LETTER

doi:10.1038/nature09560

Intercalation of a new tier of transcription regulation into an ancient circuit

Lauren N. Booth[1], Brian B. Tuch[1]† & Alexander D. Johnson[1]

Changes in gene regulatory networks are a major source of evolutionary novelty[1–3]. Here we describe a specific type of network rewiring event, one that intercalates a new level of transcriptional control into an ancient circuit. We deduce that, over evolutionary time, the direct ancestral connections between a regulator and its target genes were broken and replaced by indirect connections, preserving the overall logic of the ancestral circuit but producing a new behaviour. The example was uncovered through a series of experiments in three ascomycete yeasts: the bakers' yeast *Saccharomyces cerevisiae*, the dairy yeast *Kluyveromyces lactis* and the human pathogen *Candida albicans*. All three species have three cell types: two mating-competent cell forms (a and α) and the product of their mating (a/α), which is mating-incompetent. In the ancestral mating circuit, two homeodomain proteins, Mata1 and Matα2, form a heterodimer that directly represses four genes that are expressed only in a and α cells and are required for mating[4–6]. In a relatively recent ancestor of *K. lactis*, a reorganization occurred. The Mata1–Matα2 heterodimer represses the same four genes (known as the core haploid-specific genes) but now does so indirectly through an intermediate regulatory protein, Rme1. The overall logic of the ancestral circuit is preserved (haploid-specific genes ON in a and α cells and OFF in a/α cells), but a new phenotype was produced by the rewiring: unlike *S. cerevisiae* and *C. albicans*, *K. lactis* integrates nutritional signals, by means of Rme1, into the decision of whether or not to mate.

In *S. cerevisiae*, *K. lactis* and *C. albicans*, three cell types (**a**, α and **a**/α) are specified by transcriptional regulators (sequence-specific DNA-binding proteins) encoded at the mating type locus. An important part of this cell-type-specific circuit is the regulation of the haploid-specific genes (hsgs), a group of genes that are expressed in **a** and α cells but not in **a**/α cells[7]. The full sets of hsgs were previously identified in *S. cerevisiae*[4] and *C. albicans* (ref. 5, and B.B.T., Q. M. Mitrovich, F. M. De La Vega, C. K. Monighetti and A.D.J., unpublished observations) but not in the related species *K. lactis*. To examine the evolution of this portion of the mating circuit, we identified the genes in the *K. lactis* hsg regulon and compared them to those in *S. cerevisiae* and *C. albicans*. By profiling the expression patterns of wild-type **a**, α and **a**/α *K. lactis* cells genome-wide, we identified 12 genes that are clear hsgs under the conditions tested (Fig. 1a), two of which—*RME1* (referred to previously as *MTS1*) and *STE4*—were previously identified as hsgs in *K. lactis*[8]. Comparison of all the hsgs in the three species revealed a substantial level of turnover in the regulon; in other words, an hsg in one species is not necessarily an hsg in the other two (Fig. 1b). However, an ancestral core of four hsgs (*GPA1*, *STE4*, *STE18* and *FAR1*) share a common expression pattern in all three species. The first three genes encode the heterotrimeric G protein that, in the presence of mating pheromone, activates a downstream mitogen-activated protein kinase (MAPK) cascade required for mating[9–11]. Far1 lies further downstream in this pathway and mediates two responses needed as a prelude to mating, cell cycle arrest[12] and the formation of mating projections[13,14].

In *S. cerevisiae* and *C. albicans* **a**/α cells, all four genes of the hsg core regulon are directly repressed by the transcription regulator **a**1–α2 (refs 4, 6), a heterodimer encoded by one gene at the MAT**a** locus and one gene at the MATα locus. To determine whether this was also true in *K. lactis*, genome-wide chromatin immunoprecipitations (ChIP-chip) of **a**1 and α2 were performed in *K. lactis* **a**/α cells. In total, the upstream regions of 14 genes were observed to be occupied by both **a**1 and α2, including the *RME1* gene (Fig. 1c), which is also **a**1–α2 regulated in *S. cerevisiae*. **a**1 and α2 ChIP peaks were not observed at the promoters of any of the four core hsgs in *K. lactis*, indicating that, unlike in *S. cerevisiae* and *C. albicans*, these genes are not directly regulated by **a**1–α2.

To confirm the absence of direct **a**1–α2 regulation at *K. lactis* hsgs, we identified the **a**1–α2 recognition motif in *K. lactis* from the ChIP data, using a *de novo* motif-finding program[15]. The highest-scoring motif was similar to the **a**1–α2 motifs previously identified in *S. cerevisiae*[16] and *C. albicans*[17] (Fig. 1d). Indeed, the *S. cerevisiae* motif is efficiently recognized by the *C. albicans* **a**1–α2 protein[17], confirming that key features of this sequence have remained largely unchanged in the three species. We searched the regions 2 kilobases upstream of each *K. lactis* core hsg for the *K. lactis* **a**1–α2 motif but did not find significant matches, confirming the absence of direct **a**1–α2 regulation of these genes. (Whereas the **a**1–α2 site upstream of *RME1* had a $\log_{10}$-odds score of 4.98, the best matches at the core hsgs ranged from −0.70 to 0.93.) These results indicate that although the ancestral core hsg expression pattern is conserved in *K. lactis*, the mechanism of the regulation has changed.

To understand how the *K. lactis* hsgs are cell-type regulated we searched the upstream regions of the 12 genes identified as hsgs by expression array (Fig. 1a) for *cis*-regulatory motifs[15]. The second-highest ranking motif (the top-ranking motif was a repeat sequence) was found in 11 out of 12 of the promoters (Fig. 2a) and was similar to the *S. cerevisiae* Rme1 motif (*K. lactis* consensus, GAACCNMAA; *S. cerevisiae* consensus, GAACCTCAA[18,19]). This motif is also similar to, although longer than, the *K. lactis* Rme1 motif derived previously[20]. The Rme1 motif is absent from *S. cerevisiae* and *C. albicans* hsg promoters.

In *S. cerevisiae*, Rme1 was initially identified as a repressor of meiosis and sporulation[21,22], and was later shown to act as a transcriptional activator of other genes[19]. In *K. lactis*, Rme1 was shown to regulate mating-type interconversion[20] (the switching of **a** and α cells to the opposite cell-type by means of DNA rearrangement). We speculated that Rme1 was co-opted in the *K. lactis* lineage to positively regulate the core hsgs.

To test this hypothesis, we knocked out the *RME1* gene in *K. lactis* **a** cells and examined the gene expression profile by microarray. We found that 20 genes were downregulated in the knockout strain (Fig. 2b), including all four of the core hsgs ($P = 2 \times 10^{-10}$, hypergeometric distribution). We also observed a set of genes that was upregulated in the absence of Rme1 (Supplementary Fig. 1), including a significant number of genes orthologous to *S. cerevisiae* sporulation genes (Gene Ontology (GO): 0043934, $n = 14$, $P = 10^{-14}$, hypergeometric

[1]Department of Microbiology and Immunology and Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA. †Present address: Genome Analysis Unit, Amgen, South San Francisco, California 94080, USA.

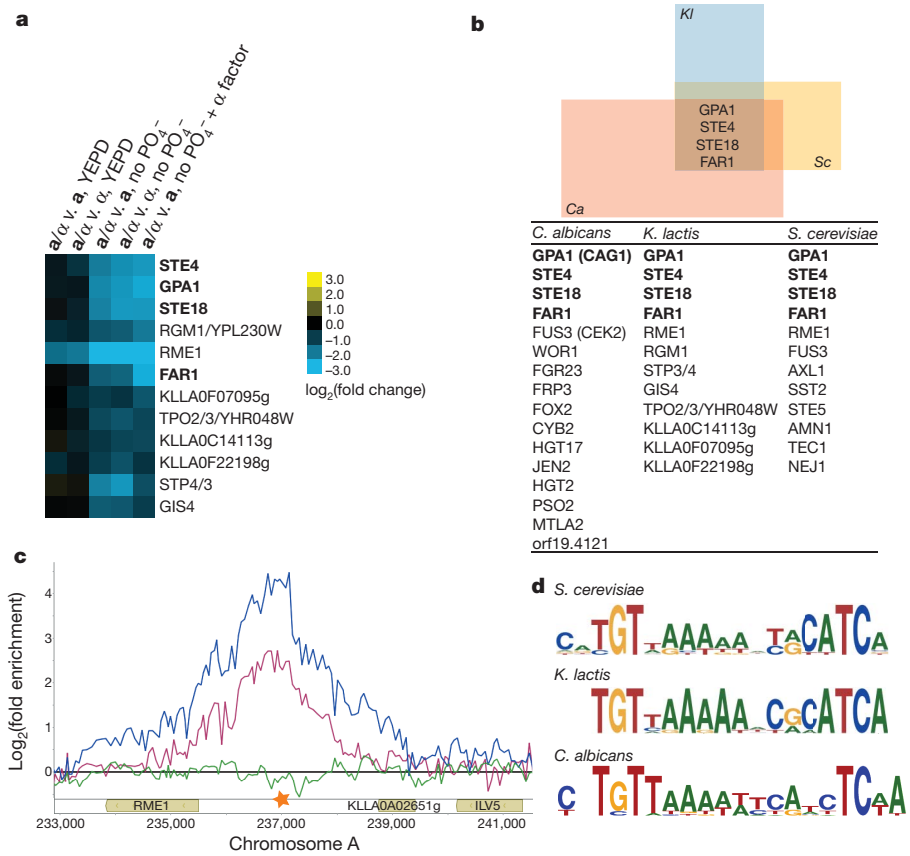


Figure 1 | The core hsgs are not directly regulated by a1–α2 in *K. lactis*. **a**, The expression profiles of the set of 12 hsgs identified in *K. lactis*. Note that phosphate starvation induces expression of the hsgs and is required to identify these genes. For example, when starved for phosphate, the heterotrimeric G protein genes are expressed in **a** and α cells at levels about fivefold higher than in **a**/α cells. **b**, A comparison of hsgs defined by transcriptional profiling in *S. cerevisiae* (*Sc*)[4], *K. lactis* (*Kl*) (panel **a**) and *C. albicans* (*Ca*) (ref. 5, and B.B.T., Q. M. Mitrovich, F. M. De La Vega, C. K. Monighetti and A.D.J., unpublished observations) shows a conserved subset of hsgs (*GPA1*, *STE4*, *STE18*, *FAR1*), which we refer to as the core hsgs (bold in **a** and **b**). **c**, ChIP enrichment profiles from experiments using haemagglutinin (HA)-tagged *MAT**a**1* **a**/α cells (magenta), HA-tagged *MATα2* **a**/α cells (blue) and, as a control, untagged **a**/α cells (green). The ChIP enrichment was determined by hybridization to a tiling microarray. The location of the **a**1–α2 motif in the *RME1* promoter is indicated by the orange star. The genes (tan boxes) are all transcribed in the reverse direction. Data were visualized with MochiView[30]. **d**, The *K. lactis* **a**1–α2 motif determined from the ChIP-chip data. For comparison, the *S. cerevisiae* and *C. albicans* motifs (derived from published ChIP data[4,6]) are also shown.

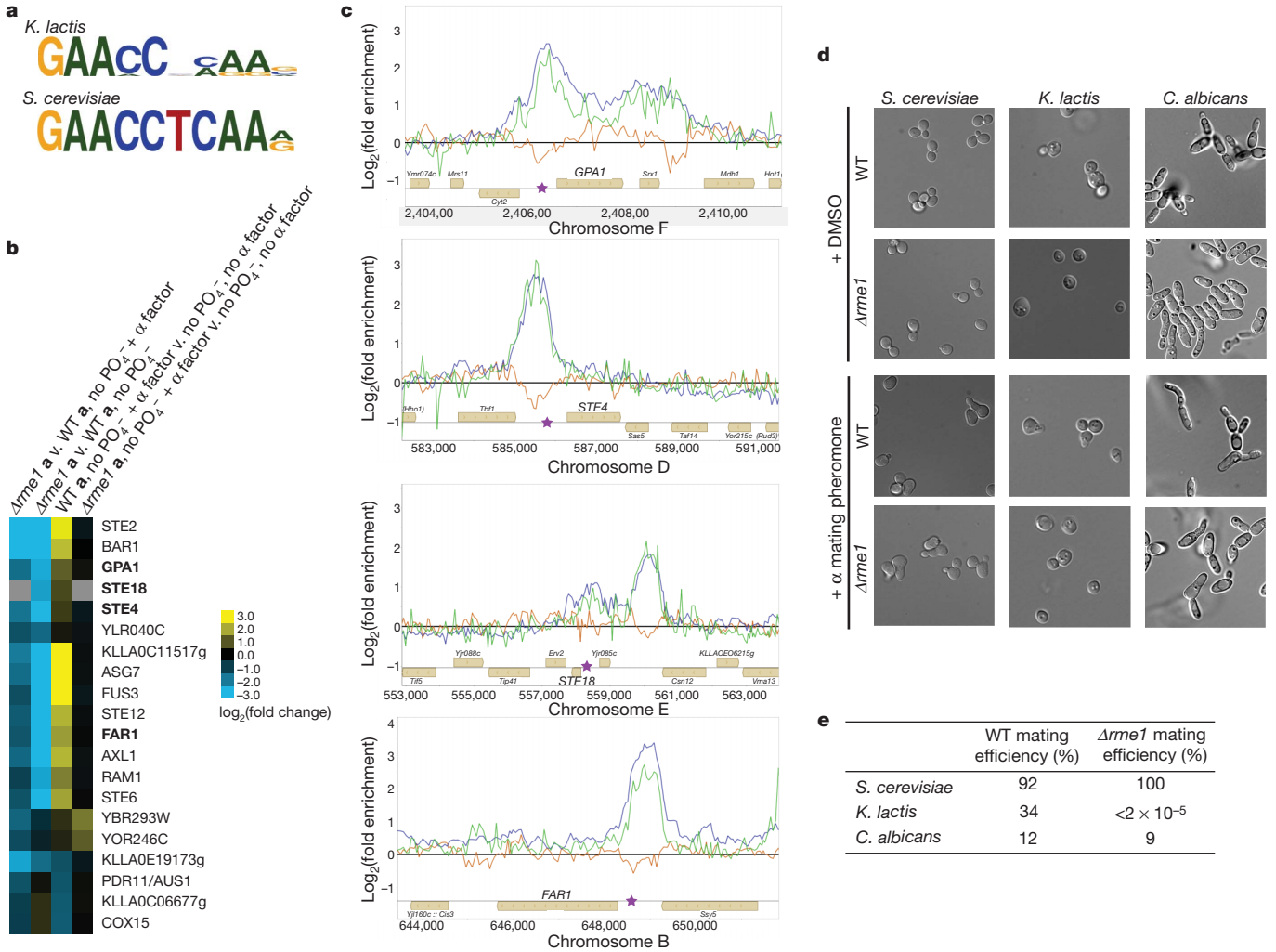
distribution) indicating that the function of Rme1 in regulating meiosis is shared with *S. cerevisiae*. Thus, whereas some of the targets of Rme1 have remained the same as in the common ancestor of *S. cerevisiae* and *K. lactis*, Rme1 has gained new targets in the *K. lactis* lineage, including the core hsgs.

To test whether Rme1 directly regulates the core hsgs in *K. lactis*, we performed a genome-wide ChIP of Rme1 in **a** cells. ChIP peaks were observed at the promoters of the four core hsgs (Fig. 2c) and were centred over the Rme1 motifs (Fig. 2c). Thus, in the *K. lactis* lineage, Rme1 was gained as a direct activator of the core hsgs by the acquisition of Rme1 *cis*-regulatory sequences at all four genes. We note that Rme1 is not the only regulator of the *K. lactis* hsgs; for example, *STE18* is repressed by Sir2 (ref. 23).

We next tested the biological role of Rme1 in mating in *K. lactis*, *S. cerevisiae* and *C. albicans* by comparing wild-type and *RME1* knockout **a** cells. In response to α pheromone, **a** cells form mating projections (polarized growth towards the source of pheromone). When *S. cerevisiae* and *C. albicans* wild-type and *Δrme1* **a** cells were exposed to α mating pheromone, both strains formed mating projections normally (Fig. 2d). In contrast, whereas *K. lactis* wild-type **a** cells produced mating projections in response to pheromone, *Δrme1* **a** cells did not, indicating that this biological response was dependent on Rme1 (Fig. 2d). As a second test of the role of Rme1, we examined mating directly using a quantitative mating assay. No difference was observed between the mating efficiencies of wild-type **a** cells and those of *Δrme1* **a** cells for *S. cerevisiae* and *C. albicans* (Fig. 2e). In contrast, the *K. lactis Δrme1* **a** cell mating efficiency was decreased, relative to the wild type, by a factor of at least 10^6 (Fig. 2e). Thus, the ability to mate is critically dependent on Rme1—but only in *K. lactis*.

Unlike *S. cerevisiae* and *C. albicans*, *K. lactis* requires a starvation signal to mate[24] and to respond to pheromone[25]. Although several different types of starvation signal can prime *K. lactis* to respond to pheromone[24,25], we found that phosphate starvation is particularly potent, and it was used in subsequent experiments. Our expression profiling experiments (Fig. 1a) revealed that *K. lactis* requires starvation to express most of its mating genes. *RME1* was also highly induced (24-fold) by phosphate starvation (Fig. 1a). We note that *S. cerevisiae RME1* transcript levels also increase tenfold under starvation conditions[18], suggesting that regulation of *RME1* by starvation may be ancestral to *S. cerevisiae* and *K. lactis*.

We next investigated in greater detail how the starvation signal is incorporated in the *K. lactis* mating regulatory circuit. The simplest model consistent with the data presented so far is that starvation upregulates *RME1*, which in turn activates transcription of the hsgs. A prediction of this model is that ectopic expression of *RME1* in *K. lactis* should override the requirement for starvation in expressing the hsgs. We created an **a** strain overexpressing *RME1* to levels that were within tenfold of the level in starved wild-type cells (using the *Kl* LAC4 promoter) and found that overexpression of *RME1* (in rich YEP-galactose medium) was sufficient to induce expression of the heterotrimeric G protein subunits (Fig. 3a). Overexpression of *RME1* is also sufficient to allow *K. lactis* to form mating projections



	WT mating efficiency (%)	*Δrme1* mating efficiency (%)
S. cerevisiae	92	100
K. lactis	34	$<2 \times 10^{-5}$
C. albicans	12	9

Figure 2 | RME1 is a direct activator of hsg expression and is required for *K. lactis* mating. a, The *K. lactis* Rme1 motif found by a *de novo* search[15] of the 12 *Kl* hsgs and the *S. cerevisiae* motif derived from two experimentally characterized binding sites[18,19]. **b**, The set of 19 genes repressed twofold or greater relative to wild type when *RME1* is absent and the cells are phosphate-starved. In bold are the core hsgs. **c**, Rme1 is a direct regulator of the core hsgs. ChIP of Rme1 was performed in *K. lactis* c-Myc-tagged *RME1* **a** cells (blue and green lines, two biological replicates) and untagged, control **a** cells (orange line). The immunoprecipitated DNA was hybridized to a tiling microarray. The genes (tan boxes) above the line are transcribed in the forward direction and those below are transcribed in the reverse direction. The location of the *Kl* Rme1 motif is indicated by a purple star. **d**, Rme1 is required only in *K. lactis* to respond to mating pheromone. Wild-type or *RME1* knockout **a** cells were exposed to either α mating pheromone or a mock treatment of dimethylsulphoxide (DMSO). Mating projections form readily when both wild-type and *Δrme1 S. cerevisiae* and *C. albicans* cells are exposed to mating pheromone. Only the *K. lactis Δrme1* **a** cells were unable to respond to the presence of α mating pheromone. **e**, Rme1 is required for mating in *K. lactis* but not in *S. cerevisiae* nor *C. albicans*. Quantitative mating assays were performed by mating wild-type or *Δrme1* **a** cells to wild-type α cells. In *S. cerevisiae* and *C. albicans* the percentage of **a** cells that was able to mate is similar for wild-type and *Δrme1*. *K. lactis Δrme1* **a** cells are mating incompetent; no mating products were isolated from the *Δrme1* **a** × wild-type α mating.

in response to pheromone in rich medium (Fig. 3b). These results strongly support the model by showing that upregulation of *RME1* is sufficient to cause biologically relevant upregulation of the heterotrimeric G proteins.

Thus, the rewiring of the *K. lactis* hsg circuit (summarized in Fig. 4) resulted in a new network configuration and a novel phenotype, relative to the ancestor. Our results suggest a possible evolutionary path for this rewiring. In the ancestor of all three yeasts, the hsgs were directly repressed by **a**1–α2. Either in an ancestor to *S. cerevisiae* and *K. lactis* or independently in each lineage, *RME1* was brought under nutritional regulation. Finally, in the *K. lactis* lineage alone, two steps occurred: the hsgs lost the *cis*-regulatory sequences for **a**1–α2 and gained the *cis*-regulatory sequences for Rme1. As described in Supplementary Fig. 2, it is possible to determine more precisely when the rewiring of the core hsgs occurred. We can infer that direct **a**1–α2 regulation of the core hsgs was probably lost several times in the ascomycete lineage, and that the *K. lactis* form of regulation of the hsgs probably arose after *K. lactis* and the closely related species *L. kluyveri* branched from their common ancestor.

Although we do not know whether acquisition of the *K. lactis* form of regulation was adaptive, this type of regulation makes logical sense given that the primary mode of growth of *K. lactis* is as a haploid[26]. The formation of spores is a strategy employed by many yeasts to survive harsh environments. For starvation to give rise to spores, *K. lactis* would first have to mate (to form the sporulation-competent **a**/α cell type), thus rationalizing the link between starvation and mating. In contrast, *S. cerevisiae* in the wild is typically at least diploid[27] and forms spores directly in response to starvation. Thus, the coupling of mating and starvation makes conceptual sense for *K. lactis* in comparison with *S. cerevisiae*.

We have described a case in which a new tier of regulation has been intercalated into an ancient transcription circuit consisting of a regulator (a homeodomain heterodimer) and a set of target genes. This change involved breaking the original connections between the regulator and its

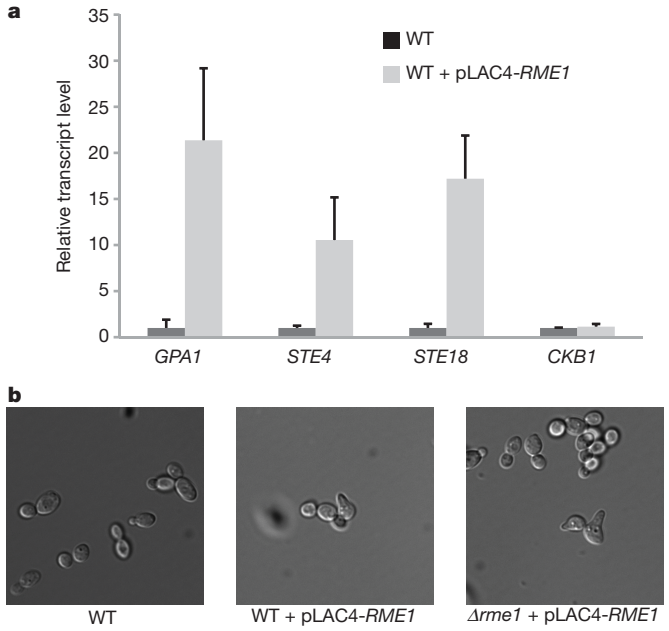


Figure 3 | Overexpression of *RME1* is sufficient for hsg expression in the absence of nutrient starvation. **a**, In the overexpression strain (pLAC4-*RME1*), *RME1* transcription is induced by galactose-containing medium, a condition that does not cause expression of the heterotrimeric G proteins or pheromone response in wild-type (WT) cells. A strain using the empty pLAC4 vector was used as a control. The transcripts were measured relative to *ACT1* transcript levels by RT–quantitative PCR (means and s.d., $n = 3$). In the absence of a starvation signal the hsgs, but not *CKB1* (a non-hsg control), are upregulated when *RME1* is overexpressed. **b**, *RME1* overexpression allows cells to respond to mating pheromone in the absence of a starvation signal. *K. lactis* **a** cells that contained only the endogenous *RME1* copy and an empty pLAC4 vector (WT), the endogenous copy of *RME1* and *RME1* driven by the pLAC4 promoter (WT + pLAC4-*RME1*) or only *RME1* driven by the pLAC4 promoter (*Δrme1* + pLAC4-*RME1*) were grown in YEP-galactose and exposed to α mating pheromone. Wild-type cells were unable to form mating projections in the absence of a starvation signal, but both strains overexpressing *RME1* (pLAC4-*RME1*) formed mating projections in the absence of a starvation signal.

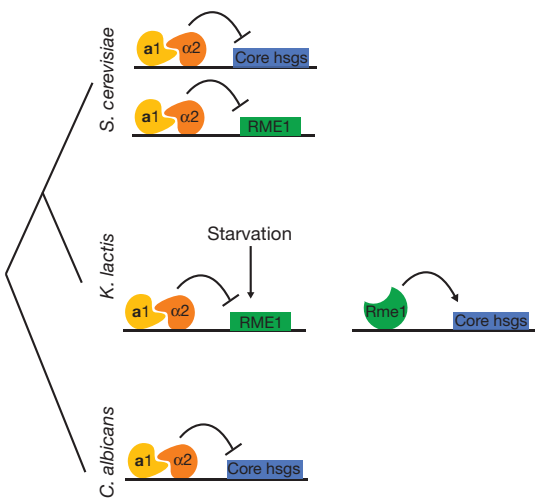


Figure 4 | A simplified model for the evolution of regulation of core hsgs in three yeasts. In all three species the core hsgs are repressed by **a**1–α2; thus, they are ON in **a** and α cells and OFF in **a**/α cells. In *S. cerevisiae* and *C. albicans* the repression is direct (**a**1–α2 binds to the promoters of these genes), but in *K. lactis* it is indirect, through Rme1. The circuit rewiring in the *K. lactis* lineage has resulted in a new mating behaviour; this species is able to mate only when starved. We show that this behaviour is due to the intercalation of Rme1, which is upregulated by starvation in *K. lactis*.

target genes and replacing them with a more complex type of hierarchy (Fig. 4). Intercalation may be a common way in which regulatory circuits evolve. This type of 'intercalary evolution' was first proposed[28] to account for a common origin of eyes. In a wide variety of species, the transcription regulator Pax6 lies at the top of the eye development hierarchy, and rhodopsins occupy the bottom. According to the proposal, different types of eye arose from evolutionary intercalation of a variety of regulatory and structural genes within this simple, deeply conserved, regulatory relationship. The change we describe here is less complex and provides a concrete example of evolutionary intercalation, one that is responsible for an important feature of modern mating behaviour in *K. lactis*. It has been known for decades that *K. lactis* (unlike its relatives *S. cerevisiae* and *C. albicans*) requires starvation to mate[24], and we have shown that this behaviour is due to the new configuration of the *K. lactis* mating circuit.

METHODS SUMMARY

Gene expression array. RNA was isolated by hot phenol extraction and reverse transcribed, and the resulting complementary DNAs were coupled to Cy5. A pooled mixture of the cDNAs was coupled to Cy3 and used as a reference. Labelled cDNAs were hybridized to Agilent arrays for visualization.

ChIP-chip. ChIP experiments were performed as described previously[29], with minor modifications.

Pheromone response assays. Exponential-phase cultures were exposed to 13-mer α-mating pheromone, and the formation of mating projections was monitored by microscopy.

Quantitative mating assay. **a** and α cultures were grown independently to exponential phase and then mixed together with α cells in fivefold excess under mating conditions. The mating products were selected for on medium that either **a** and **a**/α cells or only **a**/α cells could grow on, and efficiencies were calculated as efficiency = (**a**/α colonies)/(**a** and **a**/α colonies).

RT–quantitative PCR. Cultures were grown to exponential phase in YEP-galactose medium, and RNA was isolated by extraction with hot phenol. RNA was reverse transcribed and the cDNAs were quantified by quantitative PCR.

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

Received 27 April; accepted 6 October 2010.

1. Carroll, S. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134,** 25–36 (2008).
2. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311,** 796–800 (2006).
3. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nature Rev. Genet.* **8,** 206–216 (2007).
4. Galgoczy, D. J. *et al.* Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **101,** 18069–18074 (2004).
5. Tsong, A. E., Miller, M. G., Raisner, R. M. & Johnson, A. D. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* **115,** 389–399 (2003).
6. Srikantha, T. *et al.* TOS9 regulates white-opaque switching in *Candida albicans*. *Eukaryot. Cell* **5,** 1674–1687 (2006).
7. Herskowitz, I. A regulatory hierarchy for cell specialization in yeast. *Nature* **342,** 749–757 (1989).
8. Barsoum, E., Sjöstrand, J. O. O. & Aström, S. U. Ume6 is required for the MAT**a**/MATα-cellular identity and transcriptional silencing in *Kluyveromyces lactis*. *Genetics* **184,** 999–1011 (2010).
9. Herskowitz, I. MAP kinase pathways in yeast: for mating and more. *Cell* **80,** 187–197 (1995).
10. Dignard, D., André, D. & Whiteway, M. Heterotrimeric G-protein subunit function in *Candida albicans*: both the α and β subunits of the pheromone response G protein are required for mating. *Eukaryot. Cell* **7,** 1591–1599 (2008).
11. Coria, R. *et al.* The pheromone response pathway of *Kluyveromyces lactis*. *FEM. Yeast Res.* **6,** 336–344 (2006).
12. Chang, F. & Herskowitz, I. Identification of a gene necessary for cell cycle arrest by a negative growth factor of yeast: FAR1 is an inhibitor of a G1 cyclin, CLN2. *Cell* **63,** 999–1011 (1990).
13. Butty, A. C., Pryciak, P. M., Huang, L. S., Herskowitz, I. & Peter, M. The role of Far1p in linking the heterotrimeric G protein to polarity establishment proteins during yeast mating. *Science* **282,** 1511–1516 (1998).
14. Côte, P. & Whiteway, M. The role of *Candida albicans* FAR1 in regulation of pheromone-mediated mating, gene expression and cell cycle arrest. *Mol. Microbiol.* **68,** 392–404 (2008).
15. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 28–36 (1994).

16. Goutte, C. & Johnson, A. D. Recognition of a DNA operator by a dimer composed of two different homeodomain proteins. *EMBO J.* **13,** 1434–1442 (1994).
17. Hull, C. M. & Johnson, A. D. Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*. *Science* **285,** 1271–1275 (1999).
18. Covitz, P. A. & Mitchell, A. P. Repression by the yeast meiotic inhibitor RME1. *Genes Dev.* **7,** 1598–1608 (1993).
19. Toone, W. M. *et al.* Rme1, a negative regulator of meiosis, is also a positive activator of G1 cyclin gene expression. *EMBO J.* **14,** 5824–5832 (1995).
20. Barsoum, E., Martinez, P. & Astrom, S. U. α3, a transposable element that promotes host sexual reproduction. *Genes Dev.* **24,** 33–44 (2010).
21. Kassir, Y. & Simchen, G. Regulation of mating and meiosis in yeast by the mating-type region. *Genetics* **82,** 187–206 (1976).
22. Mitchell, A. P. & Herskowitz, I. Activation of meiosis and sporulation by repression of the RME1 product in yeast. *Nature* **319,** 738–742 (1986).
23. Hickman, M. A. & Rusche, L. N. The Sir2-Sum1 complex represses transcription using both promoter-specific and long-range mechanisms to regulate cell identity and sexual cycle in the yeast *Kluyveromyces lactis*. *PLoS Genet.* **5,** e1000710 (2009).
24. Herman, A. Interspecies sex-specific growth responses in *Kluyveromyces*. *Antonie van Leeuwenhoek* **36,** 421–425 (1970).
25. Tuch, B. B., Galgoczy, D. J., Hernday, A. D., Li, H. & Johnson, A. D. The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* **6,** e38 (2008).
26. Kurtzman, C. P. & Fell, J. W. *The Yeasts: A Taxonomic Study* 4th edn (Elsevier, 2000).
27. Ezov, T. K. *et al.* Molecular-genetic biodiversity in a natural population of the yeast *Saccharomyces cerevisiae* from 'Evolution Canyon': microsatellite polymorphism, ploidy and controversial sexual status. *Genetics* **174,** 1455–1468 (2006).
28. Gehring, W. J. & Ikeo, K. *Pax 6*: mastering eye morphogenesis and eye evolution *Trends Genet.* **15,** 371–377 (1999).
29. Nobile, C. J. *et al.* Biofilm matrix regulation by *Candida albicans* Zap1. *PLoS Biol.* **7,** e1000133 (2009).
30. Homann, O. R. & Johnson, A. D. MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol.* **8,** 49 (2010).

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

Acknowledgements We thank Q. Mitrovich, O. Homann, A. Hernday, M. Miller, C. Cain, T. Sorrells and H. Madhani for helpful discussions and technical contributions; and S. Åström for generously providing the *K. lactis* strains used in this study. The *S. cerevisiae* strains were a gift from the H. Madhani and J. Li laboratories. The work was funded by grant RO1 GM037049 from the National Institutes of Health. L.N.B. is a National Science Foundation Graduate Research Fellow.

Author Contributions L.N.B. performed all experiments. L.N.B. and B.B.T. analysed data. L.N.B., B.B.T. and A.D.J. designed the study and wrote the paper.

Author Information The gene expression array data have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE24874. For the ChIP-chip data the accession number is GSE25209. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.D.J. (ajohnson@cgl.ucsf.edu).

METHODS

Medium. Details of the medium used in the experiments presented here can be found in ref. 31. The recipe for the phosphate starvation medium can be found in ref. 25.

Strains and strain construction. The strains used in this study can be found in Supplementary Information. *S. cerevisiae* strains are S288C background and *C. albicans* strains are SC5314 background.

Gene disruption cassettes for knockouts and taggings in *K. lactis* and *L. kluyveri* were generated by fusion PCR[32] using the primers listed in Supplementary Information. Fusion PCRs were performed in a 50-µl reaction containing 0.5 µl ExTaq (Takara Bio Inc.), 0.25 mM dNTPs, 0.2 µM each primer and about 25 ng of template. The reactions were incubated as follows: 94 °C for 3 min; 35 cycles of 94 °C for 30 s, 50–55 °C (depending on primer) for 30 s and 72 °C for 1 min per kilobase; and 72 °C for 5 min. The first round of PCR consisted of three reactions that amplified the flanking homologous sequence from *K. lactis* genomic DNA with primers 1 and 3 or 4 and 6, and amplified the markers from the appropriate plasmids with primers 2 and 5. The URA3 marker was amplified from YEp24, the TRP1 marker from YEplac112, the c-Myc tagging cassette from pFA6a-13Myc-kanMX6 (ref. 33) and the 3 × HA tagging cassette from pYMN-20 (ref. 34). The products were purified with the QIAquick PCR Purification Kit (Qiagen). The second round of amplification (the fusion round) used 1 µl of each purified flank PCR product and 2 µl of the purified marker PCR product. This product was purified with the QIAquick PCR Purification Kit.

The purified fusion PCR products were transformed into *K. lactis* and *L. kluyveri* by electroporation[35,36]. Transformants were confirmed to be correct by colony PCR with the check primers listed in Supplementary Information. Tagged genes were also verified by sequencing.

To avoid mating-type switching in *K. lactis* and to increase the efficiency of tagging genes in the MAT loci, we created strains in which the silent MAT loci (HML**α** and HMR**a**) were knocked out. SAY509 and SAY572 were transformed with gene disruption cassettes for HMR**a** (yLB10b) and HML**α** (yLB11b), respectively. To generate strains with both silent cassettes deleted, these strains were mated as described[35] for 2 days, and diploids were selected for by URA/TRP drop-out SCD medium. The resulting strain (yLB12d) was sporulated on pre-SPO plates for 2 days, and haploids with both silent cassettes deleted were selected for on URA/TRP drop-out medium. The mating types were determined by colony PCR using the MAT, MAT**a** and MAT**α** check primers. These strains (yLB13a and yLB14) were mated to generate an **a**/**α** strain lacking silent MAT loci (yLB15b).

To create N-terminally 3 × HA-tagged MAT**a**1 and MAT**α**2 for ChIP experiments, yLB13a and yLB14 were transformed with the gene disruption cassettes created with the HA primers and the MAT**a**1-tag or MAT**α**-tag primers. yLB55 was mated with yLB14 to create yLB58, the **a**/**α** strain used in the MAT**a**1 ChIP experiment. yLB56a was mated with yLB13a to create yLB57a, the **a**/**α** strain used in the MAT**α**2 ChIP experiment.

RME1 was knocked-out (yLB21a) and tagged (yLB54b) in SAY572.

yDG957 is the mated product of SAY509 and SAY572.

yLB33a1 was created by sorbose-selecting TF028X, as described in ref. 37. AH136a1 was sorbose-selected from SN87, and HIS1 and LEU2 were added back as described in ref. 38.

pLAC4 (described below) was transformed into SAY572 after digestion with *Sac*II (New England Biolabs) to generate yLB61b. pLAC4-RME1 (described below) was transformed into SAY 572 and yLB21a to create yLB64a and yLB65, respectively.

pLAC4 was created by modifying pKLAC1 (New England Biolabs) as follows. The pKLAC1 vector was cut with *Hin*dIII and *Xho*I and gel-purified with the QIAquick Gel Extraction Kit to remove most of the α-mating pheromone secretory domain. The vector was dephosphorylated with APEX heat-labile dephosphatase (Epicentre), in accordance with the kit's instructions. Primers *Bam*HI *add* and *Bam*HI *add, reverse complement* were hybridized to each other at a concentration of 40 µM in 1 × T4 PNK buffer (New England Biolabs) and 40 mM additional NaCl by incubation at 94 °C for 2 min and slow cooling to 10 °C at a rate of 0.1 °C s^{-1}. The hybridized primers created sticky ends for *Hin*dIII and *Xho*I cut sites. The hybridized primers were phosphorylated by adding ATP to a final concentration of 1 mM and T4 PNK kinase (New England Biolabs) to a final concentration of 200 U ml^{-1} and incubating them at 37 °C for 10 min. The cut, dephosphorylated vector was ligated to the hybridized primers by using Fast-Link Ligase (Epicentre) at a 1:5 molar ratio and transformed into DH5α cells. The acetimidase marker in the plasmid was then replaced with kanMX6. kanMX6 was amplified from pFA6a-13myc-kanMX6 (ref. 33) by using the KAN primers. The plasmid and kanMX6 marker were cut with *Bsr*GI and *Xma*I (New England Biolabs) and ligated and transformed as described above. This plasmid was used as an empty vector control (referred to as pLAC4) in the RME1 overexpression experiments.

pLAC4-RME1 was generated by amplifying RME1 from *K. lactis* genomic DNA with the primers RME1 + *Bgl*II and RME1 + *Not*I. The *RME1* gene and the pLB12 plasmid were cut with *Bgl*II and *Not*I (New England Biolabs) and ligated and transformed as described above.

Orthologous gene set mapping. In this study we used the orthologous gene sets defined previously[25].

Gene expression arrays. Arrays were designed with OligoArray (v2.1.3)[39]. The reference sequence used was downloaded from the NCBI genome project for *Kluyveromyces lactis* NRRL Y-1140, records NC006037 to NC006042. The predicted messenger RNA sequences were used. The sequences of two shorter open reading frame (ORF)-coding transcripts, MFA1 and AGA2, that were not annotated in NCBI at the time were added to the reference. The following parameters were used: maximum number of oligonucleotides to design per input sequence, 3; size range, 60 to 60; maximum distance between the 5′ end of the oligonucleotide and the 3′ end of the input sequence, 1,500; minimum distance between the 5′ ends of two adjacent oligonucleotides, 69; T_m range, 75 to 97 °C; GC range, 15.0 to 65.0; threshold to reject secondary structures, 65.0; threshold to start to consider cross-hybridizations, 65; sequence to avoid in the oligonucleotide: GGGGG;CCCCC;TTTTT;AAAAA.

The arrays were printed by Agilent, using the 4 × 44K format.

K. lactis strains were grown in either rich medium (YEPD) to an attenuance (D_{600}) of 0.9 or in phosphate starvation medium, with or with out α-mating pheromone, as described previously[25]. The 50-ml cultures were centrifuged for 5 min at 3,700*g*, and the pellet was resuspended in 10 ml of 1 × TE and centrifuged again. The supernatant was removed and pellets were frozen in liquid nitrogen and stored at −80 °C.

RNA was isolated and reverse transcribed as described previously[38] with the exception that the RNA isolation protocol was scaled to 50-ml cultures and that SuperScript II (Invitrogen) was used. Additionally, reverse transcription reactions were performed on either individual samples or on RNA in which an equivalent amount of each RNA sample was pooled.

For the individual samples, 1.3 µg of cDNA was dried and resuspended in 5 µl of 0.1 M sodium bicarbonate. For the pooled samples, 5.9 µg of cDNA was dried and resuspended in 22.5 µl of 0.1 M sodium bicarbonate. An equivalent volume of Cy3 (pool) or Cy5 (individual) dye (Amersham) was added (dyes were resuspended in 60 µl of DMSO) and the reaction was incubated in the dark at 65 °C for 20 min. Labelled cDNAs were purified with a Clean and Concentrator -5 kit (Zymo Research).

Equal amounts of the Cy3-labelled and Cy5-labelled cDNA were hybridized overnight to the array, as described in the Agilent protocol. After hybridization, the arrays were washed as specified by Agilent. Arrays were scanned at 5 µm, averaging two lines, with an Axon GenePix 4000A scanner. Arrays were gridded with GenePix Pro v5.1. Global Lowess normalization analysis was performed for each array with a Goulphar script[40] (R Foundation for Statistical Computing). Normalized data were collapsed first by averaging the result for all duplicate probes and finally by taking the median of the probes for each ORF. Data were transformed as described for each experiment. Microarray data were clustered with Cluster version 3.0 (ref. 41) and visualized with Java TreeView v1.1.3 (ref. 42).

ChIP-chip. *K. lactis* was grown in either YEPD to D_{600} = 0.4 or in phosphate starvation medium as described previously[25]. The ChIP, DNA amplification, labelling and hybridization were performed as described previously[29]. For the **a**1 and α2 ChIPs, 2 µl of 5 mg ml^{-1} mouse anti-HA antibody clone 12CA5 (Roche) were used. The *K. lactis* tiling arrays used have been described previously[25]. Peak finding was performed with the 'Extract peaks from Data Set(s)' utility of MochiView[30]. Peak extraction was applied independently to each normalized ChIP-chip data item. Peak finding significance thresholds were kept at their default values ($P \leq 0.001$). For the **a**1 and α2 ChIPs, regions of overlap between the two ChIP-chips were determined and the overlapping chromosomal coordinates were extracted, yielding 22 regions. This set was filtered to remove peaks in the telomeres or those that fell entirely in an ORF. MEME was performed on the remaining 14 peaks.

Pheromone response assay. For *K. lactis* (SAY572 and yLB21a), the pheromone response assay was performed as described previously[25], with the exception that the medium was not supplemented with additional leucine. For *C. albicans* (AH136a1 and yLB33a1), cells were grown in SCD medium at room temperature (23–25 °C) to D_{600} = 1.0 and exposed to pheromone (or mock 10% DMSO treatment) for 4 h as described in ref. 43. For *S. cerevisiae* (BY4674 and YM1768), the cells were grown in YEPD at 30 °C to D_{600} = 0.5. α-Mating pheromone (5 µM; Sigma-Aldrich) was added and the cells were grown at 30 °C for 90 min; the presence of mating projections was monitored by microscopy.

Quantitative mating. The strains and selective medium used for these experiments are listed in Supplementary Information.

Yeasts were grown in YEPD (*S. cerevisiae*, *K. lactis*) or SCD (*C. albicans*) at 30 °C (*S. cerevisiae*, *K. lactis*) or room temperature (*C. albicans*) to D_{600} = 0.8. **a** cells (2×10^6; about 200 µl of D_{600} = 0.8) were mixed with 10^7 α cells (about 1 ml of

D_{600} = 0.8) in 10 ml of YEPD. The mixtures were filtered onto nitrocellulose (0.8 μm pore size; Millipore) with a Millipore 1225 Vacuum Manifold. Filter discs were placed onto either YEPD + 55 μg ml^{-1} adenine 2% agar plates (*S. cerevisiae*, *C. albicans*) or 2% malt extract, 3% agar plates (*K. lactis*). *S. cerevisiae* strains were allowed to mate for 5 h at 30 °C, *K. lactis* strains for 2 days at 30 °C, and *C. albicans* strains for 6 days at room temperature. The filter discs were then removed and placed in 5 ml of SD medium and the cells were dispersed by vortex-mixing. The cell suspensions were sonicated with a Branson Sonifier 450 at 30% power for 10 s. Between 1/25 and 2.5 × 10^{-6} of the cell suspension was plated onto selective medium (see Supplementary Information). *S. cerevisiae* and *K. lactis* cells were grown at 30 °C for 2 days, and *C. albicans* cells were grown at room temperature for 4 days. Mating efficiency was calculated with the limiting parental **a** cells: efficiency = (**a**/α)/(**a** + **a**/α).

RT–quantitative PCR. *K. lactis* strains were grown in YEP-galactose to D_{600} = 0.8 and centrifuged at 3,700*g* for 5 min. The pellets were washed with 1 ml of 1 × TE and centrifuged at 3,700*g* for 5 min; supernatant was removed and frozen in liquid nitrogen. RNA was isolated and reverse transcribed (using SuperScript II) as described previously[38], with all volumes scaled appropriately. cDNAs were quantified with a Bio-Rad CFX96 Real Time machine in a standard 25-μl reaction using Sybr green under standard conditions. The primers used are listed in Supplementary Information.

Hypergeometric test. The significance of the set of genes (n = 20) down twofold or greater relative to wild type in a phosphate-starved *Rme1* knockout containing all four core hsgs was calculated with a hypergeometric test. The background set of genes (n = 4,769) are defined as all genes that were detectable by our gene expression array.

GO term analysis. For the analysis of the genes that were upregulated in the absence of RME1 in *K. lactis*, we performed GO term analysis using the *S. cerevisiae* GO term finder on SGD (http://www.yeastgenome.org/cgi-bin/GO/goTermFinder.pl). Genes that were up fourfold or greater in an RME1 knockout versus wild type in the RME1 array and that had an orthologue in *S. cerevisiae* were compared with a background set of genes defined as all *S. cerevisiae* genes orthologous to *K. lactis* genes that could be detected in our gene expression array.

31. Sherman, F. Getting started with yeast. *Methods Enzymol.* **350,** 3–41 (2002).
32. Wach, A. PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast* **12,** 259–265 (1996).
33. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14,** 953–961 (1998).
34. Janke, C. *et al.* A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21,** 947–962 (2004).
35. Wolf, K., Breunig, K. & Barth, G. *Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology* (Springer, 2003).
36. Gojkovic, Z., Jahnke, K., Schnackerz, K. D. & Piskur, J. PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *J. Mol. Biol.* **295,** 1073–1087 (2000).
37. Janbon, G., Sherman, F. & Rustchenko, E. Monosomy of a specific chromosome determines L-sorbose utilization: a novel regulatory mechanism in *Candida albicans*. *Proc. Natl Acad. Sci. USA* **95,** 5150–5155 (1998).
38. Mitrovich, Q. M., Tuch, B. B., Guthrie, C. & Johnson, A. D. Computational and experimental approaches double the number of known introns in the pathogenic yeast *Candida albicans*. *Genome Res.* **17,** 492–502 (2007).
39. Rouillard, J.-M., Zuker, M. & Gulari, E. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.* **31,** 3057–3062 (2003).
40. Lemoine, S., Combes, F., Servant, N. & Le Crom, S. Goulphar: rapid access and expertise for standard two-color microarray normalization methods. *BMC Bioinformatics* **7,** 467 (2006).
41. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20,** 1453–1454 (2004).
42. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20,** 3246–3248 (2004).
43. Bennett, R. J., Uhl, M. A., Miller, M. G. & Johnson, A. D. Identification and characterization of a *Candida albicans* mating pheromone. *Mol. Cell. Biol.* **23,** 8189–8201 (2003).