

Polygenic evolution of a sugar specialization trade-off in yeast

Jeremy I. Roop¹, Kyu Chul Chang^{2,3} & Rachel B. Brem^{1,3}

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

A central goal of evolutionary genetics is to understand how organisms acquire phenotypic novelties. Such traits, if they have evolved over long timescales, can have a genetic basis quite distinct from those arisen more recently⁴. In landmark cases, single genes underlying species differences have been pinpointed and validated⁵, but the polygenic architecture of ancient traits has remained a mystery.

In hybrids formed by mating *S. cerevisiae* with other *Saccharomyces* species⁶, we noted a pattern of coherent *cis*-regulatory variation in the seven genes of the galactose metabolic pathway. During growth in medium with glucose as the sole carbon source, the *S. cerevisiae* allele at each *GAL* gene conferred low expression relative to other *Saccharomyces*, except for the repressor *GAL80*, at which the *S. cerevisiae* allele drove expression up (Fig. 1b). Likewise, purebred *S. cerevisiae* expressed *GAL* effectors at low levels in glucose, and *GAL80* at high levels, relative to other species (Fig. 1b and ref. 7). *S. paradoxus*, the sister species to *S. cerevisiae*, had an intermediate expression phenotype (Fig. 1b). Thus, the *S. cerevisiae* *GAL* program is one of heightened glucose repression relative to other species, as a product of *cis*-regulatory changes at the five loci that encode the seven *GAL* genes. Because such a pattern is unlikely under neutrality⁸, these data raised the possibility that selective pressure on the *GAL* pathway had changed along the *S. cerevisiae* lineage.

In *S. cerevisiae*, pre-expression of metabolic genes in glucose medium can boost fitness upon a switch to other carbon sources^{9–11}. We therefore expected that *GAL* expression divergence in glucose could have phenotypic correlates in other conditions. Culturing cells in 1% glucose–galactose medium, we observed a qualitative distinction between species (Fig. 2a). In *S. cerevisiae*, growth was retarded by a diauxic lag midway through the time course, reflecting the expected delay in assembling galactose metabolic machinery once glucose is exhausted^{9,10,12}. In more distantly related yeasts, we observed no lag in 1% glucose–galactose medium supplemented with galactose (Fig. 2a, b), although *S. paradoxus* had a modest lag (Fig. 2a, b) that echoed its intermediate regulatory phenotype (Fig. 1b). Glucose mixtures with maltose and raffinose engendered a lag in all members of the clade (Extended Data Fig. 1). *S. cerevisiae* strains from distinct populations all exhibited a lag in glucose–galactose cultures (Fig. 2c). These data highlight *S. cerevisiae* as an extreme among *Saccharomyces* with respect to two attributes of galactose metabolism: reduced *GAL*

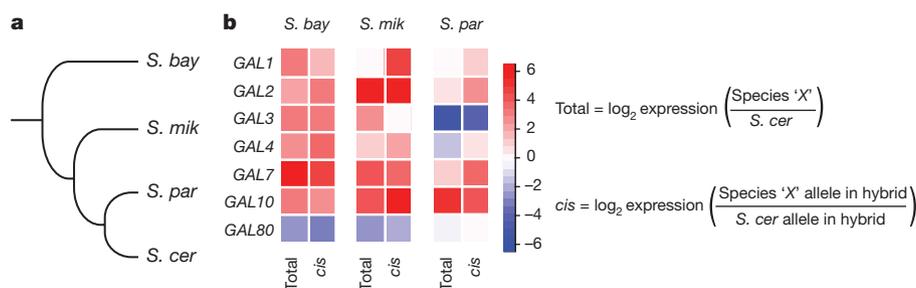


Figure 1 | Polygenic *cis*-regulatory evolution among yeast species in galactose metabolic genes. **a**, Phylogenetic tree of *Saccharomyces* species studied here¹⁴. *S. bay*, *S. bayanus*; *S. mik*, *S. mikatae*; *S. par*, *S. paradoxus*; *S. cer*, *S. cerevisiae*. **b**, Each cell reports expression, as a ratio between the indicated species and *S. cerevisiae*, of the indicated galactose metabolism

gene during culture in glucose medium⁶. Total, expression measured in purebred species; *cis*, expression from the indicated species' allele in a diploid hybrid between this species and *S. cerevisiae*, reflecting effects of *cis*-regulatory divergence.

¹Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, California 94720, USA. ²Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA. ³Buck Institute for Research on Aging, Novato, California 94945, USA.

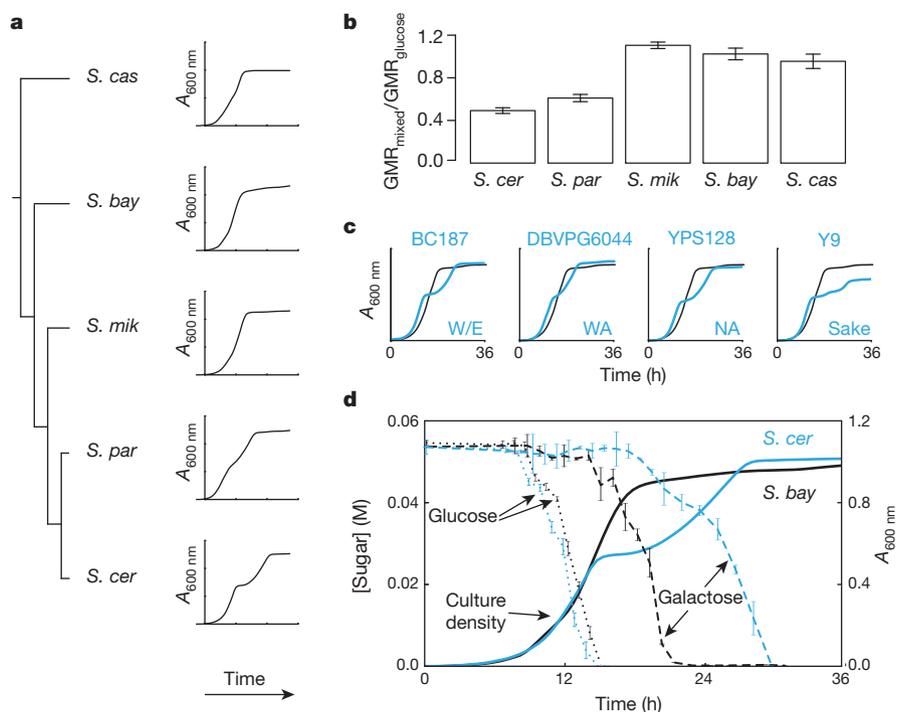


Figure 2 | Diauxic lag, in 1% glucose–galactose medium, is conserved within *S. cerevisiae* and divergent among species. **a**, Growth of *Saccharomyces* type strains inoculated into medium containing 1% glucose and 1% galactose ($n = 6$). *S. cas*, *S. castellii*. **b**, Each bar reports the geometric mean of the growth rate (GMR) of the indicated species from the time course in **a**, normalized to the analogous quantity in glucose medium. **c**, Growth of *S. cerevisiae* isolates (blue) from the indicated populations (W/E, Wine/European; WA, West African; NA, North

American) and the *S. bayanus* type strain (black), inoculated into medium containing 1% glucose and 1% galactose ($n = 6$). **d**, Growth (solid lines, $n = 4$) of *S. cerevisiae* and *S. bayanus* inoculated into medium containing 1% glucose and 1% galactose, and medium concentrations of glucose and galactose (dotted and broken lines, respectively; $n = 2$ biological replicates, each comprising 3 technical replicates). Error bars, s.e.m. Each set of data is representative of the results of two independent experiments.

gene expression during growth in pure glucose, and diauxic lag in 1% glucose–galactose.

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

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

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

negative epistasis in the seven-promoter replacement strain (Fig. 3c), and in strains harbouring intermediate *S. cerevisiae* promoter combinations (Extended Data Fig. 3).

Transgenesis of individual *S. cerevisiae* *GAL* coding regions was also sufficient for a partial lag in *S. bayanus*, in the case of *GAL1*, *GAL2*, *GAL3*, *GAL4*, *GAL7*, and *GAL10* (Fig. 3b, c). Swaps of *GAL* promoter-coding fusions revealed negative epistasis at *GAL1* and *GAL3*: for these genes, the sum of phenotypes from the respective promoter and coding transgenics was far more dramatic than the effect of the promoter-coding fusion (Fig. 3c). Combining all seven *S. cerevisiae* *GAL* coding or promoter-coding regions in *S. bayanus*, we observed an exaggerated, long-term growth delay in 1% glucose–galactose medium, distinct from the temporary lag of wild-type strains and promoter transgenics (Fig. 3b). This defect reflected dysfunction of multiple modules of the *S. cerevisiae* *GAL* pathway in *S. bayanus*, as it could be elicited by just the two regulators Gal3 and Gal4 swapped from *S. cerevisiae*, or just *S. cerevisiae* alleles of the enzymes Gal1, Gal7, and Gal10 (Extended Data Fig. 3). Coding and promoter-coding swap strains did ultimately resume active growth (Fig. 3b) and metabolize galactose from mixed-sugar medium (Extended Data Fig. 4), and their *GAL* expression induction was markedly delayed (Fig. 3d). These strains also grew poorly in pure galactose medium (Extended Data Fig. 5). Together, our data make clear that diauxic lag in 1% glucose–galactose medium can be largely recapitulated by divergent *GAL* gene promoters; *GAL* protein alleles from *S. cerevisiae* make a partial contribution to lag when tested in isolation and, when combined in *S. bayanus*, confer growth defects far exceeding those of either wild-type.

In light of the conservation of diauxic lag across *S. cerevisiae* (Fig. 2c), we hypothesized that this species had maintained its divergent galactose metabolic behaviour on the basis of a fitness benefit. Among the potential mechanisms for such an advantage, we focused on the possibility that as *S. cerevisiae* represses *GAL* genes in glucose-replete

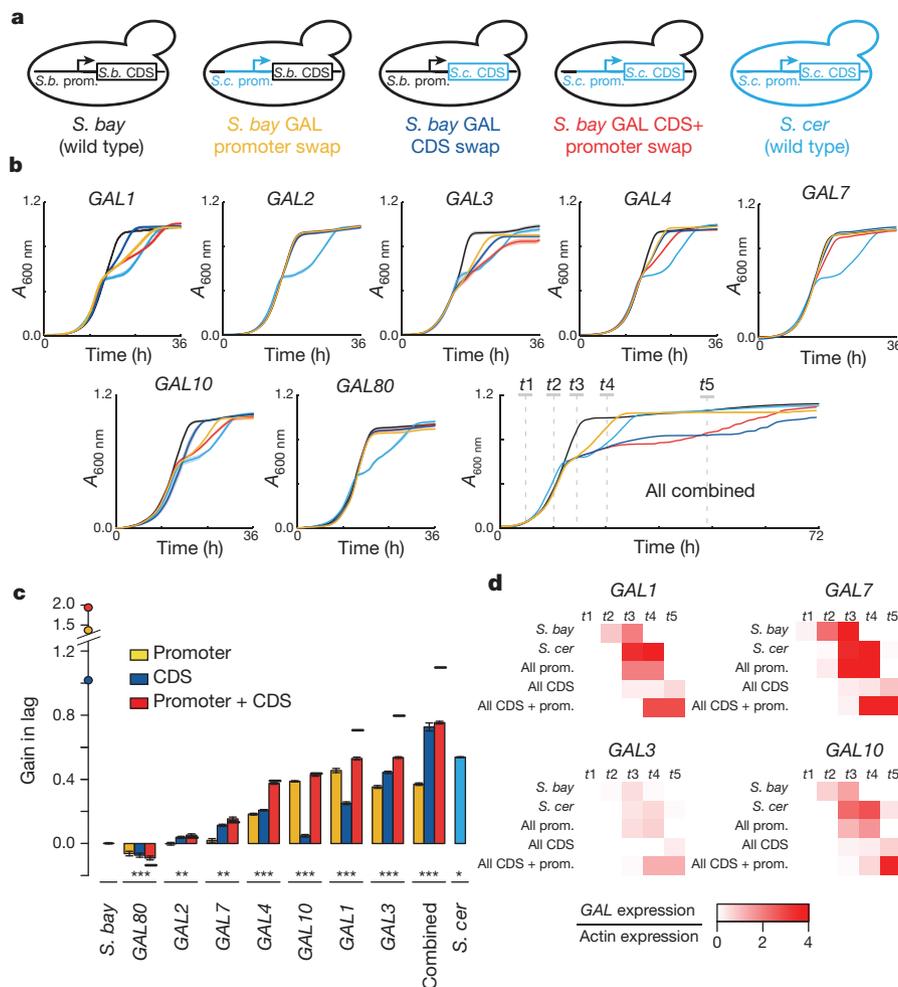


Figure 3 | *S. cerevisiae* alleles of *GAL* genes confer diauxic lag in 1% glucose-galactose medium. **a**, Replacement of *S. cerevisiae* *GAL* sequences into *S. bayanus* at the endogenous loci. **b**, Growth of *S. bayanus* harbouring *S. cerevisiae* alleles of a single *GAL* gene, or of all seven genes, inoculated into medium containing 1% glucose and 1% galactose ($n \geq 12$). **c**, Each bar reports the ratio of the GMR over the time course of the indicated strain in **b** to that of wild-type *S. bayanus*, subtracted from 1; negative values are GMRs faster than wild-type. Error bars, s.e.m.

conditions (Fig. 1), it avoids the liability of expressing unused proteins and enables rapid growth^{10,13}. When cultured in 2.5% glucose medium also containing galactose, wild-type *S. cerevisiae* exhibited a 10% faster growth rate (Fig. 4a, b), and fourfold to ninefold lower expression of *GAL* enzymes (Fig. 4c), than *S. bayanus*. Both species metabolized glucose almost exclusively across the time course (Fig. 4d, e). Replacement of all seven *S. bayanus* *GAL* promoters with *S. cerevisiae* alleles recapitulated the program of low *GAL* gene expression (Fig. 4c), and conferred a growth rate halfway between those of the wild-type species (Fig. 4a, b), in 2.5% glucose medium supplemented with galactose. The *S. bayanus* strain harbouring all seven *S. cerevisiae* *GAL* coding regions also expressed *GAL* genes at low levels (Fig. 4c), which mirrored this strain's exaggerated delay in *GAL* gene induction (Fig. 3b–d), and was associated with a partial growth benefit (Fig. 4a, b). Promoter-coding replacement conferred no additional phenotype over and above the effects of transgenesis of either region type alone (Fig. 4a–c). We conclude that *S. cerevisiae* *GAL* promoters, by shutting down expression of the galactose metabolic pathway, are adaptive in conditions of abundant glucose, and this program can be phenocopied by *S. cerevisiae* *GAL* protein alleles in *S. bayanus*. Sequence analyses revealed a high ratio of inter-specific divergence to intra-species polymorphism in *GAL* gene promoters, and not in *GAL*

Asterisks, significant differences ($P < 0.001$, Wilcoxon rank-sum) from wild-type *S. bayanus*. Also shown are expected phenotypes of promoter-CDS transgenics for a single gene (horizontal lines) or seven-locus transgenics (circles on y axis), under an additive model of contributions from the regions combined in the respective strains. **d**, *GAL* gene expression at time points indicated in the final panel of **b** ($n \geq 2$ biological replicates, each comprising 3 technical replicates). Each set of data is representative of the results of two independent experiments.

coding regions (Extended Data Tables 1–3), suggestive of a history of directional evolution at these loci.

In this work, we have dissected glucose-specialist phenotypes that distinguish *S. cerevisiae* from other members of the *Saccharomyces* clade. *S. cerevisiae* is reluctant to transition from glucose to galactose metabolism, and has a growth advantage in a high-glucose environment. Additionally, the *S. cerevisiae* program confers an increase in biomass accumulation during growth in pure galactose (Extended Data Fig. 5c) and could be beneficial when glucose availability fluctuates rapidly¹¹. As *S. cerevisiae* alleles of *GAL* gene promoters are largely sufficient for this family of traits, they may have served as an easily evolvable, and probably adaptive, origin of these characters. By contrast, the *S. cerevisiae* *GAL* proteome, which confers synthetic growth defects in modern-day *S. bayanus*, may have evolved slowly over a rugged fitness landscape, under distinct forces or at a different period. Such a model would dovetail with the *cis*-regulatory basis of a related, but genetically simple, galactose metabolism trait that evolved more recently between yeasts¹³. For any suite of divergent regulatory regions, observing *cis*-acting effects on gene expression can open a first window onto their phenotypic relevance and that of the gene products they control. With this strategy, evolutionary biologists need not be limited by polygenicity in the mapping of genotype to phenotype, even between long-diverged species.

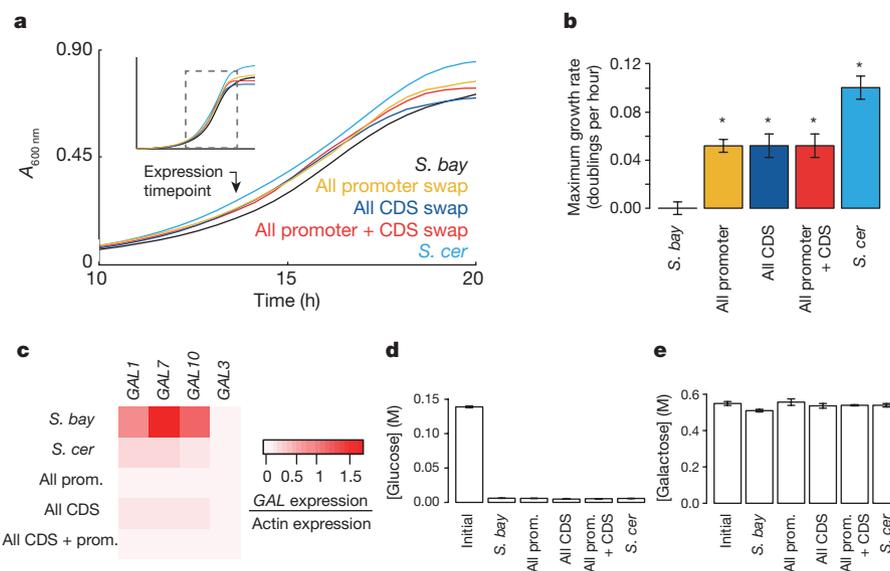


Figure 4 | *S. cerevisiae* GAL alleles confer a fitness advantage in 2.5% glucose medium supplemented with galactose. a, Growth of *S. bayanus* strains harbouring *S. cerevisiae* alleles of all seven GAL genes, and the wild-type species, inoculated into medium containing 2.5% glucose and 10% galactose. Inset shows the complete time course from which the main figure shows a narrower time window ($n \geq 136$, multiple independent experiments pooled). **b**, Each bar reports the difference in maximum growth rate between the indicated species and wild-type *S. bayanus* over the

time course of the indicated strain in **a**. Error bars, s.e.m.; asterisks indicate rates significantly different ($P < 1 \times 10^{-8}$, Wilcoxon rank-sum) from wild-type *S. bayanus*. **c**, Each row reports GAL gene expression of the indicated strain at time points indicated by the arrow in **a** ($n \geq 2$ biological replicates, each comprising 3 technical replicates). **d**, **e**, In each panel, the first bar reports sugar concentration in medium before inoculation; the remaining bars report sugar concentrations in medium after the indicated time course in **a** ($n = 3$ biological replicates, each comprising 3 technical replicates).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 March; accepted 18 December 2015.

Published online 10 February 2016.

- Orr, H. A. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* **6**, 119–127 (2005).
- Rockman, M. V. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**, 1–17 (2012).
- Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010).
- Savolainen, O., Lascoux, M. & Merilä, J.; Savolainen. Ecological genomics of local adaptation. *Nature Rev. Genet.* **14**, 807–820 (2013).
- Nadeau, N. J. & Jiggins, C. D. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet.* **26**, 484–492 (2010).
- Schraiber, J. G., Mostovoy, Y., Hsu, T. Y. & Brem, R. B. Inferring evolutionary histories of pathway regulation from transcriptional profiling data. *PLoS Comput. Biol.* **9**, e1003255 (2013).
- Caudy, A. A. *et al.* A new system for comparative functional genomics of *Saccharomyces* yeasts. *Genetics* **195**, 275–287 (2013).
- Bullard, J. H., Mostovoy, Y., Dudoit, S. & Brem, R. B. Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc. Natl Acad. Sci. USA* **107**, 5058–5063 (2010).
- Venturelli, O. S., Zuleta, I., Murray, R. M. & El-Samad, H. Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. *PLoS Biol.* **13**, e1002042 (2015).

- Wang, J. *et al.* Natural variation in preparation for nutrient depletion reveals a cost-benefit tradeoff. *PLoS Biol.* **13**, e1002041 (2015).
- New, A. M. *et al.* Different levels of catabolite repression optimize growth in stable and variable environments. *PLoS Biol.* **12**, e1001764 (2014).
- De Deken, R. H. The Crabtree effect: a regulatory system in yeast. *J. Gen. Microbiol.* **44**, 149–156 (1966).
- Peng, W., Liu, P., Xue, Y. & Acar, M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. *Nature Commun.* **6**, 6226 (2015).
- Scannell, D. R. *et al.* The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3* **1**, 11–25 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health GM087432 to R.B.B. and a Hellman Graduate Fellowship from the University of California Berkeley to J.I.R. We thank A. Arkin for his advice and resources; J. Rine for yeast strains; O. Venturelli for technical expertise; and A. Flamholz, J. Schraiber, and P. Shih for discussions.

Author Contributions J.I.R. and R.B.B. designed experiments, J.I.R. and K.C.C. conducted experiments, J.I.R. analysed data, and J.I.R. and R.B.B. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B.B. (rbrem@buckinstitute.org).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains. Strains used in this study are listed in Supplementary Table 1. Abbreviations in figures and tables are as follows: *S. cer.*, *Saccharomyces cerevisiae*; *S. par.*, *Saccharomyces paradoxus*; *S. mik.*, *Saccharomyces mikatae*; *S. bay.*, *Saccharomyces bayanus*; *S. cas.*, *Saccharomyces castellii*. Allele-swap strains constructed in haploid *S. bayanus* JRY294 or JRY296 (isogenic MAT α and MAT α derivatives of type strain CBS7001) used the MIRAGE method¹⁵ with several modifications as follows. A 1.7-kilobase region containing the *K. lactis* URA3 coding sequence and regulatory region was amplified from the pCORE-UH plasmid¹⁶ and used for each half of the inverted repeat. The *S. cerevisiae* GAL region to be swapped in was attached to one half of the inverted repeat cassette by overlap extension PCR, after which the two halves of the final cassette were ligated together. Owing to the different sizes of the two halves of the inverted repeat cassette, a second restriction digestion step as described in ref. 15 was not necessary to remove non-desired ligation products. Transformation with the cassette, followed by confirmation of positive transformants and plating onto 5-FOA medium, resulted in excision of the inverted repeat from the target genome, leaving behind a marker-less allele swap at the locus. Sanger sequencing was used to verify the correct nucleotide sequence of each swapped allele. The *S. cerevisiae* allele of each promoter and CDS was amplified from genomic DNA of YHL068 (ref. 6). Promoter, coding, and promoter-coding fusion swap strains were engineered by replacing 600 base pairs (bp) of intergenic region directly upstream of the CDS, the CDS, or these two regions combined, respectively, in *S. bayanus* with orthologous regions from *S. cerevisiae*. For allele swaps in the *S. cerevisiae* background in Extended Data Fig. 2, 720 bp of the region between the *GAL1* and *GAL10* open reading frames were amplified from *S. bayanus* strain CBS7001 and used to construct a MIRAGE cassette as above, and transformed into *S. cerevisiae* strain JRY313 (isogenic MAT α derivative of BY4743) and selected as above. For each transgenic, two or more independent transformants were used as replicates for growth profiling and sugar concentration measurements. Combining unlinked allele swaps into a single genome was accomplished by single-cell mating of single-locus swaps, followed by sporulation, tetrad dissection, and diagnostic PCR to identify segregant colonies with the allele combinations of interest. *S. cerevisiae* alleles of *GAL1*, *GAL7* and *GAL10* were combined in the *S. bayanus* background by successive allele swap transformations.

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

Growth time courses were performed as follows. Strains were grown in YP containing 2% galactose (Fig. 4 and Extended Data Fig. 5) or 2% glucose (all other figures) for 24 h with shaking at 200 r.p.m. Each strain was then back-diluted into the same medium to an absorbance (*A*) of 0.1 and grown for an additional 6 h. These log-phase cultures were then back-diluted to an absorbance of 0.02 in a 96-well plate containing 150 μ l of YP with the appropriate amount of a given carbon source. Plates were covered with a gas-permeable membrane, placed in a Tecan F200 plate reader and incubated with orbital shaking for the duration of each experiment. Measurements of $A_{600\text{ nm}}$ were made every 30 min.

Gain in lag was calculated from growth curves as $(1 - (\text{GMR of a given strain}) / (\text{GMR of } S. bayanus))$. GMR was calculated as in ref. 11, with the following differences. A window of 0.1–0.8 absorbance units was used for quantification in all figures apart from Fig. 2b. In Fig. 2b, GMR was calculated within a window bounded by 20% and 80% of the maximum final yield attained during the time course for each species. Maximum growth rate was calculated as in ref. 10 except that a window of 0.01–0.3 absorbance units was used and a geometric mean of growth rates was calculated. Final growth yield was calculated as the difference between initial and final $A_{600\text{ nm}}$ measurements as in ref. 17.

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

in Fig. 2d were as in Fig. 2a–c except that four biological replicate cultures were measured.

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

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

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

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

Data in Fig. 2d were obtained by sampling two biological replicate cultures at each time point. Data in Fig. 4d, e and Extended Data Fig. 4 were obtained by sampling three biological replicate cultures both at the start and at the endpoints of the growth time course. For all experiments, three technical replicates were assayed for each biological replicate culture sampled, and mean values are reported. All data are representative of two identical experiments conducted on different days. **Quantitative PCR.** Time point samples for quantitative PCR analysis were obtained from cultures analogously to those obtained for sugar consumption quantification detailed above. Between 2 and 15 replicate wells were harvested at each time point and pooled to have sufficient biological material for RNA isolation. RNA was isolated using an RNeasy mini kit (Qiagen) and complementary DNA was synthesized using SuperScript III (Life Technologies). DyNAmo HS SYBR green (Thermo Scientific) was used for quantitative PCR and all quantification was done on a CFX96 machine (BioRad). Gene expression levels relative to *ACT1* were calculated using the $2^{-\Delta\Delta C_t}$ method¹⁸. Three technical replicates per biological sample were assayed, and mean values are reported. Primer sequences and cycling times are in Supplementary Table 2.

Sequence analyses. Custom Python scripts were used to extract coding sequences and 600 bp promoter regions for type strains of *S. paradoxus*, *S. mikatae*, and *S. bayanus*¹⁴ for each gene that had an orthologue in each of the five *Saccharomyces sensu stricto* species as reported in ref. 14. *S. cerevisiae* population sequences were downloaded from the following sources: YJM978, UWOPS83-787, Y55, UWOPS05-217.3, 273614N, YS9, BC187, YPS128, DBVPG6765, YJM975, L1374, DBVPG1106, K11, SK1, 378604X, YJM981, UWOPS87-2421, DBVPG1373, NCYC3601, YPS606, Y12, UWOPS05-227.2, and YS2 from <http://www.yeastrc.org/g2p/home.do>; Sigma1278b, ZTW1, T7, and YJM789 from <http://www.yeast-genome.org/>; and RM11 from http://www.broadinstitute.org/annotation/genome/Saccharomyces_cerevisiae. *S. cerevisiae* sequences were aligned to each of the other

three species in turn using FSA (ref. 19), with the ‘-nucprot’ option for the coding sequence alignments. For each set of species alignments, nucleotide replacement and polymorphic sites were tabulated using Polymorphorama²⁰ for coding regions and custom Python scripts for promoter regions.

Sequence analyses of the seven genes of the *GAL* pathway (*GAL1*, *GAL2*, *GAL3*, *GAL4*, *GAL7*, *GAL10*, and *GAL80*) using *S. paradoxus*, *S. mikatae*, or *S. bayanus* as an outgroup were done as follows. For a given species comparison, we first calculated the neutrality index NI_{TG} statistic²¹ for the promoter regions of the *GAL* genes using synonymous sites in the downstream gene as putative neutral sites, as

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

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

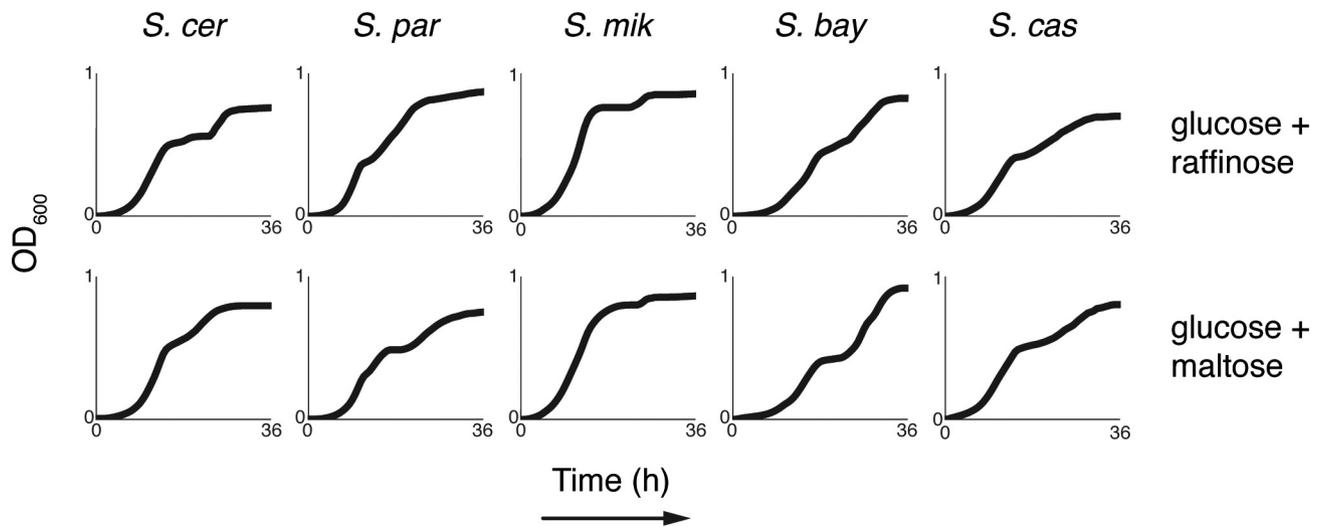
Minor alleles with a frequency of less than 0.15 were ignored^{22,23} and all site counts were corrected for multiple hits using a Jukes–Cantor model. In computing NI_{TG} for the seven genes of the *GAL* pathway, we considered that since *GAL1* and *GAL10* are adjacent genes on chromosome II sharing a 662 bp promoter region, simply counting sites within 600 bp promoter regions for each gene separately would have resulted in double-counting of 532 bp. To avoid this, we aligned the 662 bp intergenic region of these two genes and considered it one locus in the NI_{TG} calculation. Synonymous sites for this region were taken as the sum of such sites in both the *GAL1* and *GAL10* coding sequences. To evaluate significance, we generated 10,000 randomly chosen groups of six genes each and computed NI_{TG} across the promoters of each such null group as above. The empirical significance of the true NI_{TG} value for *GAL* gene promoters was then taken as the proportion of null groups whose promoters gave an NI_{TG} less than or equal to the true value. We assessed selection acting on non-synonymous sites in *GAL* coding sequences with a pipeline analogous to the above, except that *GAL1* and *GAL10* were considered separate loci, as the coding sequences do not overlap; thus our resampling test used 10,000 random groups of seven coding regions each. Tests for selection using the neutrality index statistic²⁴ [(polymorphic non-synonymous sites/polymorphic synonymous sites)/(divergent non-synonymous/divergent synonymous sites), with a similar formulation for non-synonymous sites] used the same

pipeline as above except that we tabulated the average neutrality index across genes of a group, and compared this average quantity for the *GAL* genes against resampled groups of the same size.

D_{XY} (ref. 25), the average number of pairwise differences between outgroup species and *S. cerevisiae* strains, and intrapopulation nucleotide diversity²⁶ (π), were calculated using custom Python scripts. Nucleotide diversity was ascertained on alignments containing only *S. cerevisiae* strains belonging to the European population, the most deeply sampled in our data set. Empirical significance of these statistics for *GAL* gene promoters and coding sequences was calculated using the resampling pipeline described above.

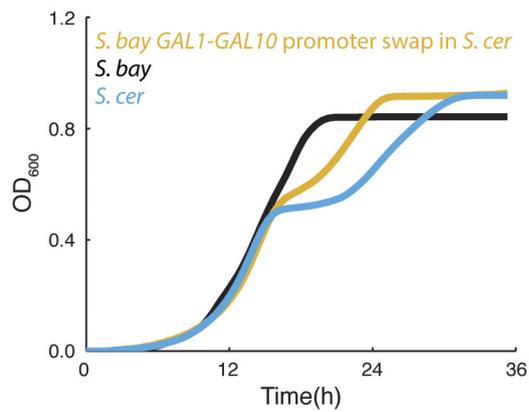
Code availability. Custom Python scripts used for data analysis are available upon request.

15. Nair, N. U. & Zhao, H. Mutagenic inverted repeat assisted genome engineering (MIRAGE). *Nucleic Acids Res.* **37**, e9 (2009).
16. Storici, F. & Resnick, M. A. The *delitto perfetto* approach to *in vivo* site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409**, 329–345 (2006).
17. Warringer, J. *et al.* Trait variation in yeast is defined by population history. *PLoS Genet.* **7**, e1002111 (2011).
18. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408 (2001).
19. Bradley, R. K. *et al.* Fast statistical alignment. *PLoS Comput. Biol.* **5**, e1000392 (2009).
20. Haddrill, P. R., Bachtrog, D. & Andolfatto, P. Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Mol. Biol. Evol.* **25**, 1825–1834 (2008).
21. Stoletzki, N. & Eyre-Walker, A. Estimation of the neutrality index. *Mol. Biol. Evol.* **28**, 63–70 (2011).
22. Charlesworth, J. & Eyre-Walker, A. The McDonald-Kreitman test and slightly deleterious mutations. *Mol. Biol. Evol.* **25**, 1007–1015 (2008).
23. He, B. Z., Holloway, A. K., Maerkl, S. J. & Kreitman, M. Does positive selection drive transcription factor binding site turnover? A test with *Drosophila* cis-regulatory modules. *PLoS Genet.* **7**, e1002053 (2011).
24. Rand, D. M. & Kann, L. M. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**, 735–748 (1996).
25. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, 1987).
26. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. USA* **76**, 5269–5273 (1979).

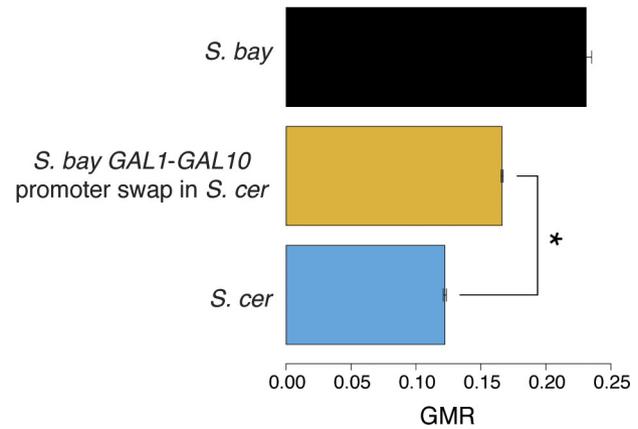


Extended Data Figure 1 | Divergence in the *S. cerevisiae* diauxic lag trait is specific to growth in glucose-galactose medium. Each trace reports growth of the indicated yeast inoculated into medium containing

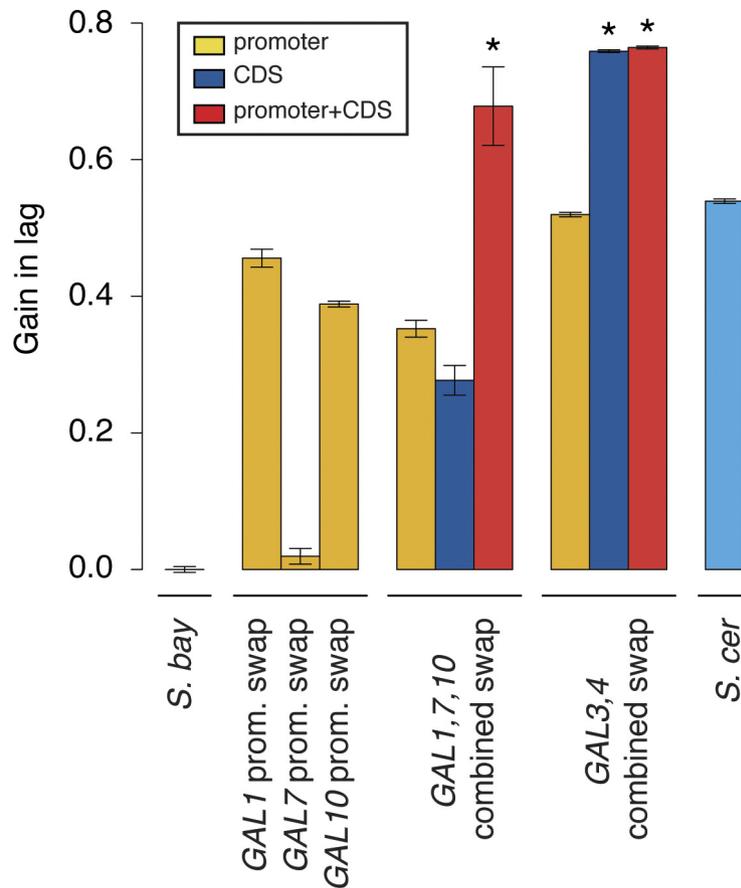
1% glucose and 1% of the indicated secondary carbon source ($n = 4$, data representative of two independent experiments).

a

Extended Data Figure 2 | The *S. bayanus* allele of the *GAL1* and *GAL10* promoters confers partial rescue of diauxic lag in *S. cerevisiae* in 1% glucose–galactose medium. **a, Data are as in Fig. 3b, except that the yellow curve reports growth of an *S. cerevisiae* strain harbouring the *GAL1* and *GAL10* promoters from *S. bayanus* ($n \geq 12$). **b**, Each bar reports**

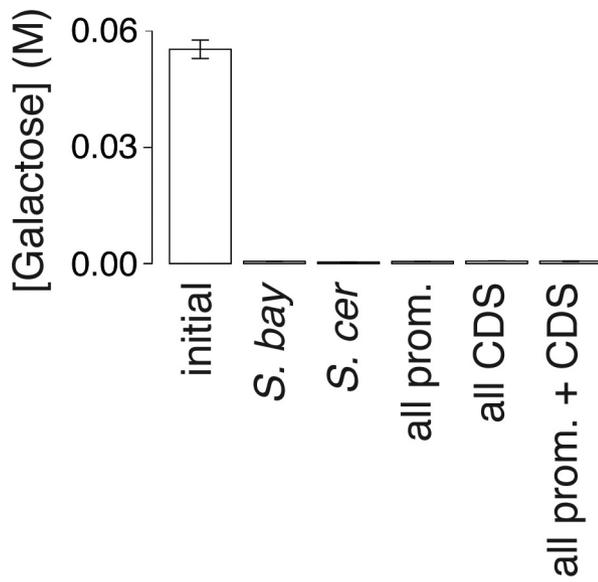
b

the GMR of the indicated strain over the time course shown in **a**. Error bars, s.e.m.; asterisk indicates a significantly different rate ($P < 1 \times 10^{-15}$, Wilcoxon rank-sum) between the transgenic swap strain and wild-type *S. cerevisiae*. Each set of data is representative of the result of two independent experiments.

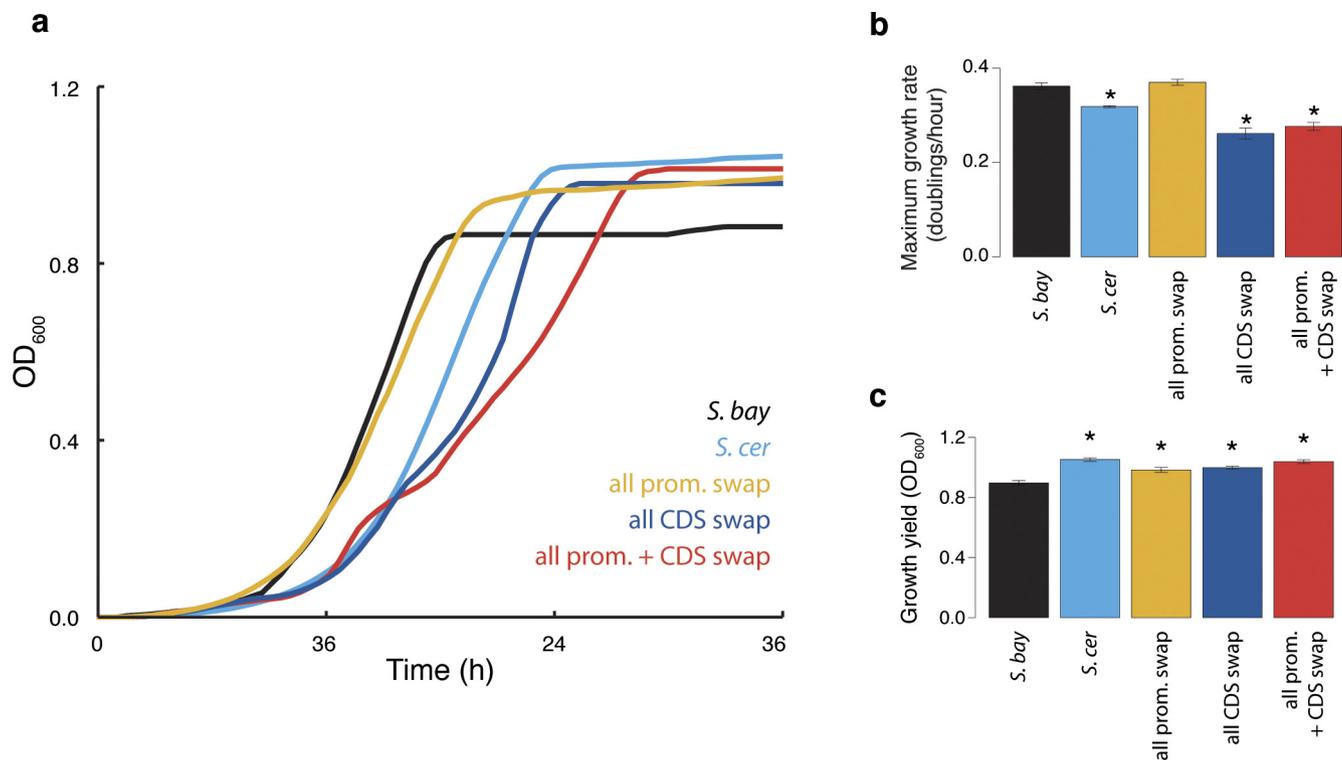


Extended Data Figure 3 | Epistasis between *S. cerevisiae* GAL gene alleles swapped into *S. bayanus*, in 1% glucose–galactose medium. Each bar reports the ratio of the GMR to that of wild-type *S. bayanus*, subtracted from 1, over a time course of the indicated strain inoculated into medium containing 1% glucose and 1% galactose. The first to fourth bars, and

the last bar, are from Fig. 3c; each of the remaining bars reports results from an *S. bayanus* strain harbouring *S. cerevisiae* alleles at the indicated combination of GAL loci ($n \geq 8$, data representative of two independent experiments). Symbols and analyses are as in Fig. 3c.



Extended Data Figure 4 | *S. bayanus* strains harbouring *S. cerevisiae* alleles of all seven *GAL* loci, inoculated into 1% glucose-galactose medium, deplete growth media of galactose. The first bar reports galactose concentration in medium before inoculation, and the remaining bars report concentrations after growth time courses, in the experiments in Fig. 3b ($n = 3$ biological replicates, each comprising 3 technical replicates).



Extended Data Figure 5 | In pure galactose medium, *S. cerevisiae* alleles of *GAL* loci are sufficient for increased biomass accumulation and, in the case of protein-coding regions, slow growth. a, Each trace reports growth of an *S. bayanus* strain harbouring *S. cerevisiae* alleles of all seven *GAL* genes, or a wild-type control, inoculated into medium containing 2% galactose ($n \geq 8$). **b**, Each bar reports maximum growth rate from the

respective time course in **a**. Error bars, s.e.m.; asterisks indicate values significantly different ($P < 0.05$, Wilcoxon rank-sum) from wild-type *S. bayanus*. **c**, Each bar reports growth yield from the respective time course in **a**. Error bars, s.e.m.; asterisks indicate values significantly different ($P < 0.05$, Wilcoxon rank-sum) from wild-type *S. bayanus*. Each set of data is representative of the result of two independent experiments.

Extended Data Table 1 | Excess of divergence relative to polymorphism in *GAL* promoter regions**Extended Data Table 1a**

Statistic	Outgroup		
	<i>S. par</i>	<i>S. mik</i>	<i>S. bay</i>
NI _{TG} <i>GAL</i> promoters	0.32	0.37	0.31
NI _{TG} genome promoters	0.72	0.91	0.73
<i>p</i> -value	0.025	0.013	0.018

Extended Data Table 1b

Statistic	Outgroup		
	<i>S. par</i>	<i>S. mik</i>	<i>S. bay</i>
NI _{TG} <i>GAL</i> non-synonymous sites	0.98	0.92	1.03
NI _{TG} genome non-synonymous sites	1.28	1.21	1.27
<i>p</i> -value	0.25	0.19	0.26

Each panel reports analyses of the NI_{TG} measure²¹ comparing polymorphism within *S. cerevisiae* with divergence between *S. cerevisiae* and the indicated outgroup species, taken across promoter sites (a) or non-synonymous coding sites (b), with normalization by the analogous measure from synonymous coding sites. In a given panel, the first row reports NI_{TG} across the seven *GAL* genes, the second row reports the mean NI_{TG} from 10,000 randomly drawn gene groups, and the bottom row reports empirical significance of the distinction between *GAL* genes and the genomic null.

Extended Data Table 2 | Divergence between species, and polymorphism within *S. cerevisiae*, at *GAL* loci**Extended Data Table 2a**

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

Extended Data Table 2b

π	promoters	CDS
<i>GAL</i>	0.00067	0.00069
genome	0.00194	0.00113
<i>p</i> -value	0.181	0.513

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

Extended Data Table 3 | Evaluation of selection acting on *S. cerevisiae* promoter and coding sequences using the neutrality index statistic**Extended Data Table 3a**

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

Extended Data Table 3b

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

Data are as in Extended Data Table 1 except that the neutrality index²⁴ was used for each test.