Repeated Cis-Regulatory Tuning of a Metabolic Bottleneck Gene during Evolution

Meihua Christina Kuang,[1,2] Jacek Kominek,[1,3] William G. Alexander,[1,3] Jan-Fang Cheng,[4] Russell L. Wrobel,[1,3] and Chris Todd Hittinger[1,2,3,*]

[1]Laboratory of Genetics, Genome Center of Wisconsin, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI

[2]Graduate Program in Cellular and Molecular Biology, University of Wisconsin-Madison, Madison, WI

[3]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI

[4]DOE Joint Genome Institute, Walnut Creek, CA

***Corresponding author:** E-mail: cthittinger@wisc.edu.

Associate editor: Patricia Wittkopp

Abstract

Repeated evolutionary events imply underlying genetic constraints that can make evolutionary mechanisms predictable. Morphological traits are thought to evolve frequently through cis-regulatory changes because these mechanisms bypass constraints in pleiotropic genes that are reused during development. In contrast, the constraints acting on metabolic traits during evolution are less well studied. Here we show how a metabolic bottleneck gene has repeatedly adopted similar cis-regulatory solutions during evolution, likely due to its pleiotropic role integrating flux from multiple metabolic pathways. Specifically, the genes encoding phosphoglucomutase activity (*PGM1*/*PGM2*), which connect *GAL*actose catabolism to glycolysis, have gained and lost direct regulation by the transcription factor Gal4 several times during yeast evolution. Through targeted mutations of predicted Gal4-binding sites in yeast genomes, we show this galactose-mediated regulation of *PGM1/2* supports vigorous growth on galactose in multiple yeast species, including *Saccharomyces uvarum* and *Lachancea kluyveri*. Furthermore, the addition of galactose-inducible *PGM1* alone 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 galactose growth phenotypes between closely related species. This repeated mode of evolution suggests that this specific cis-regulatory connection is a common way that diverse yeasts can govern flux through the pathway, likely due to the constraints imposed by this pleiotropic bottleneck gene. Since metabolic pathways are highly interconnected, we argue that cis-regulatory evolution might be widespread at pleiotropic genes that control metabolic bottlenecks and intersections.

Key words: **cis-regulatory evolution, CRISPR/Cas9, galactose, metabolism, gene network, phosphoglucomutase.**

Article

Introduction

Repeated use of the same genes to achieve similar phenotypic outcomes is thought to reflect a combination of similar selective pressures and genetic constraints (Christin et al. 2010; Stern 2013). Both coding changes (Hoekstra et al. 2006; Christin et al. 2007) and cis-regulatory changes (Sucena et al. 2003; Prud'Homme et al. 2006; Rogers et al. 2013; Rebeiz and Williams 2017) have been shown to underlie repeated phenotypic alterations. Cis-regulatory changes have been hypothesized to be the key genetic causes of morphological evolution because strong pleiotropic constraints are imposed when key developmental genes are reused spatially and temporally (Prud'Homme et al. 2006; Carroll 2008; Stern and Orgogozo 2008; Rebeiz et al. 2009; Rebeiz and Williams 2017). In contrast, physiological and metabolic traits have frequently evolved through changes in both protein-coding and cis-regulatory regions (Ihmels et al. 2005; Lin et al. 2013; Roop et al. 2016; Sood and Brickner 2017). Mutations in coding regions have been frequently shown to lead 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). Cis-regulatory rewiring has also been associated with many physiological changes, including the transition from aerobic respiration to aerobic fermentation in yeasts, even as most central metabolic functions were conserved (Ihmels et al. 2005; Jiang et al. 2010; Lin et al. 2013; Roy et al. 2013). However, it is unknown whether there are any general principles that would lead specific cis-regulatory changes to occur frequently in metabolic evolution. Here we address this question by taking advantage of trait variation among yeast species in catabolism of the sugar galactose. As a paradigm of eukaryotic molecular biology, the metabolic and regulatory pathway for *GAL*actose utilization in the budding yeast model *Saccharomyces cerevisiae* offers a suite of well-characterized molecular features. The availability of several high-quality genome assemblies from diverse

budding yeast species further allows us to determine which pathway features are conserved, which are variable, and which are associated with trait variation.

Galactose concentrations are highly variable across environments (Marsilio et al. 2001; Nierop et al. 2001), providing the opportunity for diverse yeast species to adopt different strategies for consuming this resource. Qualitative differences in galactose utilization that evolved due to the parallel losses of entire *GAL* networks have received considerable prior attention (Hittinger et al. 2004; Riley et al. 2016). Recent functional comparisons of the *GAL* network in multiple yeast species have also demonstrated quantitative variation across diverse budding yeasts, generally focusing on pairwise comparisons (Martchenko et al. 2007; Peng et al. 2015; Dalal et al. 2016; Kuang et al. 2016; Roop et al. 2016; Sood and Brickner 2017). 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) (fig. 1). In *S. cerevisiae*, the metabolic bottleneck for galactose metabolism is controlled by the enzyme phosphoglucomutase, which catalyzes the interconversion of glucose-1-phosphate into glucose-6-phosphate (Bro et al. 2005; Garcia Sanchez et al. 2010; Hong et al. 2011). Phosphoglucomutase controls the flux from the Leloir pathway into glycolysis and integrates flux from several other pathways. This enzyme is encoded by *PGM1*, or a pair of paralogs (*PGM1* and *PGM2*) in yeasts that underwent whole genome duplication (WGD) (Wolfe and Shields 1997; Marcet-Houben and Gabaldon 2015). In *S. cerevisiae*, *PGM1* encodes the minor isoform of phosphoglucomutase, whereas *PGM2* encodes the major isoform (Bevan and Douglas 1969). Overexpressing *PGM2* is among the best ways to increase flux through the pathway in *S. cerevisiae* (Bro et al. 2005; Garcia Sanchez et al. 2010). However, *PGM1* and *PGM2* are not directly regulated by Gal4 in *S. cerevisiae* (Ren et al. 2000). The expression of *PGM1* in *S. cerevisiae* is not induced by galactose, whereas *PGM2* is only mildly induced (~4 fold) in a Gal4-independent manner (Oh and Hopper 1990; Rubio-Texeira 2005). In contrast, the Leloir enzymes are highly induced (up to 1,000 fold) in a Gal4-dependent manner (Lohr et al. 1995; Rubio-Texeira 2005). Unlike these galactose-specific enzymes, phosphoglucomutase is also involved in other metabolic pathways, including the pentose phosphate pathway (Cherry et al. 2012), glycogen biosynthesis (Hirata et al. 2003; Cherry et al. 2012), and trehalose biosynthesis (Mulet et al. 2004; Cherry et al. 2012).

Here, we show that expression of the bottleneck gene *PGM1/2* has been repeatedly tuned across the budding yeast family Saccharomycetaceae, which spans about 100 My of evolution, quantitatively modulating galactose metabolism by the addition or subtraction of Gal4-binding sites in its promoter. We show that Gal4-mediated regulation is necessary for vigorous galactose metabolism in multiple yeast species and that the addition of Gal4-regulated copies of *PGM1/2* are sufficient to confer vigorous growth to species that lack this regulation, such as *S. cerevisiae*. In contrast to increasing the basal expression level, modifying this cis-regulatory connection during evolution would have provided a mechanism for *PGM1/2* expression to respond specifically to galactose, which we hypothesize resolved the constraints imposed by converging metabolic pathways. These genetic constraints and continually shifting ecological niches likely underlie the repeated evolutionary gain and loss of the Gal4-binding sites upstream of this metabolic bottleneck gene.

Results

Galactose-Mediated Regulation of Its Bottleneck Gene Is Associated with Quantitative Variation

Yeasts of the family Saccharomycetaceae display dramatic variation in their abilities to grow on galactose. Growth experiments with galactose as the sole carbon source revealed widespread quantitative variation in galactose metabolism (fig. 2A, supplementary fig. S1A, Supplementary Material online). To understand the genetic basis of this variation, we applied a comparative approach that leveraged mechanistic understandings of the *GAL* network in *S. cerevisiae* and recent functional studies on *Saccharomyces uvarum* (formerly *S. bayanus* var. *uvarum*). Prior studies showed that *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 (Caudy et al. 2013; Kuang et al. 2016; Roop et al. 2016; Sood and Brickner 2017). To determine which, if any, genetic and molecular features were associated with quantitative variation in galactose growth between species, we examined 19 DNA sequence features of the *GAL* networks of 17 species from the family Saccharomycetaceae. We chose at least two species that lacked obvious mutations in their *GAL* genes from six genera with published high-quality genome assemblies. The 19 features included the number and position of Gal4-binding sites upstream of every *GAL* gene, copy number variation, and peptide motifs in the encoded enzymes. We found that the Gal4-binding sites upstream of the bottleneck gene *PGM1/2* had the strongest association with growth on galactose among all the characteristics examined (figs. 1 and 2; supplementary table S1 and fig. S1A, Supplementary Material online, regression coefficient $R^2 = 0.81$, $P = 1.9e\text{-}7$). This correlation was observed in several different media formulations containing galactose as the sole carbon source and when the temperature was varied (supplementary fig. S4, Supplementary Material online), but it was not seen on glucose (supplementary fig. S1*B* and *C*).

Direct Regulation of *PGM1/2* by Gal4 Quantitatively Modulates Growth on Galactose in Multiple Species

Although *PGM1/2* has not been shown to be directly regulated by Gal4 in any species, including *S. cerevisiae*, *S. uvarum* offered a particularly attractive model to test the hypothesis of direct regulation for several reasons: we previously showed that *S. uvarum PGM1* was induced 18-fold by galactose, has two predicted Gal4-binding sites, and its expression was further increased in mutants lacking the corepressor pair Gal80/Gal80b (Kuang et al. 2016). To test whether *S. uvarum PGM1* is directly regulated by Gal4, we mutated one base pair of a

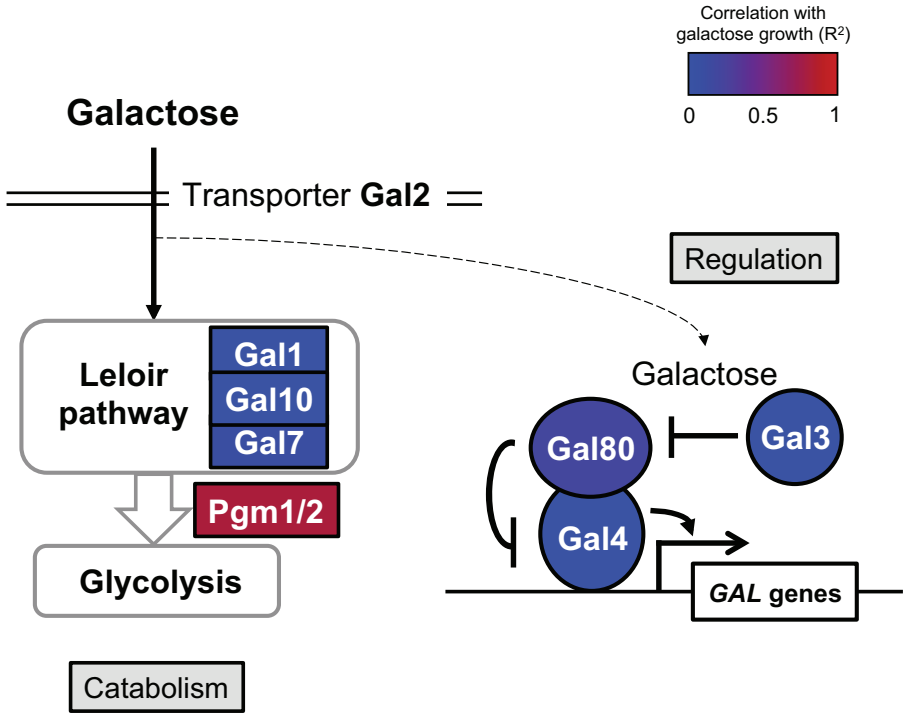

Fig. 1. Diagram of the *GAL* network in the budding yeast model *S. cerevisiae* and the role of Gal4-binding sites in explaining quantitative variation in growth on galactose. In the absence of galactose, the transcription factor Gal4 is inhibited by the corepressor Gal80, which prevents the expression of the *GAL*actose utilization genes. When galactose is present, Gal80 is sequestered by the co-inducer Gal3, allowing Gal4 to induce *GAL* gene expression. Other *GAL* genes encode the transporter Gal2 and three enzymes in the Leloir pathway that catabolize galactose. Glucose-1-phosphate, the end product of the Leloir pathway, is converted by the phosphoglucomutases Pgm1/2 into glucose-6-phosphate, which then enters glycolysis. For each protein, except for Gal2 (where paralogous hexose transporters complicated analyses), the correlation (R^2) between the number of predicted Gal4-binding sites upstream of the gene encoding it and growth on galactose across diverse yeast species (see fig. 2) is color-coded according to the key.

predicted Gal4-binding site ($CGGN_{11}CCG$) upstream of *S. uvarum PGM1* (Godecke et al. 1991; Hittinger and Carroll 2007; Robasky and Bulyk 2011). The Gal4-binding motif ($CGGN_{11}CCG$) has been shown to be conserved in both *Kluyveromyces lactis* and *S. cerevisiae*, which span about 100 My of evolution (Godecke et al. 1991; Hittinger and Carroll 2007), and the same motif is also enriched upstream of *S. uvarum GAL* genes (Kuang et al. 2016). In *S. cerevisiae*, point mutations in Gal4-binding motifs are sufficient to greatly decrease Gal4-binding strength and disrupt regulation (Giniger et al. 1985). We therefore reasoned that, if mutating the Gal4-binding motif resulted in galactose-specific growth defects, the predicted binding motif would be highly likely to be functional. Indeed, a single point mutation was sufficient to slow down growth on galactose by 20% compared with wild-type *S. uvarum* (fig. 3A), a defect that was galactose-specific (supplementary fig. S2A, Supplementary Material online). To further examine its impact on metabolic flux, we tested whether removing Gal4-mediated induction from this bottleneck gene sufficiently reduced the flux to rescue the temporary growth arrest phenotype caused by galactose metabolic overload seen in *S. uvarum* mutants lacking Gal80 corepressors (Kuang et al. 2016). As expected, the deletion of both of the predicted Gal4-binding sites upstream of *S. uvarum PGM1* rescued the temporary growth arrest phenotype, but also led to a slower maximum growth rate on galactose (supplementary fig. S3, Supplementary Material online). This evidence suggests that *S. uvarum PGM1* is a bottleneck gene that controls flux through the *GAL* pathway. Thus, 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.

To test whether Gal4-mediated induction of *PGM1/2* contributes to vigorous growth on galactose in multiple species, we examined the impact of mutating predicted Gal4-binding sites upstream of *PGM1/2*. We first developed a genome-editing approach potentially universal across yeasts by integrating a CRISPR/Cas9 system with an autonomously replicating sequence (ARS) that functions in diverse genera (Liachko and Dunham 2014). We applied this method to *PGM1* and *PGM2* genes with predicted upstream Gal4-binding sites in species with published transformation protocols. In addition to the genus *Saccharomyces*, this genome-editing system can induce targeted point mutations in at least the genera of *Lachancea* and *Kluyveromyces*, which diverged from *S. cerevisiae* about 100 Ma. Through either CRISPR/Cas9 or traditional approaches (Alexander et al. 2014), 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 [Hittinger et al. 2010]). The *L. kluyveri* mutant grew 32% more slowly on galactose

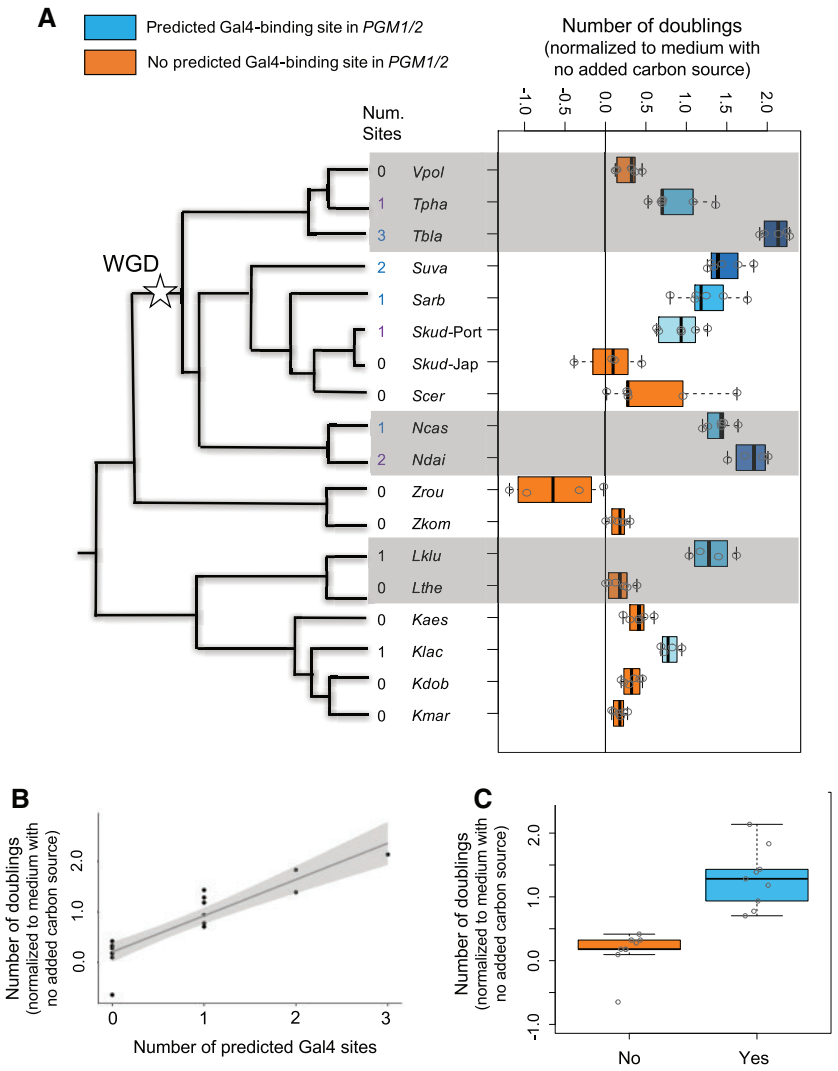

Fig. 2. Predicted Gal4-binding sites upstream of *PGM1/2* strongly correlate with growth on galactose across the yeast family Saccharomycetaceae. (*A*) The number of predicted Gal4-binding sites upstream of *PGM1* and *PGM2* strongly correlates with relative growth on galactose. The phylogeny and whole genome duplication (WGD) are shown at the left as published in prior genome-wide analyses (Shen et al. 2016). "Num. Sites" denotes the number of predicted Gal4-binding sites upstream of *PGM1* (blue for WGD species, black for non-WGD species) or *PGM2* (purple). The shades separate each monphyletic genus. Each data box is color-coded based on the number of binding sites, with the darkness of the blue color corresponding to the number of predicted binding sites, light blue indicating that the predicted binding sites had 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 h in synthetic complete medium (SC) +2% galactose, which was calculated as $\log_2[(OD_{strain} - OD_{media})/(OD_{start} - OD_{media})]$. This calculation was applied for all figures. The 15-h time point was chosen because most strains have started to initiate growth after 15 h in galactose (supplementary fig. S1*A*). Each strain is designated by a 4-letter species abbreviation (*Vpol*: *Vanderwaltozyma polyspora*, *Tpha*: *Tetrapisispora phaffii*, *Tbla*: *Tetrapisispora blattae*, *Suva*: *Saccharomyces uvarum*, *Sarb*: *Saccharomyces arboricola*, *Skud*-Port: *Saccharomyces kudriavzevii* Portuguese population, *Skud*-Jap: *Saccharomyces kudriavzevii* Japanese population (a negative control whose genome lacks a functional *GAL* network), *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*) There was a strong correlation between the number of predicted binding sites and growth on galactose. The data were extracted from 1A, and the median was used to represent each species. The gray shaded area corresponds to the 95% confidence interval. (*C*) Absence/presence of predicted binding sites (converted from 1B) revealed the same pattern ($P = 2.9e\text{-}5$, $n = 9$, $df = 13.9$, $t = -6.1$, Welch's two-sample *t*-test). Ancestral state reconstruction shows that putative Gal4-*PGM1/2* connections are evolutionarily dynamic with limited phylogenetic signal, supporting the treatment of taxa as independent (supplementary fig. S9, Supplementary Material online). Note that we were not able to obtain consistent growth with *Kazachstania africana*, so the genus *Kazachstania* was excluded. Otherwise, we included every characterized genus in this family where at least two species had published genome sequences.

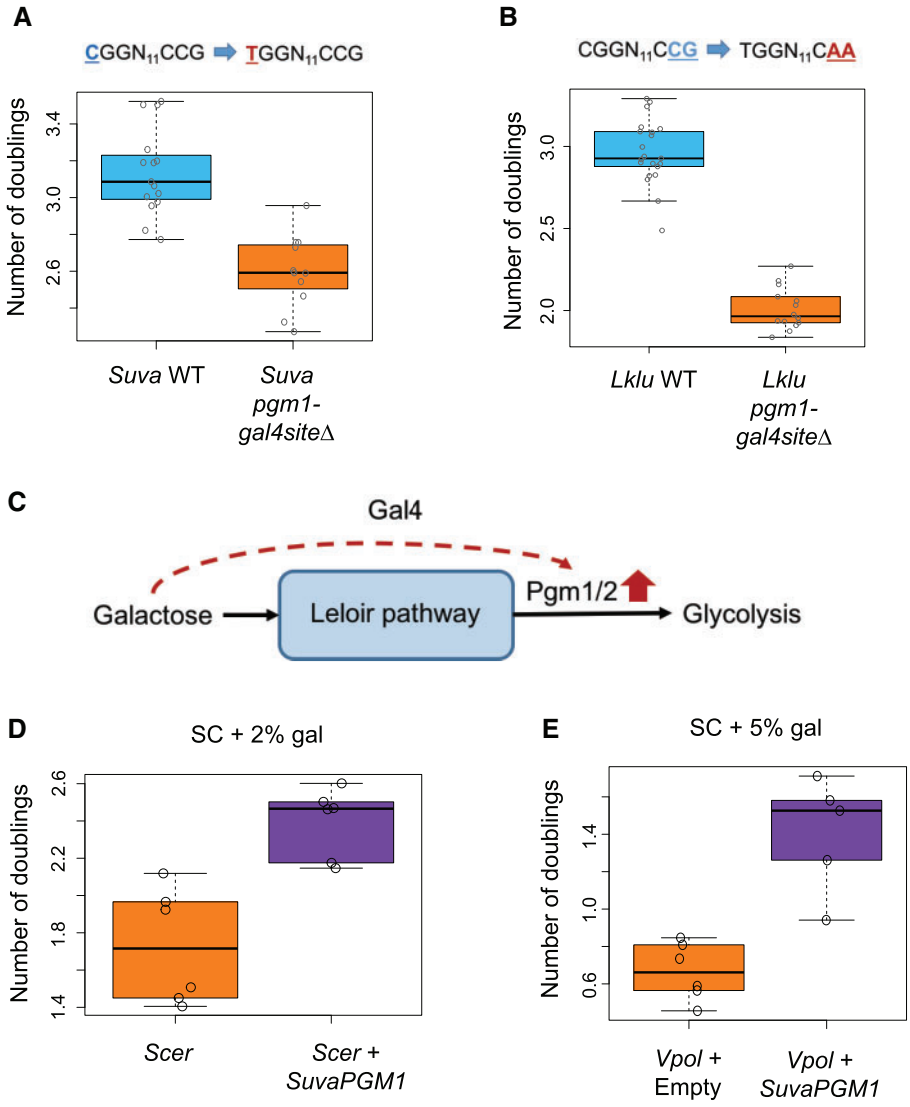

FIG. 3. Gal4-mediated regulation of phosphoglucomutase is necessary and sufficient to support vigorous growth on galactose for multiple yeast species. (*A*) Mutation of one of two predicted Gal4-binding sites upstream of *S. uvarum PGM1* reduced growth on galactose compared with wild type ($P = 6e\text{-}6$, $n_{WT} = 15$, $n_{mutant} = 11$). (*B*) Mutation of the predicted Gal4-binding site upstream of *L. kluyveri PGM1* reduced growth on galactose compared with wild type. ($P = 7.6e\text{-}9$, $n_{WT} = 21$, $n_{mutant} = 14$). Note that CCG was used as the PAM site for CRISPR/Cas9-engineering and was mutated to CAA to disrupt both the Gal4-binding site and the PAM site. (*C*) Hypothesis of *GAL* network activity tuning with a novel (dotted line) Gal4-Pgm1/2 feedforward loop. (*D*) *S. uvarum PGM1* increased growth on galactose in *S. cerevisiae* ($P = 5e\text{-}3$, $n = 6$). (*E*) *S. uvarum PGM1* increased growth on galactose in *V. polyspora* ($P = 0.01$, $n_{Empty\ vector} = 6$, $n_{SuvaPGM1} = 5$). All tests are Wilcoxon rank sum tests.

compared with wild type, but not on glucose, indicating that its Gal4-binding site is specifically required for vigorous growth on galactose (fig. 3*B* and supplementary fig. S2*B*, Supplementary Material online). However, the *K. lactis* and *S. kudriavzevii* mutants lacked observable growth defects, indicating that predicted Gal4-binding sites were not required for robust growth in the conditions we tested (supplementary fig. S5*A* and *B*, Supplementary Material online). However, full induction of *S. kudriavzevii PGM2* by galactose did require the predicted Gal4-binding site (supplementary fig. S5*E*, Supplementary Material online). Even though this site was required to reach expression levels similar to *S. uvarum PGM1*, the higher basal expression of *S. kudriavzevii PGM2* might render Gal4 induction dispensable to support its relatively modest growth (supplementary fig. S5*C* and *D*, Supplementary Material online). Thus, we conclude that direct regulation of *PGM1/2* by Gal4 supports vigorous galactose growth in some species (e.g., *S. uvarum* and *L. kluyveri*), whereas other mechanisms are important for other species.

To examine whether up-regulating phosphoglucomutase expression alone was sufficient to improve growth on galactose across yeast species, we introduced a Gal4-regulated *PGM1* gene (*S. uvarum PGM1* with both of its predicted Gal4-binding sites) 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 is predicted to enhance flux through the *GAL* pathway while minimizing pleiotropic effects when cells are not grown on galactose (supplementary figs. S2 and S5*C* and *D*, Supplementary Material online) (Kuang et al. 2016).

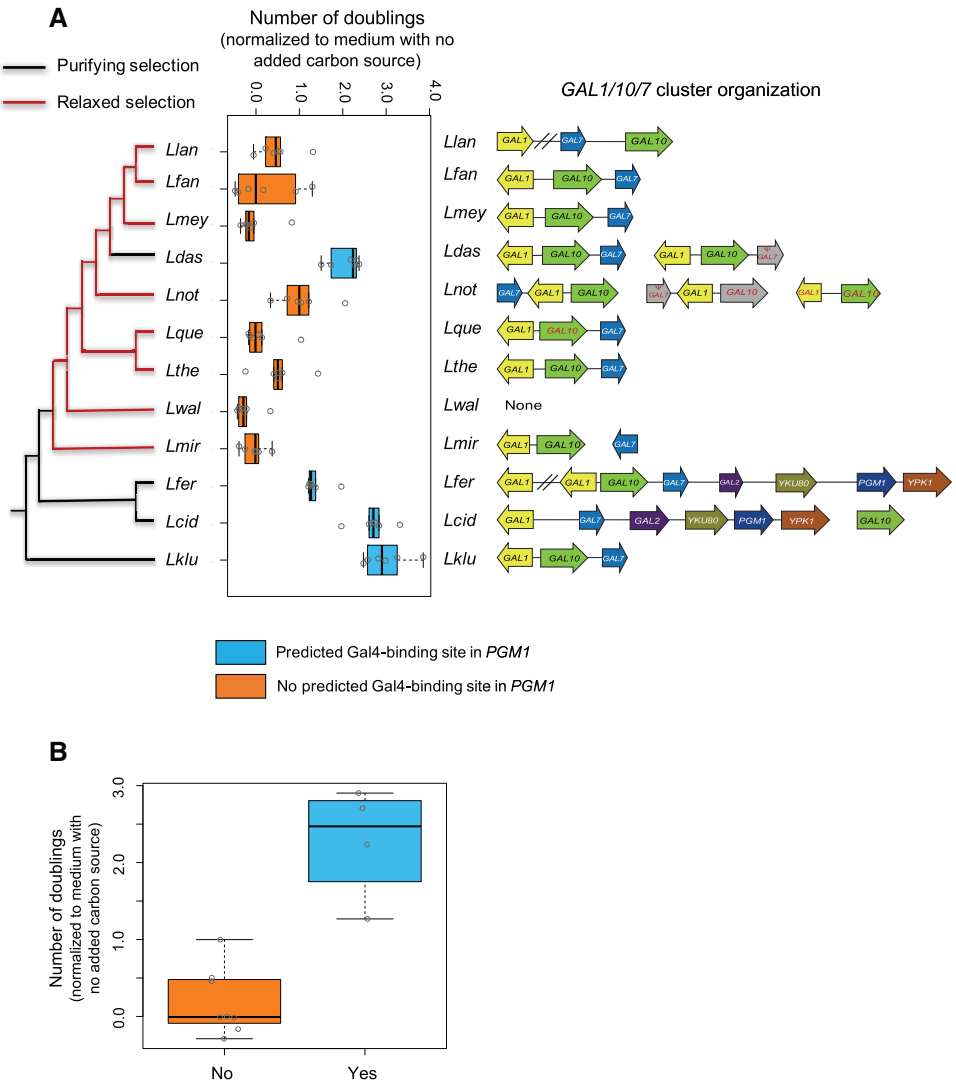

Fig. 4. Predicted Gal4-binding sites upstream of *PGM1* predicted galactose growth differences among closely related species in the genus *Lachancea*. (A) A genome-wide consensus phylogeny (Vakirlis et al. 2016) shows branches under normal purifying selection as black and those under relaxed selection as red (table 1). In the boxplot ($n = 6$), each data box is colored coded by either blue (presence of predicted Gal4-binding sites) or orange (absence). The organization of the required *GAL1/10/7* gene cluster is represented on the right. Each arrow denotes the direction of transcription. Distances are proportional, except for regions marked by two slashes. Genes on the same chromosome are connected with a black line. Each homolog is color-coded. Pseudogenes (Ψ) or likely pseudogenes 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'* nom. nud., *Llan*: *L. lanzarotensis*). All strains were cultured in SC + 2% galactose. (*B*) The presence of predicted Gal4-binding sites associates with faster growth on galactose (median of each species from 3A is plotted as individual dots) ($P = 5.8e\text{-}3$, $n_{no} = 4$ $n_{yes} = 8$, $t = -5.30$, $df = 4.08$, Welch's two-sample *t*-test).

Introduction of *S. uvarum PGM1* into *S. cerevisiae* enhanced its growth on galactose by 56% (fig. 3*D*), whereas it enhanced the growth of *Vanderwaltozyma polyspora* by 110% (fig. 3*E*). Therefore, up-regulation of *PGM1* alone is sufficient to increase the growth on galactose of multiple species.

GAL Network Atrophy Did Not Affect the Protein-Coding Sequence of the Pleiotropic Bottleneck Gene

To test whether the strong association of galactose growth with predicted Gal4 regulation of *PGM1* represents a general trend, we examined several additional closely related species in the genus *Lachancea*, which includes one of the species whose Gal4-mediated regulation we verified above (fig. 3*B*). This genus also provided a good opportunity to examine gene-trait association because all 12 known species in this family have high-quality genome assemblies available (Sarilar et al. 2015; Vakirlis et al. 2016), more than any other monophyletic budding yeast genus, including *Saccharomyces*. Similar to what we saw in our sparser sampling across the Saccharomycetaceae (fig. 2), we found that the presence of a predicted Gal4-binding site was highly correlated with

galactose growth (fig. 4 and supplementary fig. S6, Supplementary Material online). Perhaps most strikingly, the topology of the species tree and the novel location of its predicted Gal4-binding site (318 bp upstream of *PGM1* and overlapping with the neighboring predicted coding sequence) suggest that *Lachancea dasiensis* may have reacquired the ability to grow vigorously on galactose by acquiring a new Gal4-binding site (fig. 4). In fact, even though the genomes of several species of *Lachancea* were predicted to encode functional *GAL* genes, they grew so slowly on 2% galactose that taxonomists had previously scored them as nongrowing, weak, or variable (Kurtzman et al. 2011). In these cases, the presence of a predicted Gal4-binding site upstream of *PGM1* was actually a better predictor of galactose growth in the conditions we tested than the presence of *GAL* genes. We hypothesized that these slow-growing species may have experienced changes in the strength of purifying selection acting on their *GAL* genes, and indeed, we found statistically significant relaxations of the selective pressure acting against nonsynonymous substitutions in all three *GAL* genes encoding enzymes (table 1, $P = 7e\text{-}5$, Fisher's method). In contrast, we did not detect any signal of relaxed selection in the pleiotropic gene *PGM1* (table 1). Some of these species lost *GAL* genes through pseudogenization or deletion, including *GAL4* and *GAL80* in some cases, as well as experiencing translocations and gene duplications (fig. 4 and supplementary fig. S7 and supplementary notes, Supplementary Material online). Thus, we propose that the loss of the Gal4-*PGM1* regulatory connection and the relaxed selection on components dedicated to galactose metabolism may represent an early stage of *GAL* network atrophy that, in some cases, led to degeneration and complete loss.

Conditional Benefits of Direct Regulation

To further model how Gal4-*PGM1/2* regulatory connections evolved in the family Saccharomycetaceae, we performed ancestral state reconstruction using the R packages Geiger (Harmon et al. 2008) and phytools (Revell 2012). Although there was limited signal to resolve individual nodes, all likely evolutionary trajectories involve multiple gains and losses of predicted Gal4-*PGM1/2* regulatory connections during evolution (supplementary fig. S9, Supplementary Material online). Among taxa that underwent the whole genome duplication, there did not appear to be any pattern of which paralog gained or lost sites, but all of the species examined were predicted to only have Gal4-binding sites upstream of a single *PGM1/2* gene (fig. 1 and supplementary table S1, Supplementary Material online). The repeated gains and losses of Gal4-binding sites upstream of *PGM1/2* during evolution made us wonder whether the effects of Gal4-binding sites might be associated with specific galactose conditions. We hypothesized that the galactose-inducibility of *PGM1/2* could affect growth more strongly as galactose concentrations increased. Consistent with this hypothesis, when the galactose-inducible *S. uvarum PGM1* was added to *V. polyspora*, it grew much better than the wild-type strain at higher concentrations (5% galactose), but not at lower concentrations (0.5% galactose) (fig. 5*B*). Additionally, mutation of the

Table 1. Compared with Reference (Ref) Branches, Slow-Growing *Lachancea* spp. that Grew Slowly and Lacked Predicted Gal4-*PGM1* Regulation (test branches) Experienced Relaxed Selection at Dedicated *GAL* Genes, but not at the Pleiotropic Gene *PGM1*.

	GAL1 (P Value = 0.0457, K = 0.8281)		*GAL7* (P Value = 0.0005, K = 0.5429)		*GAL10* (P Value = 0.0258, K = 0.8336)		*PGM1* (P Value = 0.8985, K = 0.9934)	
Alternative Model	**Test Branches**	**Ref Branches**	**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%)	0.0101 (94.25%)	0.0098 (94.25%)
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%)	1 (4.26%)	1 (4.26%)
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%)	1.7126 (1.49%)	1.7188 (1.49%)

NOTE.—Test branches refer to internal and terminal branches leading to species growing slowly on galactose (fig. 4A). Reference (Ref) branches refer to all the remaining branches on the tree. The analysis was conducted using the RELAX method implemented in the HYPHY package, and K is the relaxation-of-selection parameter (Wertheim et al. 2015). Site classes I–III refer to groups of sites evolving under a shared selection regimen, in a descending order of the number of sites they represent (i.e., class I is the most frequently observed, and class III is the least frequently observed). Values indicate d_N/d_S ratios (ω) for each class of sites, along with the proportion of sites that belong to that class in parentheses. Note that the Test *GAL* genes are evolving significantly nearer to neutrality than the Ref *GAL* genes (i.e., ω nearer to 1 for Test than Ref), whereas Test and Ref rates are similar for *PGM1*.

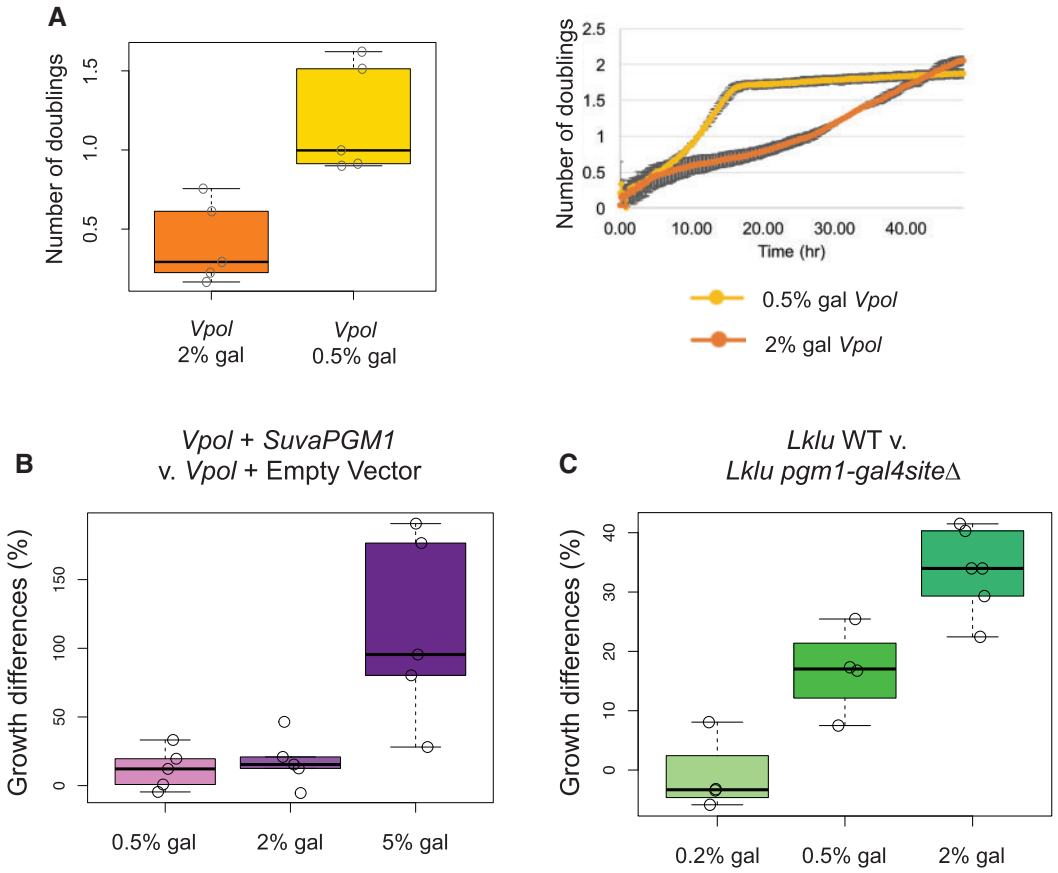

Fig. 5. The importance of Gal4-mediated regulation of *PGM1* depends on galactose concentrations. (*A*) *V. polyspora* grew better at low concentrations of galactose ($P = 1.7e\text{-}2$, $n = 5$, Wilcoxon rank sum test). Box plots are shown at the left and growth curves at the right. (*B*) When added to *V. polyspora*, *S. uvarum PGM1* only increased galactose growth at high galactose concentrations. Growth differences were calculated as: $(Div_{SuvaPGM1} - Div_{\text{Empty vector}})/Div_{\text{Empty vector}} \times 100\%$, where "*Div*" denotes the number of cell divisions at 15 h on 0.5% galactose and 30 h on 2% or 5% galactose. Note that time points were chosen immediately prior to when the wild type saturated. The gain in growth conferred by *SuvaPGM1* significantly increased as the galactose concentrations increased ($P = 1.6e\text{-}2$, $n = 5$, Jonckheere–Terpstra test). (*C*) Mutating the Gal4-binding site upstream of *PGM1* caused greater growth defects in *L. kluyveri* as galactose concentrations increased. Growth differences were calculated as: $(Div_{WT} - Div_{\text{Mutant}})/Div_{WT} \times 100\%$, where "*Div*" denotes the number of cell divisions after 15 h on 2% galactose, 10 h on 0.5% galactose, and 9 h on 0.2% galactose. Note that time points were chosen immediately prior to when the wild type saturated. The growth penalty caused by deleting the Gal4-binding site significantly decreased as the galactose concentrations decreased ($P = 2.4e\text{-}3$, $n = 4$, Jonckheere–Terpstra test). The error bars in B and C are standard deviations.

native Gal4-binding site of *L. kluyveri PGM1* caused stronger defects as galactose concentrations increased (fig. 5C). Also consistent with this model, several species of *Lachancea* that had lost Gal4-*PGM1* sites actually grew better at low concentrations than high concentrations of galactose, suggesting that they might be specialists at low concentrations of galactose (supplementary fig. S8, Supplementary Material online). This trait was not exclusive to this genus but was shared with the distantly related species *V. polyspora*, where it was particularly pronounced (fig. 5A, Supplementary Material online). These data suggest that direct induction of *PGM1/2* by Gal4 has stronger phenotypic impacts at high concentrations of galactose, but its effects are limited at low concentrations.

Discussion

Cis-Regulatory Tuning of Gene Expression and Metabolic Flux during Evolution

In summary, we have shown that, unlike the model yeast *S. cerevisiae*, several yeast species contain direct regulatory connections between Gal4 and the metabolic bottleneck gene *PGM1/2*. Galactose-mediated induction of *PGM1/2* is required for vigorous growth in at least two yeast species separated by about 100 My of evolution. Moreover, upregulation of *PGM1/2* alone is sufficient to increase galactose growth in multiple species. The addition of Gal4 regulation to *PGM1/2* provides a way to specifically increase galactose metabolism during evolution. 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 in the *GAL* network (Hittinger et al. 2004; Riley et al. 2016), we propose that variation in the number of Gal4-binding sites upstream of *PGM1/2* quantitatively tunes flux through the metabolic pathway in a condition-dependent manner (fig. 6). The repeated gains and losses of the Gal4-*PGM1/2* regulatory connection may have been driven by variation in

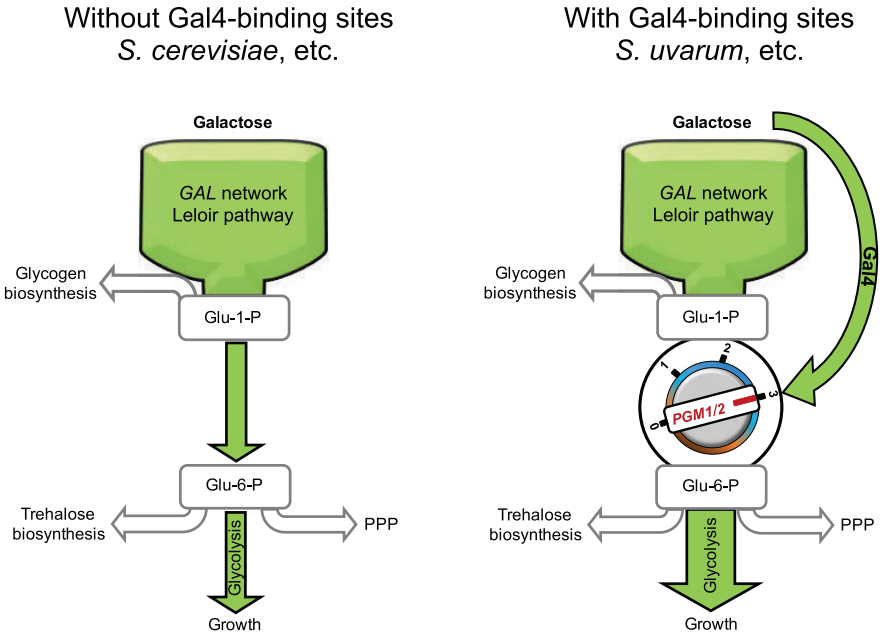

Fig. 6. The solutions to tune *GAL* network activities are limited by pleiotropic constraints on a metabolic bottleneck that integrates flux from multiple pathways. Metabolic flux from the Leloir/*GAL* pathway to glycolysis is controlled by the activity of the bottleneck gene(s) *PGM1/2*, which encode enzyme(s) that catalyze the interconversion of glucose-1-phosphate and glucose-6-phosphate. Unlike the rest of the *GAL* pathway, the activity encoded by *PGM1/2* is evolutionarily constrained by at least four converging metabolic pathways, including the *GAL* pathway, glycogen biosynthesis, trehalose biosynthesis, and pentose phosphate pathway (PPP). The left panel shows the scenario without the Gal4-binding sites, such as in *S. cerevisiae* and *V. polyspora*. The right panel shows how a "cis dial" regulating *PGM1/2* (with a varying number of Gal4-binding sites) is able to bypass the pleiotropic constraints to specifically increase flux from the *GAL* pathway to glycolysis in response to galactose, such as in *S. uvarum* and *L. kluyveri*.

galactose availability across yeast ecological niches, leading to evolutionary changes through genetic drift, adaptation, or both.

Mutations leading to transcription factor binding site gains and losses are common (Stone and Wray 2001), raising the possibility that the Gal4-*PGM1/2* regulatory connection is evolutionarily labile primarily because regulatory state changes are so easy to achieve. Nonetheless, except for *PGM1/2*, the Gal4-binding sites upstream of *GAL* genes are well conserved within the genus *Saccharomyces* (Cliften et al. 2003; Kellis et al. 2003), and our broader analyses extend this trend across the family Saccharomycetaceae (supplementary table S1, Supplementary Material online). Thus, *PGM1/2* expression has been coupled or uncoupled from Gal4-regulation multiple times to a different degree than the canonical *GAL* genes. Other authors have proposed that enzymes controlling flux, including those at beginnings and intersections of metabolic pathways, are more likely to be under strong selective pressures (Flowers et al. 2007; Wright and Rausher 2010; Olson-Manning et al. 2013). Since Pgm1/2 controls the entry point from the Leloir pathway into glycolysis and enzymes acting on glucose-6-phosphate have previously been suggested to be targets of positive selection in *Drosophila* (Flowers et al. 2007), alterations to the expression or activity of this specific metabolic bottleneck may be particularly likely to affect phenotype.

Despite the importance of direct regulation of *PGM1/2* by Gal4 in many species, other mechanisms to modify *GAL* network activity 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 that a similar mechanism may exist in *S. kudriavzevii*. We reasoned that the combination of *PGM1/2* basal expression and Gal4-independent induction may be sufficient to support low-to-moderate growth rates, perhaps because the metabolic activities upstream of phosphoglucomutase are lower such that phosphoglucomutase activity is not limiting (Hittinger et al. 2010). Thus, the 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, high-flux *GAL* networks, with the novel Gal4-*PGM1/2* feedforward loop characterized here, also tend to have retained both copies of the duplicate genes encoding homologs of the Gal80 corepressors (*Tetrapisispora blattae*, *Naumovozyma castellii*, *Naumovozyma dairenensis*, and *S. uvarum*). These dual corepressors may lead to a more robust negative feedback loop that prevents the previously characterized phenomenon of metabolic overload (Kuang et al. 2016) (table 2 and supplementary table S1, Supplementary Material online). Thus, multiple genetic changes likely coordinate with the cis-regulatory changes in *PGM1/2* to quantitatively tune *GAL* network activity.

Pleiotropic Constraints, Network Architectures, and the Predictability of Evolution

The dynamic evolution of the Gal4-*PGM1/2* regulatory connection implies that the possible ways that this bottleneck

Table 2. The Three Species Growing Fastest on Galactose have the Highest Number of Predicted Gal4-Binding Sites and Retain the Corepressor Pair Encoded by *GAL80/80B*.

Rank of Galactose Growth in This Family	Species	Number of Predicted Gal4-Binding Sites Upstream of *PGM1/2*	*GAL80* and *GAL80B*
1	*Tetrapisispora blattae*	3	Both
2	*Naumovozyma dairenensis*	2	Both
3	*Saccharomyces uvarum*	2	Both
4	*Naumovozyma castellii*	1	*GAL80* only
5	*Lachancea kluyveri*	1	*GAL80* only

gene responds to selection are constrained by pleiotropy. The growth rate on galactose is constrained by the activities of both the Leloir pathway and the phosphoglucomutase bottleneck. The enzymes of galactose catabolism are conserved from bacteria to yeasts to humans, including the *PGM1* homologs (Lu and Kleckner 1994). Unlike the rest of the *GAL* pathway, *PGM1* homologs are also involved in the pentose phosphate pathway (Cherry et al. 2012), glycogen biosynthesis (Hirata et al. 2003; Cherry et al. 2012), and trehalose biosynthesis (Mulet et al. 2004; Cherry et al. 2012) (fig. 6). Consistent with the hypothesis that the protein-coding sequence of *PGM1* is pleiotropic, yeast species that have lost Gal4-*PGM1* regulation have experienced relaxed selection on the protein-coding sequences of their dedicated *GAL* genes but not on their *PGM1* homologs (table 1). Even in yeast species that have lost dedicated *GAL* pathway genes and cannot utilize galactose, *PGM1* homologs are retained (Hittinger et al. 2004; Hittinger et al. 2010; Riley et al. 2016). The intersection of these metabolic pathways at the step controlled by *PGM1/2* likely constrains the flux of the entire *GAL* pathway. Although there are many potential ways to modify the bottleneck activity, such as by modifying basal expression or coding sequence changes in *PGM1/2*, recruiting a Gal4-binding site to specifically induce *PGM1/2* expression in response to galactose would increase flux through the *GAL* pathway, yet minimize the pleiotropic effects on other pathways in different environmental contexts.

If one envisions the rate of galactose growth as a continuous spectrum, there are many ways to marginally increase or decrease galactose growth during evolution. However, we propose that there are relatively few alternatives and many constraints to evolving a highly active *GAL* network. Up-regulating the bottleneck activity controlled by phosphoglucomutase through direct regulation of *PGM1/2* by Gal4 provides a conditional way to increase expression on galactose without pleiotropic effects on other carbon sources (fig. 6). It is likely that changes in other *GAL* genes and interacting pathways are also involved, but our analyses (supplementary table S1, Supplementary Material online) argue these changes are less repeatable, and therefore, less predictable than the novel Gal4-*PGM1/2* regulatory connection that we have characterized here. Thus, we propose that the architecture of the *GAL* network and the pleiotropy of the metabolic bottleneck gene *PGM1/2* constrain the possible outcomes and lead to the repeated evolutionary mechanisms observed (fig. 6). We further hypothesize that greater constraints may lead to higher predictability more generally. Nonetheless, the high likelihood of cis-regulatory mutations means that genetic drift probably also contributes to the repeated variation observed at this locus.

In other conserved gene networks regulating metabolism, important genes with pleiotropic roles may also display predictable evolutionary patterns for reasons that are analogous to the spatial and temporal constraints imposed by developmental regulatory networks (Carroll 2005; Carroll 2008; Stern and Orgogozo 2008; Rebeiz et al. 2009; Stern 2013). Under this model, metabolic genes that handle flux from multiple pathways would be particularly likely to 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, we argue that metabolic genes that encode enzymes that are highly interconnected and reused by multiple pathways fall under constraints that favor cis-regulatory tuning of gene expression.

Materials and Methods

Strain Construction

All strains used this study are listed in supplementary table S2, Supplementary Material online. Two genetic engineering approaches were used to introduce point mutations to mutate the predicted Gal4-binding sites: traditional methods and a CRISPR/Cas9-based approach potentially generalizable to the family Saccharomycetaceae and beyond:

Traditional methods (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. Transformations of *S. uvarum* and *S. kudriavzevii* were conducted as previously described (Hittinger et al. 2010; Alexander et al. 2014; Kuang et al. 2016).

The *SuvaPGM1* expression plasmids were constructed as follows: the *S. uvarum PGM1* coding sequence, together with 800 bps upstream and downstream, was inserted into the panARS vector pIL75 (Liachko and Dunham 2014) at the multiple cloning site digested by *Sma*I. This sequence was 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 and constructs were verified by Sanger sequencing.

CRISPR/Cas9 Approach (Lachancea kluyveri and Kluyveromyces lactis)

The backbone of two vectors, a sgRNA expression cassette (GenBank MG680559) and pKOPIS (GenBank MG680557),

were synthesized by the DOE Joint Genome Institute DNA Synthesis Science Program. The yeast sgRNA expression cassette contains the *SNR52* promoter, HDV ribozyme linked to a cloning site for the sgRNA construct, and the *SNR52-1* terminator. pKOPIS contains a *KanMX* selectable marker and encodes a Cas9 protein driven by the constitutive *RNR2* promoter and codon-optimized for expression in *S. cerevisiae*. A pXIPHOS-panARS vector (GenBank MG835323) was subsequently constructed from pKOPIS through multiple modifications: adding of an *Escherichia coli* ampicillin resistance marker, swapping the *KanMX* marker with the *NatMX* marker, and swapping the 2-μ origin with an autonomously replicating sequence (ARS) that is stable in several yeast genera and was cloned from the panARS vector pIL75 (Liachko and Dunham 2014). Through Gibson assembly, a single final vector was assembled from pXIPHOS-panARS by inserting both a target-specific sgRNA that was amplified from the sgRNA cassette and a repair template generated by PCR next to *NatMX*. This final vector thus encoded Cas9, a custom sgRNA, and the repair template needed for the gene-editing event. Vectors were electroporated into *Lachancea kluyveri* and *Kluyveromyces lactis* following previously described methods (Gojkovic et al. 2000; Kooistra et al. 2004).

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, 20 g/l glucose; fig. 2 and 3E only) and precultured for 2–3 days. The growth assays were conducted in the stated media 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 min at 595 nm. "Relative growth" in each figure denotes the number of cell divisions after 15 h or at the indicated time point, which was calculated as $\log_2[(OD_{strain} - OD_{media})/(OD_{start} - OD_{media})]$. This equation normalizes the optical density at each time point (OD_{strain}) by the starting optical density (OD_{start}) of that culture and the optical density of the medium (OD_{media}). In some cases (stated in the Y-axis label), relative growth was normalized to a 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 the division number was then calculated as $Division_{carbon\ source} - Division_{no\ carbon\ source}$. Strains from each species were tested at either room temperature or 30°C to determine their preferred growth temperatures, unless their temperature preferences were already well known. In each case, each species was cultured at its expected growth temperature (room temperature or 30°C), except when species with different optimal temperatures were cultured in the same 96-well plate. In these cases (figs. 2 and 4; supplementary figs. S1, S4*B* and *C*, S5, and S6, Supplementary Material online), strains were grown at room temperature (22–24°C). Replicates were defined as either biological replicates that were independent transformants or the same strains started from independent precultures, whereas technical replicates were where the same preculture was inoculated into independent growth assays.

Fluorescence Measurement Assays

Strains were cultured as described above. Both fluorescence levels and absorbance of each well were measured by an unshaken BMG FLUOstar Omega plate reader every 10–20 min, with the excitation filter at 485 nm, emission filter at 520 nm, and absorbance at 595 nm. The nonfluorescent wild-type strain was included as a control to correct for autofluorescence. The auto-fluorescence levels were first subtracted from the measured fluorescence levels, which were then normalized to absorbance to control for cell density variation. Replicates were defined the same as in the above growth assays.

Relaxed Selection Analysis

Nucleotide sequences of the *GAL1*, *GAL7*, *GAL10*, and *PGM1* genes were extracted from every characterized *Lachancea* species (except *Lachancea waltii*, which lost its entire *GAL* network [Hittinger et al. 2004]) and three outgroup *Kluyveromyces* species (*K. lactis, Kluyveromyces marxianus, and Kluyveromyces dobzhanskii*). In cases where duplicate genes were found, all gene copies were analyzed. All sequences 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 pseudoreplicates. 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.

Ancestral State Reconstruction

Reconstruction of ancestral states was performed by first scoring each taxon shown in figure 2 for the absence (scored as 0) or presence (scored as 1) of predicted Gal4-binding sites upstream of *PGM1/2*. We then compared the Equal Rates and the All Rates Different models of discrete character evolution using the R packages Geiger v2.0.6 (Harmon et al. 2008) and phytools v0.4.56 (Revell 2012) to determine the best-fitting model based on the Corrected Akaike Information Criterion (Akaike 1974). Finally, we simulated 1,000 stochastic character maps on the phylogeny under the best-fitting model using stochastic mutational mapping (Bollback 2006) and obtained the posterior probability (PP) of each character state at each internode of the phylogeny.

Statistical Analysis

All *P* values were two-sided and calculated using Welch's two-sample *t*-test (fig. 2C and supplementary figs. S1C and 4*B* and *C*, Supplementary Material online) or a conservative

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 the variance. To take into account the effects of decreasing galactose concentrations (fig. 5B and C, Supplementary Material online), we used the ordered Jonckheere–Terpstra test (two-sided), and *P* values from independent experiments were subsequently combined using Fisher's method. These tests were performed using Mstat software version 6.1.4 (http://mcardle.oncology.wisc.edu/mstat/; last accessed May 21, 2018). For all boxplots, the elements are defined as follows: the bottom and top of the box are the lower and upper quartiles, respectively; the band within the box is the median; and the lower and upper whiskers represent 1.5 interquartile ranges.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Kenneth H. Wolfe, Gilles Fischer, and Cletus P. Kurtzman for yeast strains; Maitreya J. Dunham for the panARS vector pIL75; Xing-Xing Shen for advice on ancestral state reconstruction; 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 and synthesis of a CRISPR/Cas9 system for *S. cerevisiae*. This work was supported by the National Science Foundation (Grant numbers DEB-1253634, DEB-1442148); by the USDA National Institute of Food and Agriculture (Hatch Project 1003258); and funded in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409 and DE-FC02-07ER64494). C.T.H. is a Pew Scholar in the Biomedical Sciences and a Vilas Faculty Early Career Investigator, supported by the Pew Charitable Trusts and the Vilas Trust Estate, respectively. 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.

Author Contributions

M.C.K. and C.T.H. designed the experiments and wrote the paper. M.C.K. conducted the experiments and performed most analyses. J.K. conducted the relaxed selection and ancestral state reconstruction analyses. W.G.A., J.-F.C., R.L.W., M.C.K., and C.T.H. designed and constructed the panARS-CRISPR/Cas9 system.

References

Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 19(6):716–723.

Alexander WG, Doering DT, Hittinger CT. 2014. High-efficiency genome editing and allele replacement in prototrophic and wild strains of *Saccharomyces*. *Genetics* 198(3):859–866.

Bevan P, Douglas HC. 1969. Genetic control of phosphoglucomutase variants in *Saccharomyces cerevisiae*. *J Bacteriol.* 98(2):532–535.

Bollback JP. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7:88.

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(11):6465–6472.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1):25–36.

Carroll SB. 2005. Evolution at two levels: on genes and form. *PLoS Biol.* 3(7):e245.

Caudy AA, Guan Y, Jia Y, Hansen C, DeSevo C, Hayes AP, Agee J, Alvarez-Dominguez JR, Arellano H, Barrett D, et al. 2013. A new system for comparative functional genomics of *Saccharomyces* yeasts. *Genetics* 195(1):275–287.

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, et al. 2012. Saccharomyces genome database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40(D1):D700–D705.

Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr Biol.* 17(14):1241–1247.

Christin PA, Weinreich DM, Besnard G. 2010. Causes and evolutionary significance of genetic convergence. *Trends Genet.* 26(9):400–405.

Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M. 2003. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* 301(5629):71–76.

Dalal CK, Zuleta IA, Mitchell KF, Andes DR, El-Samad H, Johnson AD. 2016. Transcriptional rewiring over evolutionary timescales changes quantitative and qualitative properties of gene expression. *eLife* 5:e18981.

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454(7205):762–765.

Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, Eanes WF. 2007. Adaptive evolution of metabolic pathways in *Drosophila*. *Mol Biol Evol.* 24(6):1347–1354.

Garcia Sanchez R, Hahn-Hagerdal B, Gorwa-Grauslund MF. 2010. *PGM2* overexpression improves anaerobic galactose fermentation in *Saccharomyces cerevisiae*. *Microb Cell Fact.* 9:40.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.

Giniger E, Varnum SM, Ptashne M. 1985. Specific DNA binding of *GAL4*, a positive regulatory protein of yeast. *Cell* 40(4):767–774.

Godecke A, Zachariae W, Arvanitidis A, Breunig KD. 1991. Coregulation of the *Kluyveromyces lactis* lactose permease and beta-galactosidase genes is achieved by interaction of multiple LAC9 binding sites in a 2.6 kbp divergent promoter. *Nucleic Acids Res.* 19(19):5351–5358.

Gojkovic Z, Jahnke K, Schnackerz KD, Piskur J. 2000. *PYD2* encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *J Mol Biol.* 295(4):1073–1087.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24(1):129–131.

Hirata Y, Andoh T, Asahara T, Kikuchi A. 2003. Yeast glycogen synthase kinase-3 activates Msn2p-dependent transcription of stress responsive genes. *Mol Biol Cell* 14(1):302–312.

Hittinger CT, Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449(7163):677–681.

Hittinger CT, Goncalves P, Sampaio JP, Dover J, Johnston M, Rokas A. 2010. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 464(7285):54–58.

Hittinger CT, Rokas A, Carroll SB. 2004. Parallel inactivation of multiple *GAL* pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci U S A.* 101(39):14144–14149.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313(5783):101–104.

Hong KK, Vongsangnak W, Vemuri GN, Nielsen J. 2011. Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. *Proc Natl Acad Sci U S A.* 108(29):12179–12184.

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(5736):938–940.

Jiang H, Guan W, Gu Z. 2010. Tinkering evolution of post-transcriptional RNA regulons:xPuf3p in fungi as an example. *PLoS Genet.* 6(7):e1001030.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423(6937):241.

Kooistra R, Hooykaas PJJ, Steensma HY. 2004. Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* 21(9):781–792.

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* 5:e19027.

Kurtzman CP, Fell JW, Boekhout T. 2011. The yeasts: a taxonomic study, vol. 1–3, 5th edn. Amsterdam: Elsevier.

Liachko I, Dunham MJ. 2014. An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res.* 14(2): 364–367.

Lin Z, Wang TY, Tsai BS, Wu FT, Yu FJ, Tseng YJ, Sung HM, Li WH. 2013. Identifying cis-regulatory changes involved in the evolution of aerobic fermentation in yeasts. *Genome Biol Evol.* 5(6): 1065–1078.

Lohr D, Venkov P, Zlatanova J. 1995. Transcriptional regulation in the yeast *GAL* gene family: a complex genetic network. *FASEB J.* 9(9):777–787.

Loytynoja A. 2014. Phylogeny-aware alignment with PRANK. Multiple Sequence Alignment Methods. *Methods Mol Biol.* 1079:155–170.

Lu M, Kleckner N. 1994. Molecular cloning and characterization of the *pgm* gene encoding phosphoglucomutase of *Escherichia coli*. *J Bacteriol.* 176(18):5847–5851.

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 Biol.* 13(8): e1002220.

Marsilio V, Campestre C, Lanza B, De Angelis M. 2001. Sugar and polyol compositions of some European olive fruit varieties (*Olea europaea* L.) suitable for table olive purposes. *Food Chem.* 72(4):485–490.

Martchenko M, Levitin A, Hogues H, Nantel A, Whiteway M. 2007. Transcriptional rewiring of fungal galactose-metabolism circuitry. *Curr Biol.* 17(12):1007–1013.

Mulet JM, 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(7):569–582.

Nierop KG, van Lagen B, Buurman P. 2001. Composition of plant tissues and soil organic matter in the first stages of a vegetation succession. *Geoderma* 100(1–2):1–24.

Oh D, Hopper JE. 1990. Transcription of a yeast phosphoglucomutase isozyme gene is galactose inducible and glucose repressible. *Mol Cell Biol.* 10(4):1415–1422.

Olson-Manning CF, Lee C-R, Rausher MD, Mitchell-Olds T. 2013. Evolution of flux control in the glucosinolate pathway in *Arabidopsis thaliana*. *Mol Biol Evol.* 30(1):14–23.

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

Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21(5):676–679.

Prud'Homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh S-D, True JR, Carroll SB. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* 440(7087):1050.

Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB. 2009. Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* 326(5960):1663–1667.

Rebeiz M, Williams TM. 2017. Using *Drosophila* pigmentation traits to study the mechanisms of cis-regulatory evolution. *Curr Opin Insect Sci.* 19:1–7.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* 290(5500): 2306–2309.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 3(2):217–223.

Riley R, Haridas S, Wolfe KH, Lopes MR, Hittinger CT, Goker M, Salamov AA, Wisecaver JH, Long TM, Calvey CH, et al. 2016. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A.* 113(35):9882–9887.

Robasky K, Bulyk ML. 2011. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* 39(Database issue): D124–D128.

Rogers WA, Salomone JR, Tacy DJ, Camino EM, Davis KA, Rebeiz M, Williams TM. 2013. Recurrent modification of a conserved cis-regulatory element underlies fruit fly pigmentation diversity. *PLoS Genet.* 9(8): e1003740.

Roop JI, Chang KC, Brem RB. 2016. Polygenic evolution of a sugar specialization trade-off in yeast. *Nature* 530(7590):336–339.

Roy S, Wapinski I, Pfiffner J, French C, Socha A, Konieczka J, Habib N, Kellis M, Thompson D, Regev A. 2013. Arboretum: reconstruction and analysis of the evolutionary history of condition-specific transcriptional modules. *Genome Res.* 23(6):1039–1050.

Rubio-Texeira M. 2005. A comparative analysis of the *GAL* genetic switch between not-so-distant cousins: *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. *FEMS Yeast Res.* 5(12):1115–1128.

Sarilar V, Devillers H, Freel KC, Schacherer J, Neuveglise C. 2015. Draft genome sequence of Lachancea lanzarotensis CBS 12615^T, an Ascomycetous yeast isolated from grapes. Genome. *Announcements* 3(2): e00292-15.

Shen X-X, Zhou X, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. 2016. Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. *G3* 6:3927–3939.

Sood V, Brickner JH. 2017. Genetic and epigenetic strategies potentiate Gal4 activation to enhance fitness in recently diverged yeast species. *Curr Biol.* 27(23):3591–3602.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet.* 14(11):751–764.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution?. *Evolution* 62(9):2155–2177.

Stone JR, Wray GA. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol Biol Evol.* 18(9):1764–1770.

Sucena E, Delon I, Jones I, Payre F, Stern DL. 2003. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* 424(6951):935–938.

Thomson JM, Gaucher EA, Burgan MF, De Kee DW, Li T, Aris JP, Benner SA. 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet.* 37(6):630–635.

Vakirlis N, Sarilar V, Drillon G, Fleiss A, Agier N, Meyniel JP, Blanpain L, Carbone A, Devillers H, Dubois K, et al. 2016. Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Res.* 26(7):918–932.

Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ. 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying

evolutionary innovation through gene duplication. *PLoS Biol.* 10(12): e1001446.
Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 32(3):820–832.
Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387(6634):708–713.
Wright KM, Rausher MD. 2010. The evolution of control and distribution of adaptive mutations in a metabolic pathway. *Genetics* 184(2):483–502.