided students *t*-test), and HPLC analyses (two-sided students *t*-test), were calculated using a conservative two-sided nonparametric test. Specifically, we used a Wilcoxon rank sum test that allows the rank data from multiple independent experiments to be pooled to account for day-to-day variation without making assumptions about distribution of the variance. These tests were performed using Mstat software version 6.1.4 (http://mcardle.oncology.wisc.edu/mstat/).

2.6 Acknowledgments

We thank Mick McGee for performing the HPLC experiments and analyzing the data; Audrey P. Gasch for providing access to the Guava flow cytometer; the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing Illumina and Sanger sequencing

facilities and services; Amy A. Caudy, Stephen A. Johnston, and James E. Hopper for helpful discussions; William G. Alexander for technical advice on the use of the *TK-hphMX* and *TK-kanMX* cassettes; and EmilyClare Baker and Drew T. Doering for critical reading of the manuscript. This work is based upon work supported by National Science Foundation (DEB-1253634, DEB-1442148), DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494), and National Institute of Food and Agriculture (Hatch Project 1003258), National Institutes of Health (R35 GM118110). CTH is a Pew Scholoar in Biomedical Sciences, supported by the Pew Charitable Trusts, and Alfred Toepfer Faculty Fellow, supported by the Alexander von Humboldt-Stiftung.

Chapter 3

Repeated cis-regulatory tuning of a metabolic bottleneck gene during evolution[1]

3.1 Abstract

Repeated evolutionary events imply underlying genetic constraints that can make evolutionary mechanisms predictable. Morphological traits are often thought to evolve through cis-regulatory changes because these mechanisms resolve constraints in pleiotropic genes that are reused during development. In contrast, metabolic traits are generally considered less constrained and thought to frequently evolve through protein-coding changes. Here we show how a metabolic bottleneck gene that integrates flux from multiple pathways has repeatedly adopted similar cis-regulatory solutions during evolution. Specifically, the genes encoding phosphoglucomutase activity (*PGM1/PGM2*), which connects galactose catabolism to glycolysis, have gained and lost direct regulation by the transcription factor Gal4 several times during yeast evolution. Galactose-mediated regulation of *PGM1/2* is necessary to support vigorous growth on galactose in multiple yeasts, including *Saccharomyces uvarum* and *Lachancea kluyveri*. Furthermore, the addition of galactose-inducible *PGM1* is sufficient to improve the growth on galactose of multiple species that lack this regulation, including *Saccharomyces cerevisiae*. The strong association between regulation of *PGM1/2* by Gal4 even enables remarkably accurate predictions of growth phenotypes on galactose between closely related species. Our data suggest that the repeated gain and loss of cis-regulatory changes in this metabolic bottleneck gene are driven by constraints that govern flux through converging metabolic

[1]Author contributions. Conception and design: MCK, CTH. Acquisition of data: MCK, JK, WGA, JFC, RLW, CTH. Analysis and interpretation of data: MCK, CTH, JK. Drafting or revising the article: MCK, JK, WGA, JFC, RLW, CTH.

pathways. Since metabolic pathways are highly interconnected, we argue that cis-regulatory evolution might be common at pleiotropic genes that control metabolic bottlenecks and intersections.

3.2 Introduction

Repeated use of the same genes to achieve similar phenotypic outcomes is thought to reflect a combination of genetic constraints and similar selective pressures (Stern, 2013, Christin et al., 2010). Although numerous mutational paths can lead to similar phenotypes in laboratory lines, the small subset that persists in nature can often be predicted once the genetic constraints are understood. Repeated phenotypic alterations have been shown to occur through similar amino acid substitutions (Christin et al., 2007, Hoekstra et al., 2006) while other cases have been found where similar cis-regulatory changes affect the expression of genes (Sucena et al., 2003). It has been hypothesized that cis-regulatory changes are the main genetic causes of morphological evolution because strong pleiotropic constraints are imposed when key developmental genes are reused spatially and temporally (Carroll, 2008, Stern and Orgogozo, 2008). In contrast, physiological and metabolic traits have evolved through changes in both protein-coding regions and cis-regulatory elements (Roop et al., 2016, Ihmels et al., 2005, Lin et al., 2013). These observations have been interpreted to reflect a lack of analogous constraints on the genes that control metabolic traits (Stern and Orgogozo, 2008). For example, mutations in coding regions have led to the acquisition of novel enzymatic activities and radical modifications in specificity (Thomson et al., 2005, Des Marais and Rausher, 2008, Voordeckers et al., 2012). Nonetheless, many physiological changes have been associated with dramatic cis-regulatory rewiring, including the transition from aerobic respiration to aerobic fermentation in yeasts, even as most central metabolic functions were conserved (Ihmels et al., 2005, Lin et al., 2013, Thompson et al., 2013). Thus, we hypothesized that, when metabolic pathways are constrained by pleiotropy, such as where multiple fluxes are integrated and bottlenecks occur, these tensions might also be resolved through cis-regulatory changes during evolution.

To test this hypothesis, we took advantage of the quantitative variation among yeasts in catabolism of the sugar galactose. The biochemical and regulatory pathway responsible for *GAL*actose utilization in budding yeast *Saccharomyces cerevisiae* is a paradigm of eukaryotic molecular biology, making it an attractive model to study the molecular mechanisms of evolution. Recent functional comparisons of the *GAL* network in multiple yeast species have shown that its basic regulatory architecture has been conserved with minor, but functionally consequential, variations across the family Saccharomycetaceae, which spans about 100 million years of evolution (Kuang et al., 2016, Roop et al., 2016, Peng et al., 2015, Dalal et al., 2016, Martchenko et al., 2007). The *GAL* network of *S. cerevisiae* encodes three enzymes in the galactose-specific Leloir pathway (Gal1, Gal7, and Gal10), a transporter (Gal2), and three regulators (Gal3, Gal4, and Gal80). The enzyme phosphoglucomutase, which catalyzes the conversion of glucose-1-phosphate into glucose-6-phosphate, controls the flux through the Leloir pathway and into glycolysis. In yeasts that underwent a whole genome duplication (WGD) about 100 million years ago (Wolfe and Shields, 1997, Marcet-Houben and Gabaldon, 2015), phosphoglucomutase is generally encoded by two paralogs, *PGM1* and *PGM2*, while species that did not undergo the WGD only have *PGM1* (Figure 3.1). This enzymatic step is the metabolic bottleneck for galactose metabolism (Bro et al., 2005, Garcia Sanchez et al., 2010, Hong et al., 2011). *S. cerevisiae pgm1 pgm2* double mutants cannot utilize galactose as the sole carbon source (Boles et al., 1994), and overexpressing *PGM2*, the major isoform in *S.cerevisiae*, is among the best ways to increase the rate of galactose metabolism (Bro et al., 2005, Garcia Sanchez et al., 2010).

Here we show that expression of the bottleneck gene *PGM1/2* has been repeatedly tuned across the family Sacharomycetaceae to quantitatively modulate galactose metabolism by the addition or subtraction of Gal4-binding sites in its promoter. We show Gal4-binding sites are required for optimal growth on galactose in several species of yeasts and that galactose-inducibility of *PGM1/2* is sufficient to increase galactose metabolism in species that lack direct regulation by Gal4, including *S. cerevisiae*. The repeated evolutionary gain and loss of cis-regulatory elements regulating this metabolic bottle neck provides a mechanism for *PGM1/2* expression to respond specifically to

galactose, which we hypothesize resolved the constraints imposed by converging metabolic pathways.

3.3 Results

Yeasts of the family Saccharomycetaceae display dramatic variation in their abilities to grow on galactose. Qualitative differences that evolved due to the parallel losses of entire *GAL* networks have received considerable prior attention (Hittinger et al., 2004, Riley et al., 2016), but quantitative variation in galactose metabolism is even more widespread (Figure 3.1A, Appendix B: Figure B.1A). To understand the genetic basis of this variation, we applied a comparative approach that leveraged prior mechanistic understandings of the *GAL* network in *S. cerevisiae* and recent functional studies comparing *S. cerevisiae* and *Saccharomyces uvarum* (formerly called *S. bayanus* var. *uvarum*) (Kuang et al., 2016, Roop et al., 2016). *S. uvarum* grows faster on galactose because its *GAL* network is more active than that of *S. cerevisiae*, in part due to cis-regulatory changes that affect the expression of multiple *GAL* genes (Kuang et al., 2016, Roop et al., 2016). We previously showed that *PGM1* was also induced 18-fold by galactose in *S.uvarum*, an induction not shared with *S.cerevisiae* (Kuang et al., 2016).

To test whether *PGM1* in *S. uvarum* is directly regulated by the galactose-responsive transcriptional activator Gal4 (Kuang et al., 2016), we mutated one base pair of a predicted Gal4-binding site ($CGGN_{11}CCG$) upstream of *S. uvarum PGM1*. This single point mutation was sufficient to slow galactose growth by 20% compared to wild-type *S. uvarum* (Figure 3.2A), a defect that was galactose specific (Appendix B: Figure B.2A). To further examine its impact on flux, we tested whether decoupling Gal4-induction of this bottleneck was sufficient to rescue the metabolic overload seen in *S.uvarum* strains lacking Gal80 co-repressors (Kuang et al., 2016). Indeed, the deletion of both of the predicted Gal4-binding sites upstream of *S. uvarum PGM1* rescued the growth arrest phenotype caused by metabolic overload, while leading to an even slower maximum growth rate on galactose (Appendix B: Figure B.3). This evidence suggests that *PGM1* is a bottleneck gene in *S. uvarum* and that its regulation directly controls flux through the *GAL* pathway. Thus,

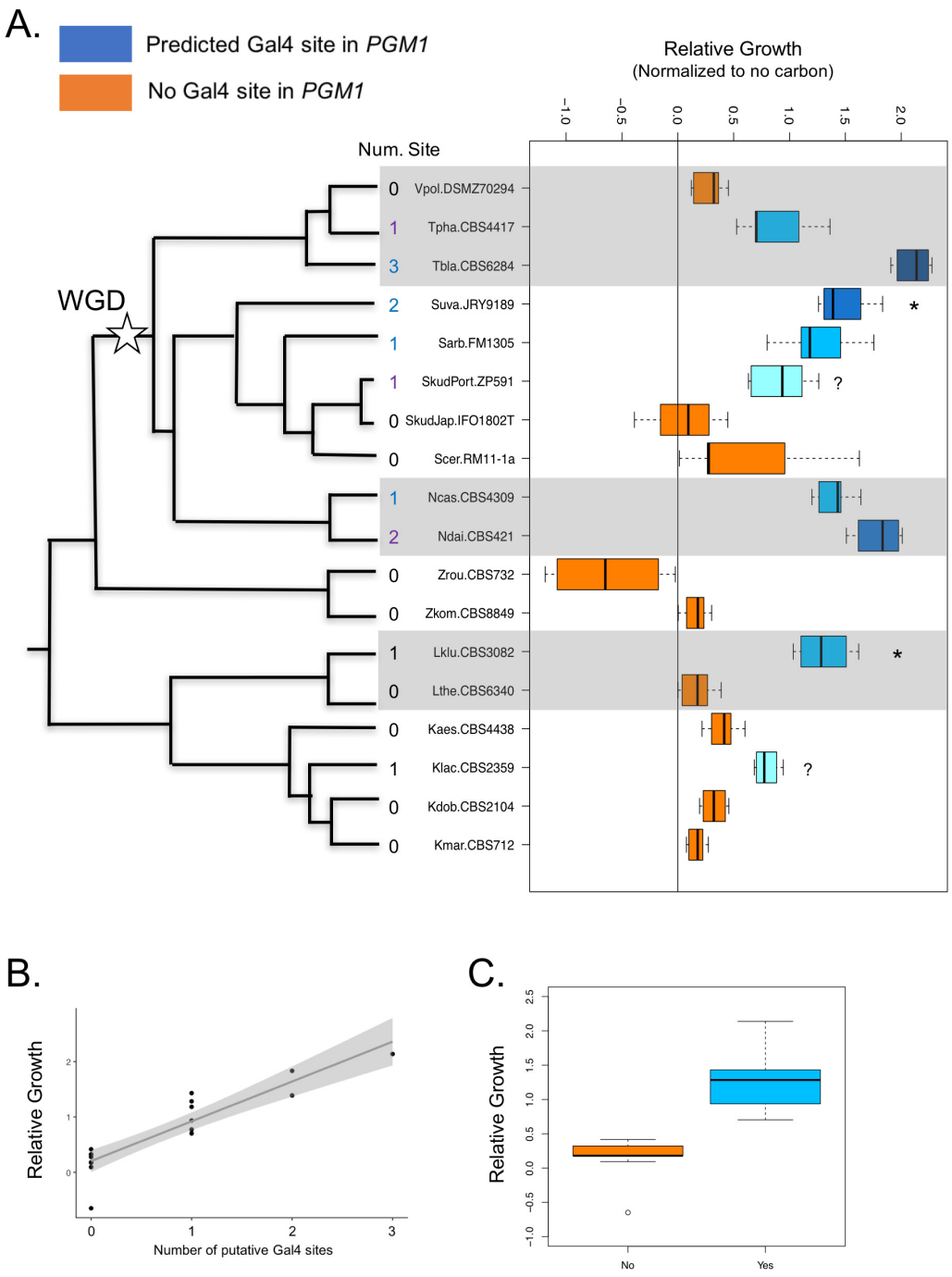


Figure 3.1: Predicted Gal4-binding sites upstream of *PGM1/2* strongly correlate with galactose growth in the family Saccharomycetaceae.

(continued) A. The number of predicted Gal4-binding sites upstream of *PGM1* or *PGM2* strongly correlates with the number of cell division after 15 hours of growth in SC + 2%galactose. 15 hours corresponds to the time when most species have started to grow. "Num. Site" denotes the number of predicted Gal4-binding sites upstream of *PGM1* or *PGM2*. The number of predicted binding sites upstream of *PGM1* are colored in blue, the number upstream of *PGM2* are in purple, and the number upstream of *PGM1* in non-WGD species in black. In the boxplot: 1) $\star$: significant phenotypic changes were observed when the binding site was mutated; 2) ?: no phenotypic changes were detected when the binding site was mutated; 3) each data box is colored coded based on the number of binding sites: the darkness of blue color corresponds to the number of binding site, with light blue (*Saccharomyces kudriavzevii* Portuguese strain and *Kluyveromyces lactis*) indicating predicted binding sites with no detected function, and orange indicating the absence of any predicted binding sites. Relative growth (n = 6) denotes the number of cell divisions after 15 hours, which was calculated as $\log_2[(\text{OD}_{strain}\text{-OD}_{media})/(\text{OD}_{start}\text{-OD}_{media})$. This calculation was applied for all figures. Each strain is designated by a 4-letter species abbreviation, a period, and its strain name (Vpol: *Vanderwaltozyma polyspora*, Tpha: *Tetrapisispora phaffii*, Tbla: *Tetrapisispora blattae*, Suva: *Saccharomyces uvarum*, Sarb: *Saccharomyces arboricola*, SkudPort: *Saccharomyces kudriavzevii* Portuguese population, SkudJap: *Saccharomyces kudriavzevii* Japanese population, Scer: *Saccharomyces cerevisiae*, Ncas: *Naumovozyma castellii*, Ndai: *Naumovozyma dairenensis*, Zrou: *Zygosaccharomyces rouxii*, Zkom: *Zygosaccharomyces kombuchaensis*, Lklu: *Lachancea kluyveri*, Lthe: *Lachancea thermotolerans*, Kaes: *Kluyveromyces aestuarii*, Klac: *Kluyveromyces lactis*, Kdob: *Kluyveromyces dobzhanskii*, Kmar: *Kluyveromyces marxianus*).

B. The strong correlation and differences between number of binding sites and growth in galactose. The data were extracted from Figure 3.1A, and the median was used to represent each species. The gray shaded area corresponds to confidence interval. C. The data from Figure 3.1B are converted to the absence and presence of predicted Gal4-binding sites (p = 2.9e-5, n = 9, Welch's two-sample t-test).

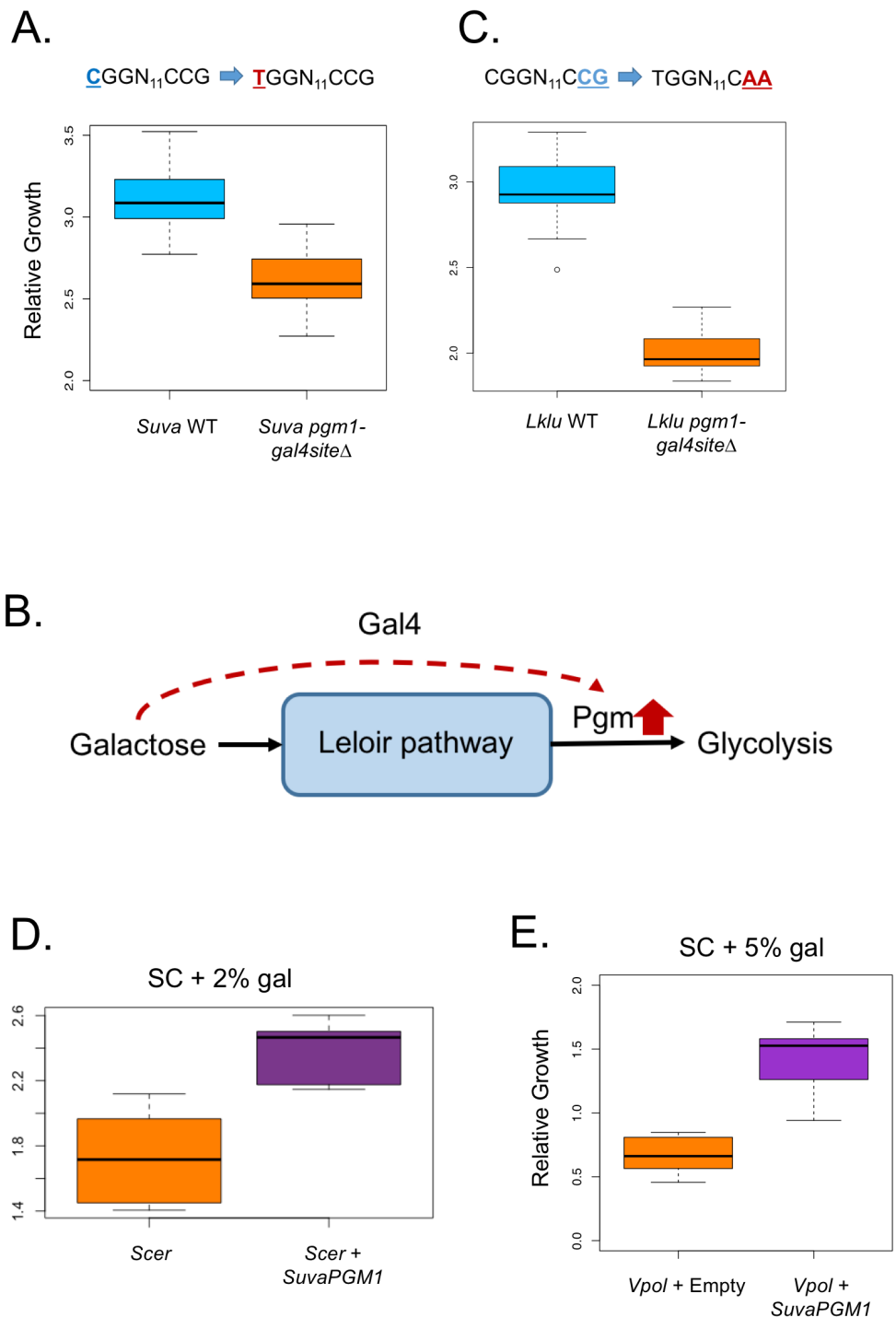


Figure 3.2: Gal4-mediated regulation of phosphoglucomutase is necessary and sufficient to support vigorous growth on galactose for multiple yeast species.

(continued) A. Mutation of one of two predicted Gal4-binding sites upstream of *S. uvarum PGM1* slowed down growth on galactose compared to wild-type (p = 6e-6, n_{WT} = 15, n_{mutant} = 11, Wilcoxon rank sum test).

B. Mutation of the predicted Gal4-binding site upstream of *L. kluyveri PGM1* slowed down growth on galactose compared to wild-type. (p = 7.6e-9, n_{WT} = 21, n_{mutant} = 14, Wilcoxon rank sum test).

C. Hypothesis of *GAL* network activity tuning with a novel (dotted line) Gal4-Pgm feedforward loop.

D. *S. uvarum PGM1* increased galactose growth in *S. cerevisiae* (p = 5e-3, n = 6).

E. *S. uvarum PGM1* increased galactose growth in *V. polyspora* (p = 0.01, $n_{Emptyvector}$ = 6, $n_{SuvaPGM1}$ = 5, Wilcoxon rank sum test).

the high metabolic flux and vigorous growth of wild-type *S. uvarum* on galactose requires direct Gal4-mediated induction of *PGM1*, a novel regulatory connection that *S. cerevisiae* lacks.

We hypothesized that Gal4-induction of *PGM1/2* might be a common strategy to support vigorous galactose utilization in yeasts (Figure 3.2B). To test this hypothesis, we examined 20 genetic characteristics of the *GAL* networks of 17 species across six genera of the family Saccharomycetaceae (Appendix B: Table B.1). We found that the number of predicted Gal4-binding sites upstream of *PGM1/2* was strongly associated with growth on galactose (Figure 3.1, Appendix B: Table B.1, Appendix B: Figure B.1A, R^2 = 0.81, p = 1.9e-7). This correlation was specific to galactose; it was observed in several different media formulations containing galactose as the sole carbon source and when growth temperature was varied (Appendix B: Figure B.4), but it was not seen on glucose (Appendix B: Figure B.1B-C).

To test whether Gal4-mediated induction of *PGM1/2* was required for vigorous growth on galactose in multiple species, we examined impact of mutating predicted Gal4-binding sites upstream of *PGM1/2*. We first developed a genome-editing approach potentially universal to the family Saccharomycetaceae by integrating a CRISPR/Cas9 system with an autonomously replicating sequence (ARS) that functions in multiple genera (Liachko and Dunham, 2014). In addition to the genus *Saccharomyces*, this genome-editing system can also induce targeted point mutations

in at least the genera of *Lachancea* and *Kluyveromyces*, which diverged from *S. cerevisiae* about 100 million years ago. Through either CRISPR/Cas9 or traditional approaches, we mutated a predicted Gal4-binding site upstream of *Lachancea kluyveri PGM1*, *Kluyveromyces lactis PGM1*, and *Saccharomyces kudriavzevii PGM2* (in a Portuguese strain capable of growth on galactose). The *L. kluyveri* mutant grew 32% slower on galactose compared to wild-type, indicating that its Gal4-binding site is required for vigorous growth on galactose, but not glucose (Figure 3.2C, Appendix B: Figure B.2B). The *K. lactis* and *S. kudriavzevii* mutants did not show any observable defects, indicating that their predicted Gal4-binding sites were not required for robust growth in the conditions we tested or that our assays were not sensitive to subtle defects (Appendix B: Figure B.5A-B). We reasoned that these species, which grew moderately well, might instead rely on high basal expression of *PGM1/2* or modulate their expression in other ways. Consistent with this idea, we found that *S. kudriavzevii PGM2* had higher basal expression than *S. uvarum PGM1* (Appendix B: Figure B.5C) and was induced by galactose to levels similar to that of *S.uvarum PGM1* (Appendix B: Figure B.5D). Interestingly, full induction of *S. kudriavzevii PGM2* required the predicted Gal4-binding site, but induction also operated partly through a Gal4-independent mechanism (Appendix B: Figure B.5E). Thus, we conclude that direct regulation of *PGM1/2* by Gal4 is necessary to support vigorous galactose growth for some species (e.g. *S. uvarum* and *L. kluyveri*), while other regulatory mechanisms are more important in other species.

To examine whether Gal4-regulated phosphoglucomutase activity was sufficient to provide vigorous growth on galactose, we introduced *S. uvarum PGM1* into multiple species that lack predicted Gal4-binding sites upstream of *PGM1/2*. *S. uvarum PGM1* has a relatively low level of basal expression and a high level of galactose induction, so it can enhance flux through the galactose utilization pathway while minimizing pleiotropic effects when cells are not grown on galactose (Appendix B: Figure B.5C) (Kuang et al., 2016). Introduction of *S. uvarum PGM1* into *S. cerevisiae* enhanced its growth on galactose by 56% (Figure 3.2D), while it enhanced the growth of *Vanderwaltozyma polyspora* by 110% (Figure 3.2E). Therefore, the acquisition of a Gal4-regulated copy of *PGM1* is sufficient to increase the growth on galactose of multiple species that lack this regulatory connection.

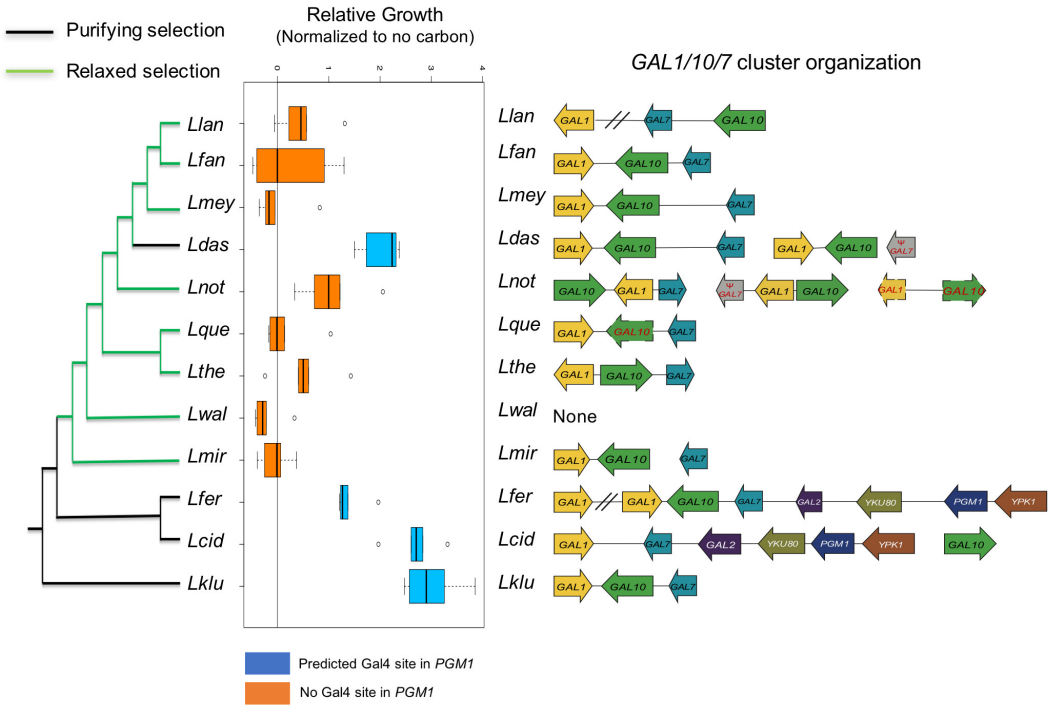


Figure 3.3: Predicted Gal4-binding sites upstream of *PGM1* predicted galactose growth differences among closely related species in the genus *Lachancea*.

(continued) Branches under purifying selection are colored as black and relaxed selection as green. Analysis is shown in Appendix B: Table B.2. In the boxplot (n = 6), each data box is colored coded by either blue (the presence of any predicted binding sites) or orange (the absence of any predicted binding sites). The *GAL1/10/7* gene cluster organization in each species is represented on the right. Each arrow denotes the coding sequence and the direction denotes the coding strand. The length of an arrow and distance in between are proportional to the actual gene lengths and distance on the chromosomes, except that distant regions will be represented by two slashes. Genes on the same chromosome are connected with a black line. Each homolog is color coded by the same color. Truncated but complete coding sequences or pseudogenes (Ψ) are represented with their gene names in red. Each species is designated by a 4-letter abbreviation (*Lklu*: *Lachancea kluyveri*, *Lcid*:*L. cidri*, *Lfer*: *L. fermentati*, *Lmir*: *L. mirantina*, *Lwal*: *L. waltii*, *Lthe*: *L. thermotolerans*, *Lque*: *L. quebecensis*, *Lnot*: *L. nothofagi*, *Ldas*: *L. dasiensis*, *Lmey*: *L. meyersii*, *Lfan*: *L. fantastica*, *Llan*: *L. lanzarotensis*).

To further assess how the gain and loss of Gal4-mediated regulation of *PGM1/2* has affected closely related species, we examined the genus *Lachancea*, which has high-quality genome assemblies for all 12 of its known species (Vakirlis et al., 2016, Sarilar et al., 2015). As we saw in our sparser sampling across the family Saccharomycetaceae (Figure 3.2), we found that the presence of a predicted Gal4-binding site was highly correlated with galactose growth in the genus *Lachancea* (Figure 3.3, Appendix B: Figure B.6). Since numerous species of *Lachancea* grew slowly in 2% galactose (and had even been previously scored as not growing on galactose by taxonomists (Kurtzman et al., 2011)), we hypothesized that these species may have experienced changes in the strength of purifying selection acting on the *GAL* genes. We used the RELAX method implemented in the HYPHY package and in the species that grew slowly, we found statistically significant relaxations in the selective pressure acting against nonsynonymous substitutions in all three genes encoding *GAL* enzymes (*GAL1*, *GAL7*, and *GAL10*) (Figure 3.3, Appendix B: Table B.2). Many of these species also lost several *GAL* genes through pseudogenization or deletion, as well as experiencing translocations and gene duplications. Even *Lachancea* species that grew vigorously on galactose experienced novel gene duplication or clustering events, suggesting

that the *GAL* pathway has had a complex evolutionary history in this genus (Figure 3.3, Appendix B: Table B.3, Supplementary notes). Perhaps most strikingly, the topology of the species tree and the novel location of its predicted Gal4-binding site upstream of *PGM1* suggest that *Lachancea dasiensis* may have reacquired the ability to grow vigorously on galactose (Figure 3.3). Under this scenario, as the *GAL* pathways of most lineages in the large clade including *Lachancea mirantina-Lachancea lanzarotensis* began to deteriorate, the decline of its function in *L. dasiensis* was partly counteracted by reacquiring a Gal4-binding site upstream of *PGM1*. In summary, these dynamic changes within the genus *Lachancea* further highlight the predictive power of the presence of Gal4-binding sites upstream of the metabolic bottleneck gene *PGM1*.

Unexpectedly but illustratively, we found that several species of *Lachancea* that grew slowly on galactose actually grew better at low concentrations of galactose than they did at high concentrations (Appendix B: Figure B.7). Better growth at low concentrations of galactose was not exclusive to this genus but was shared with the distantly related species *V. polyspora*, suggesting these yeasts might have adopted a low flux strategy (Figure 3.4A). Since all of these species lack predicted Gal4-binding sites upstream of *PGM1*, we wondered whether galactose-inducibility of *PGM1* might generally be more important at higher concentrations of galactose where there would be a stronger need to support higher metabolic flux. To test this hypothesis, we examined the impact of the galactose-inducible *S. uvarum PGM1* on *V. polyspora* at various concentrations of galactose. Although this strain did not grow any better than a wild-type strain at 0.5% galactose, it grew much better at 5% galactose (Figure 3.4B). Similarly, compared to wild-type, mutations of the Gal4-binding site of *L.kluyveri PGM1* caused limited growth defects at low concentrations of galactose, whereas these mutations caused much stronger defects at high concentrations of galactose (Figure 3.4C). These data suggest that direct Gal4-mediated induction of *PGM1/2* can enhance galactose utilization at high concentrations of galactose, but its effects are limited at low concentrations. The conditionality of the fitness benefits attached to this novel regulatory connection suggest that the repeated gains and losses of Gal4-binding sites upstream of *PGM1/2* could be driven by shifting galactose availability across yeast ecological niches and history.

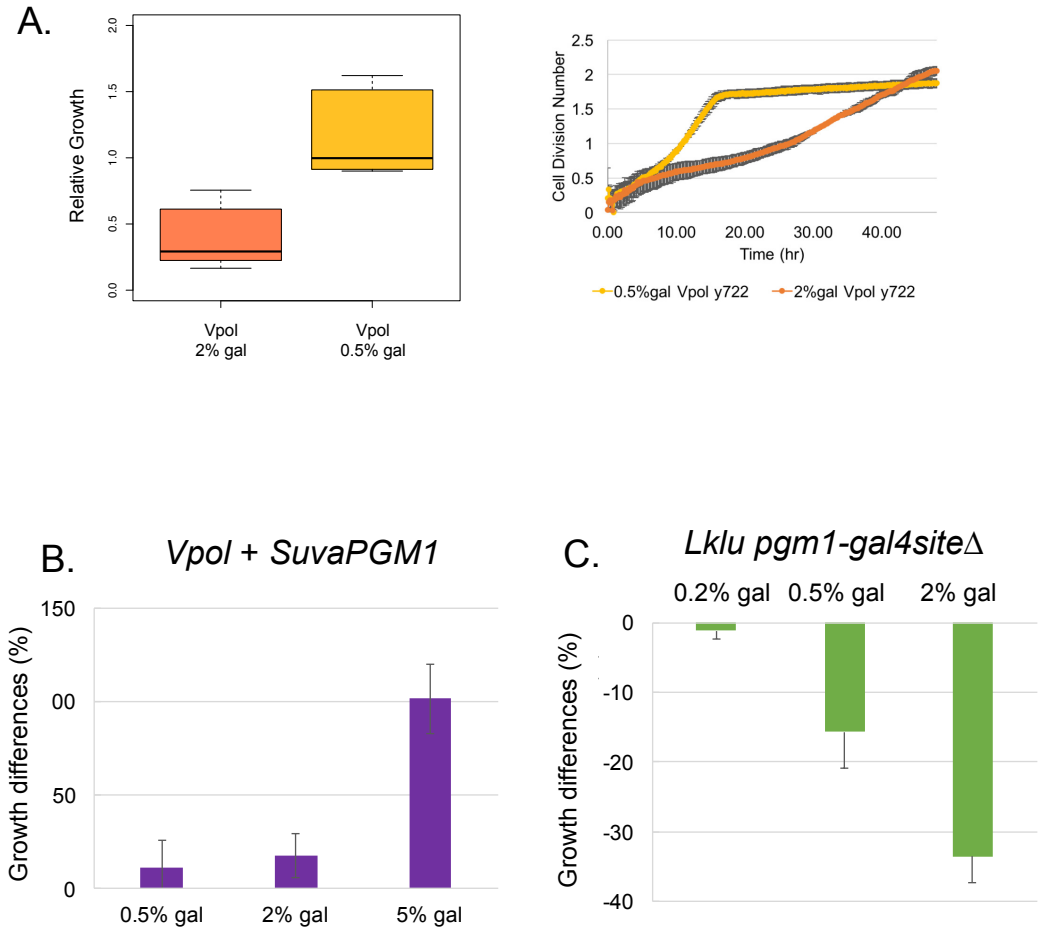


Figure 3.4: The significance of Gal4-mediated regulation of *PGM1* depends on galactose concentrations.

(continued) A. *V.polyspora* grew better at low concentrations of galactose (p = 1.7e-2, n = 5, Wilcoxon rank sum test).

B. *S. uvarum PGM1* increased galactose growth in *V. polyspora* as galactose concentrations increased. Growth differences were calculated as: ($\text{Div}_{SuvaPGM1}$-$\text{Div}_{Emptyvector}$)/$\text{Div}_{Emptyvector}$ X100%, "Div" denotes the number of cell divisions after 15 hours. Growth on 5% galactose was significantly faster than growth on 2% galactose or 0.5% galactose.

C. *PGM1* Gal4-binding site mutations caused increased growth defects in *L. kluyveri* as galactose concentrations increase. Growth differences were calculated as: (Div_{mutant}-Div_{WT})/Div_{WT} X100%, "Div" denotes the number of cell divisions after 15 hours in 2% galactose, 10 hours in 0.5% galactose, and 9 hours in 0.2% galactose. Since lower concentrations of galactose saturated much earlier before 15 hours, so the time right before wild-types saturate was chosen. Wilcoxon rank sum tests comparing the growth differences between concentrations were as follows: 1) p=0.02, $n_{0.5\%gal}$=4, $n_{2\%gal}$=6 for 2% galactose versus 0.5% galactose, 2) p=0.17, n=4 for 0.2% galactose versus 0.5% galactose.

3.4 Discussion

In summary, we have shown that, unlike the model yeast *S. cerevisiae*, many yeast species contain direct regulatory connections between Gal4 and *PGM1/2*. Galactose-mediated induction of *PGM1/2* through Gal4-binding sites is required for vigorous growth in at least two yeast species separated by over 100 million years of evolution. Moreover, the addition of *PGM1/2* genes with these novel regulatory connections is sufficient to increase galactose growth in several species that lack them. Across the family Saccharomycetaceae, the number of Gal4-binding sites upstream of *PGM1/2* is one of the best predictors of how vigorously a species grows on galactose. In addition to the well-established link between qualitative differences in galactose metabolism and presence/absence polymorphisms of the *GAL* network (Hittinger et al., 2010), we propose that variation in the number of Gal4-binding sites upstream of *PGM1/2* quantitatively tunes flux through the metabolic pathway.

Despite the importance of direct regulation of *PGM1/2* by Gal4 to quantitatively tuning *GAL* network activity across many yeast species, other mechanisms also exist. For example, galactose-mediated induction of *PGM2* occurs in *S. cerevisiae* through a mechanism that is still undetermined (Oh and Hopper, 1990). Our data suggest a similar overlapping mechanism may exist in *S. kudriavzevii*; even when the predicted Gal4-binding site upstream of *S. kudriavzevii PGM2* was mutated, it retained some galactose-dependent induction and had no detectable phenotype. In these cases, the combination of *PGM1/2* basal expression and Gal4-independent induction may be sufficient to support nearly normal growth rates, perhaps because the enzymatic activities upstream of the phosphglucomutase step are slower (Hittinger et al., 2010), and phosphglucomutase activity is not limiting. Thus, the need for and benefit of direct induction of *PGM1/2* may be strongest in cases, such as *S.uvarum*, where upstream network activities are already quite high (Kuang et al., 2016, Roop et al., 2016). Intriguingly, those yeast species that grow vigorously on galactose tend to, not only have Gal4-binding sites upstream of *PGM1/2*, but have also generally retained both copies of the duplicate genes encoding homologs of the Gal80 co-repressors (*Tetrapisispora blattae*, *Naumovozyma castellii*, *Naumovozyma dairenensis*, and *S. uvarum*). This correlation is consistent with the hypothesis that a more active network (with an additional Gal4-*PGM1/2* feed-forward loop) requires a more robust negative feedback loop (Appendix B: Table B.1) (Kuang et al., 2016). Thus, multiple genetic changes likely must coordinate with the cis-regulatory changes in *PGM1/2* to quantitatively tune *GAL* network activity.

The fact that this regulatory connection has been repeatedly gained and lost during yeast evolution suggests that the molecular mechanisms of *GAL* pathway evolution are likely constrained by several factors. The enzymes of galactose catabolism are conserved from bacteria to yeasts to humans, including the phosphoglucomutases encoded by *PGM1* homologs (Lu and Kleckner, 1994, Joshi and Handler, 1969, Oh and Hopper, 1990). The status of phosphoglucomutase as a metabolic bottleneck makes *PGM1* homologs prime targets for selection to tune metabolic flux through the pathway. In addition to galactose metabolism, *PGM1* homologs are involved in the

pentose phosphate pathway (Cherry et al., 2012), glycogen biosynthesis (Cherry et al., 2012, Hirata et al., 2003), and trehalose biosynthesis (Mulet et al., 2004, Cherry et al., 2012). Since phosphoglucomutase functions at the intersection of multiple metabolic pathways and integrates flux from these myriad sources, mutations in it are pleiotropic, and evolutionary outcomes are expected to be constrained. For example, even in species that have lost their *GAL* pathways and the ability to utilize galactose, *PGM1* homologs are retained (Hittinger et al., 2010, Hittinger et al., 2004, Riley et al., 2016). Therefore, recruiting a Gal4-binding site to specifically induce *PGM1* expression in response to galactose would allow focused selection to increase flux through the galactose utilization pathway, while minimizing the pleiotropic effects on other pathways in different environmental contexts. Thus, the pleiotropic bottleneck gene *PGM1/2* seems to have evolved on a more restricted path than many genes, including the *GAL* genes themselves. Consistent with a previous study, in three independent experimental evolved lines on galactose, no expression or coding changes were detected in any *GAL* genes but only *PGM2* is up-regulated in all three lines (Hong et al., 2011).

In other similarly small and conserved metabolic and regulatory networks, critical genes with pleiotropic roles might also display predictable evolutionary patterns. In a manner analogous to the spatial and temporal constraints imposed by the developmental regulatory networks of animals (Carroll, 2005, Carroll, 2008, Stern and Orgogozo, 2008, Stern, 2013), metabolic genes that integrate flux from multiple pathways may be particularly likely resolve conflicts between selective forces through cis-regulatory changes that enable environmentally specific responses. Indeed, decision points that integrate signals from multiple developmental pathways have been referred to as “bottleneck genes” and argued to be frequent targets of cis-regulatory changes (Stern and Orgogozo, 2008). Similarly, metabolic pathways that are highly interconnected, with many pleiotropic enzymes at the intersection of multiple pathways, may also fall under constraints that favor cis-regulatory tuning of metabolism.

3.5 Acknowledgements

We thank Hittinger Lab members for helpful discussions, Dana A. Opulente, David J. Krause, and Drew T. Doering for advice on statistics and graphics, and Samuel Deutsch from the DOE Joint Genome Institute for advice on the design synthesis of a CRISPR/Cas9 system for *S. cerevisiae*. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. This material is based upon work supported by the National Science Foundation under Grant Nos. DEB-1253634 and DEB-1442148, by USDA National Institute of Food and Agriculture Hatch Project 1003258, and funded in part by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). CTH is a Pew Scholar in the Biomedical Sciences, supported by the Pew Charitable Trusts.

3.6 Supplementary notes

The frequent genomic translocations observed in the genus *Lachancea* (Vakirlis et al., 2016) include translocations involving several *GAL* genes. In some cases, more radical *GAL* network reconfigurations also occured. Although *GAL1-GAL10-GAL7* appears as a well conserved cluster in most species in the family Saccharomycetaceae (Slot and Rokas, 2010), these genes are not clustered in four species in the genus of *Lachancea*. This cluster is duplicated and partially pseudogenized in two species (*Lachancea dasiensis* and *Lachancea nothofagi*). A third duplication of the enzyme gene cluster in *L. nothofagi* is at the end of a scaffold, so the absence of *GAL7* in this third cluster might be due to an incomplete assembly. Intriguingly, we found that *PGM1* is clustered with *GAL1*, *GAL7*, and a putative *GAL2* homolog in *Lachancea fermentati* and *Lachancea cidri*. These two clusters are not subtelomeric region, as previously found in *Torulaspora delbrueckii* (Wolfe et al., 2015), but are at least 129 kb away from the end of the chromosomes. Several additional features might explain the slow growth in galactose for three species. Homologs of *GAL4* and *GAL80* have both been lost in *Lachancea nothofagi* and *Lachancea mirantina*, even though

other *GAL* genes remain intact. Two populations of *Lachancea quebecensis* have separately acquired two different premature stop codons (PTCs) in their *GAL10* genes (Freel et al., 2016). To examine whether the maintenance of *GAL* genes might be due to the metabolism of other sugars that also require the Leloir pathway, such as lactose and melibiose, we surveyed all 12 genomes for genes encoding beta-glucosidases (breaks down lactose) and alpha-glucosidases (breaks down melibiose). We found no genes encoding beta-glucosidases, but we found genes encoding alpha-glucosidases in *L. dasiensis*, *L. nothofagi*, *L. quebecensis*, and *L. cidri* (Appendix B: Table B.3). The two species that grew slowly in galactose (*L. nothofagi*, *L. quebecensis*) also grew slowly in melibiose, so we conclude that melibiose is probably not a better induction signal for the *GAL* network than galactose.

3.7 Materials and methods

Strain constructions

Two approaches were used to introduce point mutations to mutate the predicted Gal4-binding sites: the traditional marker-based method and a CRISPR/Cas9 based approach potentially general to the family Saccharomycetaceae:

1) Traditional approach (*S.uvarum* and *S. kudriavzevii*)

We first deleted part of the promoter containing the predicted binding sites using a selectable and counterselectable marker. After removing the wild-type sequence with the marker gene, we then replaced the marker with sequences containing the desired point mutation that were introduced using PCR primers. The strain engineering was done as previously described (Kuang et al., 2016, Alexander et al., 2014). The transformation protocol for *S. uvarum* and *S. kudriavzevii* were conducted as previously described (Kuang et al., 2016, Hittinger et al., 2010).

2) CRISPR/Cas9 genetic engineering approach (*Lachancea kluyveri* and *Kluyveromyces lactis*)

The backbone of two vectors, gRNA expression cassette and pKOPIS, were synthesized by the DOE Joint Genome Institute DNA Synthesis Science Program. A pXIPHOS-panARS vector was subsequently constructed from pKOPIS by swapping the 2-micron origin with an autonomously replicating sequence (ARS) that is stable in the family Saccharomycetaceae, which was cloned

from the panARS vector pIL75 (Liachko and Dunham, 2014), and addition of an *E. coli* ampicillin resistance marker, and swapping the *kanMX* marker with the *natMX* marker. A single vector was cloned and assembled from the gRNA cassette and pXIPHOS-panARS and transformed into the strain of interest. This final vector contains Cas9, sgRNA, and the repair templates needed for the gene targeting event and was electroporated into the strain of interest. The yeast gRNA expression cassette contains *SNR52* promoter, HDV ribozyme, sgRNA, and *SNR52-1* terminator, and pKOPIS, contained the Cas9 protein modified to the yeast codon bias driven by the constitutive *RNR2* promoter with *kanMX* for the selectable marker. The gRNA expression cassette and repair template were amplified and cloned into NotI cut pXIPHOS-panARS using HiFi DNA Assembly kit (NEB). The electroporation protocol for *Lachancea kluyveri* and *Kluyveromyces lactis* followed previously described methods (Gojkovic et al., 2000, Kooistra et al., 2004).

The *SuvaPGM1* expression plasmids were constructed as follows: *S. uvarum PGM1* coding sequence together with its upstream and downstream 800 bp region were inserted into the panARS vector pIL75 (Liachko and Dunham, 2014) at the multiple cloning site digested by SmaI. This sequence were assembled using Gibson assembly (Gibson et al., 2009). GFP reporters were constructed as previously described (Kuang et al., 2016). The modified loci of all transformants or constructs were verified by Sanger sequencing.

Media and growth assays

Strains were inoculated from frozen glycerol stocks into either synthetic complete (SC) medium plus 0.2% glucose (1.72 g/L yeast nitrogen base without amino acids, 5 g/L ammonium sulfate, 2 g/L complete dropout mix, 2 g/L glucose), or YPD (10 g/L yeast extract, 20 g/L peptone, 20g/L glucose; YPD preculture was only used in Figure 3.1 and 3.2E) and cultured for2-3 days. The growth assays were conducted as previously described (Kuang et al., 2016). Briefly, the absorbance of each well was read by an unshaken BMG FLUOstar Omega plate reader every 10-20 minutes at 595 nm. "Relative growth" in each Figure denotes the number of cell divisions at 15-hour or indicated time point, which was calculated as $\log_2[(OD_{strain}\text{-}OD_{media})/(OD_{start}\text{-}OD_{media})$. This is an equation that normalized each optical density time point to its starting optical density and the optical density of the medium. In the case when division number was normalized to no carbon

source control, the division number was first calculated separately for the same strains cultured in media with or without carbon source in the same 96-well plate, and then the division number was calculated as Divsion$_{carbonsource}$ - Division$_{nocarbonsource}$. In each case, each species was cultured at its expected optimal growth temperature (room temperature or 30°C), except when species with different optimal temperature were cultured in the same 96-well plate. In these cases (Figure 3.1, Figure 3.3; Appendix B: Figure B.1, B.4B-C, B.5, B.6), strains were grown at room temperature (22-24°C).

Relaxed selection analysis

Nucleotide sequences of *GAL1/7/10* genes from every characterized *Lachancea* species (except *Lachancea waltii*, which has lost *GAL1/7/10* (Hittinger et al., 2004)) and three *Kluyveromyces* species (*Kluyveromyces lactis, Kluyveromyces marxianus*, *and Kluyveromyces dobzhanskii*) were used to obtain phylogeny-aware alignments with PRANK v150803 (Loytynoja, 2014) run in the codon mode. Codon alignments were then used to reconstruct maximum-likelihood (ML) phylogenies with RAxML v8.2.10 (Stamatakis, 2014), using the GTR model with evolutionary rate heterogeneity modeled by the gamma distribution, ML estimates of base frequencies, and 100 bootstrap pseudo-replicates. Tips and branches shared by species that showed poor growth on 2% galactose were marked as test branches. Finally, both the marked phylogenies and codon alignments were used together to run the RELAX module (Wertheim et al., 2015) implemented in the HYPHY package v2.220170606beta (Pond et al., 2005) to fit descriptive models and run the test for relaxed selection.

Statistical analysis

All *p*-values were calculated using Welch's two-sample *t*-test (Figure 3.1C; Appendix B: Figure B.1C, B.4B-C) or a conservative two-sided nonparametric test. Specifically, we used a Wilcoxon rank sum test that allows the rank data from multiple independent experiments to be pooled to account for day-to-day variation without making assumptions about distribution of the variance. These tests were performed using Mstat software version 6.1.4 (http://mcardle.oncology.wisc.edu/mstat/).

Chapter 4

Conclusions

This dissertation revealed two constraints that are general to sugar metabolism: glycolytic constraints and pleiotropic constraints at metabolic bottlenecks. These two constraints allow us to better construct a preliminary mapping between genotypes and phenotypes that are general for a variety of yeast species: any species with high activity at the bottleneck and a robust negative feedback loop in the *GAL* network is more likely to rapidly utilize galactose. Based on these two constraints, we predict that, if a species grows slowly in galactose, the addition of a Gal4-binding site upstream of *PGM1/2* and other ways to up-regulate this metabolic bottleneck allows an immediate gain in speed of growth in galactose if the upstream metabolic capacity is sufficient. However, due to the constraints of downstream glycolysis, its maximum activity is constrained by how robust its negative feedback loop is. Nonetheless, the hypothesis that robust negative feedback loops are required for a high capacity network is untested. More direct or sensitive approaches are required to directly test this hypothesis. A general scheme to confer vigorous growth in galactose in the family Saccharomycetaceae has emerged based on the studies from Chapters 2 and 3 in this dissertation, as well as previous studies: having additional galactose transporters, more active enzymes, a *GAL1* promoter with clusters of adjacent Gal4 binding sites, two co-repressors, and one or more Gal4-binding sites upstream of the bottleneck gene *PGM1/2*. In galactose, *Tetrapisispora blattae* grew the fastest among the 17 species I tested in the family Saccharomycetaceae, and has three predicted Gal4-binding sites upstream of *PGM1*. *Naumovozyma dairenensis* and *S. uvarum* are among the second and third fastest species, both of which have two predicted Gal4-binding sites upstream of either *PGM2* or *PGM1*. These three species and one more species that grew vigorously, *Naumovozyma castellii*, all have retained both co-repressors (*GAL80* and *GAL80B*). In

summary, due to the small number of genes involved, the wealth of prior knowledge base in the *GAL* network, the availability of genome sequences for dozens of yeast species, and the powerful genetic engineering techniques in non-model yeast species, this dissertation shows that the *GAL* network offers unparalleled power to dissect the evolutionary mechanisms at multiple levels in detail between genotypes and phenotypes. These results revealed the general genetic and molecular basis underlying galactose utilization trait divergence across a wide range of species, especially evolutionary constraints, and conferred predictive power to *GAL* network evolution.

Specifically, this dissertation applies comparative approaches to examine phenotypic trait divergence at multiple mechanistic levels and general constraints: 1) the functional dynamics of duplicate genes; 2) the constraints of functional divergence on duplicate genes contributing to the output of the GRN; 3) the constraints on GRN evolution with interacting GRNs; and 4) whether any recurring patterns across a broad evolutionary time scale exists. In Chapter 2, I revealed that ancient duplicate genes evolved quite dynamically. Such dynamics underlie the *GAL* network divergence between species. At the level of interaction between GRNs, I showed that high GRN activity likely requires robust negative feedback loops due to the constraints by downstream glycolysis. I further revealed that glycolytic capacity is likely a fundamental constraint to metabolism through glycolysis, and emphasized the under-appreciated role of the trehalose cycle in maintaining the metabolic balance of glycolysis in yeasts. Chapter 3 uncovered that the repeated modifications of the metabolic bottleneck gene *PGM1/2* underlie the quantitative tuning of *GAL* network activities in the family Saccharomycetaceae. The metabolic bottleneck of galactose metabolism is tuned through cis-regulatory elements, the Gal4-binding sites, in at least two distantly divergent yeast clades in this family. The number of predicted Gal4-binding sites upstream of *PGM1/2* is likely used to further tune the network activities in multiple species. The cis-regulatory evolution of *PGM1/2* likely reflects pleiotropic constraints. Since metabolic pathways are highly interconnected, pleiotropic constraints acting on enzymes at the intersecting nodes might be more prevalent than is currently realized.

4.1 The evolutionary dynamics of duplicate genes often underlie GRN evolution

The significance of duplicate genes in evolution is well established, but most studies have focused on the mechanisms of maintaining a duplicate gene or changes at particular time points. Spanning a 100-million-year time period, Chapter 2 explored how duplicate genes evolved and contributed to divergent GRN activities. After roughly eighty million years of redundancy preservation, substantial subfunctionalization or gene loss took place in one clade in the recent twenty million years, whereas functional redundancy was continuously maintained in another clade. Such functional divergence, together with other genetic changes, led to alterations in overall GRN activities. Since most yeast species that underwent a whole genome duplication (WGD-species) grow faster than non-WGD species, the whole genome duplication probably allows the WGD-species to evolve more rapid growth in galactose through two changes: the resolution of adaptive conflicts in the *GAL1* promoter (Hittinger and Carroll, 2007) and an additional co-repressor to provide robust negative feedback loops. Therefore, WGD resolves constraints in evolution to support a more active GRN.

The additional co-repressor encoded by *GAL80B* has been lost independently in multiple WGD-species. Species that lost *GAL80B* might therefore have a lower chance of evolving rapid galactose metabolism, partially constraining by the chance to re-acquire or recruit an additional co-repressor. To test this hypothesis, future studies will be required to examine two scenarios: Compared to species with one *GAL80/80B*, 1) whether species with both *GAL80* and *GAL80B* is more likely to have temporary growth arrest if deleting the two co-repressors; 2) if we artificially increase the *GAL* network activity, whether species with both *GAL80* and *GAL80B* allows higher *GAL* network activities without any metabolic or growth defects. Maintaining redundancy between duplicate genes might provide more opportunities to cope with novel ecological challenges. Once duplicate gene pairs lose redundancy or one of duplicate genes is lost, the otherwise higher tolerance to mutations will be lost as well. How functional redundancy can be maintained for such along time is not well understood, even though such redundancy has been well documented in numerous species (Dean et al., 2008, Qian et al., 2010, Gu et al., 2003, Vavouri et al., 2008). Since ancient duplicate

genes might still have more evolutionary potential than single-copy genes (Gu et al., 2004), the significance of duplicate gene divergence in GRN divergence and ecological adaptation might be even more prevalent and profound than has been appreciated thus far.

4.2 General constraints in the evolution of sugar metabolism

How GRNs co-evolve during evolution is not well characterized, but it might be part of the constraints in phenotypic evolution. Chapter 2 revealed that galactose metabolism is constrained by the downstream metabolic capacity of glycolysis. Removing two partially redundant negative feedback loops in *S. uvarum* overloads glycolysis and leads to temporary growth arrest and a series of metabolic defects in galactose utilization, such as the accumulation of reactive oxygen species. These defects can be partially rescued by decreasing network activity or decreasing the galactose concentrations. In multiple steps during glycolysis, phosphates are transferred to the glycolytic intermediates to form sugar phosphates, both in upper glycolysis (the investment phase) and lower glycolysis (the payoff phase) (van Heerden et al., 2014). When glycolysis is occurring rapidly, the intracellular phosphate pool is quickly depleted by upper glycolysis until there are not enough phosphates to support lower glycolysis. Importantly, this metabolic imbalance is alleviated through the trehalose cycle, which recycles phosphates (van Heerden et al., 2014). This will potentially constrain any GRNs regulating sugar metabolism that directly feeds into glycolysis. This constraint is not specific to microorganisms; in fact, evidence suggests that humans also require negative feedback loops to slow down glycolysis in muscles and livers (Teusinket al., 2010). In contrast, pancreatic beta cells seem to lack any negative feedback system, but they have low glucokinase activity to slow down glycolysis (Teusinket al., 2010). Importantly, overexpressing hexokinase to overload glycolysis in these human cells leads to similar metabolic defects (Iynedjian, 1998).

4.3 Evolution through cis-regulatory elements or protein coding regions

The significance of cis-regulatory elements or protein coding changes in different aspects of evolution has been debated intensely for at least the last decade (Stern and Orgogozo, 2008, Hoekstra and Coyne, 2007, Carroll, 2005). It is now well established that morphological evolution is mainly fulfilled by the alteration of cis-regulatory elements. The developmental toolkit is well conserved across the animal kingdom, from arthropods to chordates (Carroll, 2008). Since multiple genes from the developmental toolkit are re-used multiple times at different times or different places, any changes in the coding regions of the developmental toolkits are under pleiotropic constraints. Therefore, specifically modifying a demand at a particular time point or a particular position can only be accomplished through regulatory changes. Pleiotropy describes when a gene is multifunctional and affects multiple processes or multiple traits. Physiological and metabolic evolution seemingly involve more coding changes than cis-regulatory changes (Stern and Orgogozo, 2008). However, this tendency might be biased by the relative ease of detecting changes in coding sequences compared to cis-regulatory changes. Nonetheless, enzymes indeed frequently evolved through coding changes to modify their enzyme activities (Thomson et al., 2005, Voordeckers et al., 2012). Cis-regulatory evolution in metabolic pathways have been more and more frequently found (Ihmels et al., 2005, Lin et al., 2013, Roop et al., 2016). What aspects of metabolic evolution will more frequently involve cis-regulatory evolution? Chapter 3 showed that cis-regulatory evolution resolved pleiotropic constraints in a metabolic bottleneck. Phosphoglucomutase, encoded by *PGM1/2* in yeasts, lies at the bottleneck between the Leloir pathway and glycolysis, controlling the metabolic flux from the Leloir pathway into glycolysis. In addition, *PGM1/2* is pleiotropic and also required for glycogen biosynthesis (Cherry et al., 2012, Hirata et al., 2003), trehalose biosynthesis (Mulet et al., 2004, Cherry et al., 2012), and the pentose phosphate pathway (Cherry et al., 2012). Chapter 3 showed that this bottleneck has been repeatedly modified through cis-regulatory changes in the Gal4-binding sites to specifically tune the *GAL* network activities, implying constraints to modulate the bottleneck gene. Similar to developmental networks, metabolic pathways are highly interconnected. Also, many metabolic pathways share the same enzymes, such as glycolysis and

gluconeogenesis, TCA cycle, and glyoxylate cycle. Therefore, genes encoding enzymes can be pleiotropic if they function at pathway intersection or are shared by multiple metabolic pathways. Pleiotropic constraints might be more general than previously thought in metabolic evolution. I argue that using cis-regulatory elements to specifically tune a metabolic process might also be more common than previously thought.

4.4 To understand the mechanism is to predict

The fundamental task in science is to make predictions, and whether evolution is predictable is a fascinating topic to evolutionary biologists. Chapter 3 shows that the galactose growth in the family Saccharomycetaceae can be predicted merely based on the number of predicted Gal4-binding sites upstream of the bottleneck gene *PGM1/2*. This predictive power is based on mechanistic understandings of galactose metabolism, as well as the strong association across a broad phylogenetic distance. This tendency is a strong association general to 17 species, six genera in the family Saccharomycetaceae. This association even allows remarkable prediction in all 12 species in the genus of *Lachancea*. More significantly, unlike many QTL or genome-wide association studies that are based on statistics and a fragmented understanding of the molecular basis of certain traits, this association has a firm molecular and metabolic mechanistic basis with experimental support.

Since galactose is mainly metabolized through the Leloir pathway in a wide range of species, conserved from bacteria to human, and glycolytic constraints are probably also well conserved, what I learned in Chapters 2 and 3 might be general to species that metabolize galactose through the Leloir pathway. Therefore, I predict that a highly active phosphoglucomutase and a robust feedback loop are both required to rapidly metabolize galactose in a wide variety of species across life kingdoms. However, there are a number of exceptions to be considered. First, yeast species outside the family Saccharomycetaceae do not rely on Gal4 but other transcription factors, such as the Rtg1/3 in *Candida albicans* (Dalalet al., 2016). Second, not all species solely rely on galactose to switch on galactose metabolism: for instance, the signal required to turn on galactose utilization for *Schizosaccharomyces pombe* is unknown, and its subtelomeric *GAL* genes are normally silenced (Matsuzawaet al., 2011). Since galactose is seldom found as a free monosaccharide in

nature, polymers containing galactose moiety might be alternative signals to switch on the *GAL* network. Lastly, alternative pathways to catabolize galactose do exist, although they may play a minor role (Damerowet al., 2010, Fekete et al., 2004, Mojzita et al., 2012). In these above cases, the regulatory schemes might differ and thereby so might the evolutionary modes.

Chapter 3 in this dissertation has not addressed how often *PGM* relies on a Gal4-binding site to support vigorous growth and how often *PGM* relies on high enzyme activities or high basal expression. Is there any limitation in using one approach over the other? Is using either approach merely a historical contingency? The most vigorous growers in galactose all tend to have more than one Gal4-binding site upstream of *PGM1/2*, whereas the species that grew at a median rate, such as *S. kudriavzevii* and *K. lactis*, rely more on alternative approaches to Gal4 induction of *PGM1/2*. It is likely that these species are constrained by using alternative approaches to induce the bottleneck. Further understanding of such constraints might even allow us to predict galactose growth at a finer scale.

4.5 Evolutionary studies can fuel the study of metabolism and molecular mechanisms

The history of modern biology has repeatedly revealed that organism choices empower novel discoveries. Species that are particularly adept at a molecular function are more likely to lead to novel mechanistic discoveries in that area. *Caenorhabditis elegans* has an exquisite systemic RNAi pathway that likely protects it from transposons and viruses (Grishok, 2005, Sijen and Plasterk, 2003). The massive number of transposable elements present in maize genomes facilitated their discovery in this relatively challenging genetic system (Ravindran, 2012). The high number of micro-chromosomes in non-model organism *Tretrahymena* contributed to the discovery of telomeres (Blackburn et al., 2006). Because of the high network activities in *S. uvarum*, the defects of overloading glycolysis were pronounced and striking. Such defects are much milder in *S. cerevisiae* and required additional exacerbation treatments to amplify the signals so that it could be detected. Thus, novel mechanisms are often much easier to discover and molecularly dissect in species where their manifestations are exaggerated. The reasons stem in part from limitations of

experimental approaches, but also an observation bias that causes researchers to overlook inconspicuous findings. From the perspective of the necessity of adaptation and inevitability of divergence, many previously established paradigms are based on model organisms, but more exceptions are probably present in other non-model organisms. With the advance of genome sequencing techniques and the increasing usage of non-model organisms, we should be more open to investigations in a variety of organisms.

References

ACAR, M., BECSKEI, A. & VAN OUDENAARDEN, A. 2005. Enhancement of cellular memory by reducing stochastic transitions. *Nature,* 435, 228-32.

ALEXANDER, W. G., DOERING, D. T. & HITTINGER, C.T. 2014. High-Efficiency Genome Editing and Allele Replacement in Prototrophic and Wild Strains of Saccharomyces. *Genetics,* 198, 859-866.

ALON, U. 2007. Network motifs: theory and experimental approaches. *Nature Reviews Genetics,* 8, 450-461.

ARABIDOPSIS GENOME, I. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature,* 408, 796-815.

ARNOULT, L., SU, K. F. Y., MANOEL, D., MINERVINO, C., MAGRINA, J., GOMPEL, N. & PRUD'-HOMME, B. 2013. Emergence and Diversification of Fly Pigmentation Through Evolution of a Gene Regulatory Module. *Science,* 339, 1423-1426.

AURY, J. M., JAILLON, O., DURET, L., NOEL, B., JUBIN, C., PORCEL, B. M., SEGURENS, B., DAUBIN, V., ANTHOUARD, V., AIACH, N., ARNAIZ, O., BILLAUT, A., BEISSON, J., BLANC, I., BOUHOUCHE, K., CAMARA, F., DUHARCOURT, S., GUIGO, R., GOGENDEAU, D., KATINKA, M., KELLER, A. M., KISSMEHL, R., KLOTZ, C., KOLL, F., LE MOUEL, A., LEPERE, G., MALINSKY, S., NOWACKI, M., NOWAK, J.K., PLATTNER, H., POULAIN, J., RUIZ, F., SERRANO, V., ZAGULSKI, M., DESSEN, P., BETERMIER, M., WEISSENBACH, J., SCARPELLI, C., SCHACHTER, V., SPERLING, L., MEYER, E., COHEN, J. & WINCKER, P. 2006. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. *Nature,* 444, 171-8.

AVENDANO, M. S., LEIDY, C. & PEDRAZA, J. M. 2013. Tuning the range and stability of multiple phenotypic states with coupled positive-negative feedback loops. *Nature Communications,* 4.

BAKER, C. R., BOOTH, L. N., SORRELLS, T. R. & JOHNSON, A. D. 2012. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell,* 151, 80-95.

BAKER, C. R., TUCH, B. B. & JOHNSON, A. D. 2011. Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proceedings of the National Academy of Sciences of the United States of America,* 108, 7493-7498.

BENTON, M. J. 2010. Studying function and behavior in the fossil record. *PLoS Biology,* 8, e1000321.

BLACKBURN, E. H., GREIDER, C. W. & SZOSTAK, J. W. 2006. Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nature Medicine,* 12, 1133-8.

BLANC, G., BARAKAT, A., GUYOT, R., COOKE, R. & DELSENY, M. 2000. Extensive duplication and reshuffling in the Arabidopsisgenome. *Plant Cell,* 12, 1093-101.

BOLES, E., LIEBETRAU, W., HOFMANN, M. & ZIMMERMANN, F. K. 1994. A Family of Hexosephosphate Mutases in Saccharomyces-Cerevisiae. *European Journal of Biochemistry,* 220,83-96.

BRICKNER, J. H. 2010. Transcriptional memory: staying in the loop. *Current Biology,* 20, R20-1.

BRO, C., KNUDSEN, S., REGENBERG, B., OLSSON, L. & NIELSEN, J. 2005. Improvement of galactose uptake in Saccharomyces cerevisiae through overexpression of phosphoglucomutase: example of transcript analysis as a tool in inverse metabolic engineering. *Appl Environ Microbiol,* 71, 6465-72.

BROWN, S. J., HILGENFELD, R. B. & DENELL, R. E. 1994. The beetle Tribolium castaneum has a fushi tarazu homolog expressed in stripes during segmentation. *Proceedings of the National Academy of Sciences,* 91,12922-12926.

CARROLL, S. B. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature,* 376, 479-485.

CARROLL, S. B. 2005. Evolution at two levels: On genes and form. *Plos Biology,* 3, 1159-1166.

CARROLL, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell,* 134, 25-36.

CARROLL, S. B., GRENIER, J. K. & WEATHERBEE, S. D. 2005. *From DNA to diversity : molecular genetics and the evolution of animal design,* Malden, MA, Blackwell Pub.

CHALANCON, G., RAVARANI, C. N., BALAJI, S., MARTINEZ-ARIAS, A., ARAVIND, L., JOTHI, R. & BABU, M. M. 2012. Interplay between gene expression noise and regulatory network architecture. *Trends in genetics,* 28, 221-232.

CHERRY, J. M., HONG, E. L., AMUNDSEN, C., BALAKRISHNAN, R., BINKLEY, G., CHAN, E. T., CHRISTIE, K. R., COSTANZO, M. C., DWIGHT, S. S., ENGEL, S. R., FISK, D. G., HIRSCHMAN, J. E., HITZ, B. C., KARRA, K., KRIEGER, C. J., MIYASATO, S. R., NASH, R. S., PARK, J., SKRZYPEK, M. S., SIMISON, M., WENG, S. & WONG, E. D. 2012. Saccharomyces Genome Database:the genomics resource of budding yeast. *Nucleic Acids Research,* 40, D700-D705.

CHRISTIN, P. A., SALAMIN, N., SAVOLAINEN, V., DUVALL, M. R. & BESNARD, G. 2007. C4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology,* 17, 1241-7.

CHRISTIN, P. A., WEINREICH, D. M. & BESNARD, G. 2010. Causes and evolutionary significance of genetic convergence. *Trends in Genetics,* 26, 400-5.

CONANT, G. C. & WOLFE, K. H. 2006. Functional partitioning of yeast co-expression networks after genome duplication. *PLoS Biology,* 4, e109.

CONANT, G. C. & WOLFE, K. H. 2007. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Molecular systems biology,* 3, 129-129.

CONANT, G. C. & WOLFE, K. H. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nature reviews. Genetics,* 9,938.

DALAL, C. K., ZULETA, I. A., MITCHELL, K. F., ANDES, D. R., EL-SAMAD, H. & JOHNSON, A. D. 2016. Transcriptional rewiring over evolutionary timescales changes quantitative and qualitative properties of gene expression. *Elife,* 5.

DAMEN, W. G. 2002. fushi tarazu: a Hox gene changesits role. *Bioessays,* 24, 992-5.

DAMEROW, S., LAMERZ, A. C., HASELHORST, T., FUHRING, J., ZARNOVICAN, P., VON ITZSTEIN, M. & ROUTIER, F. H. 2010. Leishmania UDP-sugar pyrophosphorylase: the missing link in galactose salvage? *J Biol Chem,* 285, 878-87.

DARWIN, C. 1859. *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life,* London, J. Murray.

DE BODT, S., MAERE, S. & VAN DE PEER, Y. 2005. Genome duplication and the origin of angiosperms. *Trends Ecol Evol,* 20,591-7.

DEAN, E. J., DAVIS, J. C., DAVIS, R. W. & PETROV, D. A. 2008. Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genetics,* 4, e1000113.

DEHAL, P. & BOORE, J. L. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology,* 3, e314.

DES MARAIS, D. L. & RAUSHER, M. D. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature,* 454, 762-5.

DOBLER, S., DALLA, S., WAGSCHAL, V. & AGRAWAL, A.A. 2012. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. *Proceedings of the National Academy of Sciences of the United States of America,* 109,13040-5.

DOUGLAS, H. C. & HAWTHORNE, D. C. 1966. Regulation of genes controlling synthesis of the galactose pathway enzymes in yeast. *Genetics,* 54, 911-6.

EDWARDS, M., BIGHAM, A., TAN, J., LI, S., GOZDZIK, A., ROSS, K., JIN, L. & PARRA, E. J. 2010. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. *PLoS Genetics,* 6, e1000867.

EIZIRIK, E., YUHKI, N., JOHNSON, W. E., MENOTTI-RAYMOND, M., HANNAH, S. S. & O'BRIEN, S. J. 2003. Molecular genetics and evolution of melanism in the cat family. *Current Biology,* 13, 448-53.

FALCIANI, F., HAUSDORF, B., SCHRODER, R., AKAM, M., TAUTZ, D., DENELL, R. & BROWN, S. 1996. Class 3 Hox genes in insects and the origin of zen. *Proceedings of the National Academy of Sciences of the United States of America,* 93, 8479-84.

FARHAT, M. R., SHAPIRO, B. J., KIESER, K. J., SULTANA, R., JACOBSON, K. R., VICTOR, T. C., WARREN, R. M., STREICHER, E. M., CALVER, A., SLOUTSKY, A., KAUR, D., POSEY, J. E., PLIKAYTIS, B., OGGIONI, M. R., GARDY, J. L., JOHNSTON, J. C., RODRIGUES, M., TANG, P. K., KATO-MAEDA, M., BOROWSKY, M. L., MUDDUKRISHNA, B., KREISWIRTH, B. N., KUREPINA, N., GALAGAN, J., GAGNEUX, S., BIRREN, B., RUBIN, E. J., LANDER, E. S., SABETI, P. C. & MURRAY, M. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nature Genetics,* 45, 1183-9.

FEKETE, E., KARAFFA, L., SANDOR, E., BANYAI, I., SEIBOTH, B., GYEMANT, G., SEPSI, A., SZENTIRMAI, A. & KUBICEK, C. P. 2004. The alternative D-galactose degrading pathway of Aspergillus nidulans proceeds via L-sorbose. *Arch Microbiol,* 181, 35-44.

FORCE, A., LYNCH, M., PICKETT, F. B., AMORES, A., YAN, Y. L. & POSTLETHWAIT, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics,* 151, 1531-45.

FREEL, K. C., FRIEDRICH, A., SARILAR, V., DEVILLERS, H., NEUVEGLISE, C. & SCHACHERER, J. 2016. Whole-Genome Sequencing and Intraspecific Analysis of the Yeast Species Lachancea quebecensis. *Genome Biology and Evolution,* 8, 733-741.

FUKUSHIMA, K., FANG, X., ALVAREZ-PONCE, D., CAI, H., CARRETERO-PAULET, L., CHEN, C., CHANG, T.-H., FARR, K. M., FUJITA, T. &HIWATASHI, Y. 2017. Genome of the pitcher plant Cephalotus reveals genetic changes associated with carnivory. *Nature Ecology & Evolution,* 1,0059.

GALANT, R. & CARROLL, S. B. 2002. Evolution of atranscriptional repression domain in an insect Hox protein. *Nature,* 415, 910-913.

GARCIA SANCHEZ, R., HAHN-HAGERDAL, B. &GORWA-GRAUSLUND, M. F. 2010. PGM2 overexpression improves anaerobic galactose fermentation in Saccharomyces cerevisiae. *Microb Cell Fact,* 9, 40.

GASCH, A. P., MOSES, A. M., CHIANG, D. Y., FRASER, H.B., BERARDINI, M. & EISEN, M. B. 2004. Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biologyogy,* 2, e398.

GIBSON, D. G., YOUNG, L., CHUANG, R. Y., VENTER, J.C., HUTCHISON, C. A. & SMITH, H. O. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods,* 6, 343-U41.

GILAD, Y., BUSTAMANTE, C. D., LANCET, D. & PAABO, S. 2003. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am J Hum Genet,* 73, 489-501.

GOJKOVIC, Z., JAHNKE, K., SCHNACKERZ, K. D. & PISKUR, J. 2000. PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *Journal of Molecular Biology,* 295, 1073-1087.

GOULD, S. J. 1989. *Wonderful life : the Burgess Shale and the nature of history,* New York, W.W. Norton.

GRISHOK, A. 2005. RNAi mechanisms in Caenorhabditis elegans. *FEBS Lett,* 579, 5932-9.

GU, Z., RIFKIN, S. A., WHITE, K. P. & LI, W.-H. 2004. Duplicate genes increase gene expression diversity within and between species. *Nature genetics,* 36, 577.

GU, Z., STEINMETZ, L. M., GU, X., SCHARFE, C., DAVIS, R. W. & LI, W. H. 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature,* 421, 63-6.

HE, X. & ZHANG, J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalizationin duplicate gene evolution. *Genetics,* 169, 1157-64.

HIRATA, Y., ANDOH, T., ASAHARA, T. & KIKUCHI, A. 2003. Yeast glycogen synthase kinase-3 activates Msn2p-dependent transcription of stress responsive genes. *Molecular Biology of the Cell,* 14,302-312.

HITTINGER, C. T. & CARROLL, S. B. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature,* 449, 677-81.

HITTINGER, C. T., GONCALVES, P., SAMPAIO, J. P., DOVER, J., JOHNSTON, M. & ROKAS, A. 2010. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature,* 464, 54-8.

HITTINGER, C. T., ROKAS, A. & CARROLL, S. B. 2004. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 14144-9.

HOEKSTRA, H. E. & COYNE, J. A. 2007. The locus of evolution: Evo devo and the genetics of adaptation. *Evolution,* 61, 995-1016.

HOEKSTRA, H. E., HIRSCHMANN, R. J., BUNDEY, R. A., INSEL, P. A. & CROSSLAND, J. P. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science,* 313, 101-4.

HOEKSTRA, H. E. & NACHMAN, M. W. 2003. Different genes underlie adaptive melanism in different populations of rock pocket mice. *Mol Ecol,* 12, 1185-94.

HONG, K. K., VONGSANGNAK, W., VEMURI, G. N. & NIELSEN, J. 2011. Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. *Proceedings of the National Academy of Sciences of the United States of America,* 108, 12179-84.

HU, Y., WU, Q., MA, S., MA, T., SHAN, L., WANG, X., NIE, Y., NING, Z., YAN, L. & XIU, Y. 2017. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proceedings of the National Academy ofSciences,* 114, 1081-1086.

IHMELS, J., BERGMANN, S., GERAMI-NEJAD, M., YANAI, I., MCCLELLAN, M., BERMAN, J. & BARKAI, N. 2005. Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science,* 309, 938-40.

INNAN, H. & KONDRASHOV, F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nature reviews. Genetics,* 11, 97.

IYNEDJIAN, P. B. 1998. Glycolysis, turbo design and the endocrine pancreatic beta cell. *Trends Biochem Sci,* 23, 467-8.

JAILLON, O., AURY, J. M., BRUNET, F., PETIT, J. L., STANGE-THOMANN, N., MAUCELI, E., BOUNEAU, L., FISCHER, C., OZOUF-COSTAZ, C., BERNOT, A., NICAUD, S., JAFFE, D., FISHER, S., LUTFALLA, G., DOSSAT, C., SEGURENS, B., DASILVA, C., SALANOUBAT, M., LEVY, M., BOUDET, N., CASTELLANO, S., ANTHOUARD, V., JUBIN, C., CASTELLI, V., KATINKA, M., VACHERIE, B., BIEMONT, C., SKALLI, Z., CATTOLICO, L., POULAIN, J., DE BERARDINIS, V., CRUAUD, C., DUPRAT, S., BROTTIER, P., COUTANCEAU, J. P., GOUZY, J., PARRA, G., LARDIER, G., CHAPPLE, C., MCKERNAN, K. J., MCEWAN, P.,

BOSAK, S., KELLIS, M., VOLFF, J. N., GUIGO, R., ZODY, M. C., MESIROV, J., LINDBLAD-TOH, K., BIRREN, B., NUSBAUM, C., KAHN, D., ROBINSON-RECHAVI, M., LAUDET, V., SCHACHTER, V., QUETIER, F., SAURIN, W., SCARPELLI, C., WINCKER, P., LANDER, E. S., WEISSENBACH, J. & ROESTCROLLIUS, H. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature,* 431, 946-57.

JIANG, H., GUAN, W., PINNEY, D., WANG, W. & GU, Z. 2008. Relaxation of yeast mitochondrial functions after whole-genome duplication. *Genome research,* 18, 1466-71.

JOHNSTON, M. 1987. A model fungal gene regulatorymechanism: the GAL genes of Saccharomyces cerevisiae. *Microbiol Rev,* 51,458-76.

JOSHI, J. G. & HANDLER, P. 1969. Phosphoglucomutase VI. Purification and Properties of Phosphoglucomutases from Human Muscle. *Journal of Biological Chemistry,* 244, 3343-&.

KALIR, S., MANGAN, S. & ALON, U. 2005. A coherent feedforward loop with a SUM input function prolongs flagella expression in Escherichia coli. *Molecular systems biology,* 1.

KAYIKCI, O. & NIELSEN, J. 2015. Glucose repression in Saccharomyces cerevisiae. *FEMS YeastRes,* 15.

KEARNS, J. D., BASAK, S., WERNER, S. L., HUANG, C. S.& HOFFMANN, A. 2006. IkappaBepsilon provides negative feedback to control NF-kappaB oscillations, signaling dynamics, and inflammatory gene expression. *J Cell Biol,* 173, 659-64.

KELLEY, N. P. & PYENSON, N. D. 2015. Vertebrate evolution. Evolutionary innovation and ecology in marine tetrapods from the Triassic to the Anthropocene. *Science,* 348, aaa3716.

KOOISTRA, R., HOOYKAAS, P. J. J. & STEENSMA, H. Y. 2004. Efficient gene targeting in Kluyveromyces lactis. *Yeast,* 21, 781-792.

KUANG, M. C., HUTCHINS, P. D., RUSSELL, J. D., COON, J. J. & HITTINGER, C. T. 2016. Ongoing resolution of duplicate gene functions shapes the diversification of a metabolic network. *Elife,* 5.

KUNDU, S., HORN, P. J. & PETERSON, C. L. 2007. SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. *Genes & development,* 21, 997-1004.

KURTZMAN, C. P., FELL, J. W. & BOEKHOUT, T. 2011. The Yeasts, a Taxonomic Study Volume 1 Fifth Edition Preface. *Yeasts: A Taxonomic Study, Vols 1-3, 5th Edition*, Xiii-+.

KUTTYKRISHNAN, S., SABINA, J., LANGTON, L. L., JOHNSTON, M. & BRENT, M. R. 2010. A quantitative model of glucose signaling in yeast reveals an incoherent feed forward loop leading to a specific,transient pulse of transcription. *Proceedings of the National Academy of Sciences of the United States of America,* 107, 16743-16748.

LIACHKO, I. & DUNHAM, M. J. 2014. An autonomously replicating sequence for use in a wide range of budding yeasts. *Fems Yeast Research,* 14, 364-367.

LIN, Z., WANG, T. Y., TSAI, B. S., WU, F. T., YU, F.J., TSENG, Y. J., SUNG, H. M. & LI, W. H. 2013. Identifying cis-regulatory changes involved in the evolution of aerobic fermentation in yeasts. *Genome biology and evolution,* 5, 1065-78.

LOHR, U., YUSSA, M. & PICK, L. 2001. Drosophilafushi tarazu. a gene on the border of homeotic function. *Current Biology,* 11, 1403-12.

LOYTYNOJA, A. 2014. Phylogeny-aware alignment with PRANK. *Multiple Sequence Alignment Methods,* 1079, 155-170.

LU, M. & KLECKNER, N. 1994. Molecular-Cloning and Characterization of the Pgm Gene Encoding Phosphoglucomutase of Escherichia-Coli. *Journal of Bacteriology,* 176, 5847-5851.

MACNEIL, L. T. & WALHOUT, A. J. 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res,* 21, 645-57.

MANGAN, S., ITZKOVITZ, S., ZASLAVER, A. & ALON, U. 2006. The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli. *J Mol Biol,* 356, 1073-81.

MANGAN, S., ZASLAVER, A. & ALON, U. 2003. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of molecular biology,* 334, 197-204.

MARCET-HOUBEN, M. & GABALDON, T. 2015. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLoS Biology,* 13, e1002220.

MARTCHENKO, M., LEVITIN, A., HOGUES, H., NANTEL, A.& WHITEWAY, M. 2007. Transcriptional rewiring of fungal galactose-metabolism circuitry. *Current Biology,* 17, 1007-13.

MARTIN, A. & ORGOGOZO, V. 2013. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution,* 67, 1235-50.

MATSUZAWA, T., FUJITA, Y., TANAKA, N., TOHDA, H., ITADANI, A. & TAKEGAWA, K. 2011. New insights into galactose metabolism by Schizosaccharomyces pombe: isolation and characterization of a galactose-assimilating mutant. *J Biosci Bioeng,* 111, 158-66.

MCGRATH, C. L., GOUT, J. F., JOHRI, P., DOAK, T. G.& LYNCH, M. 2014. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res,* 24, 1665-75.

MCKEOWN, A. N., BRIDGHAM, J. T., ANDERSON, D. W., MURPHY, M. N., ORTLUND, E. A. & THORNTON, J. W. 2014. Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell,* 159, 58-68.

MOJZITA, D., HEROLD, S., METZ, B., SEIBOTH, B. & RICHARD, P. 2012. L-xylo-3-hexulose reductase is the missing link in the oxidoreductive pathway for D-galactose catabolism in filamentous fungi. *Journal of Biological Chemistry,* 287, 26010-26018.

MOORE, S. A., GARCIA, C. V., GARDNER, B. T. & LESTER, G. A. 1988. Automated spectrophotometric assay for cell division regulation in yeast. *Exp Cell Res,* 177, 399-413.

MULET, J. M., ALEJANDRO, S., ROMERO, C. & SERRANO, R. 2004. The trehalose pathway and intracellular glucose phosphates as modulators of potassium transport and general cation homeostasis in yeast. *Yeast,* 21, 569-582.

NACHMAN, M. W., HOEKSTRA, H. E. & D'AGOSTINO, S.L. 2003. The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America,* 100, 5268-73.

NELSON, D. E., IHEKWABA, A. E., ELLIOTT, M., JOHNSON, J. R., GIBNEY, C. A., FOREMAN, B. E., NELSON, G., SEE, V., HORTON, C. A., SPILLER, D. G., EDWARDS, S. W., MCDOWELL, H. P., UNITT, J. F., SULLIVAN, E., GRIMLEY, R., BENSON, N., BROOMHEAD, D., KELL, D. B. & WHITE, M. R. 2004. Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science,* 306, 704-8.

OH, D. & HOPPER, J. E. 1990. Transcription of a Yeast Phosphoglucomutase Isozyme Gene Is Galactose Inducible and Glucose Repressible. *Molecular and Cellular Biology,* 10, 1415-1422.

OHNO, S. 1970. *Evolution by gene duplication,* Berlin ; New York, Springer-Verlag.

PARKER, J., TSAGKOGEORGA, G., COTTON, J. A., LIU, Y., PROVERO, P., STUPKA, E. & ROSSITER, S. J. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. *Nature,* 502, 228-31.

PENG, W., LIU, P., XUE, Y. & ACAR, M. 2015. Evolution of gene network activity by tuning the strength of negative-feedback regulation. *Nature Communications,* 6, 6226.

POND, S. L. K., FROST, S. D. W. & MUSE, S. V. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics,* 21,676-679.

POSTLETHWAIT, J., AMORES, A., CRESKO, W., SINGER, A.& YAN, Y. L. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet,* 20, 481-90.

POUGACH, K., VOET, A., KONDRASHOV, F. A., VOORDECKERS, K., CHRISTIAENS, J. F., BAYING, B., BENES, V., SAKAI, R., AERTS, J., ZHU, B., VAN DIJCK, P. & VERSTREPEN, K. J. 2014. Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nature Communications,* 5, 4868.

QIAN, W., LIAO, B.-Y., CHANG, A. Y.-F. & ZHANG, J. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in Genetics,* 26, 425-430.

RAVINDRAN, S. 2012. Barbara McClintock and the discovery of jumping genes. *Proceedings of the National Academy of Sciences of the United States of America,* 109, 20198-9.

RIEDER, S., TAOURIT, S., MARIAT, D., LANGLOIS, B.& GURIN, G. 2001. Mutations in the agouti (ASIP), the extension (MC1R),and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). *Mammalian genome,* 12, 450-455.

RILEY, R., HARIDAS, S., WOLFE, K. H., LOPES, M. R., HITTINGER, C. T., GOKER, M., SALAMOV, A. A., WISECAVER, J. H., LONG, T. M., CALVEY, C. H., AERTS, A. L., BARRY, K. W., CHOI, C., CLUM, A., COUGHLAN, A. Y., DESHPANDE, S., DOUGLASS, A. P., HANSON, S. J., KLENK, H. P., LABUTTI, K. M., LAPIDUS, A., LINDQUIST, E. A., LIPZEN, A. M., MEIER-KOLTHOFF, J. P., OHM, R.A., OTILLAR, R. P., PANGILINAN, J. L., PENG, Y., ROKAS, A.,

ROSA, C. A., SCHEUNER, C., SIBIRNY, A. A., SLOT, J. C., STIELOW, J. B., SUN, H., KURTZMAN, C. P., BLACKWELL, M., GRIGORIEV, I. V. & JEFFRIES, T. W. 2016. Comparative genomics of biotechnologically important yeasts. *Proceedings of the National Academy of Sciences of the United States of America,* 113,9882-7.

ROOP, J. I., CHANG, K. C. & BREM, R. B. 2016. Polygenic evolution of a sugar specialization trade-off in yeast. *Nature,* 530, 336-9.

RUBIO-TEXEIRA, M. 2005. A comparative analysis of the GAL genetic switch between not-so-distant cousins: Saccharomyces cerevisiae versus Kluyveromyces lactis. *FEMS Yeast Research,* 5, 1115-1128.

SAGE, R. F. 2004. The evolution of C4 photosynthesis. *New phytologist,* 161, 341-370.

SARILAR, V., DEVILLERS, H., FREEL, K. C., SCHACHERER, J. & NEUVEGLISE, C. 2015. Draft Genome Sequence of Lachancea lanzarotensis CBS 12615T, an Ascomycetous Yeast Isolated from Grapes. *Genome Announc,* 3.

SAYOU, C., MONNIAUX, M., NANAO, M. H., MOYROUD, E., BROCKINGTON, S. F., THEVENON, E., CHAHTANE, H., WARTHMANN, N., MELKONIAN, M., ZHANG, Y., WONG, G. K., WEIGEL, D., PARCY, F. & DUMAS, R. 2014. A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science,* 343, 645-8.

SCANNELL, D. R., FRANK, A. C., CONANT, G. C., BYRNE, K. P., WOOLFIT, M. & WOLFE, K. H. 2007. Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proceedings of the National Academy of Sciences of the United States of America,* 104, 8397-402.

SCHNEIDER, A., HENEGAR, C., DAY, K., ABSHER, D., NAPOLITANO, C., SILVEIRA, L., DAVID, V. A., O'BRIEN, S. J., MENOTTI-RAYMOND, M., BARSH, G. S. & EIZIRIK, E. 2015. Recurrent evolution of melanism in South American felids. *PLoS Genetics,* 11, e1004892.

SIJEN, T. & PLASTERK, R. H. 2003. Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature,* 426, 310-4.

SLOT, J. C. & ROKAS, A. 2010. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proceedings of the National Academy of Sciences of the United States of America,* 107, 10136-10141.

SORRELLS, T. R., BOOTH, L. N., TUCH, B. B. & JOHNSON, A. D. 2015. Intersecting transcription networks constrain gene regulatory evolution. *Nature,* 523, 361-+.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30, 1312-1313.

STAUBER, M., PRELL, A. & SCHMIDT-OTT, U. 2002. A single Hox3 gene with composite bicoid and zerknullt expression characteristics in non-Cyclorrhaphan flies. *Proceedings of the National Academy of Sciences of the United States of America,* 99, 274-9.

STERN, D. L. 2013. The genetic causes of convergent evolution. *Nature Review Genetics,* 14, 751-64.

STERN, D. L. & ORGOGOZO, V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution,* 62, 2155-77.

STUART, J. J., BROWN, S. J., BEEMAN, R. W. & DENELL, R. E. 1991. A deficiency of the homeotic complex of the beetle Tribolium.*Nature,* 350, 72-4.

SUCENA, E., DELON, I., JONES, I., PAYRE, F. & STERN, D. L. 2003. Regulatory evolution of shavenbaby/ovo underlies multiplecases of morphological parallelism. *Nature,* 424, 935-8.

TEUSINK, B., WESTERHOFF, H. V. & BRUGGEMAN, F. J. 2010. Comparative systems biology: from bacteria to man. *Wiley Interdiscip Rev Syst Biol Med,* 2, 518-32.

THERON, E., HAWKINS, K., BERMINGHAM, E., RICKLEFS, R.E. & MUNDY, N. I. 2001. The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, Coereba flaveola. *Current Biology,* 11, 550-557.

THOMPSON, D. A., ROY, S., CHAN, M., STYCZYNSKY, M. P., PFIFFNER, J., FRENCH, C., SOCHA, A., THIELKE, A., NAPOLITANO, S., MULLER, P., KELLIS, M., KONIECZKA, J. H., WAPINSKI, I. & REGEV, A. 2013. Evolutionary principles of modular gene regulation in yeasts. *eLife,* 2, e00603.

THOMSON, J. M., GAUCHER, E. A., BURGAN, M. F., DE KEE, D. W., LI, T., ARIS, J. P. & BENNER, S. A. 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. *Nature Genetics,* 37, 630-5.

TSONG, A. E., TUCH, B. B., LI, H. & JOHNSON, A. D. 2006. Evolution of alternative transcriptional circuits with identical logic. *Nature,* 443, 415-20.

UJVARI, B., CASEWELL, N. R., SUNAGAR, K., ARBUCKLE, K., WUSTER, W., LO, N., O'MEALLY, D., BECKMANN, C., KING, G. F., DEPLAZES, E.& MADSEN, T. 2015. Widespread convergence in toxin resistance by predictable molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America,* 112, 11911-6.

VAKIRLIS, N., SARILAR, V., DRILLON, G., FLEISS, A., AGIER, N., MEYNIEL, J. P., BLANPAIN, L., CARBONE, A., DEVILLERS, H., DUBOIS, K., GILLET-MARKOWSKA, A., GRAZIANI, S., HUU-VANG, N., POIREL, M., REISSER, C., SCHOTT, J., SCHACHERER, J., LAFONTAINE, I., LLORENTE, B., NEUVEGLISE, C. & FISCHER, G. 2016. Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Res,* 26, 918-32.

VAN HEERDEN, J. H., WORTEL, M. T., BRUGGEMAN, F. J., HEIJNEN, J. J., BOLLEN, Y. J., PLANQUE, R., HULSHOF, J., O'TOOLE, T. G., WAHL, S. A. & TEUSINK, B. 2014. Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science,* 343, 1245114.

VAVOURI, T., SEMPLE, J. I. & LEHNER, B. 2008. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends Genet,* 24, 485-8.

VENTURELLI, O. S., EL-SAMAD, H. & MURRAY, R. M. 2012. Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. *Proceedings of the National Academy of Sciences of the United States of America,* 109, E3324-33.

VOORDECKERS, K., BROWN, C. A., VANNESTE, K., VAN DERZANDE, E., VOET, A., MAERE, S. & VERSTREPEN, K. J. 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biology,* 10, e1001446.

WAKE, D. B. 1991. Homoplasy: the result of natural selection, or evidence of design limitations? *The American Naturalist,* 138,543-567.

WALSH, J. B. 1995. How often do duplicated genes evolve new functions? *Genetics,* 139, 421-428.

WERTHEIM, J. O., MURRELL, B., SMITH, M. D., POND, S.L. K. & SCHEFFLER, K. 2015. RELAX: Detecting Relaxed Selection in aPhylogenetic Framework. *Molecular Biologyand Evolution,* 32, 820-832.

WISTOW, G. 1993. Lens crystallins: gene recruitment and evolutionary dynamism. *Trends Biochem Sci,* 18, 301-6.

WOLFE, K. H., ARMISEN, D., PROUX-WERA, E., OHEIGEARTAIGH, S. S., AZAM, H., GORDON, J. L. & BYRNE, K. P. 2015. Clade-and species-specific features of genome evolution in the Saccharomycetaceae. *Fems Yeast Research,* 15.

WOLFE, K. H. & SHIELDS, D. C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature,* 387, 708-13.

WONG, A., RODRIGUE, N. & KASSEN, R. 2012. Genomics of adaptation during experimental evolution of the opportunistic pathogen Pseudomonas aeruginosa. *PLoS Genetics,* 8, e1002928.

WOODS, R., SCHNEIDER, D., WINKWORTH, C. L., RILEY, M.A. & LENSKI, R. E. 2006. Tests of parallel molecular evolution in along-term experiment with Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America,* 103, 9107-9112.

YEAMAN, S., HODGINS, K. A., LOTTERHOS, K. E., SUREN, H., NADEAU, S., DEGNER, J. C., NURKOWSKI, K. A., SMETS, P., WANG, T., GRAY, L.K., LIEPE, K. J., HAMANN, A., HOLLIDAY, J. A., WHITLOCK, M. C., RIESEBERG, L.H. & AITKEN, S. N. 2016. Convergent local adaptation to climate indistantly related conifers. *Science,* 353, 1431-1433.

ZHEN, Y., AARDEMA, M. L., MEDINA, E. M., SCHUMER, M.& ANDOLFATTO, P. 2012. Parallel Molecular Evolution in an Herbivore Community. *Science,* 337, 1634-1637.

Appendix A: Supplementary Materials for Chapter 2

Table A.1: Amino acid identity and *GAL* gene composition between *S. uvarum* and *S. cerevisiae* *GAL* network.

S. uvarum	Identity to *S. cerevisiae* ortholog	Identity to other *S. cerevisiae* homolog	Identity to *S. uvarum* paralog	Functions of its *S. cerevisiae* ortholog
Gal1	88.6%	75.5%	78.7%	Galactokinase
Gal2	89.9%	--	92.1%	Galactose transporter
Gal2b	86.6%	--	92.1%	Galactose transporter
Gal3	83.3%	76.6%	78.7%	Co-inducer
Gal4	67.7%	--	--	Transcription factor
Gal7	83.9%	--	--	Galactose-1-phosphate uridyl transferase
Gal10	87.4%	--	--	UDP-glucose-4-epimerase
Gal80	93.6%	--	72.1%	Co-repressor
Gal80b	--	71.9%	72.1%	--

Table A.2: *S. uvarum* and *S. cerevisiae* have qualitatively similar *gal1* null phenotypes. "+" indicates growth and "-" indicates no growth. The criteria of growth and no growth are stated in the method section of Chapter 2.

	MM +2% glucose	MM +2% galactose
Scer WT	+	+
Scer gal1	+	-
*Scer gal1*Δ*::SuvaGAL3*	+	+
Suva WT	+	+
*Suva gal1*Δ	+	-

Table A.3: Metabolite ratio between *S. uvarum gal80*Δ *gal80b*Δ and wild-type in SC +2% galactose over time.

Metabolite	Mutant/WT				
	0.5 h	1 h	3 h	5 h	16 h
Gluconic Acid	0.13	0.10	0.25	0.54	2.26
Pyruvic Acid	0.69	NS	NS	NS	2.93
Succinic Acid	NS	NS	NS	0.59	0.38
Malic Acid	0.20	0.43	1.71	0.57	0.56
α-Ketoglutaric Acid	2.64	7.99	NS	NS	NS
Fumaric Acid	0.18	0.48	2.59	0.68	0.46
6-Phospho Gluconic Acid	NS	0.75	NS	NS	NS
3-Deoxy-2-keto-5-Phosphogluconic acid	NS	NS	NS	NS	NS
Sum of 2-Phosphoglyceric acid and 3-Phosphoglyceric acid	2.75	5.88	2.96	1.64	1.82
Citric Acid	NS	2.22	2.83	NS	NS
Phospho(enol)pyruvic acid	0.36	4.99	NS	2.51	2.28
Galactose-1-Phosphate	2.20	2.16	NS	NS	NS
Sum of Glucose-1-Phosphate+Mannose-1-Phosphate	2.24	1.27	na	na	NS
Trehalose-6-Phosphate*	>200	>231	>79	NS	6.43
Sum of Glycerol-2-Phosphate and Glycerol-3-Phosphate	224.79	15.74	6.80	3.15	NS
Aconitic Acid	na	NS	NS	NS	NS
Glucose-6-Phosphate	5.08	1.47	NS	NS	1.50
Fructose-6-Phosphate	NS	NS	NS	NS	NS
Mannose-6-Phosphate	NS	NS	NS	NS	NS
Seduhepulose-7-Phosphate	NS	NS	NS	NS	1.86
Fructose-1,6-bisphosphate	12.58	2.11	NS	4.59	0.60
AMP	0.81	1.21	0.77	NS	NS
ADP	0.21	0.26	NS	NS	0.64
ATP	0.20	NS	NS	2.93	NS
GMP	0.31	0.35	0.39	0.08	NS
GDP	NS	NS	2.01	NS	0.62

Note: 1) "NS" indicates that one-sided t-test does not support significant differences between mutant and wild-type.
2) "na" indicates at lease one sample is below limit of quantification by mass spectrometry so the ratio is not reported.
3) *: The fold-change at first three points of trehalose-6-phosphate is based on minimum fold-change estimation. The levels of trehalose-6-phosphate in *S. uvarum* wild-type are below limit of quantification, so we estimated the minimum fold changes based on the quantification limit normalized to the cell counts.

Table A.4: Normalized metabolite concentrations in *S. uvarum* wild-type and *gal80*Δ *gal80b*Δ in galactose over time.

Sample Name	Strain Number	Genotype	Time Point (hr)	Normalized Sample Extract Concentration (μM/CellCount)									
				Gluconic Acid	Pyruvic Acid	Succinic Acid	Malic Acid	α-Ketoglutaric Acid	Fumaric Acid	6-Phospho Gluconic Acid	3-Deoxy-2-keto-5-Phosphogluconic acid	Sum of 2-Phosphoglyceric acid and 3-Phosphoglyceric acid	Citric Acid
65-1 0.5h	yHMK 65	*Suva* WT	0.5	1.7E-09	4.8E-09	4.0E-08	2.5E-08	2.1E-09	3.5E-09	1.0E-10	8.4E-12	4.4E-10	4.7E-09
65-1 1h	yHMK 65	*Suva* WT	1	2.8E-09	3.3E-09	5.8E-08	2.2E-08	2.3E-09	2.0E-09	2.3E-10	3.5E-12	4.1E-10	4.3E-09
65-1 3h	yHMK 65	*Suva* WT	3	2.6E-09	3.9E-09	6.5E-08	8.3E-09	2.0E-09	7.7E-10	2.9E-10	4.9E-12	7.8E-10	9.6E-09
65-1 5h	yHMK 65	*Suva* WT	5	9.1E-10	4.7E-09	5.1E-08	1.1E-08	1.5E-09	1.1E-09	1.6E-10	6.2E-12	7.6E-10	1.6E-08
65-1 16h	yHMK 65	*Suva* WT	16	2.0E-10	3.4E-09	1.0E-07	8.1E-09	5.5E-10	1.5E-09	8.4E-11	1.2E-12	1.7E-10	4.2E-09
65-2 0.5h	yHMK 65	*Suva* WT	0.5	1.5E-09	5.1E-09	4.4E-08	2.4E-08	1.8E-09	1.9E-09	1.1E-10	1.6E-11	2.6E-10	3.4E-09
65-2 1h	yHMK 65	*Suva* WT	1	2.9E-09	4.0E-09	6.3E-08	2.0E-08	2.1E-09	2.0E-09	2.4E-10	4.3E-11	4.7E-10	6.0E-09
65-2 3h	yHMK 65	*Suva* WT	3	1.7E-09	3.0E-09	5.8E-08	7.9E-09	1.8E-09	5.6E-10	2.5E-10	6.6E-12	4.4E-10	8.0E-09
65-2 5h	yHMK 65	*Suva* WT	5	8.4E-10	4.7E-09	7.1E-08	9.9E-09	1.3E-09	8.2E-10	1.2E-10	3.4E-12	7.6E-10	1.6E-08
65-2 16h	yHMK 65	*Suva* WT	16	1.5E-10	2.6E-09	1.1E-07	7.3E-09	3.4E-10	1.1E-09	3.3E-11	5.1E-13	1.2E-10	3.4E-09
65-3 0.5h	yHMK 65	*Suva* WT	0.5	1.8E-09	5.2E-09	4.4E-08	3.4E-08	2.1E-09	4.3E-09	1.2E-10	6.8E-12	3.6E-10	5.1E-09
65-3 1h	yHMK 65	*Suva* WT	1	3.0E-09	2.5E-09	7.4E-08	2.5E-08	2.7E-09	1.9E-09	2.0E-10	2.5E-12	3.8E-10	6.6E-09
65-3 3h	yHMK 65	*Suva* WT	3	1.9E-09	3.9E-09	5.5E-08	6.8E-09	1.9E-09	8.8E-10	2.3E-10	3.1E-12	5.9E-10	7.5E-09
65-3 5h	yHMK 65	*Suva* WT	5	9.3E-10	8.9E-09	7.0E-08	9.7E-09	1.5E-09	8.4E-10	1.9E-10	1.7E-12	7.6E-10	1.3E-08
65-3 16h	yHMK 65	*Suva* WT	16	2.1E-10	2.1E-09	1.1E-07	1.1E-08	5.6E-10	1.6E-09	7.9E-11	1.8E-12	1.9E-10	4.1E-09
280 0.5h	yHMK 280	*Suva gal80Δ gal80bΔ*	0.5	2.2E-10	3.1E-09	3.7E-08	6.0E-09	3.9E-09	7.6E-10	1.1E-10	9.0E-12	1.2E-09	5.1E-09
280 1h	yHMK 280	*Suva gal80Δ gal80bΔ*	1	3.1E-10	4.1E-09	4.8E-08	1.1E-08	1.3E-08	1.3E-09	1.8E-10	1.7E-12	2.8E-09	1.2E-08
280 3h	yHMK 280	*Suva gal80Δ gal80bΔ*	3	4.9E-10	3.0E-09	5.1E-08	1.2E-08	3.2E-09	1.4E-09	1.9E-10	1.3E-11	1.4E-09	2.2E-08
280 5h	yHMK 280	*Suva gal80Δ gal80bΔ*	5	4.3E-10	5.6E-09	3.2E-08	4.9E-09	3.3E-09	4.6E-10	1.9E-10	3.0E-12	1.2E-09	1.4E-08
280 16h	yHMK 280	*Suva gal80Δ gal80bΔ*	16	3.4E-10	5.2E-09	4.2E-08	4.0E-09	7.0E-10	5.3E-10	3.5E-11	4.9E-13	2.3E-10	3.7E-09
281 0.5h	yHMK 281	*Suva gal80Δ gal80bΔ*	0.5	2.1E-10	3.8E-09	5.9E-08	5.0E-09	7.3E-09	5.6E-10	1.4E-10	1.1E-11	8.5E-10	5.4E-09
281 1h	yHMK 281	*Suva gal80Δ gal80bΔ*	1	3.2E-10	5.5E-09	7.1E-08	8.4E-09	2.8E-08	6.5E-10	1.6E-10	2.9E-12	2.5E-09	1.6E-08
281 3h	yHMK 281	*Suva gal80Δ gal80bΔ*	3	5.2E-10	2.7E-09	7.0E-08	1.2E-08	2.9E-08	1.8E-09	2.6E-10	1.5E-12	2.8E-09	2.9E-08
281 5h	yHMK 281	*Suva gal80Δ gal80bΔ*	5	5.4E-10	3.6E-09	4.5E-08	6.0E-09	7.0E-09	5.6E-10	2.1E-10	2.7E-12	1.3E-09	4.6E-08
281 16h	yHMK 281	*Suva gal80Δ gal80bΔ*	16	5.4E-10	1.1E-08	3.9E-08	5.4E-09	9.3E-10	6.8E-10	5.4E-11	3.4E-12	3.9E-10	1.5E-08
282 0.5h	yHMK 282	*Suva gal80Δ gal80bΔ*	0.5	2.1E-10	3.5E-09	4.5E-08	5.7E-09	4.6E-09	4.3E-10	1.2E-10	7.9E-12	8.9E-10	4.9E-09
282 1h	yHMK 282	*Suva gal80Δ gal80bΔ*	1	2.8E-10	3.3E-09	5.6E-08	1.0E-08	1.7E-08	8.7E-10	1.6E-10	3.0E-11	2.1E-09	9.7E-09
282 3h	yHMK 282	*Suva gal80Δ gal80bΔ*	3	5.4E-10	8.8E-09	5.8E-08	1.6E-08	5.0E-09	2.5E-09	2.2E-10	8.5E-12	1.2E-09	2.1E-08
282 5h	yHMK 282	*Suva gal80Δ gal80bΔ*	5	4.6E-10	9.8E-09	3.5E-08	6.2E-09	2.3E-09	8.3E-10	3.4E-10	4.9E-12	1.2E-09	1.4E-08
282 16h	yHMK 282	*Suva gal80Δ gal80bΔ*	16	4.0E-10	7.9E-09	4.0E-08	5.3E-09	5.4E-10	7.4E-10	5.2E-11	1.1E-12	2.3E-10	3.0E-09
56-1	yHMK 56	*Scer* WT	4.5h	9.0E-09	1.3E-09	3.6E-08	1.2E-08	2.1E-09	1.5E-09	6.1E-10	9.8E-12	2.6E-10	6.6E-09
56-2	yHMK 56	*Scer* WT	4.5h	9.5E-09	1.7E-09	3.9E-08	1.3E-08	2.0E-09	1.5E-09	6.2E-10	1.0E-11	2.7E-10	5.8E-09
370	yHMK 370	*Scer gal1*	4.5h	1.8E-08	7.6E-10	7.2E-09	2.1E-09	1.7E-09	1.5E-10	2.0E-09	6.5E-12	2.5E-10	4.2E-09
371	yHMK 371	*Scer gal1*	4.5h	4.5E-08	1.3E-09	9.8E-09	3.3E-09	1.8E-09	7.1E-10	4.0E-09	8.3E-12	2.2E-10	3.8E-09
CT322-1	yHCT 322	*Scer gal7Δ*	4.5h	3.5E-08	4.3E-10	4.9E-09	2.0E-09	1.8E-09	8.4E-11	3.6E-09	1.8E-12	5.1E-10	5.6E-09
CT322-2	yHCT 322	*Scer gal7Δ*	4.5h	1.8E-08	7.9E-10	7.9E-09	3.4E-09	2.6E-09	3.1E-10	1.5E-09	7.6E-12	1.6E-10	3.5E-09
CT324-1	yHCT 324	*Scer gal10Δ*	4.5h	1.9E-08	8.8E-10	8.8E-09	2.7E-09	1.6E-09	2.4E-10	2.0E-09	7.1E-12	3.0E-10	7.0E-09
CT324-2	yHCT 324	*Scer gal10Δ*	4.5h	1.9E-08	3.2E-10	7.1E-09	2.5E-09	2.3E-09	4.0E-10	2.1E-09	4.9E-12	4.8E-10	4.8E-09

Sample Name	Isocitric Acid	Phospho(enol)pyruvic acid	Galactose-1-Phosphate	Sum of Glucose-1-Phosphate+Mannose-1-Phosphate	Trehalose-6-Phosphate	Sum of Glycerol-2-Phosphate and Glycerol -3-Phosphate
65-1 0.5h	<LOQ	1.7E-10	4.9E-10	1.3E-10	<LOQ	1.6E-11
65-1 1h	<LOQ	2.4E-11	3.1E-09	3.0E-10	<LOQ	4.3E-10
65-1 3h	<LOQ	3.7E-11	4.1E-09	2.9E-10	3.7E-12	4.4E-10
65-1 5h	<LOQ	4.0E-11	3.7E-09	2.2E-10	9.5E-11	7.7E-10
65-1 16h	<LOQ	1.1E-11	2.7E-09	7.7E-10	3.4E-10	4.3E-11
65-2 0.5h	<LOQ	7.7E-11	4.0E-10	1.3E-10	<LOQ	1.4E-11
65-2 1h	<LOQ	2.8E-11	3.2E-09	3.2E-10	<LOQ	5.1E-10
65-2 3h	<LOQ	2.1E-11	3.1E-09	2.2E-10	<LOQ	3.5E-10
65-2 5h	<LOQ	3.7E-11	4.2E-09	2.3E-10	2.0E-10	6.6E-10
65-2 16h	<LOQ	7.1E-12	1.4E-10	3.6E-11	4.0E-10	1.3E-09
65-3 0.5h	<LOQ	9.3E-11	5.6E-10	1.4E-10	<LOQ	1.6E-11
65-3 1h	<LOQ	2.3E-11	2.9E-09	2.7E-10	<LOQ	3.9E-10
65-3 3h	<LOQ	3.0E-11	3.7E-09	2.3E-10	<LOQ	4.5E-10
65-3 5h	<LOQ	6.9E-11	3.4E-09	1.9E-10	1.3E-10	8.3E-10
65-3 16h	1.4E-11	1.3E-11	2.1E-10	5.0E-11	4.3E-10	1.0E-09
280 0.5h	<LOQ	4.9E-11	1.2E-09	3.1E-10	3.3E-08	3.4E-09
280 1h	<LOQ	1.2E-10	7.4E-09	4.0E-10	4.3E-08	6.9E-09
280 3h	<LOQ	5.4E-11	4.0E-09	3.2E-10	6.7E-09	2.4E-09
280 5h	<LOQ	1.4E-10	1.3E-08	3.0E-10	2.5E-08	2.8E-09
280 16h	<LOQ	2.4E-11	1.8E-09	1.5E-10	2.4E-09	7.9E-10
281 0.5h	<LOQ	3.7E-11	1.1E-09	3.1E-10	4.8E-08	3.6E-09
281 1h	<LOQ	1.6E-10	6.4E-09	3.8E-10	4.5E-08	6.8E-09
281 3h	8.4E-09	1.9E-10	3.8E-08	<LOQ	1.2E-08	4.0E-09
281 5h	9.4E-09	1.2E-10	5.3E-08	<LOQ	1.1E-09	1.9E-09
281 16h	7.5E-10	2.8E-11	1.2E-08	2.7E-10	1.2E-09	1.1E-09
282 0.5h	<LOQ	3.5E-11	9.9E-10	2.8E-10	2.6E-08	3.3E-09
282 1h	<LOQ	9.4E-11	6.1E-09	3.5E-10	2.9E-08	7.4E-09
282 3h	1.3E-09	6.2E-11	6.2E-09	3.1E-10	4.4E-09	2.1E-09
282 5h	1.5E-10	1.1E-10	1.1E-08	3.3E-10	3.1E-08	2.4E-09
282 16h	<LOQ	1.8E-11	1.1E-09	1.0E-10	3.9E-09	8.8E-10
56-1	<LOQ	3.3E-11	3.2E-09	1.8E-10	<LOQ	3.4E-10
56-2	<LOQ	3.2E-11	3.3E-09	1.8E-10	<LOQ	4.1E-10
370	<LOQ	6.4E-11	<LOQ	1.0E-10	<LOQ	5.3E-12
371	<LOQ	6.0E-11	<LOQ	1.2E-10	<LOQ	6.0E-12
CT322-1	<LOQ	1.4E-10	2.5E-08	<LOQ	<LOQ	1.8E-11
CT322-2	<LOQ	3.5E-11	1.9E-08	<LOQ	<LOQ	9.0E-12
CT324-1	<LOQ	7.7E-11	3.1E-08	<LOQ	<LOQ	8.2E-12
CT324-2	<LOQ	1.4E-10	4.0E-08	<LOQ	<LOQ	1.4E-11

Sample Name	Aconitic Acid	Glucose-6-Phosphate	Fructose-6-Phosphate	Mannose-6-Phosphate	Seduhepulose-7-Phosphate	Fructose-1,6-bisphosphate	AMP	ADP	ATP	GMP	GDP
65-1 0.5h	1.2E-10	1.4E-10	4.3E-12	6.6E-12	2.6E-10	2.6E-10	4.1E-12	8.9E-10	1.7E-08	6.9E-10	8.0E-11
65-1 1h	1.3E-10	5.4E-10	7.5E-12	2.4E-11	3.9E-10	3.8E-09	3.2E-12	1.4E-09	2.4E-08	4.1E-10	1.3E-10
65-1 3h	1.4E-10	6.8E-10	1.6E-11	5.2E-11	5.2E-10	2.4E-09	4.1E-12	1.6E-09	3.3E-08	8.3E-10	1.6E-10
65-1 5h	1.9E-10	4.4E-10	1.8E-11	2.4E-11	3.9E-10	6.4E-10	5.1E-12	8.3E-10	8.8E-09	3.2E-09	1.3E-10
65-1 16h	3.6E-11	1.9E-10	1.1E-11	1.8E-11	1.8E-10	1.6E-09	3.5E-12	3.3E-09	3.7E-08	8.9E-10	2.1E-10
65-2 0.5h	1.2E-10	1.1E-10	4.5E-12	8.6E-12	2.4E-10	1.9E-10	2.7E-12	1.0E-09	1.8E-08	5.4E-10	9.9E-11
65-2 1h	1.4E-10	6.4E-10	8.2E-12	3.5E-11	3.9E-10	4.8E-09	3.6E-12	1.5E-09	2.9E-08	4.8E-10	1.5E-10
65-2 3h	1.4E-10	4.7E-10	4.9E-12	2.3E-11	4.5E-10	1.7E-09	4.6E-12	1.3E-09	1.3E-08	1.3E-09	1.5E-10
65-2 5h	2.0E-10	4.3E-10	1.9E-11	2.5E-11	2.9E-10	6.2E-10	5.8E-12	1.3E-09	1.2E-08	4.3E-09	2.0E-10
65-2 16h	3.2E-11	1.3E-10	1.1E-12	9.9E-12	1.6E-10	1.1E-09	5.9E-12	2.9E-09	2.7E-08	1.0E-09	1.8E-10
65-3 0.5h	1.3E-10	1.6E-10	6.8E-12	1.4E-11	7.8E-10	4.7E-10	3.0E-12	1.3E-09	2.8E-08	6.2E-10	1.4E-10
65-3 1h	1.5E-10	5.9E-10	9.4E-12	3.1E-11	3.3E-10	3.3E-09	3.5E-12	1.2E-09	1.4E-08	5.1E-10	1.3E-10
65-3 3h	1.2E-10	5.1E-10	8.2E-12	3.4E-11	3.5E-10	1.9E-09	3.6E-12	1.6E-09	2.0E-08	1.1E-09	1.7E-10
65-3 5h	1.6E-10	3.6E-10	1.0E-11	2.0E-11	2.6E-10	4.3E-10	3.8E-12	1.4E-09	1.5E-08	3.2E-09	1.7E-10
65-3 16h	3.7E-11	1.7E-10	5.8E-12	1.6E-11	1.6E-10	1.8E-09	4.8E-12	3.5E-09	4.2E-08	9.5E-10	2.3E-10
280 0.5h	1.1E-10	7.7E-10	6.6E-12	3.3E-11	6.5E-10	4.4E-09	2.8E-12	2.1E-10	3.3E-09	2.1E-10	1.3E-10
280 1h	1.9E-10	9.6E-10	1.2E-11	5.8E-11	4.9E-10	1.0E-08	3.9E-12	4.0E-10	1.3E-08	1.8E-10	2.2E-10
280 3h	3.0E-10	7.7E-10	9.4E-12	4.3E-11	4.8E-10	5.1E-09	3.4E-12	1.3E-09	2.5E-08	6.2E-10	3.5E-10
280 5h	2.1E-10	4.3E-10	2.2E-11	2.5E-11	1.9E-10	2.0E-09	3.1E-12	1.0E-09	2.4E-08	2.1E-10	1.8E-10
280 16h	3.0E-11	2.2E-10	4.1E-12	1.2E-11	2.6E-10	8.3E-10	4.5E-12	2.2E-09	2.4E-08	1.0E-09	1.2E-10
281 0.5h	1.4E-10	7.8E-10	7.2E-12	3.3E-11	4.5E-10	3.7E-09	3.0E-12	2.3E-10	4.5E-09	1.7E-10	9.2E-11
281 1h	3.4E-10	9.0E-10	2.2E-11	5.4E-11	3.7E-10	8.6E-09	4.5E-12	4.2E-10	1.6E-08	1.9E-10	2.0E-10
281 3h	1.0E-09	9.5E-10	3.5E-11	5.7E-11	4.4E-10	1.2E-08	4.4E-12	9.6E-10	5.5E-08	1.4E-10	3.5E-10
281 5h	8.5E-10	5.0E-10	2.1E-11	2.4E-11	2.4E-10	3.4E-09	2.5E-12	8.3E-10	4.0E-08	1.2E-10	2.2E-10
281 16h	9.0E-11	3.1E-10	3.2E-11	1.6E-11	3.3E-10	6.4E-10	3.2E-12	1.4E-09	1.4E-08	1.4E-09	1.3E-10
282 0.5h	<LOQ	6.0E-10	2.9E-12	2.3E-11	4.1E-10	3.3E-09	2.2E-12	2.4E-10	4.8E-09	1.9E-10	1.2E-10
282 1h	1.8E-10	7.3E-10	6.5E-12	3.6E-11	4.1E-10	6.4E-09	3.9E-12	2.6E-10	1.1E-08	1.2E-10	1.5E-10
282 3h	2.6E-10	6.1E-10	2.0E-11	3.5E-11	3.9E-10	3.4E-09	1.8E-12	1.3E-09	3.2E-08	4.7E-10	2.9E-10
282 5h	2.2E-10	5.1E-10	1.1E-11	2.0E-11	3.7E-10	2.3E-09	2.9E-12	2.0E-09	4.1E-08	5.0E-10	2.5E-10
282 16h	3.2E-11	2.2E-10	7.7E-12	1.6E-11	3.4E-10	1.2E-09	3.7E-12	2.5E-09	3.7E-08	7.3E-10	1.4E-10
56-1	1.5E-10	3.6E-10	6.9E-12	2.1E-11	4.7E-10	1.9E-09	2.5E-12	2.6E-09	9.6E-08	4.0E-10	1.8E-10
56-2	1.4E-10	4.4E-10	5.6E-12	2.1E-11	4.3E-10	2.2E-09	2.1E-12	3.0E-09	1.0E-07	3.9E-10	2.1E-10
370	1.0E-10	1.3E-10	5.3E-12	1.0E-11	1.6E-10	4.1E-11	<LOQ	2.9E-09	4.8E-08	8.7E-10	2.2E-10
371	8.8E-11	2.1E-10	8.3E-12	1.6E-11	1.7E-10	1.2E-10	2.7E-12	2.9E-09	3.9E-08	1.4E-09	2.5E-10
CT322-1	1.2E-10	2.0E-10	<LOQ	1.2E-11	3.4E-11	9.9E-11	2.2E-12	3.7E-09	6.1E-08	1.7E-09	3.6E-10
CT322-2	1.2E-10	2.5E-10	9.3E-12	1.8E-11	2.6E-10	1.5E-10	1.9E-12	2.3E-09	3.0E-08	1.2E-09	2.1E-10
CT324-1	1.1E-10	1.9E-10	9.4E-12	1.8E-11	9.7E-11	1.0E-10	3.1E-12	2.5E-09	3.7E-08	1.6E-09	2.4E-10
CT324-2	1.2E-10	1.7E-10	<LOQ	9.2E-12	3.9E-11	1.1E-10	2.2E-12	2.7E-09	3.5E-08	1.5E-09	3.1E-10

Appendix B: Supplementary Materials for Chapter 3

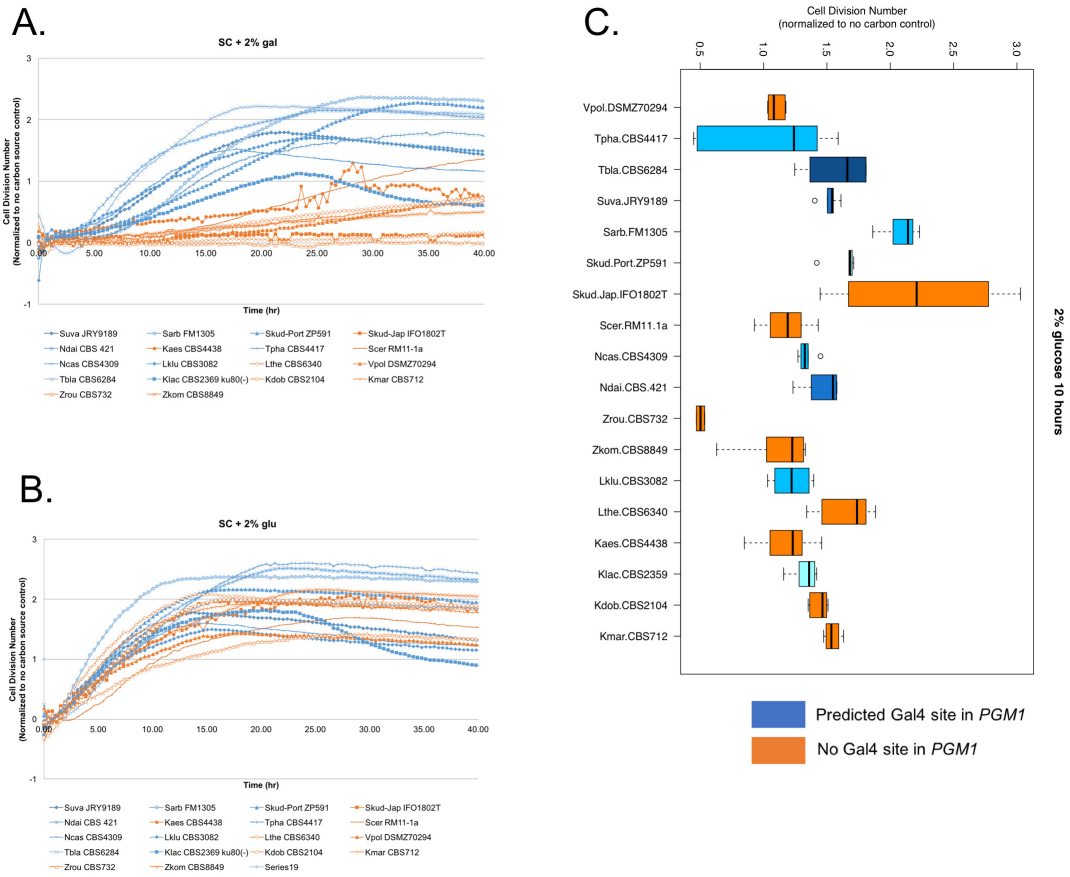


Figure B.1: The *PGM* association was specific to galactose and not related to the growth in glucose.

A. Growth curves for different species on SC+2% galactose.

B-C. There were no significant differences between species with or without predicted Gal4-binding sites on glucose. (4 hours: p=0.44, 7 hours: p=0.14, 10 hours: p=0.37, n=6, Welch's t-test).

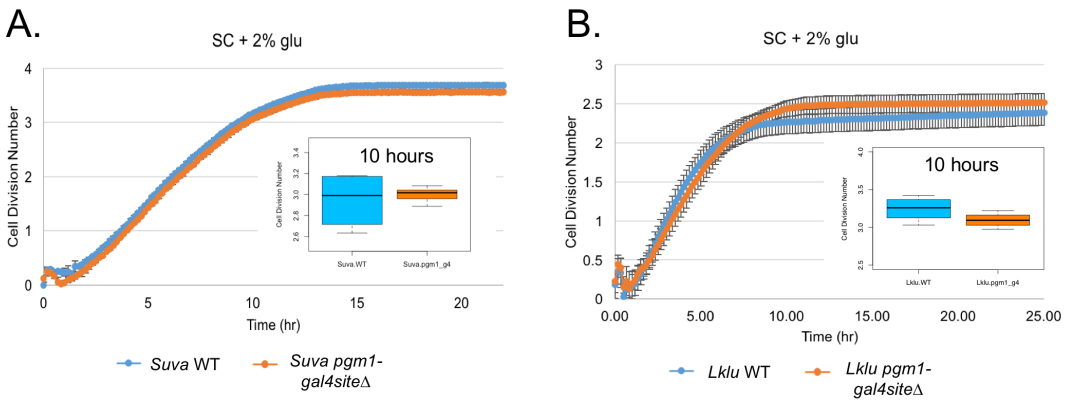


Figure B.2: Mutation of Gal4-binding sites upstream of *PGM1* did not cause any growth defects on glucose in *S. uvarum* and *L. kluyveri*.

A. Gal4-binding site mutation of *PGM1* did not cause any growth differences in SC+2% glucose for *S. uvarum*. p=0.76, $n_{WT=6}$, n_{mutant}=5.

B. Gal4-bindingsite mutation of *PGM1* did not cause any growth differences in SC+2% glucose for *L. kluyveri*. p=0.16, n_{WT}=6, n_{mutant}=4.

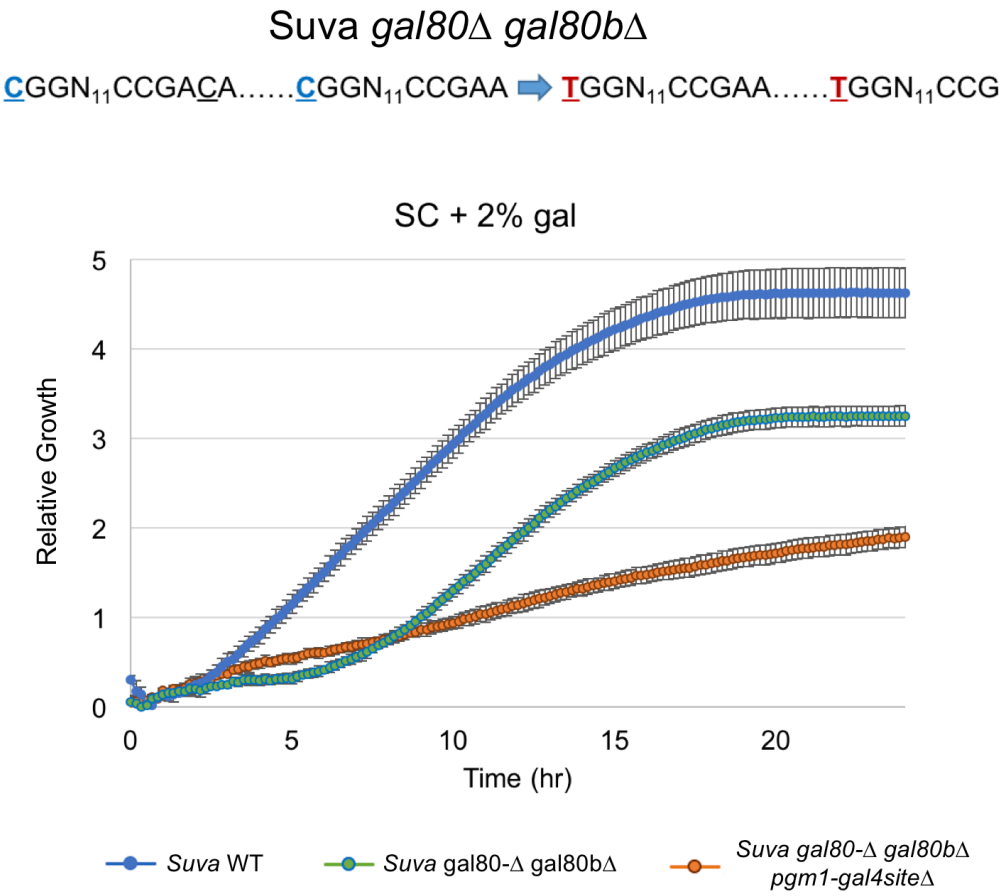


Figure B.3: Temporary Growth Arrest (a phenotype caused by metabolic overload when both Gal80 co-repressors are deleted) was completely rescued by the deletion of two predicted Gal4-binding sites upstream of *SuvaPGM1*.

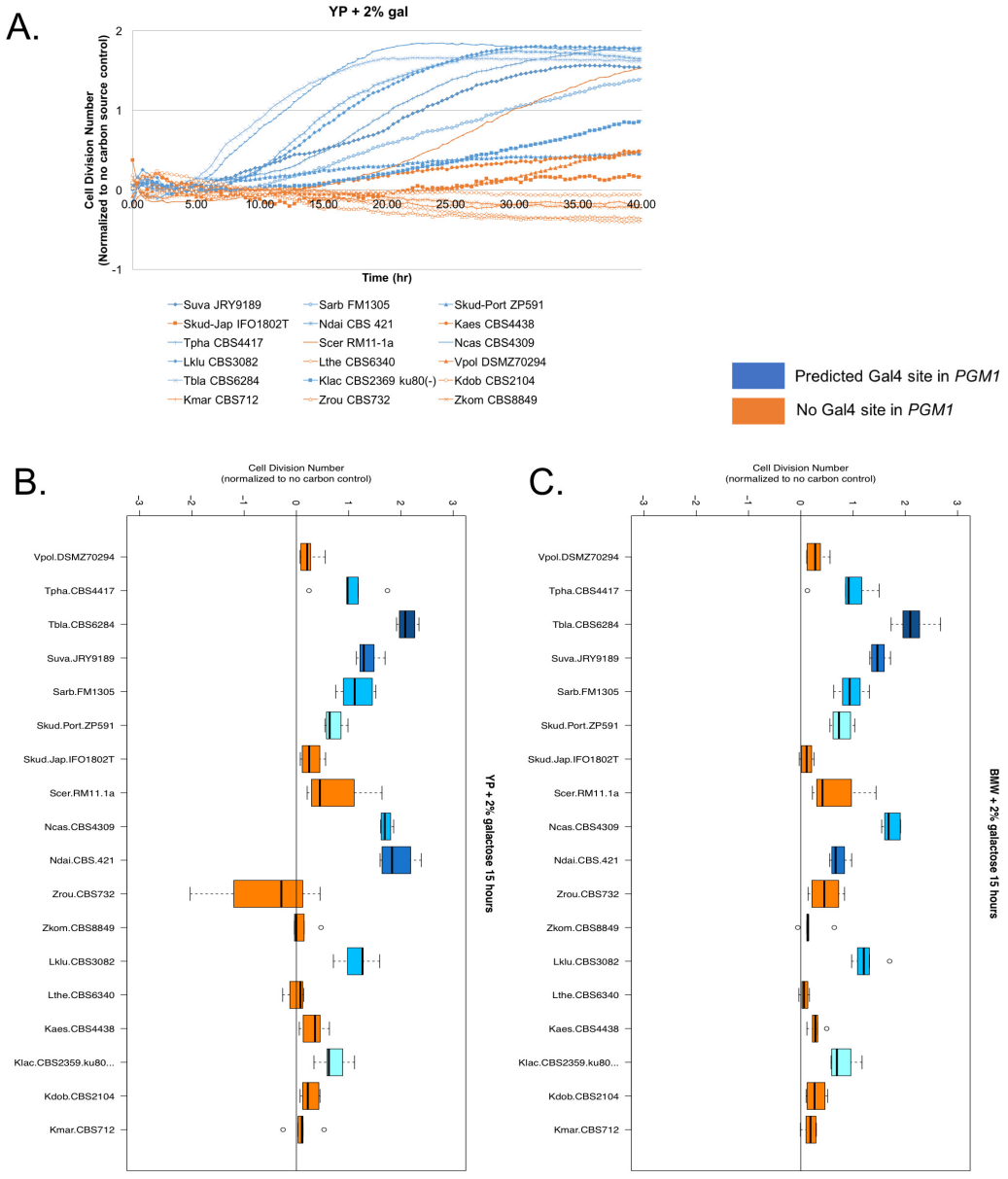


Figure B.4: The correlation between growth on galactose and the number of predicted Gal4-binding sites upstream of *PGM1/2* is maintained in different base media and temperatures.

(continued) A. The correlation between having Gal4 site or not upstream of *PGM1/2* and galactose growth is maintained at 30°C.

B-C. The correlation between having Gal4 site or not upstream of *PGM1/2* and galactose growth maintained in two different base media, YP+2% gal or BMW+2% gal gal (Thompson et al., 2013). Welch's *t*-test: B. *p*=8.4e-5, n=9; C. *p*=4.8e-4, n=9.

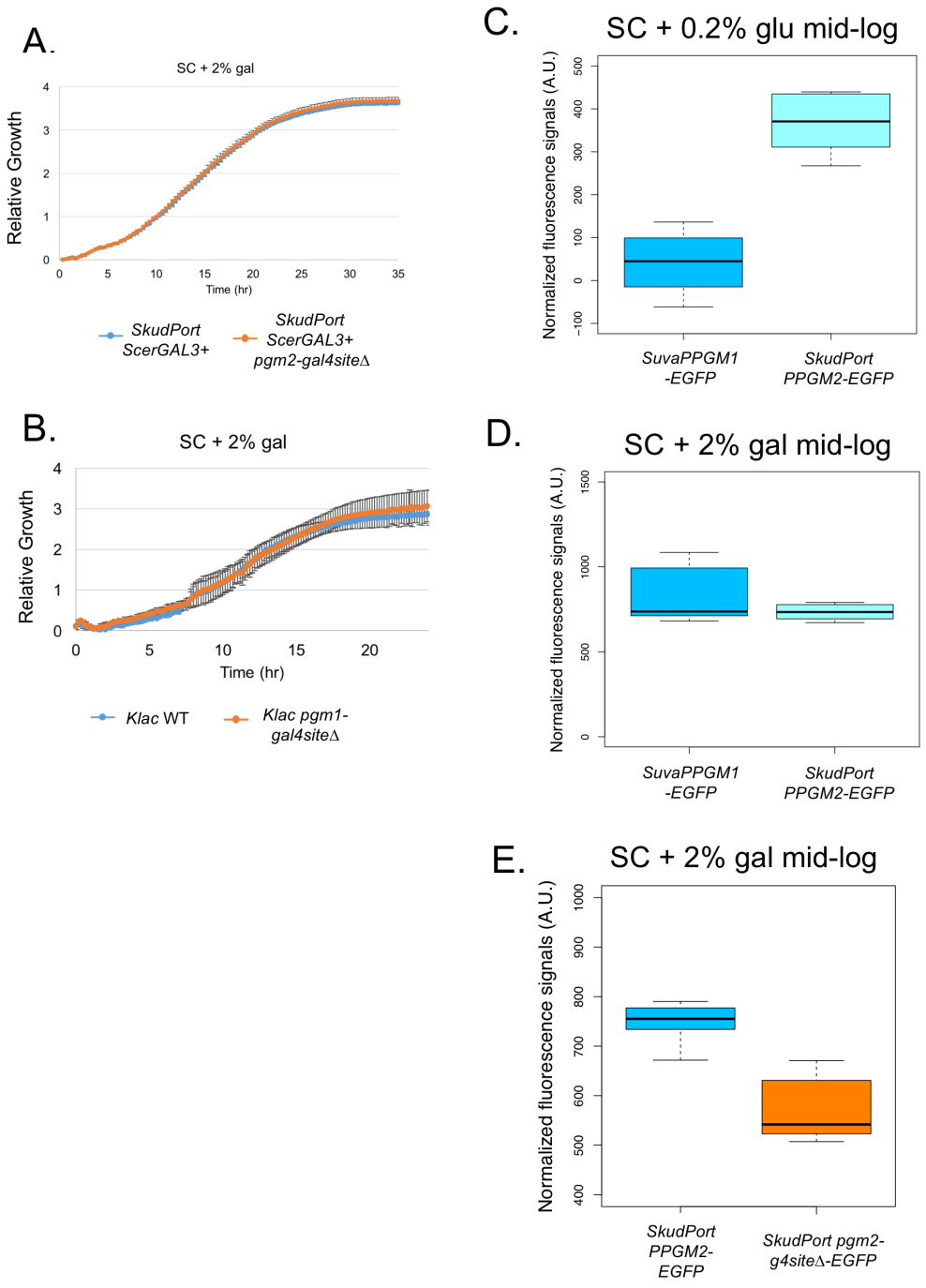


Figure B.5: The effect of Gal4 site mutation in the promoter of *PGM1/2* in *S. kudriavzevii* or *K. lactis*.

(continued) A. Mutation of the predicted Gal4-binding site upstream of *PGM2* in *S.kudriavzveii* Portuguese strain did not cause any growth differences in galactose. Note that these *S. kudriavzveii* strains were engineered to contain a *S. cerevisiae GAL3* coding region presumably to further increase upstream metabolic rate (Hittinger et al., 2010).

B. Mutation of the predicted Gal4-binding site upstream of *PGM1* in *K. lactis* did not cause any growth differences in galactose.

C. *S. kudriavzveii* Portuguese strain *PGM2* had higher basal expression than *S. uvarum PGM1*.

D. *S. kudriavzveii* Portuguese strain *PGM2* had similar expression as *S. uvarum PGM1* in galactose at mid-log phase.

E. *S. kudriavzveii* Portuguese strain *PGM2* required Gal4 binding site for full galactose induction.

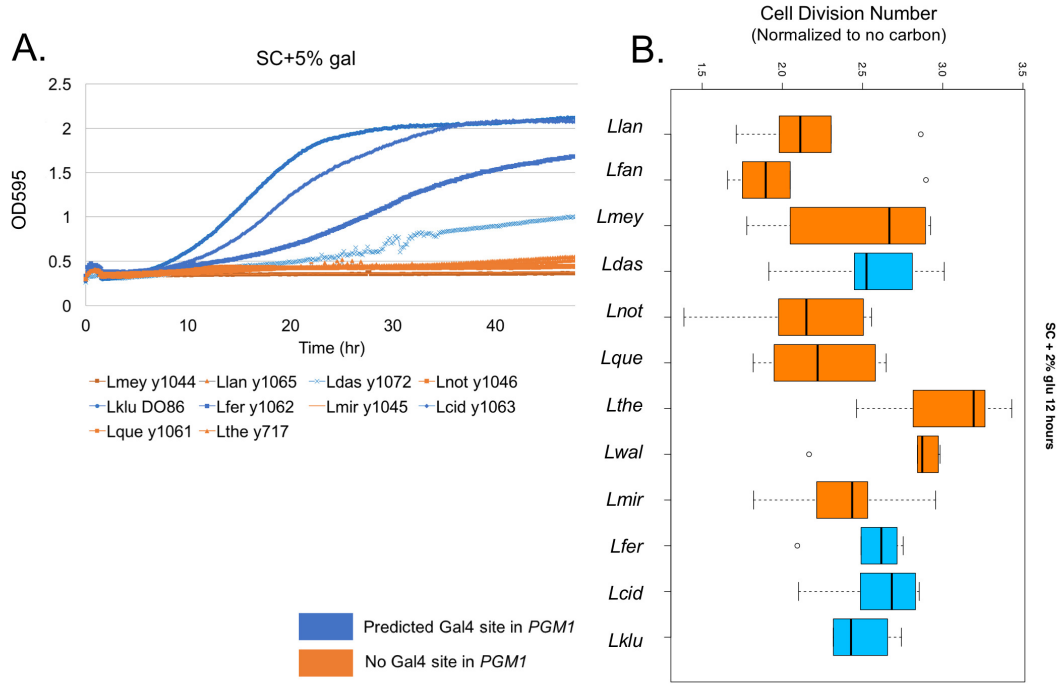


Figure B.6: *Lachancea* species showed strong association of having predicted Gal4 site upstream of *PGM1* with growth in galactose but not in glucose.

A. Growth curves for different species at SC+5% gal.

B. There were no significant differences between species with or without predicted Gal4 site in glucose.

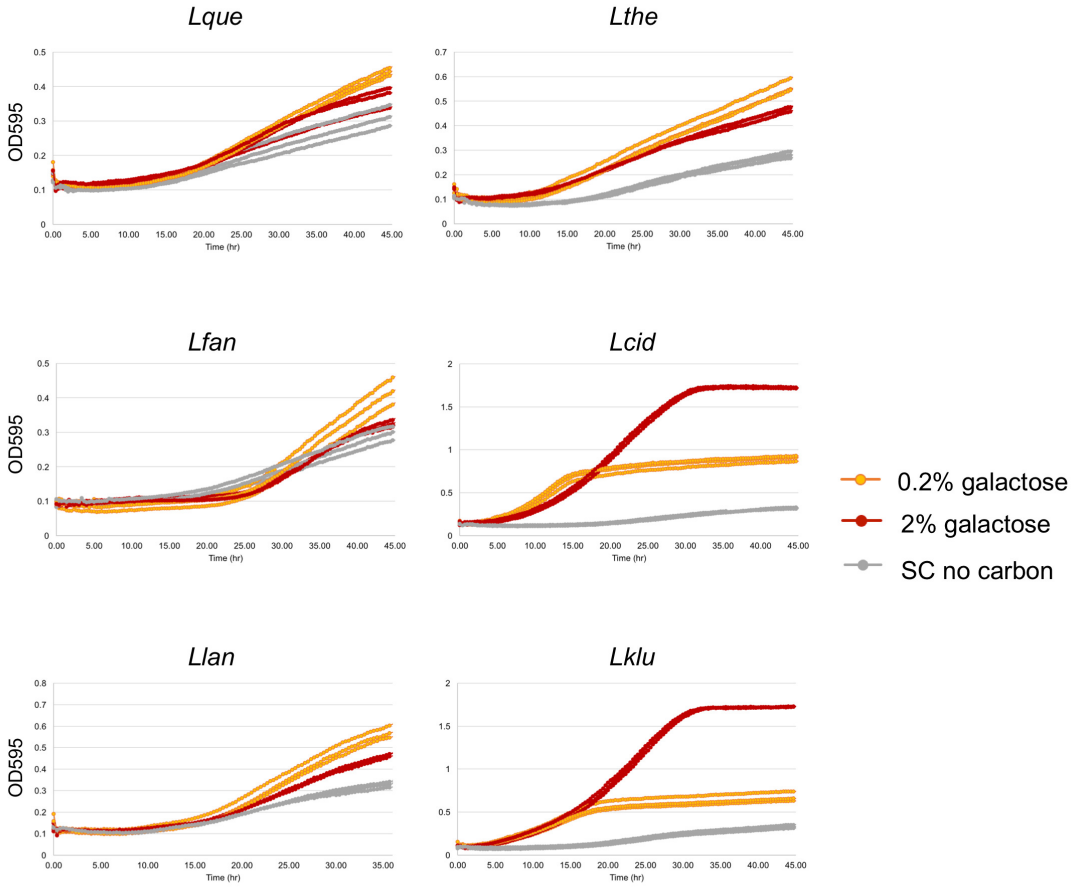


Figure B.7: Note that *L. kluyveri* served as a control to show that this pattern was specific to species that grew slowly. Three biological replicates in each condition are shown, corresponding to each individual growth curve. This experiment had been conducted twice independently.

Table B.1: A summary of 20 genetic features in 17 species of the family Saccharomycetaceae.

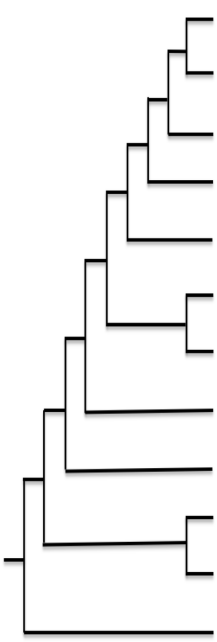
<table>
<tr><th rowspan="3"></th><th rowspan="3">Division Number in galactose at 15 hours (relative to no carbon)</th><th colspan="16">Predicted Gal4 site number and distance to start codon; [intergenic length (bp)]</th><th colspan="2">GLSSSA(A/S)</th><th colspan="2">GAL1-10-7 organization</th><th rowspan="3">Additional GAL genes</th></tr>
<tr><th colspan="2">GAL1</th><th colspan="2">GAL10</th><th colspan="2">GAL7</th><th colspan="2">GAL3</th><th colspan="2">GAL4</th><th colspan="2">GAL80</th><th colspan="2">PGM1 (all have GXXXTASHNP)</th><th colspan="2">PGM2 (all have GXXXTASHNP)</th><th rowspan="2">GAL1</th><th rowspan="2">GAL3</th><th rowspan="2">cluster</th><th rowspan="2">shared promoter</th></tr>
<tr><th># Gal4 site</th><th>Gal4 site distance to start codon; [intergenic length (bp)]</th><th># Gal4 site</th><th>Gal4 site distance to start codon; [intergenic length (bp)]</th><th># Gal4 site</th><th>Gal4 site distance to start codon; [intergenic length (bp)]</th><th># Gal4 site</th><th>Gal4 site distance to start codon; [intergenic length (bp)]</th><th># Gal4 site</th><th>Gal4 site distance to start codon; [intergenic length (bp)]</th><th># Gal4 site</th><th>Gal4 site distance to start codon; [intergenic length (bp)]</th><th># Gal4 site</th><th>Gal4 site distance to start codon; [intergenic length (bp)]</th><th># Gal4 site</th><th>Gal4 site distance to start codon; [intergenic length (bp)]</th></tr>
<tr><td>Vpol DSMZ70294</td><td>0.32</td><td>2</td><td>575, 689, (1144, 1177, 1209); [1527]</td><td>2</td><td>681, 883; [2404]</td><td>3</td><td>301, 333, 366, (821, 935); [1527]</td><td>2?</td><td>1504, 1706; [2404]</td><td>0</td><td>--; [898] (frameshift?)</td><td>1</td><td>337; [1993]</td><td>0</td><td>--; [1592]</td><td>0</td><td>-- in 1 kb; [505]</td><td>GLSSSAA</td><td>GLSSSAA</td><td>GAL3-10 ; GAL1-GAL7</td><td>GAL3-10 ; GAL1-GAL7</td><td></td></tr>
<tr><td>Tpha CBS4417</td><td>0.70</td><td>2</td><td>719, 744, (1461, 1486); [1721]</td><td>3</td><td>407, 459, 497; [1205]</td><td>2</td><td>218, 243, (960, 985); [1721]</td><td>3</td><td>691, 729, 781; [1205]</td><td>0</td><td>--; [666]</td><td>1</td><td>480; [571]</td><td>0</td><td>-- in 1 kb; [671]</td><td>1</td><td>633; [783]</td><td>GLSSSAA</td><td>GLSSSAA</td><td>GAL3-10 ; GAL1-GAL7</td><td>GAL3-10 ; GAL1-GAL7</td><td>2 GAL10 (one not syntenic?)</td></tr>
<tr><td>Tbla CBS6284</td><td>2.14</td><td>2</td><td>546, 592, (899, 945); [1571]</td><td>2</td><td>609, 655, (962, 1008); [1571]</td><td>4</td><td>446, 482, 550, 572; [971]</td><td>0</td><td>-- in 1 kb; [5944]</td><td>0</td><td>--; [850]</td><td>2 | 0</td><td>534/549, 623; [1119] | --; [2336]</td><td>3</td><td>468, 490, 652; [755]</td><td>0</td><td>--; [1633]</td><td>GLSSSAA</td><td>GLSSSAA</td><td>GAL1-10-7</td><td>GAL1-10</td><td>2 GAL80</td></tr>
<tr><td>Suva JRY9189</td><td>1.39</td><td>3</td><td>434, 452, 471; [720]</td><td>3</td><td>232, 251, 269; [720]</td><td>1</td><td>179 ; [727]</td><td>1</td><td>274, (1058, 1081); [1398]</td><td>0</td><td>--; [233]</td><td>1 | 1</td><td>155; [277] | 378; [1059]</td><td>2</td><td>323, 437; [641]</td><td>0</td><td>--; [669]</td><td>GLSSSAA</td><td>GLSSSAA</td><td>GAL1-10-7</td><td>GAL1-10</td><td>2 GAL80, 2 GAL2</td></tr>
<tr><td>Sarb FM1305</td><td>1.18</td><td>3</td><td>424, 442, 461; [712]</td><td>3</td><td>234, 253, 271; [712]</td><td>1</td><td>185; [741]</td><td>1</td><td>275; []</td><td>0</td><td>-- in 1 kb</td><td>1</td><td>177; [305]</td><td>1</td><td>303</td><td>0</td><td>--</td><td>GLSSSAA</td><td>-SA deletion</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Skud-Port ZP591</td><td>0.94</td><td>3</td><td>391, 409, 428; [673]</td><td>3</td><td>228, 247, 265; [673]</td><td>1</td><td>172; [737]</td><td>--</td><td>--</td><td>0</td><td>--; [248]</td><td>?1</td><td>79 (too close?); [204]</td><td>0</td><td>; [1072]</td><td>1</td><td>160; [681]</td><td>GLSSSAA</td><td>-SA deletion</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Scer RM11-1a</td><td>0.28</td><td>3</td><td>399, 417, 436; [667]</td><td>3</td><td>214, 233, 251; [667]</td><td>1</td><td>175; [725]</td><td>1</td><td>271; [830]</td><td>0</td><td>--; [356]</td><td>1</td><td>155; [283]</td><td>0</td><td>--; [1429]</td><td>0</td><td>--; [456]</td><td>GLSSSAA</td><td>-SA deletion</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Ncas CBS4309</td><td>1.43</td><td>2</td><td>306, 375, (1004, 1029); [1217]</td><td>1</td><td>166/176; [854]</td><td>2 | 2</td><td>188, 215; [469] | 171, 196, (825, 894); [1217]</td><td>1</td><td>661/671; [854]*(similar to GAL1)</td><td>1 | 0</td><td>193; [380] | -- in 1kb; [558]</td><td>1 | 1</td><td>301; [523] | 291; [883]</td><td>1</td><td>365; [561]</td><td>0</td><td>--; [406]</td><td>GLSSSAA</td><td>GLSSSSF</td><td>GAL1-GAL7; GAL3-GAL10-GAL7</td><td>GAL1-GAL7; GAL3-GAL10</td><td>2 GAL80, 2 GAL7, 2 GAL4</td></tr>
<tr><td>Ndai CBS421</td><td>1.83</td><td>3</td><td>421, 651, 670; [1087]</td><td>3</td><td>400, 419, 649; [1087]</td><td>2</td><td>221, 272; [1670]</td><td>0</td><td>(1381, 1432); [1670]</td><td>1 | 0</td><td>474; [1560] | -- in 1kb; [852]</td><td>1 | 2</td><td>462; [914] | 251, 318; [1054]</td><td>0?</td><td>; [3388/N]; missing start codon</td><td>2</td><td>; [603]</td><td>GLSSSAA</td><td>GLSSSSF</td><td>GAL1-10; GAL3-GAL7</td><td>GAL1-10; GAL3-GAL7</td><td>2 GAL80, 2 GAL4 (only 26% aa identity)</td></tr>
<tr><td>Zrou CBS732</td><td>-0.65</td><td>2</td><td>418, 437; [585]</td><td>2</td><td>131, 150 [585]</td><td>2</td><td>133, 157/164; [325]</td><td>--</td><td>--</td><td>0</td><td>-- in 1 kb; [577]</td><td>1</td><td>196; [979]</td><td>0</td><td>-- in 1 kb; [302]</td><td>--</td><td>--</td><td>GLSSSAA</td><td>--</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Zkom CBS8849</td><td>0.18</td><td>3</td><td>178, 212, 256; [413]</td><td>3</td><td>140, 184, 218; [413]</td><td>1</td><td>132/139; [301]</td><td>--</td><td>--</td><td>?</td><td>? (may not have GAL4?)</td><td>0</td><td>-- in 1 kb</td><td>0</td><td>-- in 1 kb</td><td>--</td><td>--</td><td>GLSSSAA</td><td>--</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Lklu CBS3082</td><td>1.28</td><td>2</td><td>123, 209; [372]</td><td>2</td><td>146, 232; [372]</td><td>1</td><td>288; [633]</td><td>--</td><td>--</td><td>0</td><td>-- in 1 kb; [722]</td><td>1</td><td>247; [935]</td><td>1</td><td>347; [448]</td><td>--</td><td>--</td><td>GLSSSAA</td><td>--</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Lthe CBS6340</td><td>0.18</td><td>3</td><td>89, 113, 155; [283]</td><td>3</td><td>111, 153, 177; [283]</td><td>1</td><td>142/157 [576]</td><td>--</td><td>--</td><td>1</td><td>573; [590]</td><td>0</td><td>-- in 1 kb; [708]</td><td>0</td><td>-- in 1 kb; [268]</td><td>--</td><td>--</td><td>GLSSSAA</td><td>--</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Kaes CBS4438</td><td>0.42</td><td>3</td><td>217, 247, 270; [634]</td><td>3</td><td>347, 370, 400; [634]</td><td>2</td><td>289, 311; [653]</td><td>--</td><td>--</td><td>0</td><td>(744)</td><td>1</td><td>122 in 1 kb</td><td>0</td><td>-- in 1 kb</td><td>--</td><td>--</td><td>GLSSSAA</td><td>--</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Klac CBS2359</td><td>0.77</td><td>4</td><td>245, 278, 300, 324; [765]</td><td>4</td><td>424, 448, 470, 503; [765]</td><td>2</td><td>435, 461; [890]</td><td>--</td><td>--</td><td>1</td><td>397; [1131]</td><td>2</td><td>286, 324; [650]</td><td>1</td><td>369; [536]</td><td>--</td><td>--</td><td>GLSSSAA</td><td>--</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Kdob CBS2104</td><td>0.32</td><td>4</td><td>264, 293, 314, 337; [729]</td><td>4</td><td>375, 398, 419, 448; [729]</td><td>2</td><td>416, 444; [836]</td><td>--</td><td>--</td><td>0</td><td>-- in 1 kb</td><td>3</td><td>322, 360, 411 in 1 kb</td><td>0</td><td>-- in 1 kb</td><td>--</td><td>--</td><td>GLSSSAA</td><td>--</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
<tr><td>Kmar CBS712</td><td>0.18</td><td>4</td><td>715, 750, 771, 796, (817); [1589]</td><td>2</td><td>755, 776, (801, 822, 857); [1589]</td><td>2</td><td>497, 531/546; [1504]</td><td>--</td><td>--</td><td>1</td><td>674 in 700 bp</td><td>2</td><td>366, 396 in 1 kb</td><td>0</td><td>-- in 1 kb</td><td>--</td><td>--</td><td>GLSSSAA</td><td>--</td><td>GAL1-10-7</td><td>GAL1-10</td><td></td></tr>
</table>

Note:
Gal4 site in the promoter examined: 1000 bp upstream or entire intergenic if more than 1000 bp; putative site more than 800 bp is considered as less likely to real and marked by parenthesis
If there are duplicate genes of other *GAL* genes, information will be separate by vertical line "|", such as an additional *GAL4* duplicate gene in Ncas

Table B.2: Signatures of relaxed selection on the *GAL* genes in the *Lachancea* species that grew slowly.

Alternative Model	***GAL1*** (p-value = 0.0457, K=0.8281)		***GAL7*** (p-value = 0.0005, K=0.5429)		***GAL10*** (p-value = 0.0258, K=0.8336)	
	Test branches	Ref branches	Test branches	Ref branches	Test branches	Ref branches
Site class I	0.0086 (79.85%)	0.0032 (79.85%)	0.0044 (77.79%)	0.0005 (77.79%)	0.0000 (68.51%)	0.0000 (68.51%)
Site class II	0.3413 (19.46%)	0.2730 (19.46%)	0.4451 (19.92%)	0.2251 (19.92%)	0.1633 (27.18%)	0.1137 (27.18%)
Site class III	56.9198 (0.69%)	131.6834 (0.69%)	2.4805 (2.29%)	5.3297 (2.29%)	1.9259 (4.31%)	2.1951 (4.31%)

Table B.3: A summary of *Lachancea GAL/MEL* gene features and growth in galactose or melibiose. The first "L" in species name column denotes the abbreviation of the genus name "Lachancea". "S" denotes Slow (slow growth), "+" denotes normal growth, "-" denotes no detectable growth, "Y" denotes "Yes" (presence of the gene), N denotes "No" (absence of the gene), "Syntenic" denotes the gene is syntenic to its ortholog in *Lachancea thermotolerans*. A cladogram is shown next to the table.



	Galactose growth	*GAL80*	*GAL4*	# Gal4 site *PGM1*	*MEL1*	Melibiose growth
Llanzarotensis	s	Syntenic	Syntenic	0	N	-
Lfantastica	s	Syntenic	Syntenic	0	N	-
Lmeyersii	-	Syntenic	Syntenic	0	N	-
Ldasiensis	+	Syntenic	Syntenic	1	Y	+
Lnothofagi	-	N	N	0	Y	s
Lthermotolerans	-	Syntenic	Syntenic	0	N	-
Lquebecensis	s	Syntenic	Syntenic	0	Y	s
Lwaltii	-	N	N	0	N	-
Lmirantina	-	N	N	0	N	-
Lfermentati	+	Syntenic	Syntenic	1	N	-
Lcidri	+	Syntenic	Syntenic	1	Y	+
Lkluyveri	+	Syntenic	Syntenic	1	N	-