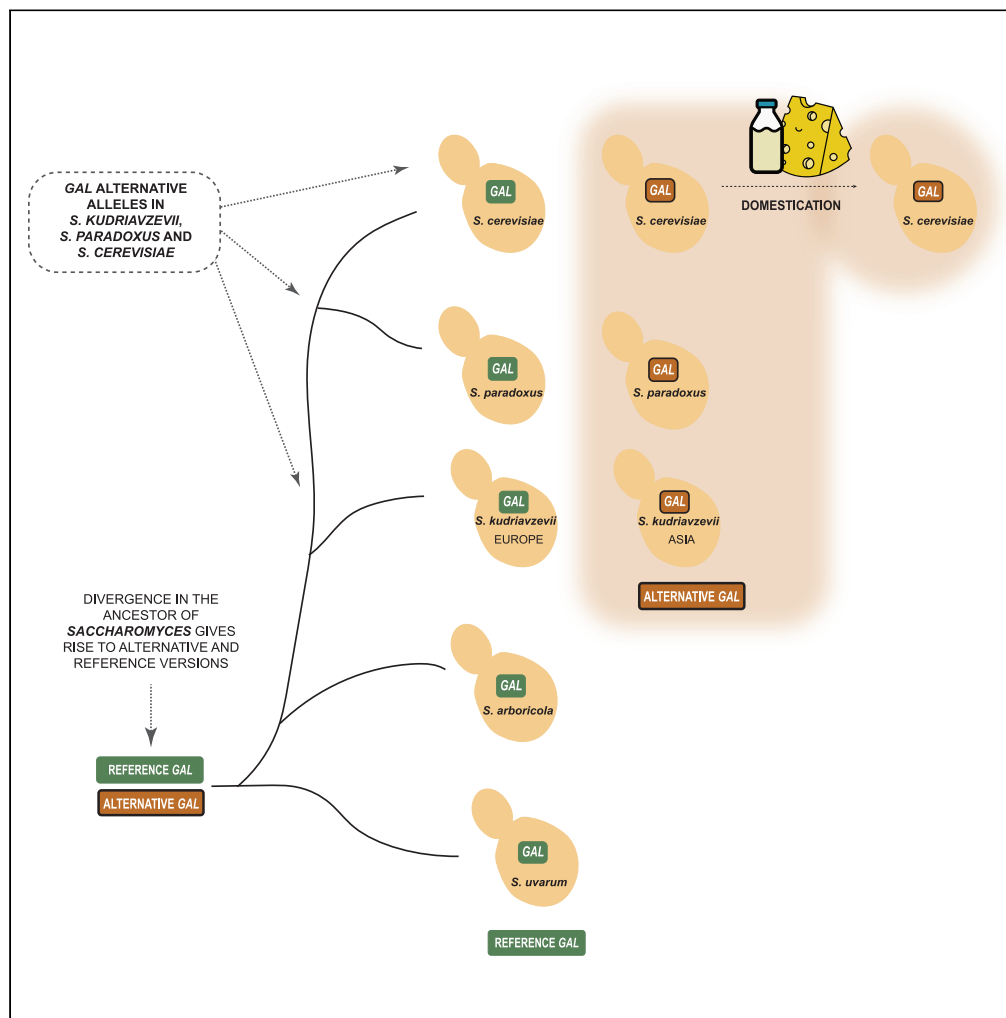


Article

Tracking alternative versions of the galactose gene network in the genus *Saccharomyces* and their expansion after domestication

Ana Pontes,
Francisca Paraíso,
Yu-Ching Liu, ...,
Chris Todd
Hittinger, Paula
Gonçalves, José
Paulo Sampaio

ap.pontes@fct.unl.pt (A.P.)
pmz@fct.unl.pt (P.G.)
jss@fct.unl.pt (J.P.S.)

Highlights

Novel alternative GAL
allele found in a wild
population of *S. cerevisiae*

Detection of alternative
GAL alleles in *S. paradoxus*
and *S. kudriavzevii*

Glucose repression is
relaxed or abolished in
strains with alternative GAL
genes

Retention of GAL
alternative alleles
promoted adaptation to
dairy environments

Pontes et al., iScience 27,
108987
February 16, 2024 © 2024 The
Authors.
[https://doi.org/10.1016/
j.isci.2024.108987](https://doi.org/10.1016/j.isci.2024.108987)

Article

Tracking alternative versions of the galactose gene network in the genus *Saccharomyces* and their expansion after domestication

Ana Pontes,^{1,2,*} Francisca Paraíso,^{1,2} Yu-Ching Liu,³ Savitree Limtong,^{4,5} Sasitorn Jindamorakot,⁶ Lene Jespersen,⁷ Carla Gonçalves,^{1,2} Carlos A. Rosa,⁸ Isheng Jason Tsai,³ Antonis Rokas,^{9,10} Chris Todd Hittinger,¹¹ Paula Gonçalves,^{1,2,*} and José Paulo Sampaio^{1,2,12,*}

SUMMARY

When *Saccharomyces cerevisiae* grows on mixtures of glucose and galactose, galactose utilization is repressed by glucose, and induction of the GAL gene network only occurs when glucose is exhausted. Contrary to reference GAL alleles, alternative alleles support faster growth on galactose, thus enabling distinct galactose utilization strategies maintained by balancing selection. Here, we report on new wild populations of *Saccharomyces cerevisiae* harboring alternative GAL versions and, for the first time, of *Saccharomyces paradoxus* alternative alleles. We also show that the non-functional GAL version found earlier in *Saccharomyces kudriavzevii* is phylogenetically related to the alternative versions, which constitutes a case of trans-specific maintenance of highly divergent alleles. Strains harboring the different GAL network variants show different levels of alleviation of glucose repression and growth proficiency on galactose. We propose that domestication involved specialization toward thriving in milk from a generalist ancestor partially adapted to galactose consumption in the plant niche.

INTRODUCTION

In nature, microorganisms typically thrive in complex and fluctuating environments and are faced with time-varying conditions that include the type of nutrients and specific stresses. Such habitat heterogeneity is understood to favor adaptive phenotypic plasticity, which can potentially lead to local adaptation.^{1,2} To cope with these challenges, biological systems evolved regulatory circuits that put in place diverse and elaborate cellular decision-making processes.³ With respect to nutrient utilization, the carbon catabolite repression systems are important and widespread regulatory mechanisms that prioritize the consumption of substrates that require the least resources and function by inhibiting pathways involved in the catabolism of carbon sources that are not preferred until the preferred ones are exhausted.^{4,5} Glucose is the preferred carbon source for the yeast *Saccharomyces cerevisiae*, as well as many other microorganisms.^{6–9} Importantly, the combination of this body of knowledge with recently acquired understanding of population structure in *S. cerevisiae*^{10–12} may foster the assessment of intra-specific phenotypic variation in the regulation of carbon metabolism and its contribution to adaptation in natural environments.

In complex and fluctuating environments, it is expected that optimization of ecological adaptation under a particular condition may have a fitness cost under other environmental conditions.^{13,14} Such evolutionary trade-offs have been well illustrated in *S. cerevisiae* growing on mixtures of carbon sources. Substrate transitions may cause growth to slow or stop for the period necessary to implement a new gene expression program suitable for the assimilation of the less preferred carbon source. The resulting growth profile is named diauxic and in such growth regimes, the overall fitness is positively correlated with the maximal growth rates on each of the carbon compounds and negatively correlated with the length of the transition phase.¹⁵ When *S. cerevisiae* is exposed to mixtures of glucose and galactose, galactose utilization is repressed

¹UCIBIO, Department of Life Sciences, Nova School of Science and Technology, Caparica 2829-516, Portugal

²Associate Laboratory i4HB, Nova School of Science and Technology, Caparica 2829-516, Portugal

³Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan

⁴Department of Microbiology Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

⁵Biodiversity Center Kasetsart University, Bangkok 10900, Thailand

⁶Microbial Diversity and Utilization Research Team, Thailand Bioresource Research Center, National Centre for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology, Development Agency (NSTDA), Pathum Thani 12120, Thailand

⁷Department of Food Science, University of Copenhagen, 1958 Frederiksberg C, Denmark

⁸Departamento de Microbiologia, ICB, C.P. 486, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil

⁹Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA

¹⁰Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN 37235, USA

¹¹Laboratory of Genetics, DOE Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, Center for Genomic Science Innovation, J.F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53726, USA

¹²Lead contact

*Correspondence: ap.pontes@fct.unl.pt (A.P.), pmz@fct.unl.pt (P.G.), js@fct.unl.pt (J.P.S.)

<https://doi.org/10.1016/j.isci.2024.108987>



by glucose and growth is typically diauxic. It is well known that the activation of the galactose (GAL) utilization pathway in this yeast is governed by a fine-tuned and strain-variable signal integration system that directs turning the GAL genes on or off, as well as determining their expression levels.^{16–18}

The genetic makeup necessary for galactose metabolism in *S. cerevisiae* and many more species of the Saccharomycotina includes GAL2 that codes for a galactose transporter and GAL7, GAL10, and GAL1 that code for the enzymes that convert galactose into glucose 1-phosphate and that are the only genes of the network that are clustered. These four genes are usually designated the GAL structural genes. PGM1 and PGM2 are accessory genes not exclusively associated with galactose utilization that code for an isomerase that generates glucose 6-phosphate that subsequently enters glycolysis. This network also includes genes that code for a transcriptional activator (GAL4), a co-repressor (GAL80), and a co-inducer (GAL3).

In *S. cerevisiae*, more than 1,000 genome sequences are available,¹² which opens the opportunity to access trait variation at the population level with an unprecedented detail. This led to the detection of distinct sets of GAL structural genes in certain populations. Besides the reference version, present in most strains and in the S288c reference genome, two alternative sets of highly divergent alleles of the GAL structural genes were recently uncovered.^{19,20} One such set of alleles, found in strains associated with dairy products,^{11,12,20,21} was found to support faster growth on galactose than on glucose, in sharp contrast to the reference alleles. The second set of alternative alleles was found in a few representatives of a wild Chinese population.^{19,22} GAL network recombination between reference and alternative alleles appears to be non-adaptive, which further highlights that the distinct galactose utilization strategies detected in *S. cerevisiae* appear to constitute an anciently established evolutionary trade-off maintained by a rare case of balancing selection acting on multiple loci instead of a single locus.²² This situation parallels the GAL network polymorphism maintained over a vast period of time in *S. kudriavzevii* and that results in Gal+ and Gal- populations.²³

Here, we take advantage of the discovery of new wild populations of *S. cerevisiae* in Southeast Asia and in South America to revisit the diversification of the GAL gene network in the genus *Saccharomyces*. We uncover an additional alternative GAL version for *S. cerevisiae* and an alternative version for *S. paradoxus* that, together with previously known alternative versions in *S. cerevisiae* and the non-functional version found earlier in *S. kudriavzevii*, represent a remarkable case of trans-specific maintenance of a highly divergent GAL gene network that encompasses the entire genus. Domestication of *S. cerevisiae* in the dairy environment co-opted the alternative GAL version of the network, and we provide evidence that supports a phenotypic specialization toward thriving in milk from a more generalist one, which is partially adapted to galactose consumption in the plant niche.

RESULTS

Alternative GAL alleles in *S. cerevisiae* and *S. paradoxus*

Recently, a large dataset of 1,276 genome sequences of *S. cerevisiae* were surveyed for the presence of alternative GAL alleles.²² Here, we expanded this survey by analyzing representatives of two novel populations that we detected in Thailand and of two other populations previously found in Brazil²⁴ (Table S1) but not yet investigated in detail. Our phylogenetic analysis comprised representatives of all known populations of *S. cerevisiae* described so far^{11,12,21,25} and included the recently described populations detected in Taiwan.¹⁰ The new Thai lineages, which we designate as Thailand 1 and Thailand 2, are entirely composed of strains isolated from natural environments (Table S1) and are closely related to other Asian lineages, namely Taiwan 2 (Thailand 1) and China III (Thailand 2). The survey of new genomes containing alternative GAL alleles allowed us to expand the inventory of alternative allelic versions from two to three, here designated as CER-A1, the alternative allele present in *S. cerevisiae* dairy strains; CER-A2, the alternative allele present in the China III population; and CER-A3, the newly found alternative allele detected in the Thailand 1 population (Figures 1A, 1B, and S2; Table S1). The first two alleles were previously found in strains associated with dairy products and in a wild Chinese population (China III),^{19,22} respectively. The novel one (CER-A3) was found in a wild Thai population (Thailand 1) that also harbored the CER-A2 and reference alleles (Figure 1A; Table S1). A phylogeny constructed based on an alignment of the concatenated GAL7-10-1 genes of all natural (non-hybrid) *Saccharomyces* species showed that all species have retained a reference GAL version and that the *S. cerevisiae* alternative versions are closely related with each other but nevertheless distinguishable (Figure 1B).

In addition, we obtained indirect evidence for the presence of an alternative GAL version in *S. paradoxus*, the closest relative of *S. cerevisiae*. This finding involved the study of a natural *S. cerevisiae* X *S. paradoxus* hybrid (strain UFMG-CM-Y5863) found in Brazil that had approximately equal contributions of the two species (Figure 1C). The existence of divergent GAL alleles in *S. paradoxus* can be seen in a sliding window analysis of the region of chromosome II that includes the GAL cluster (Figure 1D). In this analysis, we compared the divergence of the two sub-genomes of the hybrid and the divergence of an *S. paradoxus* sequence (strain YPS138 from North America) to *S. cerevisiae* S288c. Although the *S. cerevisiae* sub-genome contained the reference GAL version and showed almost no divergence to the reference *S. cerevisiae* genome, the GAL7-10-1 *S. paradoxus* sequence of the hybrid was considerably more divergent from the *S. cerevisiae* reference than the American *S. paradoxus* sequence (Figure 1D). The highly divergent region comprised only the GAL cluster, and divergence decreased sharply in the neighboring regions (Figure 1D). In line with this observation, the *S. paradoxus* sub-genome GAL sequence turned out to be phylogenetically more related to the alternative GAL versions of *S. cerevisiae* than to the reference versions of *S. cerevisiae* and remaining species in the genus (Figure 1B). This indirect finding of an alternative GAL version in *S. paradoxus* was corroborated by the detection of additional introgressed alternative *S. paradoxus* GAL versions in several Brazilian wild strains of *S. cerevisiae* (Figures 1A and 1B). The widespread dissemination of genetic material from *S. paradoxus* into the Brazilian populations WB1 and WB3 has been already reported,²⁴ and the same phenomenon was observed for two other American populations of *S. cerevisiae*, usually

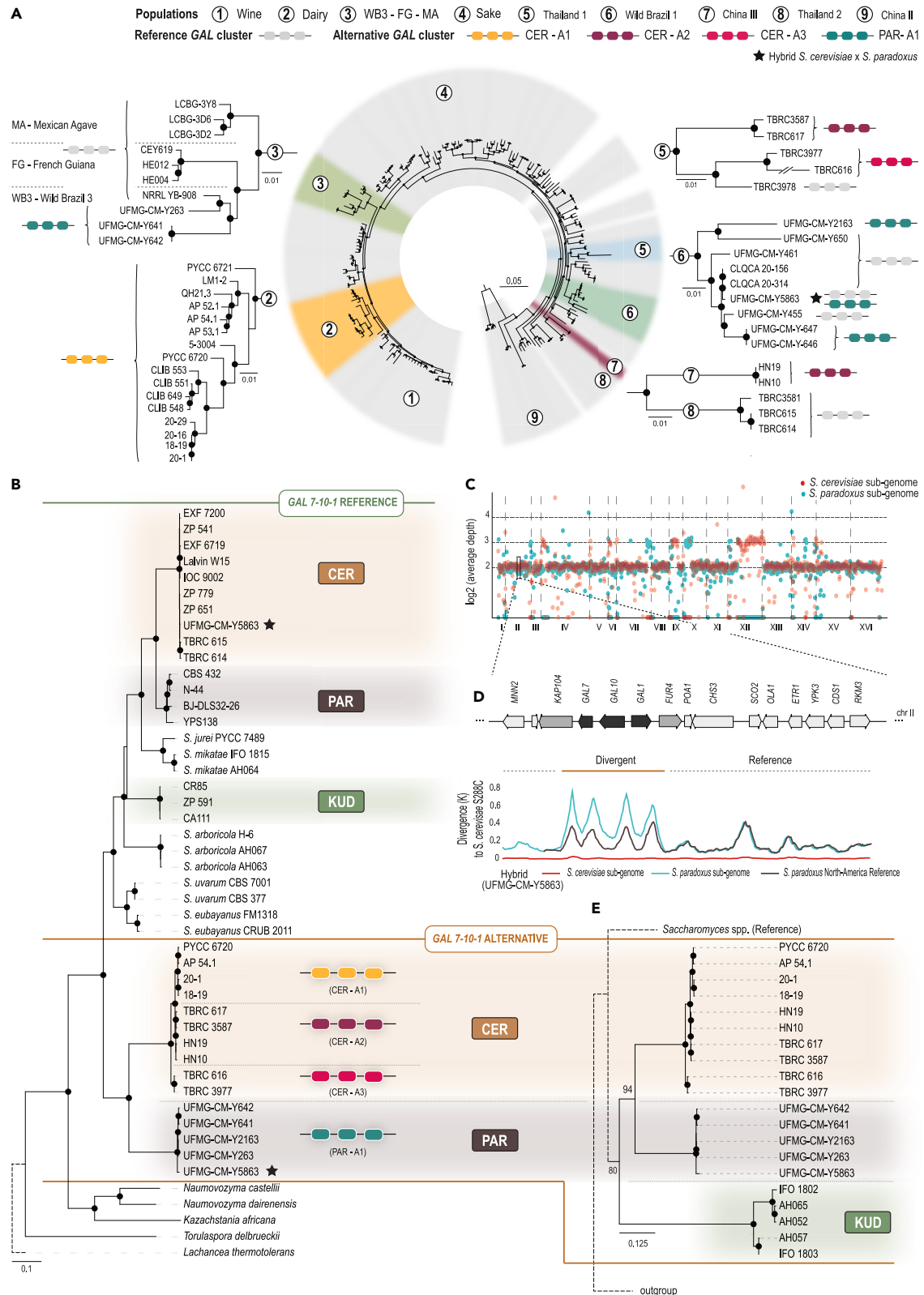


Figure 1. Reference and alternative versions of GAL genes are present in several *Saccharomyces* species and predate the formation of the genus

(A) Phylogeny representing the known lineages of *S. cerevisiae* and highlighting in color those that harbor alternative GAL versions (reference version depicted in gray). In the case of populations 3 and 6 that harbor an alternative GAL version introgressed from *S. paradoxus* (see below), the reference version is also present as depicted in the pruned trees derived from whole genome data. For the Thailand 1 population, two alternative versions (CER-A2 and CER-A3) were detected together with the reference. The phylogeny was inferred from 214 sequences and 965,534 high-quality homozygous single nucleotide polymorphisms using the TVM+F+ASC+G model of sequence evolution and the maximum-likelihood method as implemented in IQ-TREE; it was rooted with representatives from the China IX/Taiwan 1 populations. Branch lengths correspond to the expected number of substitutions per site, and black dots in tree nodes depict bootstrap support values above 95% (1,000 replicates).

(B) Concatenated maximum-likelihood phylogeny of the GAL gene cluster (GAL7-10-1) for the genus *Saccharomyces* showing the *trans*-specific organization of reference and alternative versions (CER, *S. cerevisiae*; PAR, *S. paradoxus*; KUD, *S. kudriavzevii*). The alternative GAL versions (CER-A1, CER-A2, CER-A3, and PAR-A1) are color coded. The phylogeny was rooted with *Lachancea thermotolerans*.

(C) Whole-genome normalized coverage plot of the *S. cerevisiae* X *S. paradoxus* hybrid strain UFMG-CM-Y5863 showing the approximately equal contribution of both sub-genomes. Reads that mapped to *S. cerevisiae* are shown in red and those that mapped to *S. paradoxus* are shown in blue.

(D) Divergence of the *S. cerevisiae* and *S. paradoxus* sub-genomes of UFMG-CM-Y5863 to *S. cerevisiae* S288C in the region of chromosome II that includes the GAL cluster and several flanking genes. A third sequence, that of the American *S. paradoxus* strain YPS138, was used to control for the expected divergence between *S. cerevisiae* and *S. paradoxus*. Although the *S. cerevisiae* sub-genome contained the reference GAL version and showed almost no divergence, the *S. paradoxus* sub-genome diverged more than expected by comparison with the American *S. paradoxus* strain in the GAL region but not in the flanking regions.

(E) Detail of a phylogeny of the GAL cluster including the alternative versions of *S. kudriavzevii*, which are all pseudogenes. In (D) and (E), black dots on the tree nodes indicate bootstrap values higher than 95% (2,000 replicates), and dotted lines represent branches not to scale. Scale bar corresponds to the expected number of substitutions per site.

designated as “French Guiana” and “Mexican Agaveae,”^{12,26} and that are closely related to WB3, a wild Brazilian population (Figure 1A). Except for the 1:1 hybrid strain, all other strains harboring the *S. paradoxus* alternative version lacked the *S. cerevisiae* version. However, the alternative version seems to be absent among known isolates of *S. paradoxus* itself, as shown by the surveillance of more than 50 genomes from representatives of the various populations known, spanning Europe, Far East, China, and North America (Table S2).

Alternative versions of the structural gene GAL2 were also previously detected in the Dairy and China III populations.^{19,22} In line with these findings, we also identified highly divergent versions of this gene in the newly found populations that harbored the alternative GAL7-10-1 structural genes. The alternative GAL2 genes were always duplicated, and, in case of dairy strains, the different copies resulted from recombination between reference and alternative alleles (Figure S1A). The phylogenetic placement of the remaining alternative versions was similar to what was observed for the GAL cluster (Figure S1B), suggesting that alternative GAL2 genes have the same origin as alternative structural genes, as previously suggested.²² Interestingly, the Brazilian strains harboring *S. paradoxus* GAL7-10-1 also carry GAL2 genes whose phylogenetic placement is in line with an *S. paradoxus* GAL2 alternative allele (Figure S1B). Moreover, the *Saccharomyces eubayanus* GAL2 sequences that were recently reported as being divergent²⁷ clustered at the base of the *S. eubayanus-Saccharomyces uvarum* GAL2 (reference version) clade and not with the alternative versions (Figure S1B). This suggests that the *S. eubayanus* divergent versions are the remnants of a duplication that occurred on the reference version lineage; specifically, they occurred in the ancestor of *S. eubayanus* and *S. uvarum* and do not represent the alternative version. Another interesting finding was the detection of recombinant variants of GAL7 in two strains of the African Beer population of *S. cerevisiae* (Figure S2A). These variants seem to result from the recombination of *S. cerevisiae* reference and alternative versions and have distinct contributions from each of the two versions (Figure S2B).

To further test if the diversification of the genes present in the cluster into reference and alternative versions predates the origin of the genus, we calculated the number of pairwise synonymous substitutions per synonymous sites (Ks). First, these measurements were made for the pair *S. cerevisiae*-*S. eubayanus* to have an estimate of divergence across the genus. Then, equivalent measurements were made between *S. cerevisiae* and *S. kudriavzevii* and, for an estimate of intraspecific divergence, between two *S. cerevisiae* strains from distinct populations. Then, these values were used as reference points to which estimates of *S. cerevisiae* and *S. paradoxus* reference versus alternative GAL cluster genes Ks estimates were compared. The Ks values calculated between reference and alternative GAL alleles were higher (Ks ~1.5–2.5) than the value for *S. cerevisiae*-*S. eubayanus* (Ks~1.1), which suggests that the divergence of the reference-alternative pairs is older than that of the genus *Saccharomyces* (Figure S3), as already reported for *S. cerevisiae*.²²

GAL pseudogenes of *S. kudriavzevii* correspond to the alternative GAL alleles

Using the GAL7-10-1 sequences retrieved from the genomes of Asian *S. kudriavzevii* isolated in Taiwan, we observed that those genes appear to be phylogenetically more closely related to the alternative versions of *S. cerevisiae* and *S. paradoxus* than to the European version of *S. kudriavzevii* (Figures 1E and S4A). This result suggests that the European version corresponds to the reference version, and the Asian version might correspond to the alternative version of the GAL genes. Given that the Asian strains contain only GAL pseudogenes whose sequences are shorter than the normal sequences, we trimmed the remaining sequences in the alignment to the length of the corresponding pseudogenes for the phylogenies shown in Figures 1E and S4A. Despite their short lengths and degrees of pseudogenization, these sequences occupy positions in the phylogeny that suggest a shared evolutionary history with the *S. cerevisiae* and *S. paradoxus* alternative GAL version. We considered four tree topologies of alternative evolutionary scenarios and all yielded significantly lower likelihoods (Figure S4B, approximately unbiased test, p value < 0.05). In addition, we observed that the sequence divergence pattern is similar in the three species (*S. cerevisiae*, *S. paradoxus*, and *S. kudriavzevii*), with a parallel divergence built up in the flanking genes KAP104 and FUR4

(Figure S4C), which suggests a similar and ancient process of balancing selection in the three species. Furthermore, the *S. kudriavzevii* GAL2 allele found in the Asian strains appears to correspond also to the alternative version, as sequences from these versions clustered with alternative versions of *S. cerevisiae* and *S. paradoxus*, albeit between these two species and with low statistical support (Figure S1C).

Other genes related to galactose metabolism

PGM1/2 code for phosphoglucomutases that convert glucose-1-phosphate to glucose-6-phosphate, thereby connecting galactose metabolism with glycolysis. The genus-level phylogenetic reconstruction of the evolution of PGM1 hinted at a trajectory distinct from that of the GAL cluster because a duplication of PGM1 seems to have occurred more recently, only for *S. cerevisiae* and *S. paradoxus* and after speciation (Figure 2). Therefore, the *trans*-specific effect seen for the GAL cluster was not observed for PGM1. However, the number of alternative alleles we could recover was similar for the GAL cluster and PGM1 in *S. cerevisiae* and *S. paradoxus* (Figure 2). For *S. kudriavzevii*, reference PGM1 versions were found in the genomes that harbored the alternative and inactive GAL7-10-1 version (Figure 2), contrary to *S. cerevisiae* and *S. paradoxus*.

It was shown that, in the Saccharomycetaceae, the promoter region of PGM1/2 was the best predictor of robust growth on galactose and that improved growth was linked to the presence of one or more Gal4-binding site(s) in the promoter of either of the two paralogs.²⁸ In *S. cerevisiae* strains with the GAL reference version, PGM2 responds modestly to galactose through an unidentified mechanism independent of Gal4.^{29,30} More recently, it was shown that *S. cerevisiae* strains with alternative GAL alleles also harbored a Gal4-binding site in the PGM1 promoter region, contrary to the strains carrying the reference GAL allele.²² In line with these findings, we observed that, in genomes with CER-A1, CER-A2, and CER-A3 alternative versions of GAL7-10-1, alternative versions of PGM1 with one Gal4-binding site could also be found (Figure 2). Alternative *S. paradoxus* PGM1 versions with promoters harboring Gal4-binding sites were also detected in *S. cerevisiae* strains carrying *S. paradoxus* alternative GAL alleles (Figure 2). In *S. kudriavzevii*, like in *S. cerevisiae* GAL reference strains,³¹ PGM2 seems to be the most important of the two paralogs for galactose metabolism because PGM2 harbors a functional Gal4-binding site in this species, although only in the European strains. However, no alternative versions of either PGM1 or PGM2 were found in *S. kudriavzevii*, and reference PGM1 versions were found in genomes that harbored either the reference or the alternative and inactive GAL7-10-1 version (Figure 2). Moreover, we detected the Gal4-binding site in *Saccharomyces arboricola*, *S. uvarum*, and *S. eubayanus*, three species for which we only found the reference version of GAL7-10-1. For *S. kudriavzevii*, *Saccharomyces jurei*, and *Saccharomyces mikatae*, the Gal4-binding site in the promoter region of PGM1 was always absent. It is interesting to note that the alternative versions of PGM1 always harbor the Gal4-binding site in their promoters (Figure 2; Table S1), an observation already made but for a more limited dataset.²²

The MEL1 gene is yet another Gal4-regulated gene. This gene codes for an α -galactosidase that hydrolyzes oligo- or disaccharides like raffinose or melibiose, which are natural sources of galactose. The MEL1 gene is very rare in *S. cerevisiae*,¹¹ and it was therefore striking to observe that it is present in most of the wild *S. cerevisiae* strains that possess the alternative GAL cluster (Table S1). This difference of MEL1 occurrence in GAL reference and alternative wild backgrounds is statistically significant for the approximately 200 strains listed in Table S1 (t test, p value < 0.001).

Growth rates on glucose and galactose

It has been reported that strains with the alternative CER-A1 and CER-A2 alleles grow faster on galactose than strains with the reference alleles.^{21,22} Here, we expanded the analysis to include the novel alternative alleles presently uncovered. To this end, growth rates on glucose and galactose were first measured for strains harboring the reference allele. Those measurements varied considerably, as illustrated by the difference between growth rates on glucose and galactose (Figure 3A; Table S1), in line with previous studies.^{18,19} When the analysis was extended to strains with alternative versions, average growth rates on galactose were significantly higher for strains with alternative alleles in comparison with the reference GAL version (35%–50% higher; Kruskal-Wallis test, p value < 0.001; Figure 3B). On glucose, wild strains harboring alternative alleles (CER-A2, CER-A3, and PAR-A1) exhibited growth rates similar to those of strains carrying the reference GAL versions (Figure 3C), but domesticated strains carrying the CER-A1 allele had significantly lower median growth rates when compared with strains with the reference version (Kruskal-Wallis test, p value < 0.001), as previously observed.¹⁹

Glucose repression of galactose utilization for alternative versus reference alleles

Growth of *S. cerevisiae* on mixtures of glucose and other sugars usually leads to a diauxic growth pattern that reflects repression by glucose of the utilization of less favored carbon sources. Accordingly, growth was diauxic in strains harboring reference GAL7-1-10 alleles growing on 96-well plates in medium containing 1% (w/v) each of glucose and galactose. However, none of the strains carrying any of the alternative alleles exhibited this growth pattern (Figure 4A). To confirm that this difference is due to a disruption in the usual pattern of sequential consumption of glucose and galactose, larger scale cultures were set up on medium containing equal amounts of glucose and galactose, and sugar consumption was quantified using HPLC. The results in Figures 4B and S5 show that, for strains carrying alternative GAL7-1-10 alleles, the pattern of sugar consumption is clearly different from that observed for strains carrying the reference allele. Whereas in the latter case galactose consumption is initiated only after glucose is exhausted, strains with alternative versions consume glucose and galactose simultaneously or galactose consumption, although delayed, is initiated before glucose is totally consumed (Figures 4B and S5). For strains carrying the CER-A1 allele, it was previously reported that the combined effect of the lack of binding sites for the Mig1 repressor in the promoters of GAL4 (activator of the GAL regulon) and of GAL1-10, combined with a mutation in the repressor GAL80, probably preventing its binding to Gal4, resulted in constitutive GAL structural gene expression in dairy strains.¹⁹ In addition, these strains lost the HXT6 and HXT7



Reference and alternative versions of *PGM1* are highlighted, together with presence or absence of the binding site for Gal4 in the promoter region of *PGM1* and *PGM2*. Maximum likelihood phylogeny inferred from 52 *PGM1* coding sequences using the GTR+F+I+G4 model of sequence evolution. The phylogeny was rooted with *Lachancea thermotolerans*, and the scale bar corresponds to the expected number of substitutions per site. Black dots on the tree nodes mark bootstrap values higher than 95% (2,000 replicates). The dotted line represents branches not to scale.

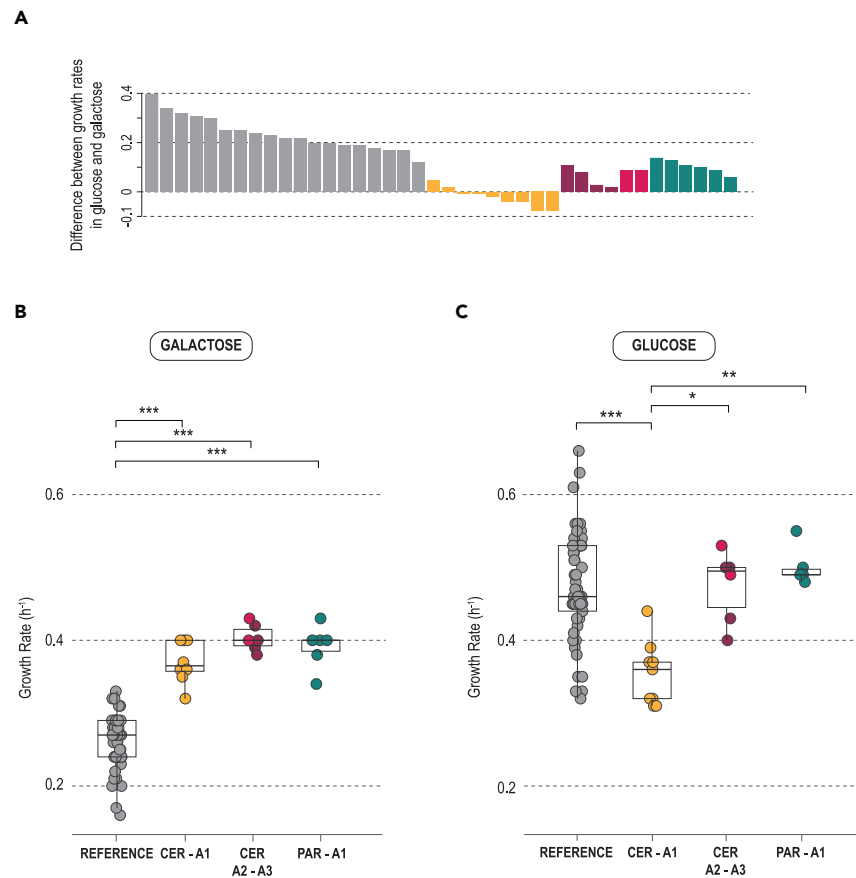


Figure 3. Comparison of growth rates in glucose and galactose among strains carrying the reference and alternative GAL versions

(A) Bar plot depicting, for each strain, the difference between growth rate in glucose and growth rate in galactose in synthetic media supplemented with 2% (w/v) of carbon source. GAL allele types are color coded. For strains with the reference GAL allele, the results of a representative group are depicted, and full results are shown in Table S1.

(B and C) Growth rates in galactose and glucose, respectively, measured using the same conditions as in (A). All groups were compared against each other with Dunn's post hoc test. Levels of significance (p value): ***0.001, **0.01, and *0.05.

genes, which negatively impacts glucose consumption rates. All these alterations resulted in co-consumption of glucose and galactose, with both sugars starting to be used at the same time and at very similar rates in dairy strains. The wild strains, in addition to having the Gal4-binding site in the *PGM1* promoter, also lacked a Mig1-binding site, either both in the GAL 1–10 promoter and the *GAL4* gene or only in the former. However, no relevant mutations were present in *GAL80* and *GAL3*. This gave rise to a pattern of sugar consumption in wild strains that was partially derepressed and was intermediate between the completely derepressed dairy strains and the fully repressed reference strains (Figure S5). Also, the extent of the delay of galactose assimilation relative to glucose and the rates of glucose consumption vary between the different alternative versions, raising the possibility that additional unidentified factors also impact the co-assimilation phenotype in strains carrying different classes of alleles. Because the *HXT6* and *HXT7* genes were present and seemingly functional in all wild strains, glucose transport is probably not responsible for differences in glucose consumption rates, although deficient expression of these transporters and/or the activity of other transporters cannot be excluded at this stage as possible explanations.

DISCUSSION

Ancient origin and trans-specific distribution of GAL reference and alternative versions

Heretofore, the existence of highly divergent reference and alternative versions of *GAL* structural genes had only been demonstrated for *S. cerevisiae*.¹⁹ These versions have an ancient origin that predates the formation of the species and were maintained by balancing selection.²² Here, we show that a similar situation might exist in *S. paradoxus* as we detected an alternative *GAL* allele that groups closely to *S. cerevisiae* alternative *GAL* alleles. The increase in divergence seen in the two genes that flank *GAL7-1-10* is similar to what has been previously reported for the same genes in *S. cerevisiae*²² and *S. kudriavzevii*²³ and supports the long-term maintenance of divergent states by balancing selection in the three species.

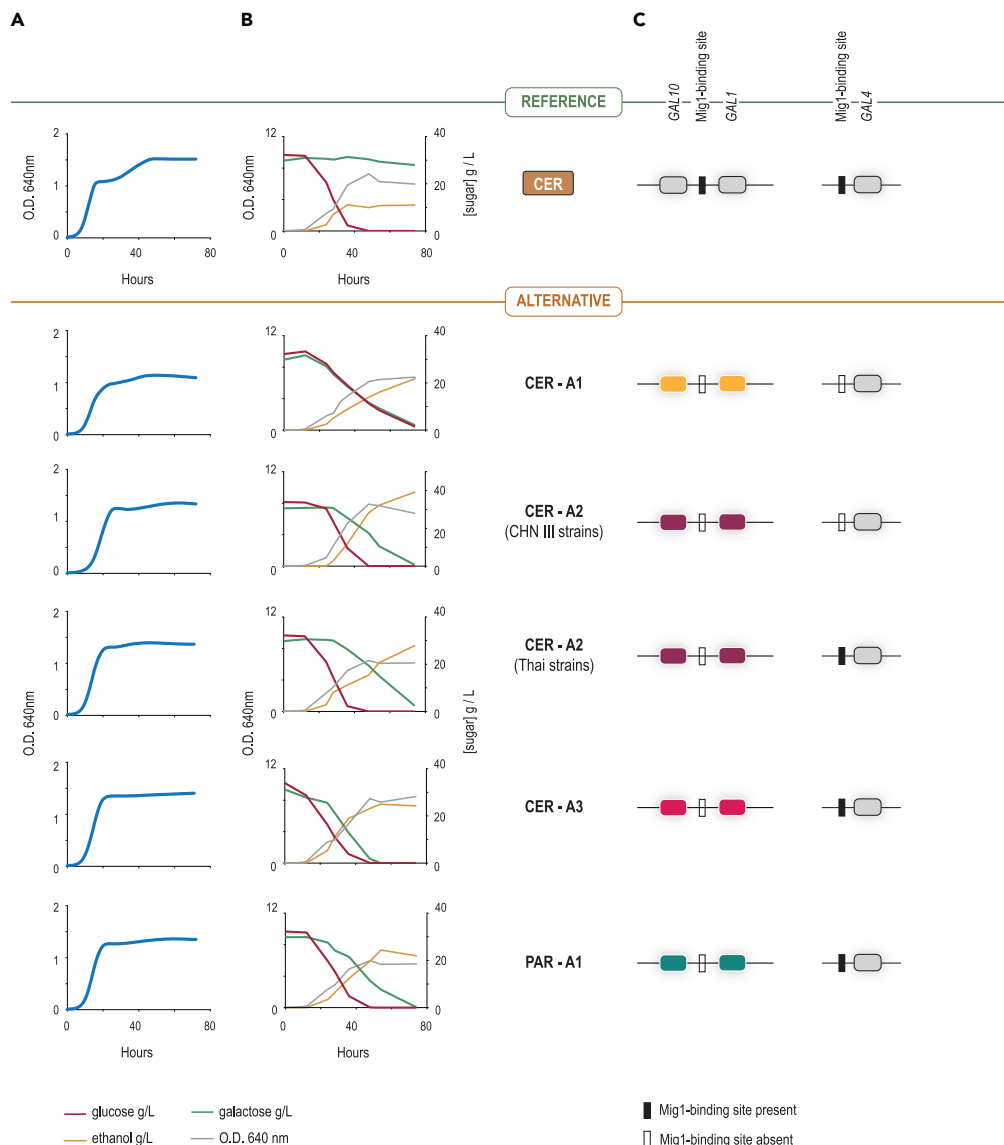


Figure 4. Strains with alternative GAL alleles have different levels of glucose repression

(A) Growth curves in glucose and galactose medium (1% w/v each) of representative strains harboring either the reference or each of the various GAL7-10-1 alternative alleles.

(B) Profiles of glucose and galactose consumption, ethanol production, and growth for the strains shown in (A) in glucose and galactose medium (3% w/v each). Median values for all variables are shown from the replicas depicted in Figure S5.

(C) Identification of presence or absence of Mig1-binding sites in the promoter regions of GAL10-7-1 and GAL4 across the spectrum of alternative versions.

We did not detect the alternative alleles directly in existing *S. paradoxus* genomes but from *S. cerevisiae* X *S. paradoxus* hybrid or introgressed ones. We find it possible that the alternative *S. paradoxus* version is present in “clean” *S. paradoxus* genomes mainly for two reasons. First, the number of cultures and genome sequences available for this species is substantially lower than that of *S. cerevisiae*. Secondly, the frequency of the alternative version in *S. cerevisiae* is very low, especially in wild strains, as will be discussed below. Therefore, the relatively low number of *S. paradoxus* genomes surveyed is compatible with the current observations. Specifically, although we did not detect the alternative version in 50 genomes representative of the different populations of *S. paradoxus*, and another study that analyzed the GAL gene network of 108 *S. paradoxus* genomes also did not report cases of high sequence divergence,²⁷ more than 1,000 genomes have been inspected in *S. cerevisiae*. We suggest that the finding of the PAR-A1 allele only in introgressed or hybrid *S. cerevisiae* strains might be seen as accidental and not linked with the interspecies hybridization process. At this stage we see no reason, apart from poor sampling, for the absence of this allele in “clean” *S. paradoxus* strains, a situation that might change in the future.

In line with previous findings of incompatibility between alternative and reference GAL genes,²² our observations suggest the same level of genetic conflict in *S. paradoxus*. Only alternative *S. paradoxus* GAL7-1-10, GAL2, and PGM1 were detected in Brazilian *S. cerevisiae* genomes retaining *S. paradoxus* introgressions. Moreover, our analyses in *S. kudriavzevii* suggest that the pseudogenized GAL alleles known in this species^{23,32} correspond to the alternative versions of *S. cerevisiae* and *S. paradoxus*.

Together, these findings support a scenario of an ancient divergence of GAL alleles in the ancestor of the genus *Saccharomyces* that become *trans*-specific through multiple speciation events, including the most recent one, that gave rise to *S. cerevisiae*. The MHC locus in mammals is a well-known case of an ancient *trans*-specific polymorphism that spans species boundaries.³³ Although an introgression from an unknown species could, conceptually, yield the observed results, forward simulations models rejected the alternative possibilities of neutral introgression or introgression followed by maintenance of the alternative and reference alleles with multi-locus balancing selection.²² Given the most likely scenario of ancient divergence, the absence of the alternative version in most of the existing *Saccharomyces* species remains to be explained. We searched the available genomes of *S. mikatae*, *S. jurei*, *S. arboricola*, *S. uvarum*, and *S. eubayanus*, but we only found the reference versions of GAL7-10-1. Moreover, a recent study that included an expanded genome dataset of all species of the genus *Saccharomyces* also did not report additional divergent GAL alleles, except for divergent GAL2 alleles,²⁷ which, as our results show, were more closely related to the reference version. As already discussed for *S. paradoxus*, it remains to be elucidated if the alternative version has been lost in these species or if it has not yet been found, as genome sampling in *Saccharomyces* species other than *S. cerevisiae* is considerably lower.

For *S. cerevisiae*, the disproportion between the frequencies of reference and alternative GAL alleles is notorious. From more than 1,300 genomes that have been examined, approximately 50 (less than 4%) contain the alternative versions. Our findings in *S. paradoxus*, in which the alternative version was only indirectly found, and in *S. kudriavzevii*, in which the alternative version is non-functional, suggest that the reference version is currently more advantageous and hint at the alternative version being on the verge of extinction. It is however remarkable that the mechanisms that in the past favored the maintenance of the two versions acted on a similar way in multiple species. We hypothesize that the large population sizes and the wide ranges of *Saccharomyces* species, together with their heterogeneous environments, favored the retention of extensive polymorphisms such as the GAL multi-locus divergence. Such polymorphisms would be hard to maintain in macroscopic organisms with smaller population sizes and experiencing, by virtue of their size and smaller number, more homogeneous environments.

Ecological relevance of galactose

Galactose is frequently associated with milk and milk products, which are a direct consequence of livestock domestication.^{34,35} However, this sugar is also a common building block of many oligosaccharides of plant origin that constitute natural galactose stores.^{36,37} For example, whereas the trisaccharide raffinose (galactose, glucose, and fructose) contains one galactose moiety, stachyose, verbascose, and ajugose contain 2, 3, and 4 galactose moieties, respectively. Thus, galactose of plant origin is ubiquitous^{38,39} and can exceed glucose when released from plant oligosaccharides.⁴⁰ Given the primal association of *Saccharomyces* to the arboreal niche,^{41–43} it is conceivable that efficient galactose utilization was among the favored traits present in the ancestor of the genus. For example, *S. uvarum*, an early derived species in the genus, has an active GAL gene network^{39,40,44} that, among other features, includes the induction of PGM1 by galactose²⁸ (Figure 5). We postulate that a more generalist strategy toward sugar utilization is an ancestral trait in the genus and that, gradually, a specialist glucose-prioritizing lifestyle gained prevalence, with *S. cerevisiae* being the epitome of this ecological behavior (Figure 5).

Phenotypic plasticity at the population level

Evolution of galactose utilization in budding yeasts is remarkably dynamic, particularly with respect to regulatory mechanisms.²⁹ Moreover, in *S. cerevisiae*, various studies have contributed to reveal an extensive phenotypic plasticity of galactose utilization strategies,^{16–18} in which the strains carrying the alternative GAL alleles^{19,22} represent one of the ends of this spectrum. Here, we detected an additional version of an alternative GAL allele (CER-A3) in a wild Thai population that also harbored the CER-A2 alternative allele and the reference one. The coexistence of alternative and reference alleles in the same population or in groups of populations that share the same environment, like in the case of Chinese sympatric wild populations, can be seen as a strategy to maximize ecological fitness in a changing environment.⁴⁵ Given that selection maintains the alternative and reference alleles, it is conceivable that, in the habitats explored by these lineages, the coexistence of cells with different phenotypes is adaptive, which would allow for a more thorough exploration of the nutritional landscape. Thus, by maintaining cells with different optimization programs, a population could explore both low and high galactose concentrations, which can be seen as a case of niche complementation⁴⁶ that maximizes optimality under a particular condition and reduces adjustment between conditions.^{3,47} However, even if this was the reason that promoted the maintenance of reference and alternative alleles, their asymmetric representation in *S. cerevisiae* and *S. paradoxus* and inactivation in *S. kudriavzevii* (Figure 5) might be an indication that the selective advantages of maintaining the alternative GAL system have diminished over time. Alternatively, it is of relevance to note that the maintenance of the two states by balancing selection does not require that their frequencies are similar. Finally, the reason behind the loss of the alternative version by gene inactivation in *S. kudriavzevii* might be particular to this species, given that the two versions are spatially separated at a continental scale.^{23,27}

Functional evolution of GAL reference and alternative alleles

Numerous studies have analyzed the function and regulation of the (reference) GAL network in *S. cerevisiae*.^{7,16,17,48} Overall, this network encodes a regulatory program of slow induction and strong repression of galactose utilization by glucose. This program leads to a fitness gain when glucose is abundant but to a fitness defect during transition from glucose to galactose.⁴⁸ By contrast, *S. uvarum* is more active toward galactose utilization and has better induction and less-stringent repression mechanisms.^{28,44,49} Even in *S. paradoxus*, the closest relative of



Domestication recovered the alternative and rare *S. cerevisiae* GAL version

The transition from the plant niche to the dairy environment during microbe domestication appears to be a common feature and encompasses other yeasts,⁵¹ filamentous fungi in cheeses,^{52,53} and lactic acid bacteria in milk.^{54,55} Similarly, to what we observed here, strong metabolic specialization toward the milk environment transformed plant niche generalists into dairy specialists.

Limitations of the study

Although one of the goals of this research was to search for the presence of alternative versions of the GAL gene network among the species of the genus *Saccharomyces*, this task was not fully accomplished because some species have been poorly sampled. Given the observed low frequency of detection of alternative versions among well-represented species, inferences for taxa with low numbers of isolates are difficult to make.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Genome sequencing, read alignment, and genotype calling
 - Phylogenetic analyses and survey of specific genes
 - Divergence analysis and hybrid identification
 - Search for binding sites of transcription factors
 - Growth assays and sugar consumption determination
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.108987>.

ACKNOWLEDGMENTS

Feng-Yan Bai is gratefully acknowledged for making available two *S. cerevisiae* cultures from a wild Chinese population. This study was supported by Fundação para a Ciência e a Tecnologia (Portugal) grants UIDP/04378/2020, UIDB/04378/2020 (UCIBIO), LA/P/0140/2020 (i4HB), PTDC/BIA-EVL/1100/2020, and PTDC/BIA-EVL/0604/2021. The computation work was carried out with support of INCD funded by FCT and FEDER under grants 22153-01/SAICT/2016 and 2023.09581.CPCA.A1. Research in the Rokas lab is supported by the National Science Foundation (DEB-2110404), the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R01 AI153356), and the Burroughs Wellcome Fund. Research in the Hittinger Lab was supported by the National Science Foundation under Grant No. DEB-2110403, the USDA National Institute of Food and Agriculture (Hatch Project 1020204 and 7005101), in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409) and an H. I. Romnes Faculty Fellowship (Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation). I. Tsai was supported by National Science and Technology Council (111-2628-b-001-021), Taiwan. C. Rosa was funded by "INCT Yeasts: Biodiversity, preservation and biotechnological innovation," funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, grant #406564/2022-1, and CNPq process numbers 0457499/2014-1, 313088/2020-9, and 408733/2021, and Fundação do Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG, process numbers APQ-01525-14 and APQ-02552-15). Field work in West Africa by L. Jespersen was supported by DANIDA, the Danish International Development Assistance.

AUTHOR CONTRIBUTIONS

Conceptualization, A.P., P.G. and J.P.S.; resources, C.G., I.J.T., S.L., S.J., L.J., and C.A.R.; investigation, A.P. and F.P.; data curation, A.P.; formal analysis, A.P., F.P., Y-C.L., and I.J.T.; visualization, A.P. and J.P.S.; writing—original draft, review, and editing, A.P., I.J.T., A.R., C.T.H., P.G., and J.P.S.; supervision, I.J.T., P.G., and J.P.S.; funding acquisition, C.G, P.G., and J.P.S.; project administration, J.P.S.

DECLARATION OF INTERESTS

A.R. is a scientific consultant for LifeMine Therapeutics, Inc. The authors declare no other competing interests.

Received: September 5, 2023

Revised: November 27, 2023

Accepted: January 17, 2024

Published: January 23, 2024

REFERENCES

- Kawecki, T.J., and Ebert, D. (2004). Conceptual issues in local adaptation. *Ecol. Lett.* 7, 1225–1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>.
- Rees, J.S., Castellano, S., and Andrés, A.M. (2020). The Genomics of Human Local Adaptation. *Trends Genet.* 36, 415–428. <https://doi.org/10.1016/j.tig.2020.03.006>.
- Perkins, T.J., and Swain, P.S. (2009). Strategies for cellular decision-making. *Mol. Syst. Biol.* 5, 326. <https://doi.org/10.1038/MSB.2009.83>.
- Gancedo, J.M. (1998). Yeast Carbon Catabolite Repression. *Microbiol. Mol. Biol. Rev.* 62, 334–361. <https://doi.org/10.1128/mmr.62.2.334-361.1998/asset/5efb087b-7f00-4b86-b6b7-bf8c07f411cb/assets/graphic/mr0280016005.jpeg>.
- Magasanik, B. (1961). Catabolite repression. *Cold Spring Harbor Symp. Quant. Biol.* 26, 249–256. <https://doi.org/10.1101/SQB.1961.026.01.031>.
- Görke, B., and Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* 6, 613–624. <https://doi.org/10.1038/nrmicro1932>.
- Johnston, M. (1999). Feasting, fasting and fermenting: glucose sensing in yeast and other cells. *Trends Genet.* 15, 29–33. [https://doi.org/10.1016/S0168-9525\(98\)01637-0](https://doi.org/10.1016/S0168-9525(98)01637-0).
- Kayikci, Ö., and Nielsen, J. (2015). Glucose repression in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 15, fov068. <https://doi.org/10.1093/femsyr/fov068>.
- Broach, J.R. (2012). Nutritional Control of Growth and Development in Yeast. *Genetics* 192, 73–105. <https://doi.org/10.1534/genetics.111.135731>.
- Lee, T.J., Liu, Y.C., Liu, W.A., Lin, Y.F., Lee, H.H., Ke, H.M., Huang, J.P., Lu, M.Y.J., Hsieh, C.L., Chung, K.F., et al. (2022). Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and evolution of predomesticated lineages. *Genome Res.* 32, 864. <https://doi.org/10.1101/GR.276286.121>.
- Pontes, A., Hutzler, M., Brito, P.H., and Sampaio, J.P. (2020). Revisiting the taxonomic synonyms and populations of *Saccharomyces cerevisiae* – phylogeny, phenotypes, ecology and domestication. *Microorganisms* 8, 903.
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freil, K., Llored, A., et al. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* 556, 339–344. <https://doi.org/10.1038/s41586-018-0030-5>.
- Bennett, A.F., and Lenski, R.E. (2007). An experimental test of evolutionary trade-offs during temperature adaptation. *Proc. Natl. Acad. Sci. USA* 104, 8649–8654. <https://doi.org/10.1073/PNAS.0702117104/asset/fd888273-a806-459d-9bd2-e7b6b7280936/assets/graphic/zpq0160760220002.jpeg>.
- Garland, T. (2014). Trade-offs. *Curr. Biol.* 24, R60–R61. <https://doi.org/10.1016/j.cub.2013.11.036>.
- New, A.M., Cerulus, B., Govers, S.K., Perez-Samper, G., Zhu, B., Boogmans, S., Xavier, J.B., and Verstrepen, K.J. (2014). Different levels of catabolite repression optimize growth in stable and variable environments. *PLoS Biol.* 12, e1001764. <https://doi.org/10.1371/journal.pbio.1001764>.
- Escalante-Chong, R., Savir, Y., Carroll, S.M., Ingraham, J.B., Wang, J., Marx, C.J., and Springer, M. (2015). Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proc. Natl. Acad. Sci. USA* 112, 1636–1641. https://doi.org/10.1073/PNAS.1418058112/suppl_file/pnas.1418058112.sapp.pdf.
- Ricci-Tam, C., Ben-Zion, I., Wang, J., Palme, J., Li, A., Savir, Y., and Springer, M. (2021). Decoupling transcription factor expression and activity enables dimmer switch gene regulation. *Science* 372, 292–295. https://doi.org/10.1126/science.aba7582/suppl_file/aba7582-ricci-tam-sm.pdf.
- Wang, J., Atolia, E., Hua, B., Savir, Y., Escalante-Chong, R., and Springer, M. (2015). Natural variation in preparation for nutrient depletion reveals a cost-benefit tradeoff. *PLoS Biol.* 13, e1002041. <https://doi.org/10.1371/journal.pbio.1002041>.
- Duan, S.-F., Shi, J.-Y., Yin, Q., Zhang, R.P., Han, P.-J., Wang, Q.-M., and Bai, F.Y. (2019). Reverse Evolution of a Classic Gene Network in Yeast Offers a Competitive Advantage. *Curr. Biol.* 29, 1126–1136.e5. <https://doi.org/10.1016/j.cub.2019.02.038>.
- Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., et al. (2018). Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. *Mol. Biol. Evol.* 35, 1712–1727. <https://doi.org/10.1093/molbev/msy066>.
- Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., and Bai, F.-Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nat. Commun.* 9, 2690. <https://doi.org/10.1038/s41467-018-05106-7>.
- Boocock, J., Sadhu, M.J., Durvasula, A., Bloom, J.S., and Kruglyak, L. (2021). Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast. *Science* 371, 415–419. <https://doi.org/10.1126/science.aba0542>.
- Hittinger, C.T., Gonçalves, P., Sampaio, J.P., Dover, J., Johnston, M., and Rokas, A. (2010). Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 464, 54–58. <https://doi.org/10.1038/nature08791>.
- Barbosa, R., Almeida, P., Safar, S.V.B., Santos, R.O., Morais, P.B., Nielly-Thibault, L., Leducq, J.B., Landry, C.R., Gonçalves, P., Rosa, C.A., and Sampaio, J.P. (2016). Evidence of natural hybridization in Brazilian wild lineages of *Saccharomyces cerevisiae*. *Genome Biol. Evol.* 8, 317–329. <https://doi.org/10.1093/gbe/evv263>.
- Han, D.Y., Han, P.J., Rumbold, K., Koricha, A.D., Duan, S.F., Song, L., Shi, J.Y., Li, K., Wang, Q.M., and Bai, F.Y. (2021). Adaptive Gene Content and Allele Distribution Variations in the Wild and Domesticated Populations of *Saccharomyces cerevisiae*. *Front. Microbiol.* 12, 631250. <https://doi.org/10.3389/fmicb.2021.631250>.
- D'Angioli, M., De Chiara, M., Yue, J.X., Irizar, A., Stenberg, S., Persson, K., Llored, A., Barré, B., Schacherer, J., Marangoni, R., et al. (2020). A yeast living ancestor reveals the origin of genomic introgressions. *Nature* 587, 420–425. <https://doi.org/10.1038/s41586-020-2889-1>.
- Peris, D., Ubbelohde, E.J., Kuang, M.C., Kominek, J., Langdon, Q.K., Adams, M., Koshalek, J.A., Hulfachor, A.B., Oplente, D.A., Hall, D.J., et al. (2023). Macroevolutionary diversity of traits and genomes in the model yeast genus *Saccharomyces*. *Nat. Commun.* 14, 690. <https://doi.org/10.1038/s41467-023-36139-2>.
- Kuang, M.C., Kominek, J., Alexander, W.G., Cheng, J.F., Wrobel, R.L., and Hittinger, C.T. (2018). Repeated cis-regulatory tuning of a metabolic bottleneck gene during evolution. *Mol. Biol. Evol.* 35, 1968–1981. <https://doi.org/10.1093/molbev/msy102>.
- Oh, D., and Hopper, J.E. (1990). Transcription of a yeast phosphoglucosyltransferase isozyme gene is galactose inducible and glucose repressible. *Mol. Cell Biol.* 10, 1415–1422. <https://doi.org/10.1128/MCB.10.4.1415>.
- Rubio-Teixeira, M. (2005). A comparative analysis of the GAL genetic switch between not-so-distant cousins: *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. *FEMS Yeast Res.* 5, 1115–1128. <https://doi.org/10.1016/j.femsyr.2005.05.003>.
- Bro, C., Knudsen, S., Regenber, B., Olsson, L., and Nielsen, J. (2005). Improvement of galactose uptake in *Saccharomyces cerevisiae* through overexpression of phosphoglucosyltransferase: Example of transcript analysis as a tool in inverse metabolic engineering. *Appl. Environ. Microbiol.* 71, 6465–6472. <https://doi.org/10.1128/AEM.71.11.6465-6472.2005/asset/76340492-8397-49e2-91eb-7248d0361683/assets/graphic/zam0110559590003.jpeg>.
- Hittinger, C.T., Rokas, A., Carroll, S.B., and Howard, (2004). Parallel Inactivation of Multiple GAL Pathway Genes and Ecological Diversification in Yeasts.
- Klein, J., Sato, A., Nagl, S., and O'hugin, C. (1998). Molecular Trans-species Polymorphism.
- Helmer, D., and Gourichon, L. The Development of the Exploitation of Products from Capra and Ovis (Meat, Milk and Fleece) from the PPNB to the Early Bronze in the Northern Near East (8700 to 2000 BC cal.) Emmanuelle VILA
- Vigne, J.D., Carrère, I., Briois, F., and Guilaïne, J. (2011). The early process of mammal domestication in the Near East: New evidence from the Pre-Neolithic and Pre-Pottery Neolithic in Cyprus. *Curr. Anthropol.* 52, S255–S271. <https://doi.org/10.1086/659306>.
- Sengupta, S., Mukherjee, S., Basak, P., and Majumder, A.L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Front. Plant Sci.* 6, 656. <https://doi.org/10.3389/fpls.2015.00656/bibtext>.
- Gangli, R., and Tenhaken, R. (2016). Raffinose family oligosaccharides act as galactose stores in seeds and are required for rapid germination of Arabidopsis in the dark. *Front. Plant Sci.* 7, 1115. <https://doi.org/10.3389/fpls.2016.01115/full>.
- Marsilio, V., Campestre, C., Lanza, B., and 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, 485–490. [https://doi.org/10.1016/S0308-8146\(00\)00268-5](https://doi.org/10.1016/S0308-8146(00)00268-5).

39. Saura-Calixto, F., Cañellas, J., and García-Raso, A. (1984). Gas chromatographic analysis of sugars and sugar-alcohols in the mesocarp, endocarp, and kernel of almond fruit. *J. Agric. Food Chem.* 32, 1018–1020. https://doi.org/10.1021/jf00125a017/asset/jf00125a017.fp.png_v03.
40. Nierop, K.G., Van Lagen, B., and Buurman, P. (2001). Composition of plant tissues and soil organic matter in the first stages of a vegetation succession. *Geoderma* 100, 1–24. [https://doi.org/10.1016/S0016-7061\(00\)00078-1](https://doi.org/10.1016/S0016-7061(00)00078-1).
41. Wang, Q.M., Liu, W.Q., Liti, G., Wang, S.A., and Bai, F.Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol. Ecol.* 21, 5404–5417. <https://doi.org/10.1111/j.1365-294X.2012.05732.x>.
42. Libkind, D., Hittinger, C.T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., Gonçalves, P., and Sampaio, J.P. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc. Natl. Acad. Sci. USA* 108, 14539–14544. <https://doi.org/10.1073/pnas.1105430108>.
43. Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., Albertin, W., Durrans, P., Sherman, D.J., Marullo, P., et al. (2014). A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat. Commun.* 5, 4044. <https://doi.org/10.1038/ncomms5044>.
44. Kuang, M.C., Hutchins, P.D., Russell, J.D., Coon, J.J., and Hittinger, C.T. (2016). Ongoing Resolution of Duplicate Gene Functions Shapes the Diversification of a Metabolic Network. <https://doi.org/10.7554/eLife.19027.001>.
45. Chevin, L.M., Lande, R., and Mace, G.M. (2010). Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory. *PLoS Biol.* 8, e1000357. <https://doi.org/10.1371/journal.pbio.1000357>.
46. Cordero, O.X., and Polz, M.F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12, 263–273. <https://doi.org/10.1038/nrmicro3218>.
47. Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., and Sauer, U. (2012). Multidimensional optimality of microbial metabolism. *Science* 336, 601–604. https://doi.org/10.1126/science.1216882/suppl_file/schuetz.sm.pdf.
48. Roop, J.I., Chang, K.C., and Brem, R.B. (2016). Polygenic evolution of a sugar specialization trade-off in yeast. *Nature* 530, 336–339. <https://doi.org/10.1038/nature16938>.
49. Caudy, A.A., Guan, Y., Jia, Y., Hansen, C., DeSevo, C., Hayes, A.P., Agee, J., Alvarez-Dominguez, J.R., Arellano, H., Barrett, D., et al. (2013). A new system for comparative functional genomics of *Saccharomyces* yeasts. *Genetics* 195, 275–287. <https://doi.org/10.1534/genetics.113.152918>.
50. Peng, W., Liu, P., Xue, Y., and Acar, M. (2015). Evolution of gene network activity by tuning the strength of negative-feedback regulation. *Nat. Commun.* 6, 1–9. <https://doi.org/10.1038/ncomms7226>.
51. Varela, J.A., Puricelli, M., Ortiz-Merino, R.A., Giacomobono, R., Braun-Galleani, S., Wolfe, K.H., and Morrissey, J.P. (2019). Origin of Lactose Fermentation in *Kluyveromyces lactis* by Interspecies Transfer of a Neo-functionalized Gene Cluster during Domestication. *Curr. Biol.* 29, 4284–4290.e2. <https://doi.org/10.1016/j.cub.2019.10.044>.
52. Ropars, J., Didiot, E., Rodríguez de la Vega, R.C., Benetot, B., Coton, M., Poirier, E., Coton, E., Snirc, A., Le Prieur, S., and Giraud, T. (2020). Domestication of the Emblematic White Cheese-Making Fungus *Penicillium camemberti* and Its Diversification into Two Varieties. *Curr. Biol.* 30, 4441–4453.e4. <https://doi.org/10.1016/j.cub.2020.08.082>.
53. Cheeseman, K., Ropars, J., Renault, P., Dupont, J., Gouzy, J., Branca, A., Abraham, A.L., Ceppi, M., Conseiller, E., Debuchy, R., et al. (2014). Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nat. Commun.* 5, 2876. <https://doi.org/10.1038/ncomms3876>.
54. Siezen, R.J., Starrenburg, M.J.C., Boekhorst, J., Renckens, B., Molenaar, D., and Van Hylckama Vlieg, J.E.T. (2008). Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. *Appl. Environ. Microbiol.* 74, 424–436. https://doi.org/10.1128/aem.01850-07/suppl_file/supplementary_table_s3.doc.
55. Bachmann, H., Starrenburg, M.J.C., Molenaar, D., Kleerebezem, M., and Van Hylckama Vlieg, J.E.T. (2012). Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. *Genome Res.* 22, 115–124. <https://doi.org/10.1101/gr.121285.111>.
56. Deprieto, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. <https://doi.org/10.1038/ng.806>.
57. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
58. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>.
59. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
60. Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. <https://doi.org/10.1093/nar/gkab301>.
61. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Pribelski, A.D., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
62. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
63. Vilella, A.J., Blanco-Garcia, A., Hutter, S., and Rozas, J. (2005). VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* 21, 2791–2793. <https://doi.org/10.1093/bioinformatics/bti403>.
64. Langdon, Q.K., Peris, D., Kyle, B., and Hittinger, C.T. (2018). Spidder: A species identification tool to investigate hybrid genomes with high-throughput sequencing. *Mol. Biol. Evol.* 35, 2835–2849. <https://doi.org/10.1093/molbev/msy166>.
65. Monteiro, P.T., Oliveira, J., Pais, P., Antunes, M., Palma, M., Cavalheiro, M., Galocha, M., Godinho, C.P., Martins, L.C., Bourbon, N., et al. (2020). YEASTRACT+: A portal for cross-species comparative genomics of transcription regulation in yeasts. *Nucleic Acids Res.* 48, D642–D649. <https://doi.org/10.1093/nar/gkz859>.
66. Okonechnikov, K., Golosova, O., Fursov, M., UGENE team, Vaskin, Y., Efremov, I., German Grehov, O.G., Kandrov, D., Rasputin, K., Syabro, M., et al. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>.
67. Sprouffske, K., and Wagner, A. (2016). Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinf.* 17, 172. <https://doi.org/10.1186/s12859-016-1016-7>.
68. Poplin, R., Ruano-Rubio, V., Deprieto, M.A., Fennell, T.J., Carneiro, M.O., Van Der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at bioRxiv 1. <https://doi.org/10.1101/201178>.
69. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* 51, 492–508. <https://doi.org/10.1080/10635150290069913>.
70. Warringer, J., Zörgö, E., Cubillos, F.A., Zia, A., Gjuvsland, A., Simpson, J.T., Forsmark, A., Durbin, R., Omholt, S.W., Louis, E.J., et al. (2011). Trait variation in yeast is defined by population history. *PLoS Genet.* 7, e1002111. <https://doi.org/10.1371/journal.pgen.1002111>.
71. Shumate, A., and Salzberg, S.L. (2021). LiftOff: Accurate mapping of gene annotations. *Bioinformatics* 37, 1639–1643. <https://doi.org/10.1093/bioinformatics/btaa1016>.
72. Emms, D.M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. <https://doi.org/10.1186/s13059-019-1832-y>.
73. Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2003). Multiple Sequence Alignment Using ClustalW and ClustalX. <https://doi.org/10.1002/0471250953.bi0203s00>.
74. Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. <https://doi.org/10.1093/molbev/msm088>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Yeast Nitrogen Base	BD	Cat#239210
D-(+)-Glucose	Sigma-Aldrich	Cat#G8270; CAS no: 50-99-7
D-(+)-Galactose	Sigma-Aldrich	Cat# G0750, CAS no: 59-23-4
Ethanol	Honeywell	Cat#32221; CAS no: 64-17-5
D-(+)-Raffinose pentahydrate	Sigma-Aldrich	Cat#83400; CAS no: 17629-30-0
D-(+)-Glucosamine hydrochloride	Sigma-Aldrich	Cat#G4875; CAS no: 66-84-2
Critical commercial assays		
MiSeq Reagent Kit v2 (500-cycles)	Illumina	MS-102-2003
Nextera XT	Illumina	FC-131-1096
NextSeq 500/550 Mid Output Kit v2 (300 cycles)	Illumina	20024905
Deposited data		
Whole Genome sequencing data	DDBJ/ENA/GenBank	PRJEB63270
GAL/PGM gene sequence	DDBJ/ENA/GenBank	OR393451-OR393572
Experimental models: Organisms/strains		
See Table S1 for list of strains	This paper	N/A
Software and algorithms		
GATK	Depristo et al. ⁵⁶	https://gatk.broadinstitute.org/hc/en-us
BWA (v.0.7.17)	Li et al. ⁵⁷	https://github.com/lh3/bwa
Picard (v.2.22.8)	Github	http://broadinstitute.github.io/picard/
IQ-TREE (v.16.11)	Nguyen et al. ⁵⁸	http://www.iqtree.org/
MAFFT (v7.407)	Katoh et al. ⁵⁹	https://github.com/DomBennett/om..mafft
catfasta2phym	Github	https://github.com/nylander/catfasta2phym
iTOL (v5)	Letunic et al. ⁶⁰	https://itol.embl.de/
SPAdes (v3.13.1)	Bankevich et al. ⁶¹	https://github.com/ablab/spades
trimmomatic (v0.36)	Bolger et al. ⁶²	https://github.com/usadellab/Trimmomatic
variscan (v2.0)	Vilella et al. ⁶³	http://www.ub.edu/softevol/variscan/
sppIDER (v1)	Langdon et al. ⁶⁴	https://github.com/GLBRC/sppIDer
YEASTRACT+	Monteiro et al. ⁶⁵	http://www.yeasttract.com/
UGENE v43.0	Okonechnikov et al. ⁶⁶	https://github.com/ugeneunipro/ugene
Growthcurver	Sprouffske et al. ⁶⁷	https://cran.r-project.org/web/packages/growthcurver/index.html
Other		
P680 HPLC Pump	Dionex	N/A
ASI 100 Automated Sample Injector	Dionex	N/A
Tecan Spark Microplate reader	Tecan Trading	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, José Paulo Sampaio (jss@fct.unl.pt).

Materials availability

No unique reagents were generated in this study.

Data and code availability

- Whole-genome data has been deposited in DDBJ/ENA/GenBank and is publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#). The sequences of the GAL/PGM genes of the strains isolated in Taiwan have been deposited in GenBank and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Saccharomyces cerevisiae strains used in this study are listed in [Table S1](#). The selected strains were grown in YNB supplemented either with 0.2% (w/v) glucose, 2% (w/v) of either glucose or, galactose –or with 1%–3% (w/v) glucose + 1%–3% (w/v) of galactose. All growth experiments were performed at 25°C with no agitation.

METHOD DETAILS

Genome sequencing, read alignment, and genotype calling

DNA was extracted from overnight grown cultures and paired-end Illumina MiSeq (500 cycles) and NextSeq (300 cycles) reads were obtained. Additional genomic data used in this study were retrieved from public databases as indicated in [Table S1](#).

Single nucleotide polymorphisms (SNPs) from whole-genome sequencing data were extracted following an adapted GATK germline short variant discovery pipeline.⁶⁸ Reads from each genome were mapped to a combined reference genome that included *S. cerevisiae* S288c (version R64-1-1) and *S. paradoxus* CBS 432 (ASM207905v1) using BWA v.0.7.17.⁵⁷ Duplicated reads were marked with Picard v.2.22.8 (<http://broadinstitute.github.io/picard/>). SNPs, INDELS (insertions and deletions), and genotype determination were performed on all samples simultaneously using a local re-assembly of haplotypes (using GATK HaplotypeCaller, Genomics DBImport, and Genotype GVCFs). Standard hard filtering parameters of variant quality score recalibration were adjusted according to GATK best practice recommendations (GATK VariantFiltration with parameter values QD < 2.0, QUAL < 30.0, SOR > 3.0, FS > 60.0, MQ < 40.0, MQRankSum < –12.5, ReadPosRankSum < –8.0).⁵⁶

Phylogenetic analyses and survey of specific genes

The whole-genome phylogeny was inferred using the maximum-likelihood method as implemented in IQ-TREE v.16.11,⁵⁸ using the best-fit model of sequence evolution (TVM+F+ASC+G4) and ultrafast bootstrap approximation with 1,000 replicates.⁵⁹ Strains from the populations CHN IX and Taiwan 1 were used to root the tree based on previous studies.¹⁰

For gene phylogenies, nucleotide sequences were aligned with MAFFT v7.407.⁵⁹ The aligned sequences were used to reconstruct a maximum-likelihood phylogeny using IQ-TREE v.16.11,⁵⁸ the best-fit model of sequence evolution inferred by the software (–m TEST), and ultrafast bootstrap with 2,000 replicates. For the concatenated phylogeny of the GAL gene cluster (GAL7-10-1) the same strategy was used but, after the alignment of individual genes, the PERL script catfasta2phym (https://github.com/nylander/catfasta2phym) was used to concatenate the files. To reconstruct the phylogeny with the concatenated file, a partition flag (–spp) and a file with information on the partitions was used in the command line of IQ-TREE. The software iTOL v5⁶⁰ was used for all tree visualizations. Tree topology tests were performed using an approximately unbiased (AU) test⁶⁹ in IQ-TREE v.16.11⁵⁸ against the original phylogeny.

Specific genes were investigated in whole-genome *de novo* assemblies generated with SPAdes v3.13.1.⁶¹ Prior to assembly, reads were processed with trimmomatic v0.36⁶² to remove adapter sequences. In order to retrieve genes of interest, a local BLAST database was created for each genome and ORFs of interest were searched with blastn using query sequences from GAL7, GAL10, GAL1, GAL2, GAL3, GAL4, GAL80, and PGM1 retrieved from the SGD database; the MEL1 sequence was retrieved from strain UWOPS 03–461.4.⁷⁰ Annotations of *S. cerevisiae* strain Lalvin W15 and ZP 779 were transferred from the S288c reference using liftOff v.6.1.⁷¹

Divergence analysis and hybrid identification

Divergence of the GAL gene cluster and flanking regions was estimated using VARISCAN v2.0⁶³ for each selected strain by comparison with the reference genome of *S. cerevisiae* using RunMode 21. The results were processed using a 1,000- bp sliding window with 100- bp step increments. For smaller regions (analyses of individual genes), the same approach was used with a 100- bp sliding window with 20- bp step increments. For detection and analysis of hybrid genomes and introgressions, the program spider v1⁶⁴ was used. Single copy orthologs between *S. cerevisiae* strain Lalvin W15 and ZP 779, *S. cerevisiae* Lalvin W15 – *S. eubayanus* FM1-318 and *S. cerevisiae* Lalvin W15 – *S. kudriavzevii* ZP 591 species pairs were inferred from the proteomes using Orthofinder v2.5.4⁷² and aligned using CLUSTALW v2.1.⁷³ Synonymous substitutions (Ks) of *Saccharomyces* species ortholog-pair or *S. cerevisiae* strain gene-pair were calculated from the alignments using PAML v4.10.6.⁷⁴

Search for binding sites of transcription factors

The searches for the Gal4-binding site motif (CGGN₁₁CCG) in the *PGM1* promoter region and for the Mig1-binding site (SYGGRG) in the promoter regions of the *GAL4* and *GAL1/GAL10* genes were performed using YEASTRACT+.⁶⁵ For the Mig1-binding sites, the results were manually curated using UGENE v43.0⁶⁶ to verify and retrieve sequences.

Growth assays and sugar consumption determination

Cultures were pre-grown in 200 μ L of yeast nitrogen base (YNB) medium (Difco) supplemented with 0.2% (w/v) glucose in a 96-well plate overnight at 25°C. Then, 5 μ L of the cultures were inoculated in 200 μ L of YNB supplemented with 2% (w/v) of either, glucose or, galactose –or with 1% (w/v) glucose +1% (w/v) of galactose. In all cases, at least two independent experiments were performed, growth was followed by measuring the absorbance at 640 nm for 72 h in a microplate reader (Tecan Spark, Tecan Trading, Männedorf, Switzerland) at 25°C. Growth rates and carrying capacity were calculated using the R package Growthcurver.⁶⁷

For measurements of glucose and galactose in mixtures of both sugars, cultures were pre-grown overnight in YNB with 0.2% (w/v) glucose at 25°C. Cultures were then inoculated in 50 mL of YNB with 3% (w/v) glucose +3% (w/v) galactose, at an initial optical density 640nm of 0.015 and incubated at 25°C for 74 h. Growth was monitored and samples were harvested at various time points for glucose, galactose, and ethanol quantification. Samples were analyzed by high-performance liquid chromatography (HPLC) using a carbohydrate analysis column (300 mm by 7.8 mm, Aminex HPX-87P; Bio-Rad) and a differential refractometer (Shodex R-101). The column was kept at 80°C, and water was used as the mobile phase at 0.6 mL min⁻¹. To account for strain variability, two independent experiments and measurements were performed for at least two representative strains of each *GAL* allele.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed in RStudio with custom scripts under available packages. The Kruskal-Wallis and t-test were performed in R using `kruskal.test` and `t.test` function (respectively), after testing for the normality of the data with the Shapiro-Wilk normality test (`shapiro.test`). For all statistical tests, '***' corresponds to a p value of 0.001, '**' p value 0.01, and '*' p value 0.05.