ARTICLES

Gene duplication and the adaptive evolution of a classic genetic switch

Chris Todd Hittinger[1]† & Sean B. Carroll[1]

How gene duplication and divergence contribute to genetic novelty and adaptation has been of intense interest, but experimental evidence has been limited. The genetic switch controlling the yeast galactose use pathway includes two paralogous genes in *Saccharomyces cerevisiae* that encode a co-inducer (*GAL3*) and a galactokinase (*GAL1*). These paralogues arose from a single bifunctional ancestral gene as is still present in *Kluyveromyces lactis*. To determine which evolutionary processes shaped the evolution of the two paralogues, here we assess the effects of precise replacement of coding and non-coding sequences on organismal fitness. We suggest that duplication of the ancestral bifunctional gene allowed for the resolution of an adaptive conflict between the transcriptional regulation of the two gene functions. After duplication, previously disfavoured binding site configurations evolved that divided the regulation of the ancestral gene into two specialized genes, one of which ultimately became one of the most tightly regulated genes in the genome.

Gene duplication has long been recognized as a major source of new genes and functions. Until recently, it was generally assumed that duplicate genes were free to evolve new functions ('neofunctionalization') because the original function was maintained by the other copy[1–5]. However, several recent case studies and comparisons of genome content have suggested that most new genes do not have novel functions[6–10]. Instead, paralogous gene pairs are often 'subfunctionalized' with two or more functions being partially or completely subdivided between the two genes after gene duplication. Various evolutionary models have been proposed to account for the fates of paralogues during subfunctionalization.

The duplication–degeneration–complementation (DDC) model explains the preservation of duplicate genes by a neutral mechanism in which each paralogue accumulates loss-of-function mutations (degeneration) that are complemented by the other copy[11]. Examples of DDC include the complementary losses of spatial domains of expression following gene duplication in multicellular organisms[6,11,12]. DDC may explain the initial retention of most duplicate genes, but DDC alone is not adaptive. Gene duplication is only a necessary contributor to adaptation if it allows for otherwise prohibited adaptive mutations.

One set of circumstances under which some adaptive mutations can be prohibited is in the case of multifunctional genes. The performance of multiple functions by a single gene or 'gene sharing' is prevalent in nature[10] but presents the possibility that mutations that optimize one function may compromise the other. Such an 'adaptive conflict' may be resolved by gene duplication followed by subfunctionalization and positive selection, thereby creating two optimized specialist genes in a manner forbidden during gene sharing[13–15].

Adaptive conflict resolution has been suggested to have been important to the specialization of several duplicated genes[13–15], but demonstration of specific molecular changes involved in adaptive conflict has proven elusive. Here we dissect the evolution of a pair of duplicate genes from the galactose (GAL)-use pathway from their bifunctional ancestor and show that they are nearly completely subfunctionalized in *Saccharomyces cerevisiae*. With one function being performed by its paralogue, *GAL3*, *GAL1* was free to evolve previously disfavoured adaptive changes in its upstream regulatory element that provided a more dynamic and tightly regulated transcriptional response. We propose that an adaptive conflict in the regulation of these two functions was resolved along the lineage leading to *S. cerevisiae* by gene duplication and the subfunctionalization of *GAL1* and *GAL3*.

Gene duplication and the GAL genetic switch. The co-inducer (Gal3) of the classic genetic switch[16–18] of the GAL pathway of *S. cerevisiae* is encoded by a member of a duplicate gene pair created by a whole-genome duplication about 100 million years ago[19–23]. Its paralogue encodes the first enzyme of the GAL pathway, the galactokinase Gal1, the upstream regulatory elements of which are the DNA component of this switch. *GAL1* is one of the most tightly regulated genes in the genome, being repressed in the absence of galactose and induced about 1,000-fold in its presence[24,25]. In contrast, *GAL3* expression is induced a modest three- to fivefold from its basal expression[26]. Gal3 co-induces the pathway by sequestering the co-repressor Gal80 in the cytoplasm in a galactose- and ATP-dependent fashion[27–30]. The transcription factor Gal4 is then free to activate the GAL pathway, including *GAL1*. The switch works similarly in *Kluyveromyces lactis*, a yeast species that did not undergo the whole-genome duplication, except that a single bifunctional *K. lactis* protein encoded by a *GAL1* homologue performs both the co-induction and galactokinase functions[17,27,31,32].

Protein subfunctionalization. Gal1, Gal3 and *K. lactis* Gal1 (which we define hereafter as KlacGal1) are similar in sequence[26,33] and structure[34], and each protein is sufficient to act as the sole co-inducer in plasmid complementation and *in vitro* assays[27,29,35]. Addition of a Ser-Ala (SA) dipeptide to the degenerated active site of Gal3 has been shown to be necessary and sufficient to confer galactokinase activity[35]. These two residues are leading candidates to account for the functional divergence of Gal1 and Gal3[15,17,32,35].

Complementation and *in vitro* assays may demonstrate the sufficiency of a protein to perform a function, but natural selection can operate on very small fitness differences. Optimization of a single subfunction is expected to be quantitative, so more sensitive tests are required to detect functional differences among proteins and

[1]Howard Hughes Medical Institute, Laboratory of Genetics, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706, USA. †Present address: Center for Genome Sciences, School of Medicine, Washington University in St Louis, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA.

possible instances of adaptive conflict. Therefore, we have employed a competitive growth assay capable of detecting fitness differences among strains in the order of 10^{-3} offspring per generation, well beyond the detection limit of previous assays. In this assay, genetically manipulated green fluorescent protein (GFP)-tagged strains are competed against an otherwise identical blue fluorescent protein (BFP)-tagged strain in liquid culture, cells are counted by flow cytometry, and the effect on fitness is calculated. Targeted gene replacement allowed us to make any desired genetic change and provided the control over genetic background necessary to detect small fitness differences. To address the contribution of coding changes to the divergence of these paralogues, we used this assay to measure the fitness effects of targeted replacements of the coding regions of *GAL3* and *GAL1*.

We first examined the ability of various coding sequences to perform the galactokinase function of Gal1 *in vivo*. There was no effect of any coding replacement in non-inducing conditions, but we observed strong fitness defects in galactose when other proteins were required to perform the function of Gal1 (Table 1). This was expected for Gal3 (known to have no galactokinase activity[26,35]) and for proteins designed to mimic the SA dipeptide deletion in the would-be active site of Gal3 (Gal1-ΔSA and KlacGal1-ΔSA). In contrast to complementation assays[27,35], our quantitative assay detected strong fitness defects for even the galactokinases Gal3-∇SA (a version of Gal3 with the SA dipeptide inserted) and KlacGal1. Gal3-∇SA has a much slower enzymatic rate than Gal1[35], and this seems to be a substantial handicap under competitive conditions. The KlacGal1 defect was largely due to toxicity associated with expressing this heterologous protein at high levels in *S. cerevisiae* and not poor enzyme activity itself (Supplementary Table 1).

We next tested the ability of these coding sequences to perform the co-induction function of Gal3. Gal3-∇SA performed poorly as a co-inducer (Fig. 1), a defect not detectable in previous complementation assays[35]. Most significantly, this result raises the possibility that the SA residues could be a source of adaptive conflict between the galactokinase and co-induction functions. For example, tight binding of galactose and ATP by Gal3 in a complex with Gal80 might be inconsistent with efficient galactose phosphorylation. If this were the case, then deleting these two residues from Gal1 and KlacGal1 should improve their co-induction functions and make them function more like Gal3. Instead, we found that deleting these residues further impairs their co-induction function relative to wild-type Gal1 and KlacGal1 (Fig. 1).

Because the removal of the SA residues from Gal1 and KlacGal1 did not improve their co-induction function, the SA residues were not a source of adaptive conflict; nor are the galactokinase and co-induction protein functions inherently antagonistic. Instead, the presence of the SA residues has opposite effects on the co-induction capabilities of Gal1 and Gal3. The differential effect of the SA residues must be due to coding differences at other sites within the proteins. These opposite effects mean that the SA residues are a case of intragenic 'sign epistasis',[36] which is said to be present when the direction of the fitness effect of a mutation depends on residue(s) present at other site(s) within the protein. The SA residues of Gal1/3 are the first case of sign epistasis reported between paralogues, but sign epistasis may be a general feature of duplicated proteins that have diverged in function and accumulated interacting changes.

Despite the divergence between Gal3 and Gal1 (74% identity) or Gal3 and KlacGal1 (58% identity) in addition to the SA dipeptide, Gal1 and KlacGal1 were fairly effective at performing the co-induction function of Gal3 at higher galactose concentrations (≈2–3% fitness defect; Fig. 1). Even under more challenging conditions of low galactose concentrations, strains expressing Gal1 and KlacGal1 from the endogenous *GAL3* upstream regulatory elements (hereafter, the *GAL3* promoter or P_{GAL3}) still performed relatively well. The performance of these strains contrasts sharply with the *gal3-Δ* strain where Gal1 is driven only by its native *GAL1* promoter (Fig. 1). Taken together, these experiments suggest that, aside from the complete degeneration of the galactokinase activity of Gal3, most of the functional divergence between these paralogues has been regulatory. This inference led us to focus on the upstream regulatory elements as a potential source of adaptive conflict.

Most adaptive divergence is regulatory. To assess the divergence in the upstream regulatory elements, we swapped the promoters of the two paralogues and assessed their capabilities to drive proper expression of *GAL1* and *GAL3*. Competition assays revealed that P_{GAL1} and P_{GAL3} were both inferior at performing the function of the other promoter (Fig. 2; a competition assay in which the tested promoter drives the expression of the paralogous coding sequence in a null mutant background). The inability of P_{GAL3} to drive sufficient levels of galactokinase is due to its comparatively low messenger RNA expression level when fully induced (Table 2)[26]. In contrast, whereas P_{GAL1} is strongly induced in the presence of galactose, it is more

Table 1 | Effects of the targeted replacement of the *GAL1* coding region

Coding region replacing GAL1	Growth	Fitness defect	s.d.
GAL1$^+$	+	−0.000	±0.002
GAL3$^+$	−	−0.726	±0.007
KlacGAL1$^+$	+ (s.)	−0.482	±0.017
gal1-ΔSA	−	−0.639	±0.007
gal3-∇SA	+ (v.s.)	−0.623	±0.005
klacgal1-ΔSA	−	−0.697	±0.010
gal1-atg1ctg	−	−0.624	±0.007
gal1-Δ	−	−0.724	±0.007

s., slow; v.s., very slow; s.d., standard deviation of three independent lines. Growth was scored manually in liquid culture using a single strain. Cultures contained 2% galactose as the sole carbon source. All data are normalized to the *GAL1*$^+$ control. *gal1-atg1ctg* (a mutation of the start codon) and *gal1-Δ* are both null mutations, but *gal1-Δ* causes an additional induction-dependent fitness defect, the basis of which is unknown. Note the poor performance of *KlacGAL1*$^+$ and *gal3-∇SA* relative to *GAL1*$^+$.

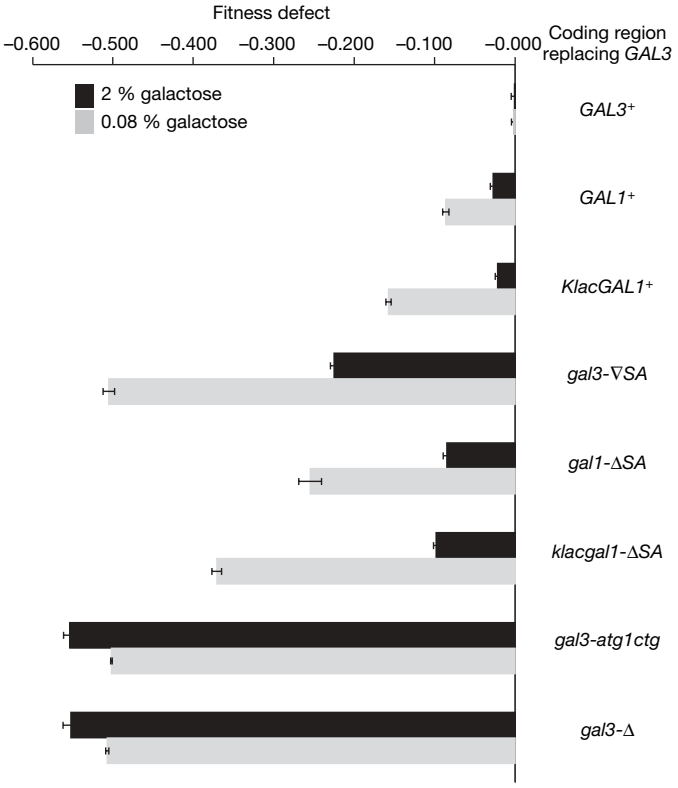

Figure 1 | *GAL3*$^+$ encodes the best co-inducer. The fitness effects of the targeted replacement of the coding region of *GAL3* under two different culture conditions are plotted as bar graphs. Error bars are standard deviations of three independent lines. All data are normalized to the *GAL3*$^+$ control. Note that *GAL3*$^+$ outperforms *GAL1*$^+$ and *KlacGAL1*$^+$, but the defect is moderate. Further note that the effect of the presence of the SA dipeptide depends on the identity of the coding region being tested.

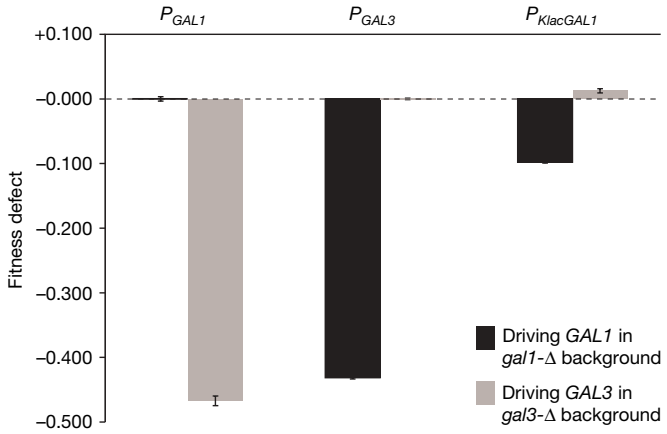

Figure 2 | Most adaptive divergence between *GAL1* and *GAL3* occurred in the promoters. Fitness effects of promoter replacements are shown as bar graphs. Black, competitions between strains with the stated promoter driving the expression of the *GAL1* coding region in a *gal1*-Δ background and a BFP-labelled P_{GAL1} strain in 2% galactose; data normalized to the P_{GAL1} strain. Grey, competitions between strains with the stated promoter driving the expression of the *GAL3* coding region in a *gal3*-Δ background and a BFP-labelled strain in 0.08% galactose; data normalized to the P_{GAL3} strain. Error bars are standard deviations of three independent lines. Note that P_{GAL1} and P_{GAL3} are both inferior at performing the function of the other promoter, and $P_{KlacGAL1}$ is not optimized for driving *GAL1* expression.

efficiently repressed in non-inducing conditions than P_{GAL3} (Table 2)[24,25]. When the co-inducer is driven by this tightly regulated promoter, basal expression is apparently too low to allow for efficient induction of the pathway when cells are presented with galactose. Quantitative comparisons of the fitness defects of these strains to the defects of strains harbouring null alleles suggests that the regulated expression of these two paralogues was almost completely subdivided between a loosely regulated promoter providing the basal expression necessary for pathway induction and a tightly regulated promoter driving high levels of the enzyme on induction.

Analysis of $P_{KlacGAL1}$ in *S. cerevisiae* provides additional support for the view that P_{GAL1} and P_{GAL3} were subfunctionalized from an ancestral bifunctional state. $P_{KlacGAL1}$ drives strong basal expression and performs well in the role of driving expression of the co-inducer Gal3, as measured by both fitness (Fig. 2) and mRNA expression (Table 2)[32,33]. Although it induces much more strongly than P_{GAL3}, this expression level still falls well short of the more optimal P_{GAL1} promoter (Table 2). This approximately 50% reduction in the level of *GAL1* induction driven by $P_{KlacGAL1}$ reduces organismal fitness by about 10% (Fig. 2), suggesting that P_{GAL1} has acquired adaptive mutations that allow it to induce more fully in the presence of galactose. We therefore wondered whether the demands on $P_{KlacGAL1}$ to serve as a bifunctional promoter driving the expression of a protein performing two distinct functions, one regulatory and one enzymatic, might disallow concurrent optimization of the expression of each function.

Adaptive conflict in phasing of binding sites. One striking difference between P_{GAL1} and $P_{KlacGAL1}$ is the spacing or helical phasing of binding sites for the transcriptional activator Gal4 (refs 16, 17, 24, 25, 33). *S. cerevisiae* and all closely related *Saccharomyces sensu stricto* species possessing P_{GAL1} have a core of three almost immediately adjacent Gal4 binding sites[37], which places the three bound Gal4 dimers on approximately the same side of the double helix. In contrast, the three core Gal4 binding sites of $P_{KlacGAL1}$ are spaced approximately half a helical turn apart such that the central Gal4 dimer is bound on the opposite side of the double helix (Fig. 3a)[33]. Previous analysis has demonstrated that Gal4 dimers that are nearby and on the same helical phase are cooperatively bound and more efficiently inhibited by the co-repressor Gal80 than either single binding sites or sites bound by Gal4 dimers on opposite helical phases[38].

To test whether the helical phasing of Gal4 binding sites might be responsible for some of the difference in expression between P_{GAL1} and $P_{KlacGAL1}$, we altered the spacing of the Gal4 binding sites in $P_{KlacGAL1}$ such that they were in phase, like in P_{GAL1}. Indeed, this alteration decreased non-induced $P_{KlacGAL1}$ expression by about 27% ($P < 10^{-4}$; presumably because of more efficient binding by Gal80 (ref. 38)) but also increased the induced expression by about 13% ($P < 10^{-2}$; perhaps owing to cooperative binding of co-activators by Gal4) (Fig. 3b; Supplementary Fig. 1b). Furthermore, these Gal4 binding site changes in $P_{KlacGAL1}$ increased fitness when driving expression of the enzyme (by about 1% at higher galactose concentrations; $P < 10^{-4}$), but they decreased fitness when driving

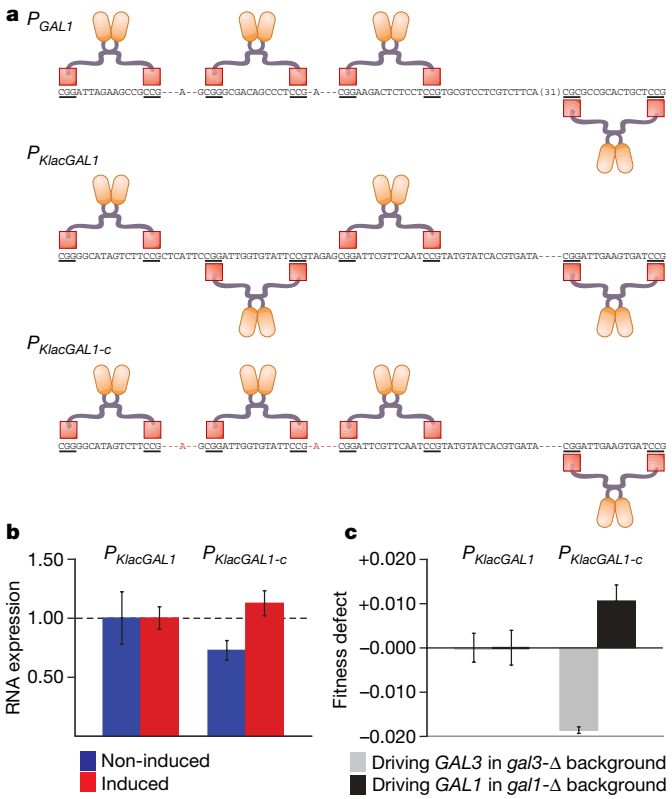

Figure 3 | Adaptive conflict in the configuration of Gal4 binding sites. **a**, Schematic of P_{GAL1}, $P_{KlacGAL1}$ and the mutant (nucleotide changes in red) $P_{KlacGAL1\text{-}c}$. **b**, *GAL1* mRNA expression normalized to $P_{KlacGAL1}$. Error bars are standard deviations of 12 replicates. **c**, Fitness effects of promoter replacement normalized to $P_{KlacGAL1}$ strains. Grey, competitions between a *gal3*-Δ strain with the stated promoter driving *GAL3* and a BFP-labelled strain in 0.08% galactose. Black, competitions between a *gal1*-Δ strain with stated promoter driving *GAL1* and a BFP-labelled P_{GAL1} strain in 2% galactose. Error bars are standard deviations of 3 independent lines and 12 cultures (4 each of three independent lines), respectively. Note that altering the phasing of $P_{KlacGAL1}$ to mimic P_{GAL1} improves fitness when driving *GAL1* as a result of an increase in the induced expression, but impairs fitness when driving *GAL3* owing to a decrease in the non-induced expression.

Table 2 | Normalized expression of promoters driving *GAL1*

Promoter	Non-induced	s.d.	Induced	s.d.
P_{GAL1}	1.00	±0.25	1.00	±0.13
P_{GAL3}	2.12	±0.38	0.13	±0.02
$P_{KlacGAL1}$	9.95	±1.64	0.54	±0.09

s.d., standard deviation of nine replicates. *GAL1* mRNA levels (measured by quantitative RT–PCR) driven from each promoter were normalized to *ACT1* mRNA levels and to the experimental value from P_{GAL1} for each condition. P_{GAL1} induces to ~1,000-fold above non-induced conditions[24,25].

expression of the co-inducer (by about 2% at lower galactose concentrations; $P < 10^{-2}$) (Fig. 3c; Supplementary Fig. 1c). Thus, alteration of binding site configuration caused opposing effects on the promoter's ability to perform its two distinct functions. These results suggest that an adaptive conflict was resolved by subdividing the transcriptional regulation of each function between two specialist genes during the evolution of the lineage that gave rise to *S. cerevisiae*.

The *K. lactis GAL1* promoter is not optimized. These results imply that some mutations that might have improved the performance of the promoter of the ancestral bifunctional *GAL1* gene were disfavoured before the gene duplication. We tested this idea further by examining the fitness impact of various promoters by placing them upstream of the bifunctional *K. lactis GAL1* gene in *K. lactis*.

We found that P_{GAL1} of *S. cerevisiae* actually outperformed $P_{KlacGAL1}$ at high galactose concentrations in *K. lactis* (Supplementary Tables 2–4). Similarly, $P_{KlacGAL1\text{-}\sigma}$ in which the Gal4 binding sites are placed on the same side of the helix, also outperformed $P_{KlacGAL1}$ at high galactose and lactose concentrations in *K. lactis*. Lactose is a common substrate for *K. lactis*, so it seems likely that *K. lactis* could benefit from *cis*-phasing of Gal4 binding sites in $P_{KlacGAL1}$ under some conditions. Yet, this configuration failed to evolve in *K. lactis*. The data from the fitness experiments in *S. cerevisiae* indicate that constraints imposed by the co-induction function of *KlacGAL1* may explain the suboptimal configuration of $P_{KlacGAL1}$.

The path to specialization. We have dissected the evolution and divergence of the paralogues *GAL1* and *GAL3* from a bifunctional ancestral gene by using precise genetic manipulation and a sensitive direct competition assay. In *S. cerevisiae*, these genes are nearly completely subfunctionalized, and both genes are required for efficient growth when galactose is the sole carbon source. Subfunctionalization of the coding regions was asymmetric: Gal3 completely lost galactokinase activity, whereas Gal1 retained substantial co-induction capability. Subfunctionalization of the promoters was nearly complete with P_{GAL1} and P_{GAL3} each unable to effectively perform the function of the other regulatory element. Comparison with $P_{KlacGAL1}$ suggests that P_{GAL1} has evolved tighter regulation and become more inducible. We showed that about 10% of this change (with respect to both fitness and gene expression level) was due to the

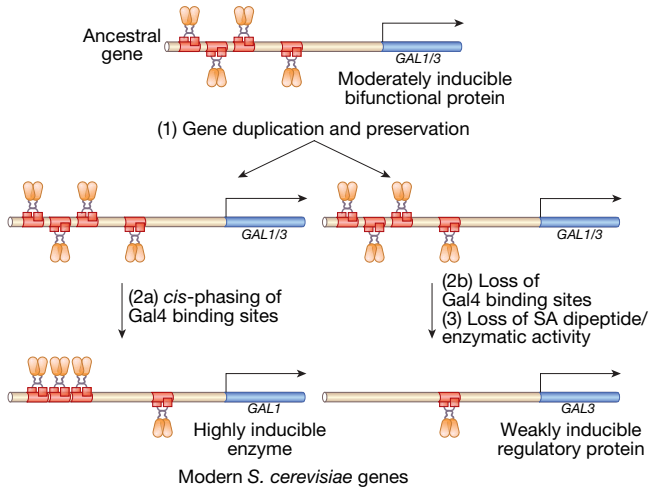

Figure 4 | Model of the evolution of the GAL genetic switch. Step wise subfunctionalization of the ancestral bifunctional *GAL1/3* gene. (1) Whole-genome duplication created two identical copies[19–23], which were probably preserved by DDC[11]. (2) The loss of Gal4 binding sites in P_{GAL3} and the evolution of *cis*-phasing of Gal4 binding sites in P_{Gal1} occurred in the lineage leading to the common ancestor of all *Saccharomyces sensu stricto*. *cis*-phasing of Gal4 binding sites helped optimize *GAL1* expression in a manner inconsistent with efficient co-induction in *S. cerevisiae*, resolving the putative adaptive conflict. (3) the loss of the SA dipeptide and the final loss of Gal3's enzymatic activity occurred within the *sensu stricto* clade[22].

helical phasing of Gal4 binding sites, although other important changes must also have occurred. The altered helical phasing of P_{GAL1} also decreased non-induced gene expression and conferred a fitness penalty on the promoter's ability to perform the co-induction function in *S. cerevisiae*. These findings suggest a scenario for the evolution of this genetic switch and the ultimate resolution of the underlying adaptive conflict (Fig. 4).

After the whole-genome duplication, *S. cerevisiae GAL1* and *GAL3* were integrated into a more complex and, in some ways, more optimal genetic pathway. Consistent with the DDC model, degeneration of complementary functions has been the major consequence. The most striking degeneration is the complete loss of galactokinase activity by Gal3, in which the final inactivating event (the loss of the SA dipeptide) occurred recently within the *sensu stricto* clade[22]. Our data demonstrate that the loss of enzymatic activity was not inherently adaptive but may have simply been the final step in a gradual degeneration that took nearly 100 million years to complete.

GAL3 may or may not have acquired any adaptive mutations of its own, but its retention as the co-inducer of the pathway allowed *GAL1* to evolve a more specialized regulatory profile (Fig. 4). We have demonstrated that the configuration of Gal4 binding sites upstream of *GAL1* is an important adaptive component of this regulatory logic that allows for stronger induction. However, this configuration is ill-suited to driving expression of the co-inducer in *S. cerevisiae* such that it is only adaptive when another co-inducer is present in the form of *GAL3*. We propose that these previously disfavoured mutations were adaptive for *GAL1* after the duplication of the ancestral *GAL1/3* and that subdividing the enzymatic and regulatory functions allowed this specific adaptive conflict to be resolved in the lineage that gave rise to *S. cerevisiae* (Fig. 4).

Optimization of regulation after duplication. Our demonstration of the optimization of *GAL1* expression after duplication raises the general questions of: how common are adaptive conflicts within regulatory elements, and what mechanisms may resolve them? Examination of whole-genome expression profiles has suggested that duplicated genes do generally evolve divergent expression patterns[39–43]. Some of the best-documented cases of neofunctionalization of coding sequences have also involved changes in gene expression[3,5,44]. The widespread divergence of gene expression following duplication events raises the possibility that some of the ancestral *cis*-regulatory elements may also have harboured adaptive conflicts.

In addition to the conflicts presented by the control of a bifunctional protein, we can envision two additional circumstances where an adaptive conflict may exist in regulatory elements. First, there are well-documented cases of 'enhancer-sharing', such as in the *Hox* gene complex, where a single enhancer governs the expression of two adjacent genes[45]. This phenomenon also occurs on a genome-wide level between adjacent and nearby genes in yeast[46]. Second, an enhancer may be bifunctional and drive the expression of one gene in two spatial domains within a multicellular organism. Like $P_{GAL1/3}$, such regulatory elements have the potential for adaptive conflict when the optimization of the expression of two different genes or of one gene in two different places may be in conflict. In such cases, duplication of regulatory sequences (or the entire gene) might allow for the evolution of novel gene expression profiles through previously forbidden mutations.

METHODS SUMMARY

Fluorescently labelled yeast strains were engineered to be genetically identical, except for tested mutations, and were competed against a standard strain. GFP/non-GFP ratios of live cells were determined by flow cytometry, and malthusian selection coefficients were calculated[47]. mRNA levels were monitored by quantitative PCR with reverse transcription (RT–PCR). See Methods for details.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 June; accepted 8 August 2007.

1. Ohno, S. *Evolution by Gene Duplication* (Springer, New York, 1970).
2. Taylor, J. S. & Raes, J. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* **38,** 615–643 (2004).
3. Irwin, D. M., Prager, E. M. & Wilson, A. C. Evolutionary genetics of ruminant lysozymes. *Anim. Genet.* **23,** 193–202 (1992).
4. Yokoyama, S. Molecular evolution of color vision in vertebrates. *Gene* **300,** 69–78 (2002).
5. Zhang, J., Zhang, Y. P. & Rosenberg, H. F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genet.* **30,** 411–415 (2002).
6. Prince, V. E. & Pickett, F. B. Splitting pairs: the diverging fates of duplicated genes. *Nature Rev. Genet.* **3,** 827–837 (2002).
7. van Hoof, A. Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics* **171,** 1455–1461 (2005).
8. Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440,** 341–345 (2006).
9. Schilke, B. *et al.* Evolution of mitochondrial chaperones utilized in Fe–S cluster biogenesis. *Curr. Biol.* **16,** 1660–1665 (2006).
10. Piatigorsky, J. *Gene Sharing and Evolution: The Diversity of Protein Functions* (Harvard Univ. Press, Cambridge, Massachusetts, 2007).
11. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151,** 1531–1545 (1999).
12. Tumpel, S., Cambronero, F., Wiedemann, L. M. & Krumlauf, R. Evolution of *cis* elements in the differential expression of two *Hoxa2* coparalogous genes in pufferfish (*Takifugu rubripes*). *Proc. Natl Acad. Sci. USA* **103,** 5419–5424 (2006).
13. Piatigorsky, J. & Wistow, G. The recruitment of crystallins: new functions precede gene duplication. *Science* **252,** 1078–1079 (1991).
14. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* **256,** 119–124 (1994).
15. Hughes, A. L. *Adaptive Evolution of Genes and Genomes* (Oxford Univ. Press, Oxford, 1999).
16. Johnston, M. A model fungal gene regulatory mechanism: the *GAL* genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **51,** 458–476 (1987).
17. Bhat, P. J. & Murthy, T. V. Transcriptional control of the GAL/MEL regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. *Mol. Microbiol.* **40,** 1059–1066 (2001).
18. Ptashne, M. & Gann, A. *Genes and Signals* (Cold Spring Harbor Laboratory Press, Woodbury, New York, 2001).
19. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387,** 708–713 (1997).
20. Dietrich, F. S. *et al.* The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304,** 304–307 (2004).
21. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428,** 617–624 (2004).
22. Hittinger, C. T., Rokas, A. & Carroll, S. B. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc. Natl Acad. Sci. USA* **101,** 14144–14149 (2004).
23. Scannell, D. R. *et al.* Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl Acad. Sci. USA* **104,** 8397–8402 (2007).
24. Johnston, M. & Davis, R. W. Sequences that regulate the divergent *GAL1–GAL10* promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4,** 1440–1448 (1984).
25. West, R. W. Jr, Yocum, R. R. & Ptashne, M. *Saccharomyces cerevisiae GAL1–GAL10* divergent promoter region: location and function of the upstream activating sequence *UASG*. *Mol. Cell. Biol.* **4,** 2467–2478 (1984).
26. Bajwa, W., Torchia, T. E. & Hopper, J. E. Yeast regulatory gene *GAL3*: carbon regulation; *UASGal* elements in common with *GAL1*, *GAL2*, *GAL7*, *GAL10*, *GAL80*, and *MEL1*; encoded protein strikingly similar to yeast and *Escherichia coli* galactokinases. *Mol. Cell. Biol.* **8,** 3439–3447 (1988).
27. Zenke, F. T. *et al.* Activation of Gal4p by galactose-dependent interaction of galactokinase and Gal80p. *Science* **272,** 1662–1665 (1996).
28. Yano, K. & Fukasawa, T. Galactose-dependent reversible interaction of Gal3p with Gal80p in the induction pathway of Gal4p-activated genes of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **94,** 1721–1726 (1997).
29. Platt, A. & Reece, R. J. The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex. *EMBO J.* **17,** 4086–4091 (1998).
30. Peng, G. & Hopper, J. E. Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein. *Proc. Natl Acad. Sci. USA* **99,** 8548–8553 (2002).
31. Meyer, J., Walker-Jonah, A. & Hollenberg, C. P. Galactokinase encoded by *GAL1* is a bifunctional protein required for induction of the *GAL* genes in *Kluyveromyces lactis* and is able to suppress the *gal3* phenotype in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11,** 5454–5461 (1991).
32. Rubio-Texeira, M. A comparative analysis of the *GAL* genetic switch between not-so-distant cousins: *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. *FEMS Yeast Res.* **5,** 1115–1128 (2005).
33. Webster, T. D. & Dickson, R. C. The organization and transcription of the galactose gene cluster of *Kluyveromyces lactis*. *Nucleic Acids Res.* **16,** 8011–8028 (1988).
34. Thoden, J. B., Sellick, C. A., Timson, D. J., Reece, R. J. & Holden, H. M. Molecular structure of *Saccharomyces cerevisiae* Gal1p, a bifunctional galactokinase and transcriptional inducer. *J. Biol. Chem.* **280,** 36905–36911 (2005).
35. Platt, A., Ross, H. C., Hankin, S. & Reece, R. J. The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase. *Proc. Natl Acad. Sci. USA* **97,** 3154–3159 (2000).
36. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evol. Int. J. Org. Evol.* **59,** 1165–1174 (2005).
37. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254 (2003).
38. Melcher, K. & Xu, H. E. Gal80–Gal80 interaction on adjacent Gal4p binding sites is required for complete *GAL* gene repression. *EMBO J.* **20,** 841–851 (2001).
39. Wagner, A. Asymmetric functional divergence of duplicate genes in yeast. *Mol. Biol. Evol.* **19,** 1760–1768 (2002).
40. Makova, K. D. & Li, W. H. Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* **13,** 1638–1645 (2003).
41. Gu, Z., Rifkin, S. A., White, K. P. & Li, W. H. Duplicate genes increase gene expression diversity within and between species. *Nature Genet.* **36,** 577–579 (2004).
42. Huminiecki, L. & Wolfe, K. H. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. *Genome Res.* **14,** 1870–1879 (2004).
43. Li, W. H., Yang, J. & Gu, X. Expression divergence between duplicate genes. *Trends Genet.* **21,** 602–607 (2005).
44. Goodman, M., Moore, G. W. & Matsuda, G. Darwinian evolution in the genealogy of haemoglobin. *Nature* **253,** 603–608 (1975).
45. Gould, A., Morrison, A., Sproat, G., White, R. A. & Krumlauf, R. Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping *Hox* expression patterns. *Genes Dev.* **11,** 900–913 (1997).
46. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26,** 183–186 (2000).
47. Hartl, D. L. & Clark, A. G. in *Principles of Population Genetics* 216 (Sinauer Associates, Sunderland, Massachusetts, 1997).
48. Brachmann, C. B. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14,** 115–132 (1998).
49. Kooistra, R., Hooykaas, P. J. & Steensma, H. Y. Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* **21,** 781–792 (2004).
50. Guldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* **24,** 2519–2524 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. L. Williams for sharing unpublished data, reagents and the competition assay; J. E. Selegue for technical support; M. Johnston and Washington University in St Louis for access to laboratory space and equipment; Collection de Levures d'Intérêt Biotechnologique for *K. lactis* strain CLIB210; H. Y. Steensma for the NHEJ-deficient *K. lactis* strain and transformation protocol; J. H. Hegemann for the *kanMX* cassette; B. L. Williams, P. M. van Wynsberghe and Carroll laboratory members for technical advice; L. M. Olds for artwork; and B. L. Williams, B. Prud'homme and M. Johnston for critical reading of the manuscript. The Howard Hughes Medical Institute supported this work through an investigatorship (S.B.C.) and predoctoral fellowship (C.T.H.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.B.C. (sbcarrol@wisc.edu) and C.T.H. (cthittinger@wustl.edu).

METHODS

Strain construction. *MAT**a** ura3*-Δ *lys2*-Δ BY4724 strains of *S. cerevisiae*[48] were genetically engineered to contain only the desired mutations by standard aptamer-*URA3*-based selection and counter-selection. GFP was expressed from a P_{TDH3}-*yEGFP*-T_{CYC1} or P_{TDH3}-*yEBFP*-T_{CYC1} construct integrated into the genome between base pairs 199,270 and 199,271 of chromosome I. GFP- and BFP-labelled strains differed by a single base pair that encoded a single substitution in the GFP/BFP chromophore (Y66F) and were equally fit in competition assays. Coding regions were replaced from start codon to stop codon in the GFP-labelled strain, including control constructs that were engineered back to wild type. Promoter manipulations were carried out by inserting the desired promoter (defined as the entire upstream intergenic region) upstream of the desired coding region, with the entire 3′ intergenic region and no intervening sequence, between base pairs 199,230 and 199,231 of chromosome I.

*MAT**a** ku80*-Δ CBS2359 NHEJ-deficient strains of *K. lactis*[49] were genetically engineered to contain only the desired mutations. First, a *klacgal1*Δ::*kanMX* strain was constructed by transforming *K. lactis*[49] with a PCR product derived from a pUG6 template and with 90-mers that contained the *kanMX* amplification sites[50] fused to portions of the *K. lactis* genome necessary to replace the coding region of *KlacGAL1* with the *kanMX* cassette in the forward orientation from stop codon to start codon. A single kan^+ gal^- strain was transformed in parallel with PCR products constructed with 90-mers designed to target each promoter fused to the complete coding and 3′ intergenic region of *KlacGAL1* to the *URA3* locus. These products replaced the coding region of *URA3* in the reverse orientation from stop codon to start codon (*MAT**a** ku80*-Δ *klacgal1*Δ::*kanMX ura3*Δ::P_X-*KlacGAL1* CBS2359). Putative transformants were replica-plated from galactose plates to YPD plates containing 5-fluoro-orotic acid and screened by PCR. The *K. lactis* GFP-labelled competitor also contained the GFP construct from the *S. cerevisiae* GFP strains fused immediately downstream of the inserted 3′ UTR of *KlacGAL1* in the forward orientation (*MAT**a** ku80*-Δ *gal1*Δ::*kanMX ura3*Δ::$P_{KlacGAL1}$-*KlacGAL1*::P_{TDH3}-*yEGFP*-T_{CYC1} CBS2359). The *K. lactis* GFP-labelled strain used contained 14 consecutive A's instead of 15 A's in the region covered by the forward targeting primer at the 5′ end of the promoter. This strain produced data similar to another *K. lactis* GFP-labelled strain that had 15 consecutive A's but was missing a base pair in the former promoter of *URA3* in the region covered by the reverse targeting primer.

All experimental strains from both species were engineered in triplicate, and each manipulated region was sequence-verified. All primer sequences are available on request.

Competition assays. For each *S. cerevisiae* competition experiment, frozen stocks were plated to YPD. Individual colonies were cultured in autoclaved SC (made according to the manufacturer's instructions with Complete Supplement Mixture (Qbiogene) and Yeast Nitrogen Base without Amino Acids (Amresco)) containing 2% autoclaved raffinose (non-inducing) for 2 days at 30 °C on a rotor wheel. Saturated cultures of each strain were added to aliquots from a single BFP culture. These approximately 50% GFP mixtures were used to inoculate fresh SC with 2% raffinose cultures, which were grown for 2 days. A portion of each culture was then used to inoculate SC with X% galactose at a known concentration, which was varied so that the number of generations was between eight and ten. A portion of this culture was also used to determine the proportion of GFP cells at the start of the experiment (GFPstart) by counting 2×10^5 live cells (those not staining with propidium iodide) and scoring for the presence or absence of GFP, using flow cytometry. After 2 days in SC with X% galactose, a portion of this culture was also used to determine the proportion of GFP cells at the end of the experiment (GFPend) in the same manner. The exact number of generations (t) for each experiment was determined using values from cell counts for a standard GFP versus BFP competition and the known dilution factors. Malthusian selection coefficients (m) were then calculated as $m = \ln(10^{\log[(\mathrm{GFPend}/\mathrm{BFPend})/(\mathrm{GFPstart}/\mathrm{BFPstart})]/t})$ derived from equation (6.4) in ref. 47: $A(t)/B(t) = A(0)/B(0)\ e^{mt}$. For a given experiment performed in parallel against aliquots of the same culture of the stated BFP competitor strain, selection coefficients were normalized by subtracting the measured selection coefficient of the stated GFP control strain. Experiments were carried out in similar fashion in *K. lactis*, except that filter-sterilized carbon sources were used, sucrose was used as the non-inducing carbon source, and the GFP-labelled strain was the competitor strain. For *K. lactis*, at least 10^5 live cells were counted, except for in 0.08% and 0.2% galactose where at least 10^4 cells were counted.

Quantification of mRNA expression. Total RNA was obtained from early log phase cells grown in SC with 2% raffinose before (non-induced) or after (induced) a 60-min induction, with galactose added to a final concentration of 2%. The strains used were the same used in the competition assays with the tested promoter driving the expression of *GAL1* in a *gal1-Δ* background. DNase-treated total RNA (5–25 ng) was used in custom probe-based Taqman One-Step RT–PCR assays using a 7900HT Fast Real-Time PCR System (Applied Biosystems) according to the manufacturer's instructions. The experimental primer/probe set monitored *GAL1* mRNA expression from various promoters, whereas *ACT1* mRNA was a control for RNA quantity in separate reactions. *GAL1* primers were CGTTTATTATGCCAGATATCACAACA (forward; 1,400 nM) and CCGTTCGATGCCGGATT (reverse; 1,400 nM) with a 5′ 6-FAM- and 3′ TAMRA-labelled probe (200 nM), TTCCACACCCTGGAACGGCGAT (Integrated DNA Technologies). *ACT1* primers were GTGATGTCGATGTCCGTAAGGAA (forward; 1,400 nM) and TCTGGAGGAGCAATGATCTTGAC (reverse; 1,400 nM) with a 5′ 6-FAM- and 3′ TAMRA-labelled probe, ATCACCGCTTTGGCTCCATCTTCCAT (100 nM or 200 nM). Twenty microlitre reaction conditions were 30 min at 48 °C, 10 min at 95 °C, and 40 cycles of alternating 15 s at 95 °C and 60 s at 60 °C. After applying the recommended background correction from cycles 3 through 15, threshold cycle (C_t) values were obtained with a threshold of 0.2 using the manufacturer's software. Failed reactions, whose C_t was undetermined or greater than one C_t from the next nearest reading, were excluded. Each biological replicate was supported by at least two RT–PCR replicates. C_t values were converted to relative quantities, normalized to *ACT1* mRNA expression and normalized to the stated promoter.

Statistics. All P values given are from one-tailed non-parametric tests performed using Mstat version 4.01 (http://mcardle.oncology.wisc.edu/mstat/). Gene expression data from several experiments were pooled before analysis with Wilcoxon rank sum tests. For *S. cerevisiae*, fitness values varied (in magnitude, but not relative rankings) with the number of generations and batch of culture media, especially for highly negative values. Therefore, fitness data reported are for a representative experiment employing each of the three independently engineered strains in parallel, but all data were incorporated into statistical tests using Lehman's test, a multiple experiment permutation test based on the Wilcoxon rank sum test. Wilcoxon rank sum tests were used on the *K. lactis* fitness data.