



Published in final edited form as:

Science. 2021 January 22; 371(6527): 415–419. doi:10.1126/science.aba0542.

## Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast

James Boocock<sup>1,2,3,4</sup>, Meru J. Sadhu<sup>1,2,3,4,\*</sup>, Arun Durvasula<sup>1</sup>, Joshua S. Bloom<sup>1,2,3,4,†</sup>, Leonid Kruglyak<sup>1,2,3,4,†</sup>

<sup>1</sup>Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA

<sup>2</sup>Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA

<sup>3</sup>Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA, USA

<sup>4</sup>Institute for Quantitative and Computational Biology, University of California, Los Angeles, Los Angeles, CA, USA

### Abstract

Metabolic pathways differ across species but are expected to be similar within a species. We discovered two functional, incompatible versions of the galactose pathway in *Saccharomyces cerevisiae*. We identified a three-locus genetic interaction for growth in galactose, and used precisely engineered alleles to show that it arises from variation in the galactose utilization genes *GAL2*, *GAL1/10/7*, and phosphoglucomutase (*PGM1*), and that the reference allele of *PGM1* is incompatible with the alternative alleles of the other genes. Multiloci balancing selection has maintained the two incompatible versions of the pathway for millions of years. Strains with alternative alleles are found primarily in galactose-rich dairy environments, and they grow faster in galactose but slower in glucose, revealing a trade-off on which balancing selection may have acted.

Variation in nutrient availability between environments has led to the evolution of diverse metabolic pathways. In humans, mutations in these pathways give rise to diseases known as inborn errors of metabolism (1). The budding yeast *Saccharomyces cerevisiae* is commonly used for studying eukaryotic metabolism (2). A classic well-studied pathway for galactose metabolism includes a galactose transporter, encoded by the gene *GAL2*, and the enzymes encoded by *GAL1*, *GAL10*, and *GAL7*, which convert galactose to glucose-1-phosphate (3). Phosphoglucomutase, encoded by *PGM1* and *PGM2*, then converts glucose-1-phosphate to glucose-6-phosphate—the substrate for glycolysis.

† Corresponding author. j bloom@mednet.ucla.edu (J.S.B.); lkruglyak@mednet.ucla.edu (L.K.).

\* Present address: Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

**Author contributions:** J.B. performed experiments with assistance from M.J.S. and J.S.B.; J.B. analyzed data with assistance from J.S.B. and A.D.; L.K. and J.S.B. supervised the project. J.B., A.D., J.S.B., and L.K. wrote the manuscript. All authors discussed and agreed on the final version of the manuscript.

**Competing interests:** The authors declare no competing financial interests.

Within the same genus, some strains of *Saccharomyces kudriavzevii* can metabolize galactose, whereas others have lost this ability through pseudogenization of multiple genes in the pathway, and it was proposed that the two versions of the pathway have been maintained through multiloci balancing selection (4). Balancing selection maintains genetic diversity against the forces of genetic drift and has typically been observed to act on single loci (5). Multiloci balancing selection is expected to be extremely rare because it has to overcome the independent segregation of alleles at the different loci.

We studied growth in galactose in a large set of crosses in *S. cerevisiae* (6) and observed a genetic interaction among three loci in crosses involving the soil strain CBS2888 (three-way effect size 0.19 SD units, chi-square test,  $P < 10^{-15}$ ) (Fig. 1A, figs. S1 and S2, and tables S1 to S3). The nonadditive nature of the effects of the three loci is best illustrated by the phenotype of segregants that inherit the CBS2888 allele at the loci on chromosome II (ChrII) and ChrXII and the non-CBS2888 allele at the locus on ChrXI; these segregants grow much slower in galactose than those with any other combination of alleles (Fig. 1A). The three loci contain genes that encode components of galactose metabolism: *GAL1*, *GAL10*, and *GAL7* on ChrII; *PGMI* on ChrXI; and *GAL2* on ChrXII (Fig. 1B and fig. S1) (3). The CBS2888 alleles of these genes were highly diverged from the reference (fig. S3 and table S4). We hereafter refer to the divergent galactose alleles found in CBS2888 as the alternative alleles and alleles observed in the other strains as the reference alleles.

We used CRISPR-Cas9 to engineer strains with all eight possible combinations of the three alternative and three reference galactose alleles in a common genetic background (fig. S4 and tables S5 to S7) (7, 8). We then measured the growth rates of the eight engineered strains in galactose and recapitulated the mapping results (Fig. 1C, fig. S5, and tables S1 and S2), demonstrating that variants in the coding and intergenic regions of *GAL1/10/7* and *GAL2* and in the promoter region of *PGMI* are responsible for the observed genetic interaction. In particular, the strain with the reference *PGMI* promoter allele and the alternative *GAL1/10/7* and *GAL2* alleles exhibited a severe growth defect in galactose, confirming that the components of the reference and alternative pathways are incompatible.

To better understand cis-acting regulatory differences between the alternative and reference galactose alleles (9), we grew a diploid hybrid strain (CBS2888xBY) in glucose, transferred it to galactose medium, and sequenced RNA from samples collected throughout a growth time course (7). In glucose, the expression of the CBS2888 allele of *PGMI* was slightly lower than that of the reference allele (fig. S6). By contrast, 1 hour after the switch to galactose, the expression of the CBS2888 allele of *PGMI* was 15.5 times greater than that of the reference allele (binomial test,  $P < 10^{-100}$ ), and this difference persisted for the rest of the time course (fig. S6 and table S8).

The alternative *PGMI* promoter allele contains a *GAL4* upstream activating sequence (UAS) (10), whereas the reference allele does not (fig. S4). We engineered a point mutation disrupting the UAS in a strain with all three alternative galactose alleles (7). This single mutation recapitulated the growth defect in galactose observed in a strain with a combination of the reference allele of the *PGMI* promoter and alternative alleles of the other *GAL* genes

(fig. S7). We conclude that the induction of *PGMI* in galactose, mediated through a *GAL4* UAS, is critical for the proper functioning of the alternative galactose pathway.

We searched for the alternative and reference galactose alleles in worldwide collections of sequenced *S. cerevisiae* isolates comprising 1276 strains (11, 12) and found three common combinations: only reference alleles (1213 strains), only alternative alleles (49 strains), and 8 strains from China with the alternative *GAL1/10/7* allele and alleles of *GAL2* and the *PGMI* promoter that differ from both the reference and the alternative alleles (Fig. 2A, fig. S8, and table S9) (7). No strains carried the reference *PGMI* promoter allele and the alternative *GAL1/10/7* and *GAL2* alleles, suggesting that this combination causes a fitness disadvantage in natural environments and has been purged by selection. This hypothesis is further supported by a high linkage disequilibrium index ( $r = 0.59$ ) for the three loci (fig. S9) (13).

The alternative galactose alleles are fixed in two lineages of strains found in dairy products, including Camembert cheese from France, kefir grains from Japan, and fermented yak and goat milk from China (table S9). These environments are rich in lactose, a disaccharide of glucose and galactose. *S. cerevisiae* relies on the activity of other fungi and bacteria to break down lactose into glucose and galactose, which it then metabolizes (14). This observation suggests that the alternative alleles are maintained by natural selection in dairy environments.

We dated the split between the alternative and reference galactose alleles to ~3.2 billion generations ago (95% confidence interval = 2.5 to 4.5 billion generations), which predates the most recent common ancestor of the *Saccharomyces* genus (figs. S10 and S11 and tables S4 and S10) (7, 15). Phylogenetic clustering placed the alternative galactose alleles outside the *Saccharomyces* genus and supports an ancient origin of the alternative alleles (Fig. 2B and figs. S12 and S13) (7, 16).

One force that can maintain highly diverged alleles within a species is balancing selection. This process is expected to generate a signature of elevated sequence divergence at linked neutral sites that decays with genetic distance from the selected variant (5). We examined the rate of synonymous substitutions per site (dS) across the CBS2888 genome relative to the reference and observed a strong signature of ancient balancing selection at all three galactose loci (Fig. 3 and figs. S14 to S18) (7).

The strains with the alternative or Chinese alleles contain *GAL2* genes duplicated in tandem, and *GAL2* is also duplicated in two other yeast species: *Saccharomyces uvarum* and *Saccharomyces eubayanus*. We aligned all the *GAL2* paralogs and observed that the N-terminal cytosolic regions (amino acids 1 to 67) were highly dissimilar within species and phylogenetically clustered across species (fig. S19). These results suggest that the N-terminal regions of the *GAL2* paralogs are functionally distinct and maintained by selection, and they also provide evidence that the alternative alleles have an ancient origin in *Saccharomyces* (fig. S20).

It has been proposed that the alternative galactose alleles arose through introgression around the time humans domesticated milk-producing animals, but no species that could have

donated the alleles has been identified (7, 17). A relatively recent introgression would generate a sharp boundary between dS at the *GAL* genes and the rest of the genome. Instead, our data suggest that the variation at these loci has accumulated within *S. cerevisiae* over time.

We performed forward genetic simulations to distinguish between scenarios that could have given rise to the observed signatures of balancing selection (figs. S21 to S23) (7). Models of recent introgression (<50 million generations ago), with or without balancing selection, were not well-supported when compared with a model of ancient balancing selection (figs. S24 to S27 and table S11) (7).

Balancing selection can act on fitness trade-offs, in which alleles with higher fitness in one environment have lower fitness in another (5). Although all strains grow faster in glucose than in galactose [ $t$  statistic ( $T$ ) = 7.80,  $t$  test,  $P < 10^{-5}$ ], the strains with the alternative alleles grow faster in galactose than the strains with the reference alleles (Figs. 1C and 4A). *S. cerevisiae* encounters and metabolizes a wide variety of sugars (18) but prefers glucose (19). In glucose, the strains with the reference alleles grow 2% faster than strains with the alternative alleles ( $T = -3.12$ ,  $t$  test,  $P = 0.017$ ) (Fig. 4B and fig. S28). This faster growth provides an explanation for the maintenance of the reference alleles in the strains that do not frequently encounter galactose.

In the strains with reference alleles, the *GAL* genes are robustly repressed by glucose and induced by galactose (3). This leads to a pause in growth known as the diauxic shift, when yeast switch from metabolizing glucose to metabolizing galactose. Strains with the three alternative galactose alleles do not undergo a diauxic shift (Fig. 4C and fig. S29). RNA sequencing showed that in glucose, the reference alleles are repressed, whereas the alternative *GAL* alleles are constitutively expressed (fold change = 40.6, binomial test,  $P < 10^{-16}$ ) (Fig. 4D, fig. S30, and table S8) (7). The constitutive expression of the *GAL* genes eliminates the diauxic shift (20), providing a fitness benefit when galactose is encountered. However, gene expression can be costly (21), and this could explain why the alternative galactose pathway leads to a growth disadvantage in glucose.

The incompatible allele combinations we identified may provide a model for classical galactosemia, an inborn error of metabolism caused by recessive mutations in *GALT*, the human homolog of *GAL7* (22), that can lead to life-threatening symptoms if galactose is not eliminated from diet. The precise molecular mechanisms of galactosemia are not well understood (23), but yeast models of galactose toxicity suggest that the incompatibility observed in this work arises from the same metabolic defect that underlies galactosemia. Finally, our results go beyond previous findings (4) in showing that balancing selection can preserve two alternate, functionally distinct states of a multiloci genetic network, providing a general mechanism for the maintenance of complex, interacting genetic variation at coadapted alleles.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thank O. Schubert, E. Ben-David, L. Guo, and S. Zdraljevic for helpful manuscript feedback and edits.

**Funding:** This work was supported by funding from the Howard Hughes Medical Institute and an NIH grant (2RO1GM102308-06 to L.K.). A.D. is supported by the NSF Graduate Research Fellowship (DGE-1650604).

## Data and materials availability:

Sequencing data are available under NCBI BioProject PRJNA575066. The processed datasets are available on Dryad (24). The following code is available on Zenodo: custom scripts used to analyze processed data and create figures (25), a custom Python package used with our population genetic analysis (26), and the code used to perform and analyze the forward genetic simulations (27).

## REFERENCES AND NOTES

1. Saudubray JM, Garcia-Cazorla À, *Pediatr. Clin. North Am.* 65, 179–208 (2018). [PubMed: 29502909]
2. Nielsen J, *Annu. Rev. Biochem.* 86, 245–275 (2017). [PubMed: 28301739]
3. Sellick CA, Campbell RN, Reece RJ, *Int. Rev. Cell Mol. Biol.* 269, 111–150 (2008). [PubMed: 18779058]
4. Hittinger CT et al., *Nature* 464, 54–58 (2010). [PubMed: 20164837]
5. Charlesworth D, *PLOS Genet* 2, e64 (2006). [PubMed: 16683038]
6. Bloom JSet al., *eLife* 8, e49212 (2019). [PubMed: 31647408]
7. Materials and methods and supplementary text are available as supplementary materials.
8. Sadhu MJ et al., *Nat. Genet.* 50, 510–514 (2018). [PubMed: 29632376]
9. Albert FW, Kruglyak L, *Nat. Rev. Genet.* 16, 197–212 (2015). [PubMed: 25707927]
10. Traven A, Jelcic B, Sopta M, *EMBO Rep.* 7, 496–499 (2006). [PubMed: 16670683]
11. Duan S-F et al., *Nat. Commun.* 9, 2690 (2018). [PubMed: 30002370]
12. Peter Jet et al., *Nature* 556, 339–344 (2018). [PubMed: 29643504]
13. Okada Y, *Hum. Genome Var.* 5, 29 (2018). [PubMed: 30374405]
14. Legras JLet et al., *Mol. Biol. Evol.* 35, 1712–1727 (2018). [PubMed: 29746697]
15. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES, *Nature* 423, 241–254 (2003). [PubMed: 12748633]
16. Scannell DRet et al., *Genes Genomes Genet.* 1, 11–25 (2011).
17. Duan SF et al., *Curr. Biol.* 29, 1126–1136.e5 (2019). [PubMed: 30905601]
18. Turcotte B, Liang XB, Robert F, Soontornngun N, *FEMS Yeast Res.* 10, 2–13 (2010). [PubMed: 19686338]
19. Kim JH, Roy A, Jouandot D II, Cho KH, *Biochim. Biophys. Acta* 1830, 5204–5210 (2013). [PubMed: 23911748]
20. Roop JI, Chang KC, Brem RB, *Nature* 530, 336–339 (2016). [PubMed: 26863195]
21. Lang GI, Murray AW, Botstein D, *Proc. Natl. Acad. Sci. U.S.A.* 106, 5755–5760 (2009). [PubMed: 19299502]
22. Coelho AI, Rubio-Gozalbo ME, Vicente JB, Rivera I, *J. Inherit. Metab. Dis.* 40, 325–342 (2017). [PubMed: 28281081]
23. Lai K, Elsas LJ, Wierenga KJ, *IUBMB Life* 61, 1063–1074 (2009). [PubMed: 19859980]
24. Boocock J, Sadhu MJ, Durvasula A, Bloom JS, Kruglyak L, Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast (Figure creation), Dryad (2020); 10.5068/D14370.
25. Boocock J, theboocock/ancient\_bal\_scripts: submission, Zenodo (2020); 10.5281/zenodo.4132713.

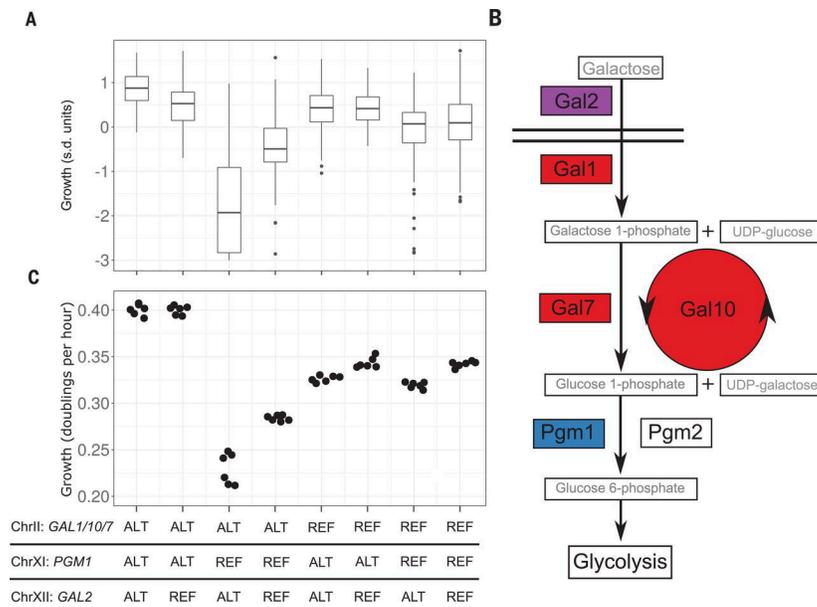
26. Boocock J, theboocock/popgen\_utilities: submission, Zenodo (2020); 10.5281/zenodo.4131787.
27. Boocock J, theboocock/gal\_bal: submission, Zenodo (2020); 10.5281/zenodo.4107954.

Author Manuscript

Author Manuscript

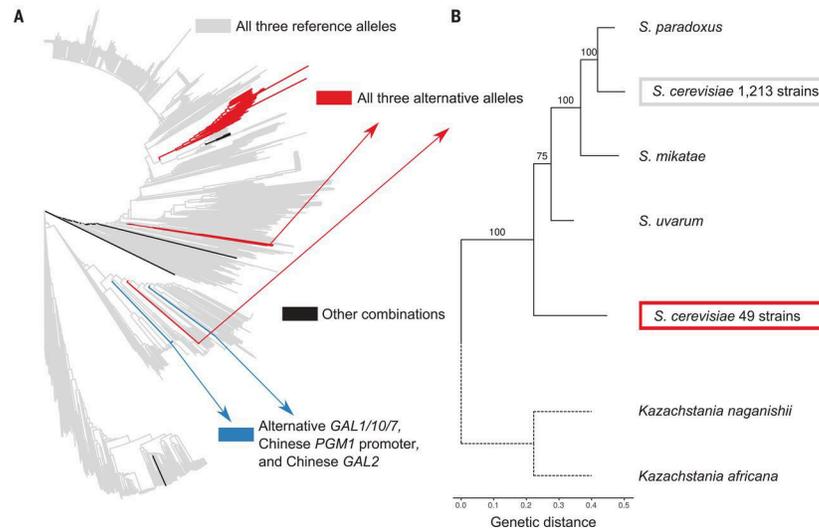
Author Manuscript

Author Manuscript



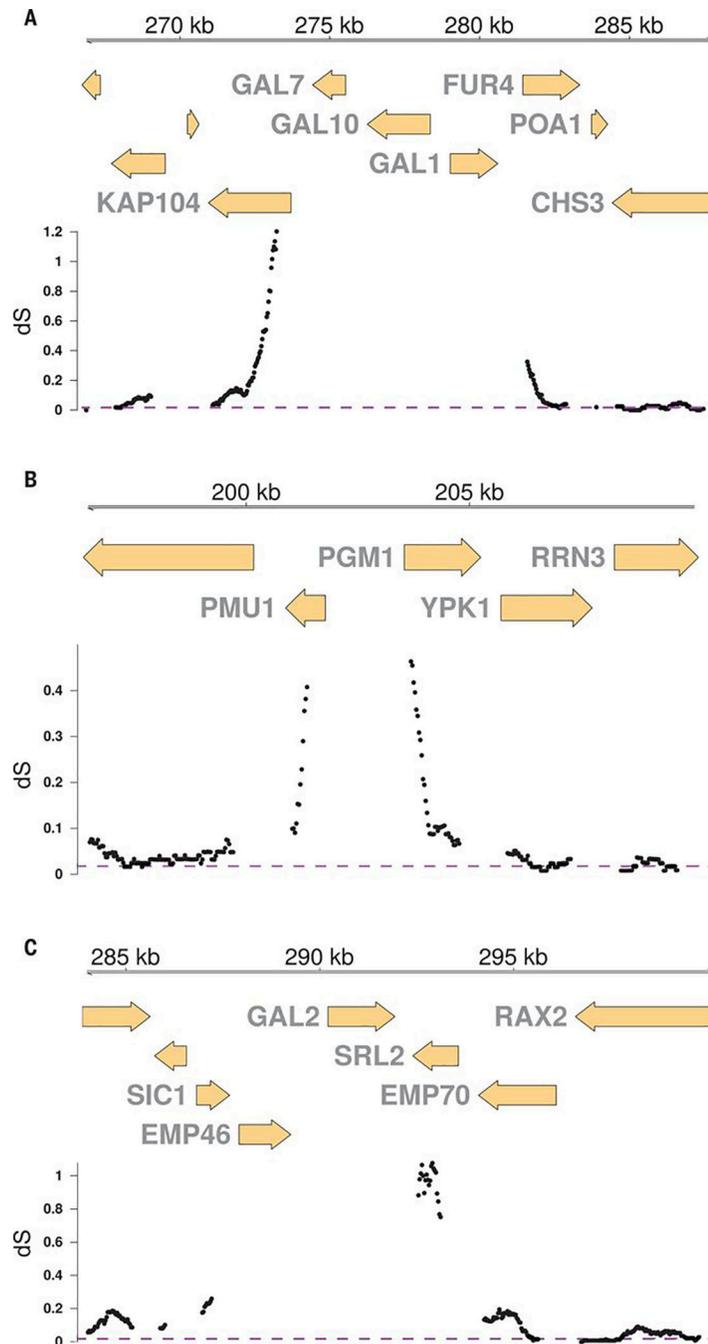
**Fig. 1. Three-locus genetic interaction for growth in galactose.**

(A) Boxplots show growth in galactose of yeast segregants ( $n = 867$ ) derived from a cross between CBS2888 and YJM981. Each boxplot corresponds to segregants with one of eight distinct combinations of alleles at the three loci (ChrII, *GAL1/10/7*; ChrXI, *PGM1*; and ChrXII, *GAL2*). (B) Galactose metabolic pathway. Components of the pathway corresponding to the three loci are shown in different colors. UDP, uridine 5'-diphosphate. (C) Growth of allele replacement strains in galactose. BY alleles [reference (REF)] and CBS2888 alleles [alternative (ALT)].



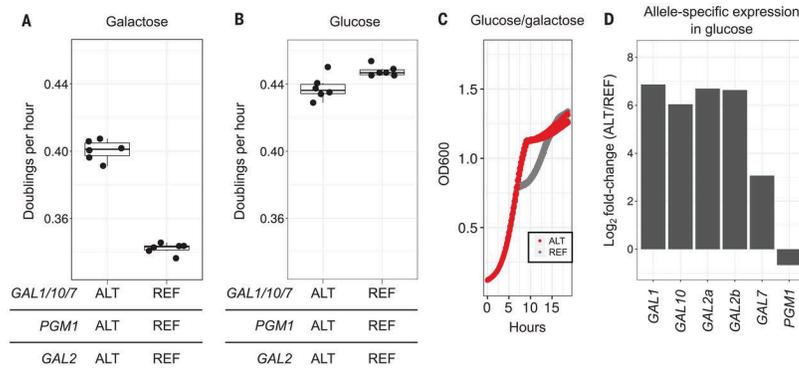
**Fig. 2. The alternative galactose alleles are broadly distributed and fall outside the *Saccharomyces* genus.**

(A) Genome-wide neighbor-joining tree of 1276 sequenced yeast isolates. (B) Bootstrapped maximum likelihood phylogenetic tree of the *GAL1/10/7* alleles from CBS2888 (alternative), BY (reference), other species in the *Saccharomyces* genus, and two outgroup species. The outgroup branches (dotted lines) were rescaled to the average branch length.



**Fig. 3. A signature of ancient balancing selection.**

Estimated rate of synonymous substitutions per site (dS) in 200-codon windows stepped every 10 base pairs between CBS2888 (alternative) and BY (reference) genes surrounding the galactose loci. (A to C) Genes adjacent to *GAL1/10/7* (A), genes adjacent to the *PGM1* promoter (B), and genes adjacent to *GAL2* (C). The purple dashed line shows the genome-wide average dS of 0.014. The moderately elevated dS in some genes (*SIC1* and *KAP104*) provides evidence of ongoing balancing selection (4).



**Fig. 4. Trade-offs between the alternative and reference alleles in the galactose pathway.** (A) Growth rates of allele replacement strains ( $n = 6$ ) with all three reference (right) or all three alternative (left) alleles in galactose as a sole carbon source. (B) As in (A), but for cells grown in glucose. (C) Growth curves of allele replacement strains with all three alternative (red) or reference (gray) galactose alleles in mixed glucose and galactose medium. OD600, optical density at 600 nm. (D) Allele-specific expression of the galactose genes from a diploid hybrid (CBS2888xBY) strain grown in glucose.