A New System for Comparative Functional Genomics of *Saccharomyces* Yeasts

Amy A. Caudy,[1,*] Yuanfang Guan,[†,2] Yue Jia,[†] Christina Hansen,[†] Chris DeSevo,[†] Alicia P. Hayes,[†] Joy Agee,[†] Juan R. Alvarez-Dominguez,[†] Hugo Arellano,[†] Daniel Barrett,[†] Cynthia Bauerle,[†] Namita Bisaria,[†] Patrick H. Bradley,[†] J. Scott Breunig,[†] Erin Bush,[†] David Cappel,[†] Emily Capra,[†] Walter Chen,[†] John Clore,[†] Peter A. Combs,[†] Christopher Doucette,[†] Olukunle Demuren,[†] Peter Fellowes,[†] Sam Freeman,[†] Evgeni Frenkel,[†] Daniel Gadala-Maria,[†] Richa Gawande,[†] David Glass,[†] Samuel Grossberg,[†] Anita Gupta,[†] Latanya Hammonds-Odie,[†] Aaron Hoisos,[†] Jenny Hsi,[†] Yu-Han Huang Hsu,[†] Sachi Inukai,[†] Konrad J. Karczewski,[†] Xiaobo Ke,[†] Mina Kojima,[†] Samuel Leachman,[†] Danny Lieber,[†] Anna Liebowitz,[†] Julia Liu,[†] Yufei Liu,[†] Trevor Martin,[†] Jose Mena,[†] Rosa Mendoza,[†] Cameron Myhrvold,[†] Christian Millian,[†] Sarah Pfau,[†] Sandeep Raj,[†] Matt Rich,[†] Joe Rokicki,[†] William Rounds,[†] Michael Salazar,[†] Matthew Salesi,[†] Rajani Sharma,[†] Sanford Silverman,[†] Cara Singer,[†] Sandhya Sinha,[†] Max Staller,[†] Philip Stern,[†] Hanlin Tang,[†] Sharon Weeks,[†] Maxwell Weidmann,[†] Ashley Wolf,[†] Carmen Young,[†] Jie Yuan,[†] Christopher Crutchfield,[†] Megan McClean,[†] Coleen T. Murphy,[†] Manuel Llinás,[†] David Botstein,[†] Olga G. Troyanskaya,[1,†,‡] and Maitreya J. Dunham[1,§]

*Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada, [†]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, [‡]Department of Computer Science, Princeton University, Princeton, New Jersey 08540, and [§]Department of Genome Sciences, University of Washington, Seattle, Washington 98195

ABSTRACT Whole-genome sequencing, particularly in fungi, has progressed at a tremendous rate. More difficult, however, is experimental testing of the inferences about gene function that can be drawn from comparative sequence analysis alone. We present a genome-wide functional characterization of a sequenced but experimentally understudied budding yeast, *Saccharomyces bayanus* var. *uvarum* (henceforth referred to as *S. bayanus*), allowing us to map changes over the 20 million years that separate this organism from *S. cerevisiae*. We first created a suite of genetic tools to facilitate work in *S. bayanus*. Next, we measured the gene-expression response of *S. bayanus* to a diverse set of perturbations optimized using a computational approach to cover a diverse array of functionally relevant biological responses. The resulting data set reveals that gene-expression patterns are largely conserved, but significant changes may exist in regulatory networks such as carbohydrate utilization and meiosis. In addition to regulatory changes, our approach identified gene functions that have diverged. The functions of genes in core pathways are highly conserved, but we observed many changes in which genes are involved in osmotic stress, peroxisome biogenesis, and autophagy. A surprising number of genes specific to *S. bayanus* respond to oxidative stress, suggesting the organism may have evolved under different selection pressures than *S. cerevisiae*. This work expands the scope of genome-scale evolutionary studies from sequence-based analysis to rapid experimental characterization and could be adopted for functional mapping in any lineage of interest. Furthermore, our detailed characterization of *S. bayanus* provides a valuable resource for comparative functional genomics studies in yeast.

ANALYSIS of the genome sequences of related species has provided tremendous insight into the key functional elements of genomes as revealed by patterns of DNA sequence conservation. The *Saccharomyces* yeasts have been particularly well sampled by sequencing projects over the past decade (reviewed in Dujon 2010), and comparative analyses have revealed a history of gene duplication (Dietrich *et al.* 2004; Kellis *et al.* 2004), conservation at DNA binding sites (Cliften *et al.* 2003; Kellis *et al.* 2003), and coevolution of binding sites with regulators (Gasch *et al.* 2004). However, to enable more thorough understanding of the underlying biology, sequence-based studies must be

complemented by the experimental study of functional divergence. Within *Saccharomyces cerevisiae*, comprehensive analysis of gene expression, protein levels, and metabolite levels demonstrates the ability of gene expression rather than raw sequence data to predict phenotype (Guan *et al.* 2008). In the yeasts, studies of promoter usage (Borneman *et al.* 2007), transcription factor binding (Doniger *et al.* 2005), stress sensitivity (Kvitek *et al.* 2008), transcriptional network changes (Tsong *et al.* 2006; Tuch *et al.* 2008), mating (Zill and Rine 2008), replication timing (Muller and Nieduszynski 2012), protein levels (Khan *et al.* 2012), and nucleosome occupancy (Guan *et al.* 2011; Tsankov *et al.* 2010) demonstrate that interesting evolutionary features emerge when processes are compared in detail within these eukaryotes.

Despite this foundational work, no studies have yet attempted to experimentally characterize gene function on a systematic scale in nonmodel newly sequenced species. An ideal study of gene function in a new species would establish precise functions for all species-specific genes and allow a systematic comparison of gene function and regulation for orthologs between species. Such a study can form the groundwork for connecting functional and regulatory differences to the sequence variants that have accumulated over evolutionary time. Conversely, genes with conserved function and regulation can be used to infer DNA sequence changes that are either neutral or that coevolved to maintain the selected characters. Gene-expression analysis fits these requirements, as genes of shared functions are highly correlated in their expression, and, conversely, gene-expression correlations are highly predictive of gene function (Stuart *et al.* 2003; van Noort *et al.* 2003; Hibbs *et al.* 2007; Huttenhower *et al.* 2007).

Limited comparative analyses of gene expression among different species have already been attempted and show how rapidly networks can evolve (reviewed in Whitehead and Crawford 2006). Comparisons between extremely divergent systems can discover core pathways shared over vast evolutionary differences (Stuart *et al.* 2003; Bergmann *et al.* 2004), while focusing on species that are less diverged permits study of more rapidly adapting processes and facilitates identification of the specific sequence changes that might be driving these differences. Furthermore, observing a phenomenon in multiple species provides solid evidence that it is not specific to a laboratory-adapted model organism but is instead an evolutionarily conserved biological response (Hess *et al.* 2006; Zill and Rine 2008; Airoldi *et al.* 2009).

To examine the conservation and divergence of gene function, we selected the yeast *Saccharomyces bayanus* var. *uvarum* (henceforth referred to as *S. bayanus* for simplicity) for comparison with *S. cerevisiae*. The two species diverged ~20 million years ago and have a comparable level of DNA sequence divergence as mouse and human (80% conserved in coding regions and 62% conserved in intergenic regions as compared to *S. cerevisiae*). We have recently used next-generation sequencing to create a high-quality assembly and gene model prediction of the *S. bayanus* genome, and we created an extendable genome browser to facilitate its use (Scannell *et al.* 2011). Importantly, sequence conservation of functional elements is still detectable (for example, noncoding RNAs; Kavanaugh and Dietrich 2009). Like *S. cerevisiae*, *S. bayanus* is a species used in winemaking, and recent studies of its genome content and relationship to lager yeasts have clarified taxonomic confusion (Libkind *et al.* 2011). The phylogenetic proximity and shared natural history with *S. cerevisiae* also make it possible to select specific experimental conditions for *S. bayanus* by reference to the vast literature available for *S. cerevisiae*, one of the most popular model organisms. The two species can make interspecific hybrids, allowing complementation tests with *S. cerevisiae* alleles. However, with a few exceptions (Serra *et al.* 2003; Talarek *et al.* 2004; Jones *et al.* 2008; Zill and Rine 2008; Gallagher *et al.* 2009; Zill *et al.* 2010), little experimental work has been performed in *S. bayanus* and even less at genome scale (Bullard *et al.* 2010; Tsankov *et al.* 2010; Busby *et al.* 2011; Guan *et al.* 2011; Muller and Nieduszynski 2012).

We first compared the basic growth characteristics of the two species and developed genetic tools and protocols to facilitate experimental manipulations of *S. bayanus*. Following this characterization of the species, we then produced a gene-expression compendium of over 300 microarrays in *S. bayanus*, guided by a machine-learning analysis of the entire *S. cerevisiae* literature that predicts an optimal set of conditions for expression analysis (Guan *et al.* 2010), and assembled a set of published expression experiments in *S. cerevisiae* for comparison. Similar to comparative sequence analysis, comparing the gene-expression responses of different species allows the identification of programs of conserved gene regulation and of alterations in gene-expression response. In comparing the *S. bayanus* and *S. cerevisiae* data, we noted a number of examples of divergence in gene expression between the species (Guan *et al.* 2013). Also, because genes of like function typically have correlated gene expression (Eisen *et al.* 1998), patterns of coexpression can be used to predict the functional roles of genes (Sharan *et al.* 2007).

Our analysis of these data sets reveals both regulatory change and evolution of gene function amid overall

doi: 10.1534/genetics.113.152918
Manuscript received June 5, 2013; accepted for publication July 4, 2013
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152918/-/DC1.
[1]Corresponding authors: University of Toronto, 160 College St., Room 308, Toronto, ON M5S3E1, Canada. E-mail: amy.caudy@utoronto.ca; Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544 and Department of Computer Science, Princeton University, Princeton, New Jersey 08540. Email: ogt@genomics.princeton.edu; Department of Genome Sciences, Foege Bldg. S403B, Box 355065, 3720 15th Ave. NE, Seattle, WA 98195-5065. E-mail: maitreya@uw.edu.
[2]Present address: Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI 48109

conservation. Specific examples include expression rewiring in the pathways controlling meiosis and galactose utilization, oxidative stress driving expression of a species-specific network, and evidence for divergence of specific functional groups.

Materials and Methods

The strains used in this study are described in Supporting Information, Table S1. Custom oligonucleotide probes specific for *S. bayanus* genes were designed and printed using a pin-style arraying robot. *S. bayanus* cells were grown and exposed to a variety of stimuli and RNA was harvested and labeled by direct incorporation of fluorescent nucleotides into cDNA. Deletion and insertion mutants were produced in diploids by homologous recombination using adaptations of standard methods for *S. cerevisiae* and haploids were obtained by sporulation and dissection. *S. bayanus* data and a compendium of *S. cerevisiae* data were processed for gene function prediction using support vector machines. As there were no existing biological process annotations in *S. bayanus*, we adopted the annotations from *S. cerevisiae* for training.

The microarray expression data are available from GEO as GSE16544 and GSE47613. The interactive network view of the expression data and searchable prediction results are available at http://bayanusfunction.princeton.edu. Complete methods information is included as File S1.

Results

Developing *S. bayanus* into a new model system required an initial characterization of its growth habits and preferences, along with the development of genetic tools to enable the types of studies that are routine in established model systems.

Phenotypic analysis and genetic tools

We began our work in *S. bayanus* by measuring its growth and physiology. As previously reported (Goncalves *et al.* 2011; Salvado *et al.* 2011), in minimal media at 20°, *S. bayanus* grows faster than *S. cerevisiae* (Figure 1A). The species grew at nearly equal rates at 25°, and at 30° *S. bayanus* grew more slowly than *S. cerevisiae* (Figure 1, B and C). Accordingly, *S. bayanus* was more sensitive to heat shock than *S. cerevisiae*; transfer to 40° slowed growth of *S. bayanus* more than it did in *S. cerevisiae* (Figure 1, D and E). This heat sensitivity precludes efficient lithium acetate transformation using heat shock at 42°, so we modified our procedure to use a milder 37° heat shock for *S. bayanus*.

When grown on glucose medium to the point of glucose depletion, *S. bayanus* underwent a diauxic shift marked by a growth arrest followed by a shift to ethanol consumption and a slower growth rate (Figure 1F), consistent with its natural history and qualitatively similar to the behavior of *S. cerevisiae*. We also measured the growth inhibition by a variety of transition metals, salts, and oxidants (Figure 1G). The survival of *S. bayanus* and *S. cerevisiae* was similar during starvation for the essential nutrients sulfate and phosphate (Figure 1H). Finally, we analyzed our *S. bayanus* strain for the presence of the 2μ plasmid and observed that it does not carry detectable levels of the plasmid, although a hybrid with *S. cerevisiae* prepared in our laboratories maintains this DNA element (Figure 1I).

We constructed a Tn7 insertion library (Kumar *et al.* 2004) to create a collection of *S. bayanus* mutant strains. We built a Tn7 transposon carrying a ClonNat resistance marker selectable in both bacteria and yeast. The transposed marker carries stop codons in all reading frames near both termini and so is expected to produce truncations when inserted within genes. Our library contained ~50,000 unique genomic insertions, and we have used it to screen for a variety of phenotypes including auxotrophies, drug resistance, and copper resistance (see below). By transforming the library into *MAT*α strains and using a ClonNat resistance marker, mutants isolated from this Tn7 set can be used directly in complementation assays by mating to *S. cerevisiae* strains from the widely used *MAT***a** deletion set that carries complementary G418 drug resistance. Insertion mutations can also be mapped using microarray or sequencing technologies (see below). We expect that this mutant collection will be a valuable resource for mutation screening in this new species.

Gene-expression data set

Just as lessons learned from early whole-genome sequencing projects led to more efficient sequencing of related genomes in subsequent projects, we can leverage the thousands of microarray experiments performed in the yeast *S. cerevisiae* to direct efficient expression profiling in a related organism. Given the shared history of these species, we reasoned that experiments with high predictive value of gene function in *S. cerevisiae* were also likely to be useful in related yeasts. We also assumed that most of these treatments were likely to target similar ranges of functional categories in the two species. With these ideas in mind, we developed a data-driven experiment recommendation system to identify the minimal set of maximally informative experiments for functional characterization of the *S. bayanus* genome based on the *S. cerevisiae* gene-expression literature (Guan *et al.* 2010).

We carried out 304 microarray measurements in 46 experimental manipulations (detailed in Table S2). Because of the many practical similarities with *S. cerevisiae*, the experiments were effectively prototyped for us by their original *S. cerevisiae* publications, in many cases needing only minor modification to adapt them for *S. bayanus*. Our computationally selected treatments perturbed the majority of the genes in the cell: 4828 of the 4840 *S. bayanus* genes measured by our array show twofold or greater change in at least one treatment.

Hierarchical clustering of this *S. bayanus* gene-expression compendium revealed a number of groups of genes coexpressed under a variety of conditions (Figure 2, numerical data in Table S3). Although clustering was performed solely

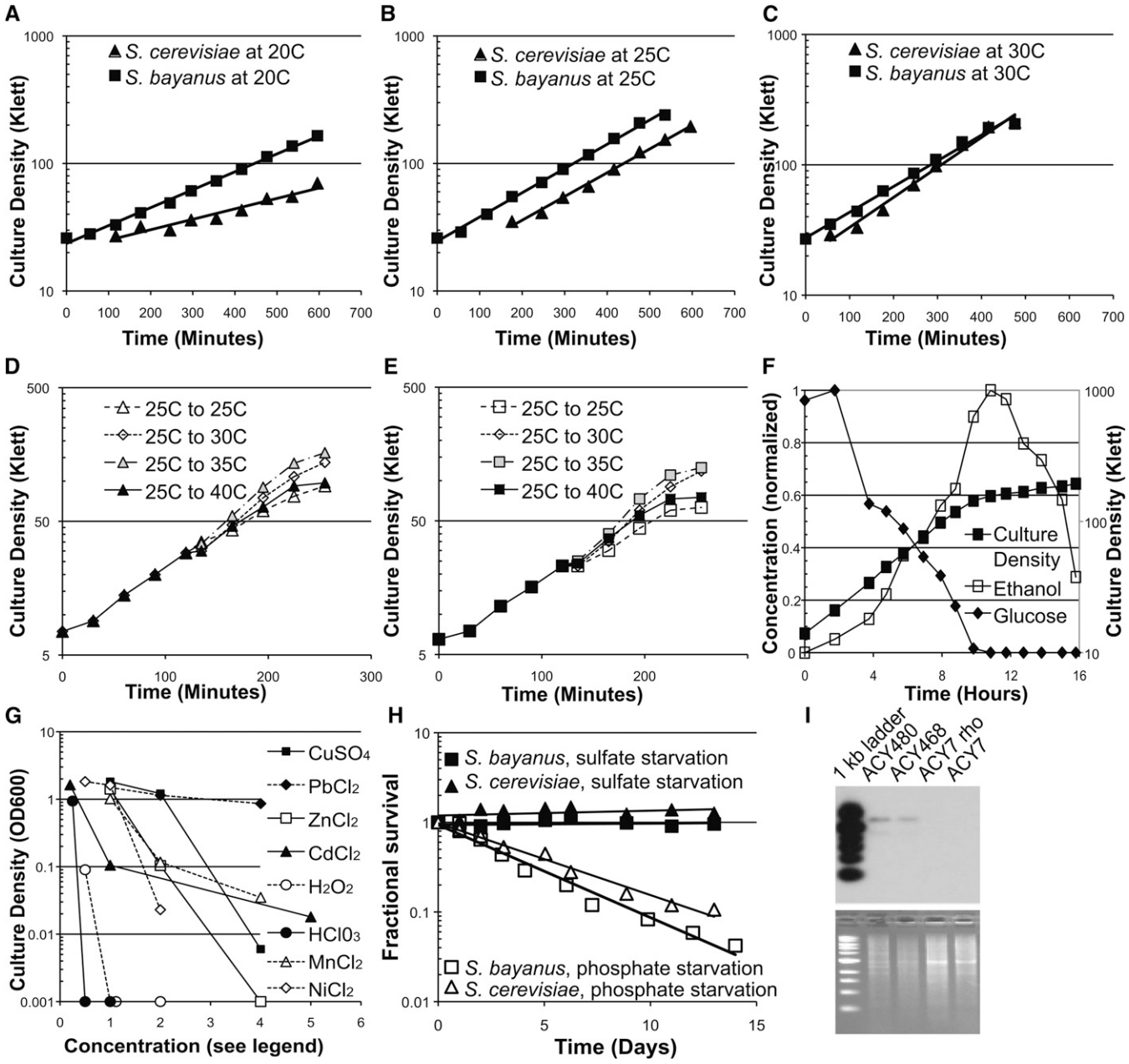


Figure 1 Growth, physiology, and molecular characteristics of *S. bayanus*. (A–C) Growth of *S. cerevisiae* and *S. bayanus* at different temperatures. (D) Temperature shifts in *S. bayanus*. Cultures were shifted at the 120-min timepoint. (E) Temperature shifts in *S. cerevisiae*, as in D. (F) Diauxic shift in *S. bayanus*. Culture density, ethanol concentration, and glucose concentration were monitored. (G) Response of *S. bayanus* to stress. Culture density was measured as absorbance at 600 nm in an overnight culture. Concentrations are in millimolar, except for cadmium (10× μM), ammonium (M), and bleach (%). (H) Survival during nutrient starvation of *S. bayanus* and *S. cerevisiae*. Strains were growth to saturation in chemostat medium with the indicated limiting nutrient (phosphate or sulfate), and viability was measured over time. (I) 2μ plasmid is absent in a pure *S. bayanus* strain but present in a hybrid. Genomic DNA was restriction digested and hybridized with a probe corresponding to the 2μ plasmid.

on the *S. bayanus* data and was not informed by the evolutionary relationships between *S. bayanus* and *S. cerevisiae* genes, we noted many groups of *S. bayanus* genes nevertheless showing expression patterns similar to those in *S. cerevisiae*. Most strikingly, two large cohorts of genes responded coordinately to multiple stresses, with one group repressed and the other induced. This large-scale response indicates that *S. bayanus* shows the canonical environmental stress response identified in *S. cerevisiae* (Gasch *et al.* 2000) and other yeasts (Gasch 2007). Other treatments elicited gene-expression responses from smaller groups of genes. For instance, a group of genes was strongly upregulated in response to alpha-factor pheromone. This pheromone response declined as cells were released from alpha-factor arrest into the cell cycle. As another example, two other groups of genes were expressed periodically

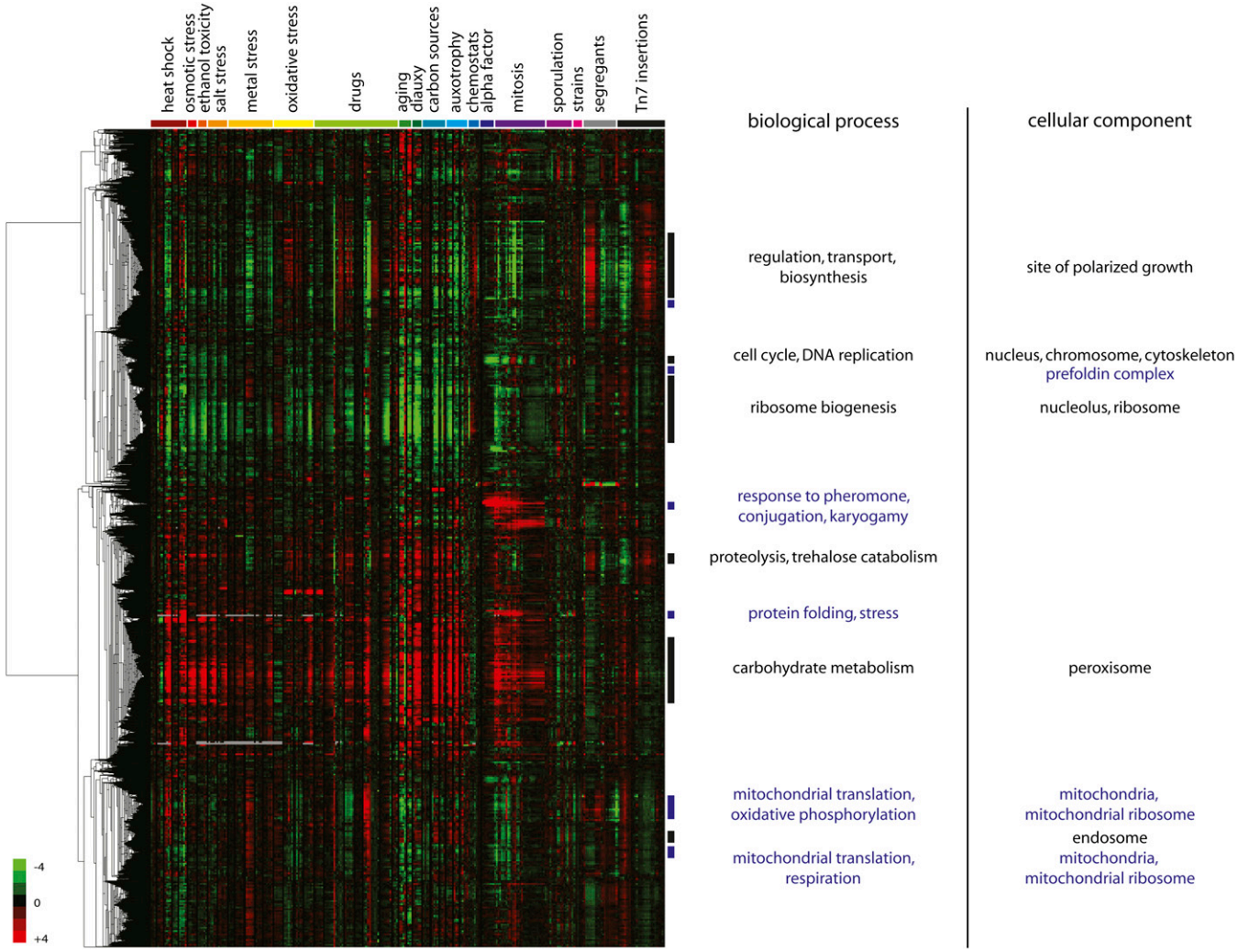


Figure 2 *S. bayanus* gene-expression megacluster. Forty-six *S. bayanus* gene-expression data sets are shown as indicated by color-coded experiment labels with genes hierarchically clustered along the other dimension. Each data set was either zero transformed or mean centered to remove the reference. By assuming that all *S. bayanus* genes carry the annotations of their *S. cerevisiae* orthologs, statistically significant GO term enrichments for clusters of 25 or more genes with a correlation coefficient above 0.7 were determined using the program GOTermFinder, using a background distribution of only orthologous genes. These are indicated with vertical bars and labeled with related terms collapsed for simplicity. Data sets are ordered to group similar conditions. The data as plotted are available in Table S3.

during the cell cycle with different phases of peak gene expression.

As an initial test of whether these expression clusters reflect functional gene groupings in both species, we started with the simplest—and almost certainly incorrect—assumption that all genes in *S. bayanus* have the same functions as their orthologs in *S. cerevisiae*. Using these inferred annotations, we calculated the Gene Ontology (GO) term enrichment for correlated clusters, and we observe significant enrichment for genes of like biological process and cellular component among the clusters of genes with coherent expression (Figure 2). Further, the expression patterns in these clusters showing compartment-specific or biological process enrichment are consistent with the expression patterns of genes involved in the same biological process in similar *S. cerevisiae* experiments. For instance, the cluster of genes activated by mating pheromone was enriched for genes whose *S. cerevisiae* orthologs have experimentally validated roles in response to pheromone, conjugation, and karyogamy.

Gene-expression patterns diverge in subtle ways

Although many aspects of gene expression are conserved, we noted a number of instances of gene-expression patterns different from those observed in *S. cerevisiae* orthologs in response to similar treatments. In *S. cerevisiae*, the galactose metabolism genes were induced only to detectable levels in the presence of galactose (Gasch *et al.* 2000). However, in *S. bayanus*, the orthologs of the galactose structural genes *GAL1*, *GAL10*, *GAL7*, and *GAL2* were detectably induced not only when cells were exposed to galactose, but also when cells were switched from glucose to other less-preferred

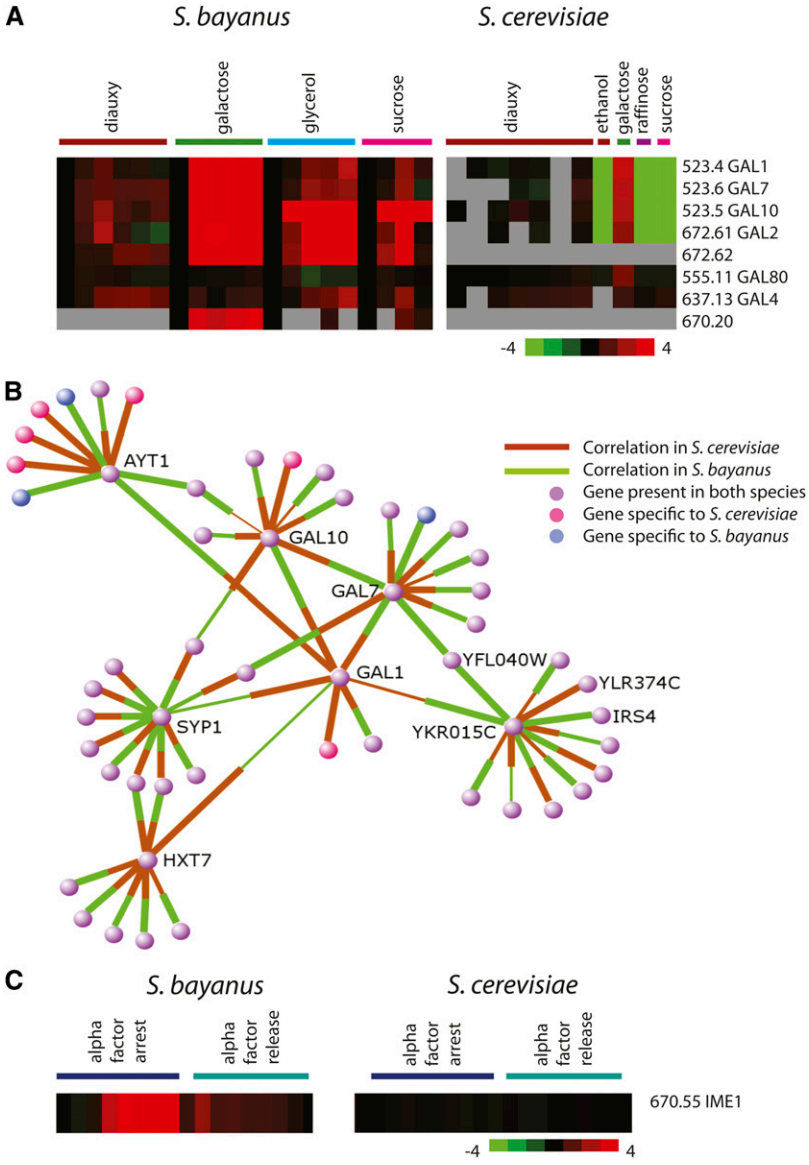


Figure 3 Altered expression of genes in *S. bayanus*. (A) Galactose structural genes are induced by multiple carbon sources in *S. bayanus* but not in *S. cerevisiae*. *S. cerevisiae* carbon source data (Gasch *et al.* 2000) and diauxic shift data (Brauer *et al.* 2005) are as published. (B) A network view of gene-expression correlations with *GAL1* comparing the pairwise expression correlation over all data in both species. The nodes indicate genes, and the thickness of lines indicates expression correlation. In cases where expression correlation is negative, no line is shown for that species, and the target gene is labeled. The node color indicates whether each gene is present in one species (pink or blue) or both (purple). An interactive network viewer for all genes is available at http://bayanusfunction.princeton.edu. (C) Expression of *SbayIME1* and *ScerIME1* are compared under conditions of alpha-factor arrest and alpha-factor release. *S. cerevisiae* alpha-factor arrest (Roberts *et al.* 2000) and release (Pramila *et al.* 2006) are as published.

carbon sources including ethanol, raffinose, sucrose, and glycerol (Figure 3A). The derepression of galactose metabolism genes on nonglucose carbon sources has been previously described in detail in *S. cerevisiae* (Matsumoto *et al.* 1981; St John and Davis 1981; Yocum *et al.* 1984), but the magnitude of this increase in gene expression on nonglucose carbon sources is much greater in *S. bayanus*. We verified this expression difference between *S. bayanus* and *S. cerevisiae* using quantitative PCR for *GAL1* (Figure S1). This activation of the galactose structural genes by multiple carbon sources suggests that *S. bayanus* might have evolved in an environment in which galactose becomes available at the same time as other nonglucose carbon sources.

We created a resource that presents a network view comparing gene expression between *S. cerevisiae* and *S. bayanus* (http://bayanusfunction.princeton.edu). The gene-expression network around *GAL1* showed that *GAL1*, *GAL10*, and *GAL7* have a correlation of 0.99 in both species under all expression conditions (Figure 3B). However, the correlation of the *GAL* genes with other genes revealed differences in regulation between species. For instance, the ortholog of the hexose transporter *HXT7* had a correlation of 0.98 with the galactose genes in *S. bayanus* because this and other hexose transporters were upregulated whenever glucose was low. In contrast, in *S. cerevisiae* the correlation between *HXT7* and *GAL1* was only 0.19 because *HXT7* was upregulated in response to declining glucose concentration while *GAL1* was not.

Transcription factors as a group showed higher than expected divergence in expression between *S. bayanus* and *S. cerevisiae*, and the *S. bayanus* ortholog of *IME1* (*670.55*, which we will refer to as *SbayIME1*) in particular showed exceptions to the diploid-specific expression observed in *S. cerevisiae*. In *S. cerevisiae*, *IME1* expression is primarily limited to diploid cells (Kassir *et al.* 1988), but in haploid *MAT***a** *S. bayanus*, *SbayIME1* was induced over 10-fold by

alpha-factor pheromone (Figure 3C). As observed in *S. cerevisiae*, *SbayIME1* is required for sporulation (data not shown), and although *SbayIME1* was strongly induced by alpha factor we did not observe significant changes in the pheromone response of *Sbayime1* mutant cells (Figure S2). Chromatin immunoprecipitation experiments observed twofold higher levels of the pheromone response transcription factor SbaySte12 (570.3) at the *SbayIME1* promoter as compared to Ste12 occupancy at the *IME1* promoter in *S. cerevisiae* (Borneman *et al.* 2007), supporting our observation of differential pheromone activation of *SbayIME1* in *S. bayanus* as compared to *ScerIME1*. In *S. cerevisiae* Ime1 is subject to translational regulation (Sherman *et al.* 1993), and the lack of an effect on transcription in response to pheromone in the *Sbayime1* mutant could similarly be explained by post-transcriptional regulation. *IME1* has been observed to be under selective pressure in *S. cerevisiae* (Gerke *et al.* 2009), and the altered expression here may suggest that it is evolving to take on additional roles.

S. bayanus gene function predictions via machine learning are confirmed by mutational analysis

By comparing gene expression between orthologs under known conditions we were able to find examples of changes in gene expression and use these changes to infer functional differences between species. Such inferences are limited by existing knowledge of the link between expression and biological function and by the availability of directly comparable data sets in both species. These limits can be overcome using computational interpretation of expression data, which accurately predicts gene function over much larger data sets than a human can process (Huttenhower and Troyanskaya 2008). Using a support vector machine (SVM) learning method trained using the GO biological process annotations of *S. cerevisiae* orthologs, we predicted the functional roles of *S. bayanus* genes (Table S4).

Many gene functions are preserved over vast evolutionary distance, as evidenced by the many examples of mammalian genes that can complement deletion mutations in yeast (reviewed in Osborn and Miller 2007). Accordingly, we found that many genes were predicted to have the same function in *S. bayanus* and *S. cerevisiae* even though the SVM does not reference protein sequence homology while making predictions. For example, we predicted a role in oxidative phosphorylation for *643.11*, the ortholog of *RPM2*, the mitochondrial RNAseP required for processing mitochondrial tRNAs from transcripts. Consistent with this prediction, an insertion mutant in *SbayRPM2* was respiratory deficient (Figure S3). Similarly, we predicted a role in cell morphogenesis for *678.66*, the ortholog of *AMN1*. A knockout mutant of *Sbayamn1* lost daughter cell adhesion ("clumpiness," Figure S4), as has been observed for the *amn1* deletion allele in *S. cerevisiae* (Yvert *et al.* 2003). As a third example, we predicted a role for telomeric silencing and protein acetylation for *668.17*, the ortholog of the protein acetyltransferase *ARD1*. In a *MAT***a** insertion mutant of *Sbayard1*, we observed repression of *MAT***a** haploid-specific genes, as reported for *ard1* mutants (Whiteway *et al.* 1987) (Figure S5A) and note that the mutation causes genome-wide expression changes (Figure S5B). For the whole-genome duplicate serine/protein kinases *642.24* (*DBF2*) and *636.21* (*DBF20*), we predicted roles in the regulation of mitosis and the regulation of DNA damage checkpoints, similar to the established roles of the *S. cerevisiae* orthologs in regulating cytokinesis and mitotic exit. As in *S. cerevisiae*, mutations in these genes are synthetic lethal (data not shown).

The functional predictions also can predict gene functions not yet known in *S. cerevisiae*. We carried out a screen for Tn7 mutants resistant to copper sulfate and identified a resistant mutant (Figure 4, A and B). Using an array-based method (Gabriel *et al.* 2006), we mapped the insertion upstream of *610.13*, the ortholog of *OPT1* (Figure 4C). Deletion analysis of *SbayOPT1* and the divergently transcribed neighboring gene *SbayPEX2* (*610.12*) confirmed that mutation of *SbayOPT1* was responsible for resistance to copper (Figure 4D). The functional predictions for *SbayOPT1* include cation homeostasis, the GO parent term that includes copper ion homeostasis (our functional predictions did not include GO terms with few members). *ScerOPT1* (also named *HGT1*) has been characterized as a high-affinity glutathione transporter induced by sulfur starvation (Bourbouloux *et al.* 2000; Srikanth *et al.* 2005). Copper resistance had not been investigated in this mutant, although sensitivity to cadmium had been noted (Serero *et al.* 2008). The *OPT1* mutant in *S. cerevisiae* also showed increased resistance to copper (Figure 4E). Of note, *S. bayanus* is more sensitive to copper than the laboratory strain of *S. cerevisiae*; our screen in the sensitized background of *S. bayanus* likely provided added sensitivity to detect genes involved in the response to copper (Figure 4E). These results suggest the potential for a relationship between glutathione transport and copper resistance and demonstrate how the predictions of gene function in *S. bayanus* provide information about conserved gene function in *S. cerevisiae*.

Different rates of functional divergence characterize different gene groups

Just as genes involved in different biological pathways have been observed to evolve at the sequence level at different rates (Aris-Brosou 2005; Wolf *et al.* 2006), certain classes of genes may show more rapid functional divergence. We examined our predictions of gene function in both species and identified cases in which a pair of orthologs showed very large changes in predicted function between species (Table 1, full data in Table S5). We observed the smallest number of changes in ribosomal biogenesis and in electron transport, and many core metabolic processes showed few changes, consistent with these genes' typical conservation at the sequence level. Processes showing the highest amount of change included response to osmotic stress, autophagy, and organelle inheritance. Although it was not immediately obvious why these processes are changing so quickly, these results

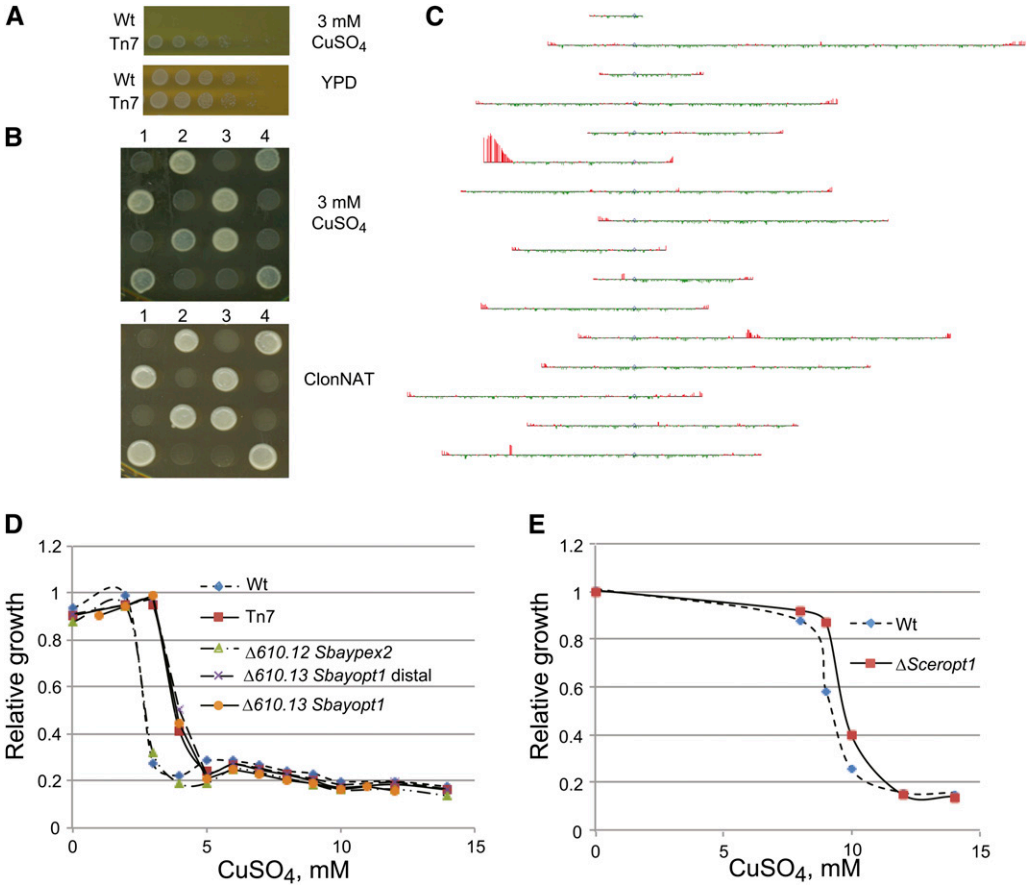


Figure 4 *610.13/OPT1* mediates resistance to copper toxicity in *S. bayanus* and *S. cerevisiae*. (A) A Tn7 insertion mutant was identified in a screen for mutants resistant to copper sulfate; a series of dilutions of cells were plated on YPD and YPD with 3 mM $CuSO_4$. (B) The resistance phenotype cosegregates with the ClonNAT resistance marker carried by the transposon; the mutant strain was backcrossed to wild type, and tetrads (in columns, indicated by numbers) were phenotyped for resistance to ClonNAT and for growth on YPD with 3 mM $CuSO_4$. (C) The site of the insertion was mapped by enriching genomic DNA for transposon DNA and using an array hybridization technique (Gabriel *et al.* 2006). Data are mapped onto the chromosomes, which are aligned by the centromeres. Subsequent PCR amplification using primers specific to the transposon and flanking regions mapped the insertion site between the genes *SbayPEX2 (610.12)* and *SbayOPT1 (610.13)*. (D) Mutation of the *S. bayanus* gene *SbayOPT1* confers copper resistance; *SbayOPT1* is divergently transcribed from *SbayPEX2*, so a deletion of only the 3′ distal portion of *SbayOPT1* was also tested to exclude any effect on *SbayPEX2*. (E) Mutation of the *610.13* ortholog *OPT1* in *S. cerevisiae* confers copper resistance.

will help to guide future experiments. We also observed significant change in small GTPase mediated signal transduction and hypothesize that this may reflect the constitutive signaling through the mating pathway caused by a mutation common in laboratory strains of *S. cerevisiae* (Lang *et al.* 2009) not present in the *S. bayanus* strains used here.

Annotations for species-specific genes

Genome sequence analysis allows comparison of gene content in different species, which can suggest the evolutionary pressures that shape specific lineages (Gordon *et al.* 2009). Similarly, examining the functional roles predicted for genes found in one species but not another can suggest potential functions for these species-unique genes, revealing species-specific adaptations. We examined the expression data of *S. bayanus* genes that do not have orthologs in *S. cerevisiae* and found a prominent cluster of 25 genes that includes 13 genes specific to *S. bayanus*—including 8 with no orthologs in any surveyed yeast (Gordon *et al.* 2009) (Figure 5A). These genes were induced 16- to 32-fold by peroxide stress, bleach, and MMS but not other stresses or any other conditions tested in our compendium. Peroxide, bleach, and MMS all increase reactive oxygen levels (Winter *et al.* 2008; Kitanovic *et al.* 2009), so we propose that this group of genes responds specifically to oxidative stress. Two DNA sequence motifs are enriched in the promoters of the *S. bayanus* genes in this cluster, and these motifs are very similar (Table S6, $P < 7 \times 10^{-5}$, Mahony *et al.* 2007) to motifs established by analysis of sequence conservation among the *sensu stricto* yeasts (Kellis *et al.* 2003). Furthermore, one of the motifs is similar to that of *S. cerevisiae CAD1* (Harbison *et al.* 2004), a transcription factor with a role in stress response (Wu *et al.* 1993). As the *CAD1* ortholog in *S. bayanus* has been annotated as a pseudogene (Scannell *et al.* 2011), it is likely that some other transcription factor may be activating these genes. The stress-responsive gene *YAP1* has a similar binding site in *S. cerevisiae* and is a candidate for the oxidative stress activation we observe. The number of genes specific to *S. bayanus* annotated to oxidative stress suggests that *S. bayanus* may encounter a different spectrum of stresses.

Our functional predictions for genes in our oxidative stress cluster included response to toxin (GO:0009636), sulfur metabolic process (GO:0006790), and response to temperature stimulus (GO:0009266) (Figure 5A). Many of these functions have been demonstrated for the 12 genes that have *S. cerevisiae* orthologs, and 10 of the 12 *S. cerevisiae* orthologs are induced by hydrogen peroxide (Gasch *et al.*

Table 1 Most conserved and diverged biological processes

GO term	Fraction genes changing functional prediction rank by 75% or more
Least change	
GO:0016072(rRNA metabolism)	0.02
GO:0006118(electron transport)	0.03
GO:0006508(proteolysis and peptidolysis)	0.04
GO:0007001(chromosome organization and biogenesis)	0.05
GO:0006807(nitrogen compound metabolism)	0.06
Greatest change	
GO:0006914(autophagy)	0.31
GO:0006970(response to osmotic stress)	0.32
GO:0007264(small GTPase mediated signal transduction)	0.34
GO:0048308(organelle inheritance)	0.43
GO:0007568(aging)	0.46

The top and bottom five GO SLIM biological process terms for changes in gene function. The fraction of genes with change in ranked prediction scores of 75% or more is shown for each term.

2000; Causton *et al.* 2001). Five of the *S. cerevisiae* orthologs of this cluster have been assigned the GO biological process of response to toxin (GO enrichment, $P < 4.07 \times 10^{-9}$, Bonferroni corrected), and two of the *S. cerevisiae* orthologs in this cluster have roles in sulfur metabolism: *GTT2* is a glutathione *S*-transferase, and *YCT1* is a cysteine transporter. The predicted role in toxin response is consistent with the activation by oxidative stress, because in *S. cerevisiae*, genes assigned to this biological process are induced by the mycotoxin citrinin, which causes oxidative stress (Iwahashi *et al.* 2007). Also, the sulfur metabolic process includes genes involved in sulfur assimilation, a biochemical process that consumes reducing equivalents. Of the 12 proteins in this cluster that have *S. cerevisiae* homologs, 5 are proteins of unknown function. These functional predictions from *S. bayanus* may help to inform functional experiments on the *S. cerevisiae* orthologs.

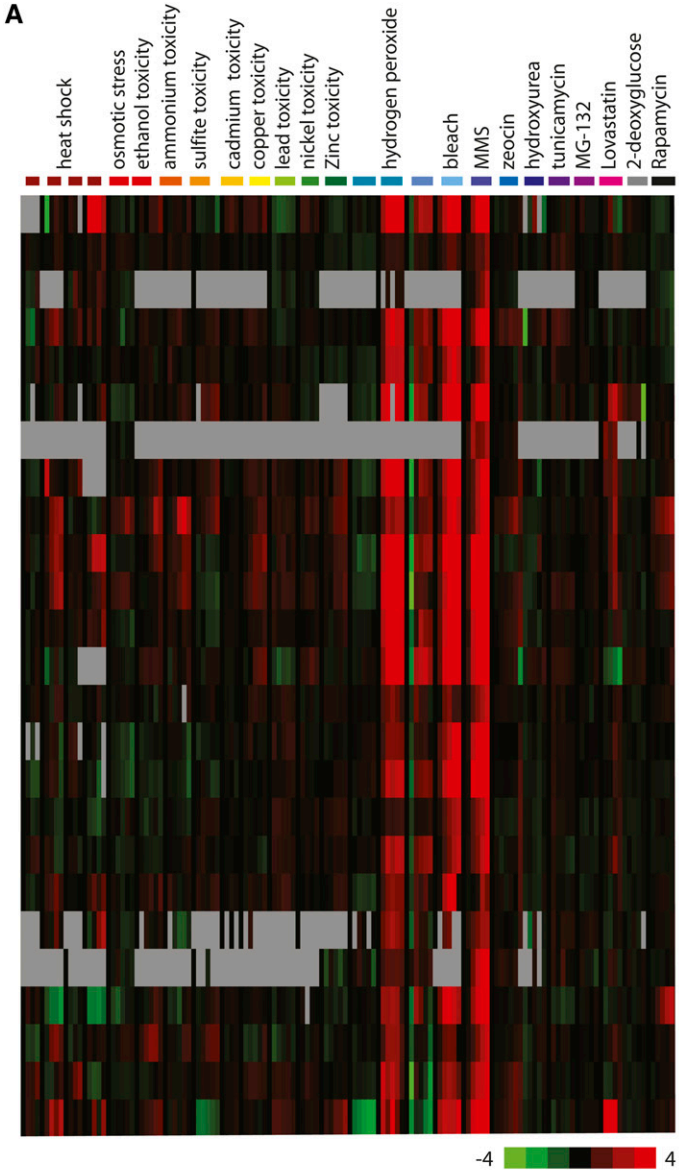
Gene duplicates are known to play a prominent role in yeast genome evolution. Among our functional predictions for the *S. bayanus* genome, we examined the seven genes present in duplicate in *S. bayanus* but not in *S. cerevisiae* and noted that our expression data had yielded a prediction of a role in galactose metabolism for one of these genes (Table S4), which had also been previously noted on the basis of comparative homology (Hittinger *et al.* 2004, 2010; Cliften *et al.* 2006; Gordon *et al.* 2009; Scannell *et al.* 2011). Both duplicates of the ancestral *GAL80* gene are retained in *S. bayanus*, but only *GAL80* is present in *S. cerevisiae*. The *S. bayanus GAL80* ortholog *555.11* retains its function as a repressor of galactose genes, as GAL genes were no longer repressed when *Sbaygal80* mutant cells were grown in glucose (Figure S6), a derepression known in *Scergal80* mutants (Douglas and Hawthorne 1966; Yocum and Johnston 1984). In addition, *670.20*, the ohnolog of *SbayGAL80*, which itself has no ortholog in *S. cerevisiae*, was predicted to function in galactose metabolism by our SVM. Indeed, we observed activation of *670.20* in response to galactose (Figure 3A), and Gal4 binding sites are present upstream of the gene. The galactose-specific activation of *670.20* differs from the response of the other *S. bayanus* GAL family genes, which are activated by growth on multiple nonglucose carbon sources. We also noted that *670.20* was derepressed in the *Sbaygal80* mutant, as were other GAL genes (Figure S6).

To more directly study the role of *670.20* in galactose metabolism, we measured the gene-expression response of *670.20* mutants to a shift from raffinose to galactose and observed a set of genes that failed to be activated by galactose in the *670.20* mutant (Figure 5B). These four genes are also members of the oxidative stress cluster shown in Figure 4. Notably, the genes regulated by the *S. bayanus*-specific *670.20* are themselves present only in *S. bayanus*, forming a species-specific network.

Discussion

Although the genomes of many nonmodel organisms are now sequenced, this flood of data has not been matched by functional experimental data in these species. Much of this can be attributed to the difficulty of working with unfamiliar organisms, but many other species lend themselves to laboratory study for comparative work. For example, the fly species sequenced by the 12 *Drosophila* species consortium (Consortium *et al.* 2007) can all be lab reared, as can several sequenced species of nematodes (Cutter *et al.* 2009). Yeast are of course another taxa with many lab-amenable species.

Using gene-expression data we functionally annotated all the genes in *S. bayanus* (Table S4) and demonstrated the accuracy of our predictions using targeted mutational analysis. A sufficiently complex gene-expression data set can be used not only to compare strategies of gene regulation but also to predict biological function (Guan *et al.* 2008). Identifying regulatory changes across different species provides interesting insight into selection and adaptation. For



S. bayanus gene	*S. bayanus* Inferred Biological Process Annotation	*S. cerevisiae* ortholog	*S. cerevisiae* Biological process Annotations
658.54	vitamin metabolism, sulfur metabolism	OYE3	response to toxin
380.1	mRNA metabolism, vesicle-mediated transport		
584.3	vesicle-mediated transport, establishment of protein localization		
610.15	organic acid transport, response to inorganic substance		
601.6	sulfur metabolism, amine transport	YLL056C	response to toxin
673.33	sulfur metabolism, cofactor metabolism		
493.2	meiotic cell cycle, M phase		
601.1	cofactor metabolism, vitamin metabolism	GTT2	sulfur metabolic process,cofactor metabolic process, response to toxin
453.5	response to toxin, proteolysis and peptidolysis	YMR090W	
588.13	proteolysis and peptidolysis, cofactor metabolism		
602.5	response to toxin, transcription from RNA polymerase II promoter		
617.2	sulfur metabolism, ion homeostasis		
641.4	sulfur metabolism, response to toxin	AAD14	response to toxin, aldehyde metabolic process
490.5	cofactor metabolism, lipid metabolism	YGL114W	
640.6	response to temperature, transcription from RNA polymerase II promoter		
621.2	response to temperature, proteolysis and peptidolysis		
664.13	sulfur metabolism, proteolysis and peptidolysis	AIM2	
643.28	sulfur metabolism, amino acid and derivative metabolism		
601.7	sulfur metabolism, ion homeostasis	YCT1	amine transport, organic acid transport
626.1	cell wall organization and biogenesis, sporulation		
602.1	cofactor metabolism, M phase		
547.2	drug transport, response to DNA damage stimulus	BSC5	
518.16	vitamin metabolism, cofactor metabolism	MRS4	biogenesis, ion transport, mitochondrial transport, rna splicing
591.11	drug transport, transcription from RNA polymerase II promoter	YDR132C	
565.14	cofactor metabolism, proteolysis and peptidolysis	YLR346C	response to toxin

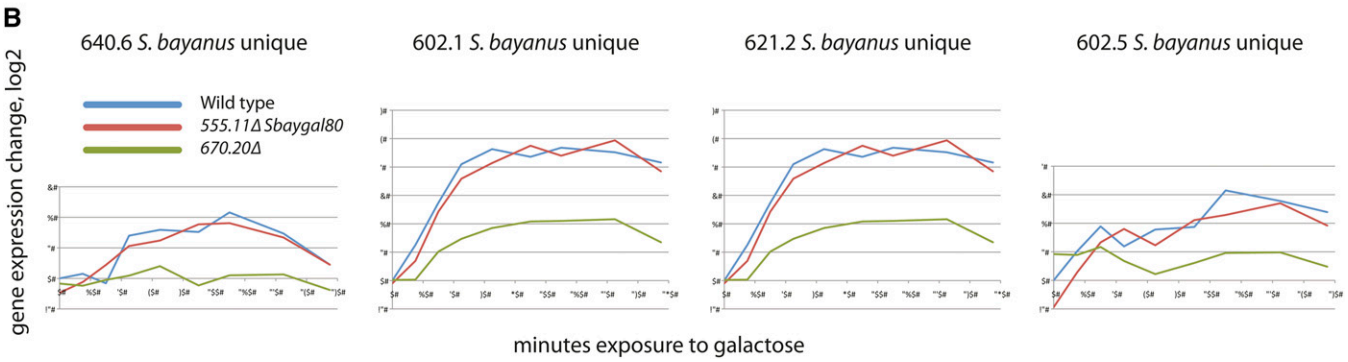


Figure 5 Functional roles of a set of genes specific to *S. bayanus*. (A) A group of genes is strongly induced by oxidative stresses but not other stresses. We predicted biological process annotations for this cluster of genes and show the *S. cerevisiae* orthologs and their experimentally based biological process annotations as assigned by the *Saccharomyces* Genome database. The predicted annotations are the two highest scoring annotations; Table S4 contains a complete list, for all genes and Figure S7 describes the quality of gene function predictions across all terms. (B) The *GAL80* ohnolog *670.20* regulates a set of genes unique to *S. bayanus* when cells are shifted from raffinose to galactose. Graphs show $\log_2$ ratios of expression data from microarrays, zero transformed to the initial timepoint in wild type cells.

instance, comparing the protein sequences encoded in bacterial genomes has helped to predict the metabolic capabilities of different lineages (Downs 2006). Our measurements of gene expression under well-characterized conditions directly relevant to defined biological processes illustrate examples of altered gene regulation that suggest functional differences between species.

Conversely, evidence of gene function in other species may be used to generate hypotheses about the functions of the orthologous genes of model systems, many of which still lack annotations (Peña-Castillo and Hughes 2007). Our study demonstrates the potential of computationally predicted annotations for both functional characterization and evolutionary analysis of new species.

The tools we have developed are generic and could easily be applied to other nonmodel organism species of interest. Application of our comparative approach to other groups of related species, such as *Candida* yeasts, *Drosophila* species, worms, or mammals, could extend the evolutionary observations made here. Since our experimental and analytical framework are agnostic to species and platform, they should be easily transferable to other systems. This new style of comparative functional genomics will ultimately allow better understanding of conservation and divergence in gene function and regulation and allow rapid adoption of experimental systems beyond the traditional model organisms.

Acknowledgments

We thank Jessica Buckles and Donna Storton in the Princeton Microarray Facility for assistance with production and processing of arrays, John Matese at the Princeton Microarray Database for support of data processing, John Wiggins for technical support of classroom computing resources, Dannie Durand and Manolis Kellis for helpful conversations, Jasper Rine for information on the *bar1* mutation in the *S. bayanus* strain, Doug Koshland and Yixian Zheng for purchase of microarray oligonucleotides, Zhenjun Hu (developer of VisAnt) for assistance with gene network visualization, Zeiss for donating use of tetrad dissection microscopes, and Molecular Devices for donating use of Genepix software. O.G.T. is supported by the National Science Foundation CAREER award DBI-0546275, and by National Institutes of Health (NIH) R01 grants GM071966 and HG005998. M.J.D. is supported in part by a grant from the National Institute of General Medical Sciences (8 P41 GM103533-17) from the NIH. A.A.C. is supported in part by grants from the Canadian Institutes for Health Research. All authors were supported by the National Institute of General Medical Sciences (NIGMS) Center of Excellence P50 GM071508, and by donations from the A. V. Davis Foundation and Princeton University for funding of QCB301, Experimental Project Laboratory. M.J.D. is a Rita Allen Scholar and a Canadian Institute for Advanced Research Fellow. All authors participated in the design and execution of experiments through the educational activities of the NIGMS Center for Quantitative Biology at Princeton University. A.A.C., Y.G., M.J.D., and O.G.T. analyzed the data and wrote the manuscript.

Literature Cited

Airoldi, E. M., C. Huttenhower, D. Gresham, C. Lu, A. Caudy *et al.*, 2009 Predicting cellular growth from gene expression signatures. PLoS Comput. Biol. 5: e1000257.

Aris-Brosou, S., 2005 Determinants of adaptive evolution at the molecular level: the extended complexity hypothesis. Mol. Biol. Evol. 22: 200–209.

Bergmann, S., J. Ihmels, and N. Barkai, 2004 Similarities and differences in genome-wide expression data of six organisms. PLoS Biol. 2: E9.

Borneman, A., T. Gianoulis, Z. Zhang, H. Yu, J. Rozowsky *et al.*, 2007 Divergence of transcription factor binding sites across related yeast species. Science 317: 815–819.

Bourbouloux, A., P. Shahi, A. Chakladar, S. Delrot, and A. K. Bachhawat, 2000 Hgt1p, a high affinity glutathione transporter from the yeast *Saccharomyces cerevisiae*. J. Biol. Chem. 275: 13259–13265.

Brauer, M. J., A. J. Saldanha, K. Dolinski, and D. Botstein, 2005 Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures. Mol. Biol. Cell 16: 2503–2517.

Bullard, J. H., Y. Mostovoy, S. Dudoit, and R. B. Brem, 2010 Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. Proc. Natl. Acad. Sci. USA 107: 5058–5063.

Busby, M. A., J. M. Gray, A. M. Costa, C. Stewart, M. P. Stromberg *et al.*, 2011 Expression divergence measured by transcriptome sequencing of four yeast species. BMC Genomics 12: 635.

Causton, H. C., B. Ren, S. S. Koh, C. T. Harbison, E. Kanin *et al.*, 2001 Remodeling of yeast genome expression in response to environmental changes. Mol. Biol. Cell 12: 323–337.

Cliften, P., P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton *et al.*, 2003 Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. Science 301: 71–76.

Cliften, P. F., R. S. Fulton, R. K. Wilson, and M. Johnston, 2006 After the duplication: gene loss and adaptation in *Saccharomyces* genomes. Genetics 172: 863–872.

Consortium, D. G., A. G. Clark, M. B. Eisen, D. R. Smith, C. Bergman *et al.*, 2007 Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203–218.

Cutter, A. D., A. Dey, and R. L. Murray, 2009 Evolution of the *Caenorhabditis elegans* genome. Mol. Biol. Evol. 26: 1199–1234.

Dietrich, F. S., S. Voegeli, S. Brachat, A. Lerch, K. Gates *et al.*, 2004 The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. Science 304: 304–307.

Doniger, S. W., J. Huh, and J. C. Fay, 2005 Identification of functional transcription factor binding sites using closely related *Saccharomyces* species. Genome Res. 15: 701–709.

Douglas, H. C., and D. C. Hawthorne, 1966 Regulation of genes controlling synthesis of the galactose pathway enzymes in yeast. Genetics 54: 911–916.

Downs, D. M., 2006 Understanding microbial metabolism. Annu. Rev. Microbiol. 60: 533–559.

Dujon, B., 2010 Yeast evolutionary genomics. Nat. Rev. Genet. 11: 512–524.

Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein, 1998 Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95: 14863–14868.

Gabriel, A., J. Dapprich, M. Kunkel, D. Gresham, S. C. Pratt *et al.*, 2006 Global mapping of transposon location. PLoS Genet. 2: e212.

Gallagher, J. E. G., J. E. Babiarz, L. Teytelman, K. H. Wolfe, and J. Rine, 2009 Elaboration, diversification and regulation of the Sir1 family of silencing proteins in *Saccharomyces*. Genetics 181: 1477–1491.

Gasch, A. P., 2007 Comparative genomics of the environmental stress response in ascomycete fungi. Yeast 24: 961–976.

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen *et al.*, 2000 Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11: 4241—4257.

Gasch, A. P., A. Moses, D. Y. Chiang, H. B. Fraser, M. Berardini *et al.*, 2004 Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol. 2: e398.

Gerke, J., K. Lorenz, and B. Cohen, 2009 Genetic interactions between transcription factors cause natural variation in yeast. Science 323: 498–501.

Goncalves, P., E. Valerio, C. Correia, J. M. de Almeida, and J. P. Sampaio, 2011 Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. PLoS ONE 6: e20739.

Gordon, J. L., K. P. Byrne, and K. H. Wolfe, 2009 Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. PLoS Genet. 5: e1000485.

Guan, Y., C. Myers, D. Hess, Z. Barutcuoglu, A. Caudy *et al.*, 2008 Predicting gene function in a hierarchical context with an ensemble of classifiers. Genome Biol. 9(Suppl. 1): S3.

Guan, Y., M. Dunham, A. Caudy, and O. Troyanskaya, 2010 Systematic planning of genome-scale experiments in poorly studied species. PLoS Comput. Biol. 6: e1000698.

Guan, Y., V. Yao, K. Tsui, M. Gebbia, M. J. Dunham *et al.*, 2011 Nucleosome-coupled expression differences in closely-related species. BMC Genomics 12: 466.

Guan, Y., M. J. Dunham, O. Troyanskaya, and A. Caudy, 2013 Comparative gene expression between two yeast species. BMC Genomics 14: 33.

Harbison, C. T., D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. MacIsaac *et al.*, 2004 Transcriptional regulatory code of a eukaryotic genome. Nature 431: 99–104.

Hess, D., W. Lu, J. Rabinowitz, and D. Botstein, 2006 Ammonium toxicity and potassium limitation in yeast. PLoS Biol. 4: e351.

Hibbs, M. A., D. C. Hess, C. L. Myers, C. Huttenhower, K. Li *et al.*, 2007 Exploring the functional landscape of gene expression: directed search of large microarray compendia. Bioinformatics 23: 2692–2699.

Hittinger, C. T., A. Rokas, and S. B. Carroll, 2004 Parallel inactivation of multiple *GAL* pathway genes and ecological diversification in yeasts. Proc. Natl. Acad. Sci. USA 101: 14144–14149.

Hittinger, C. T., P. Goncalves, J. P. Sampaio, J. Dover, M. Johnston *et al.*, 2010 Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature 464: 54–58.

Huttenhower, C., and O. Troyanskaya, 2008 Assessing the functional structure of genomic data. Bioinformatics 24: i330.

Huttenhower, C., A. I. Flamholz, J. N. Landis, S. Sahi, C. L. Myers *et al.*, 2007 Nearest neighbor networks: clustering expression data based on gene neighborhoods. BMC Bioinformatics 8: 250.

Iwahashi, H., E. Kitagawa, Y. Suzuki, Y. Ueda, Y. Ishizawa *et al.*, 2007 Evaluation of toxicity of the mycotoxin citrinin using yeast ORF DNA microarray and oligo DNA microarray. BMC Genomics 8: 95.

Jones, E., P. Berget, J. M. Burnette, C. Anderson, D. Asafu-Adjei *et al.*, 2008 The spectrum of Trp- mutants isolated as 5-fluoroanthranilate-resistant clones in *Saccharomyces bayanus*, *S. mikatae* and *S. paradoxus*. Yeast 25: 41–46.

Kassir, Y., D. Granot, and G. Simchen, 1988 *IME1*, a positive regulator gene of meiosis in *S. cerevisiae*. Cell 52: 853—862.

Kavanaugh, L. A., and F. S. Dietrich, 2009 Non-coding RNA prediction and verification in *Saccharomyces cerevisiae*. PLoS Genet. 5: e1000321.

Kellis, M., B. W. Birren, and E. S. Lander, 2004 Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428: 617–624.

Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423: 241–254.

Khan, Z., J. S. Bloom, S. Amini, M. Singh, D. H. Perlman *et al.*, 2012 Quantitative measurement of allele-specific protein expression in a diploid yeast hybrid by LC-MS. Mol. Syst. Biol. 8: 602.

Kitanovic, A., T. Walther, M. O. Loret, J. Holzwarth, I. Kitanovic *et al.*, 2009 Metabolic response to MMS-mediated DNA damage in *Saccharomyces cerevisiae* is dependent on the glucose concentration in the medium. FEMS Yeast Res. 9: 535–551.

Kumar, A., M. Seringhaus, M. C. Biery, R. J. Sarnovsky, L. Umansky *et al.*, 2004 Large-scale mutagenesis of the yeast genome using a Tn7-derived multipurpose transposon. Genome Res. 14: 1975–1986.

Kvitek, D. J., J. L. Will, and A. P. Gasch, 2008 Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates. PLoS Genet. 4: e1000223.

Lang, G. I., A. W. Murray, and D. Botstein, 2009 The cost of gene expression underlies a fitness trade-off in yeast. Proc. Natl. Acad. Sci. USA 106: 5755–5760.

Libkind, D., C. T. Hittinger, E. Valério, C. Gonçalves, J. Dover *et al.*, 2011 Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc. Natl. Acad. Sci. USA 108: 14539–14544.

Mahony, S., P. Auron, and P. Benos, 2007 DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. PLoS Comput. Biol. 3: e61.

Matsumoto, K., A. Toh-e, and Y. Oshima, 1981 Isolation and characterization of dominant mutations resistant to carbon catabolite repression of galactokinase synthesis in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 1: 83–93.

Muller, C. A., and C. A. Nieduszynski, 2012 Conservation of replication timing reveals global and local regulation of replication origin activity. Genome Res. 22: 1953–1962.

Osborn, M. J., and J. R. Miller, 2007 Rescuing yeast mutants with human genes. Brief. Funct. Genomics Proteomics 6: 104–111.

Peña-Castillo, L., and T. R. Hughes, 2007 Why are there still over 1000 uncharacterized yeast genes? Genetics 176: 7–14.

Pramila, T., W. Wu, S. Miles, W. S. Noble, and L. L. Breeden, 2006 The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle. Genes Dev. 20: 2266–2278.

Roberts, C. J., B. Nelson, M. J. Marton, R. Stoughton, M. R. Meyer *et al.*, 2000 Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science 287: 873–880.

Salvado, Z., F. N. Arroyo-Lopez, J. M. Guillamon, G. Salazar, A. Querol *et al.*, 2011 Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. Appl. Environ. Microbiol. 77: 2292–2302.

Scannell, D. R., O. A. Zill, A. Rokas, C. Payen, M. J. Dunham *et al.*, 2011 The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. G3: Genes, Genomes, Genetics 1: 11–25.

Serero, A., J. Lopes, A. Nicolas, and S. Boiteux, 2008 Yeast genes involved in cadmium tolerance: identification of DNA replication as a target of cadmium toxicity. DNA Repair (Amst.) 7: 1262–1275.

Serra, A., P. Strehaiano, and P. Taillandier, 2003 Characterization of the metabolic shift of *Saccharomyces bayanus* var. *uvarum* by continuous aerobic culture. Appl. Microbiol. Biotechnol. 62: 564–568.

Sharan, R., I. Ulitsky, and R. Shamir, 2007 Network-based prediction of protein function. Mol. Syst. Biol. 3: 88.

Sherman, A., M. Shefer, S. Sagee, and Y. Kassir, 1993 Post-transcriptional regulation of *IME1* determines initiation of meiosis in *Saccharomyces cerevisiae*. Mol. Gen. Genet. 237: 375—384.

Srikanth, C. V., P. Vats, A. Bourbouloux, S. Delrot, and A. K. Bachhawat, 2005 Multiple cis-regulatory elements and the yeast sulphur regulatory network are required for the regulation of the yeast glutathione transporter, Hgt1p. Curr. Genet. 47: 345–358.

St John, T. P., and R. W. Davis, 1981 The organization and transcription of the galactose gene cluster of *Saccharomyces*. J. Mol. Biol. 152: 285–315.

Stuart, J. M., E. Segal, D. Koller, and S. Kim, 2003 A gene-coexpression network for global discovery of conserved genetic modules. Science 302: 249–255.

Talarek, N., E. J. Louis, C. Cullin, and M. Aigle, 2004 Developing methods and strains for genetic studies in the *Saccharomyces bayanus* var. *uvarum* species. Yeast 21: 1195–1203.

Tsankov, A. M., D. A. Thompson, A. Socha, A. Regev, and O. J. Rando, 2010 The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol. 8: e1000414.

Tsong, A. E., B. B. Tuch, H. Li, and A. D. Johnson, 2006 Evolution of alternative transcriptional circuits with identical logic. Nature 443: 415–420.

Tuch, B. B., D. J. Galgoczy, A. D. Hernday, H. Li, and A. D. Johnson, 2008 The evolution of combinatorial gene regulation in fungi. PLoS Biol. 6: e38.

van Noort, V., B. Snel, and M. A. Huynen, 2003 Predicting gene function by conserved co-expression. Trends Genet. 19: 238–242.

Whitehead, A., and D. L. Crawford, 2006 Variation within and among species in gene expression: raw material for evolution. Mol. Ecol. 15: 1197–1211.

Whiteway, M., R. Freedman, S. Van Arsdell, J. W. Szostak, and J. Thorner, 1987 The yeast *ARD1* gene product is required for repression of cryptic mating-type information at the *HML* locus. Mol. Cell. Biol. 7: 3713–3722.

Winter, J., M. Ilbert, P. Graf, D. Ozcelik, and U. Jakob, 2008 Bleach activates a redox-regulated chaperone by oxidative protein unfolding. Cell 135: 691–701.

Wolf, Y. I., L. Carmel, and E. V. Koonin, 2006 Unifying measures of gene function and evolution. Proc. Biol. Sci. 273: 1507–1515.

Wu, A., J. A. Wemmie, N. P. Edgington, M. Goebl, J. L. Guevara *et al.*, 1993 Yeast bZip proteins mediate pleiotropic drug and metal resistance. J. Biol. Chem. 268: 18850–18858.

Yocum, R. R., and M. Johnston, 1984 Molecular cloning of the *GAL80* gene from *Saccharomyces cerevisiae* and characterization of a *gal80* deletion. Gene 32: 75–82.

Yocum, R. R., S. Hanley, R. West, Jr., and M. Ptashne, 1984 Use of lacZ fusions to delimit regulatory elements of the inducible divergent *GAL1–GAL10* promoter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 4: 1985–1998.

Yvert, G., R. B. Brem, J. Whittle, J. M. Akey, E. Foss *et al.*, 2003 Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. Nat. Genet. 35: 57–64.

Zill, O. A., and J. Rine, 2008 Interspecies variation reveals a conserved repressor of alpha-specific genes in *Saccharomyces* yeasts. Genes Dev. 22: 1704–1716.

Zill, O. A., D. Scannell, L. Teytelman, and J. Rine, 2010 Co-evolution of transcriptional silencing proteins and the DNA elements specifying their assembly. PLoS Biol. 8: e1000550.

Communicating editor: M. Johnston

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152918/-/DC1

A New System for Comparative Functional Genomics of *Saccharomyces* Yeasts

Amy A. Caudy, Yuanfang Guan, Yue Jia, Christina Hansen, Chris DeSevo, Alicia P. Hayes, Joy Agee, Juan R. Alvarez-Dominguez, Hugo Arellano, Daniel Barrett, Cynthia Bauerle, Namita Bisaria, Patrick H. Bradley, J. Scott Breunig, Erin Bush, David Cappel, Emily Capra, Walter Chen, John Clore, Peter A. Combs, Christopher Doucette, Olukunle Demuren, Peter Fellowes, Sam Freeman, Evgeni Frenkel, Daniel Gadala-Maria, Richa Gawande, David Glass, Samuel Grossberg, Anita Gupta, Latanya Hammonds-Odie, Aaron Hoisos, Jenny Hsi, Yu-Han Huang Hsu, Sachi Inukai, Konrad J. Karczewski, Xiaobo Ke, Mina Kojima, Samuel Leachman, Danny Lieber, Anna Liebowitz, Julia Liu, Yufei Liu, Trevor Martin, Jose Mena, Rosa Mendoza, Cameron Myhrvold, Christian Millian, Sarah Pfau, Sandeep Raj, Matt Rich, Joe Rokicki, William Rounds, Michael Salazar, Matthew Salesi, Rajani Sharma, Sanford Silverman, Cara Singer, Sandhya Sinha, Max Staller, Philip Stern, Hanlin Tang, Sharon Weeks, Maxwell Weidmann, Ashley Wolf, Carmen Young, Jie Yuan, Christopher Crutchfield, Megan McClean, Coleen T. Murphy, Manuel Llinás, David Botstein, Olga G. Troyanskaya, and Maitreya J. Dunham

DOI: 10.1534/genetics.113.152918

File S1

Supporting Methods

Strains and Media: Strains used are listed in Table S1. We note the taxonomic controversy regarding the precise species designation of the sequenced isolate, a derivative of which we have used for the majority of our experiments. Unless otherwise indicated, we used derivative of the type strain, CBS 7001, which is typically regarded to be *S. bayanus* var. *uvarum* or, by some nomenclature, *S. uvarum* (RAINIERI *et al.* 2006). For simplicity, we use the terminology "*S. bayanus*" throughout since this is the terminology still associated with the sequencing projects and other recent studies. Unless otherwise noted, cultures were grown at 25ºC. YNB minimal and YPD rich media were made according to standard recipes with 2% glucose. Additives are as noted below for each dataset. Chemostat media was made as described (sulfur and glucose (SALDANHA *et al.* 2004), phosphate (GRESHAM *et al.* 2008), potassium (HESS *et al.* 2006)). Knockouts were prepared by standard homologous recombination techniques with drug resistance markers.

Microarray design and production: The *S. cerevisiae* and *S. bayanus* ORF and contig sequences were downloaded from the *Saccharomyces* Genome Database and concatenated to create a hybrid genome. The design process was based on the hybrid in anticipation of utilizing these arrays in *S. cerevisiae*/*S. bayanus* interspecific hybrids, although only *S. bayanus*-specific probes are reported in this study.

We used the joint assembled sequence created at Washington University (CLIFTEN *et al.* 2003; CLIFTEN *et al.* 2006; ROBERTS *et al.* 2000) and the program Array Oligo Selector (BOZDECH *et al.* 2003) to design a microarray of 70mers specific to each open reading frame in both genomes. Under the default stringency settings, 711 genes were too similar to another sequence in the combined genomes for a sufficiently unique oligonucleotide to be designed. For these cases, the program was rerun in the context of each single genome in order to provide more complete coverage of the purebred genomes. 485 genes were still too similar to other sequences in the single genomes to pass this test and were left off the array. The resulting 4840 *S. bayanus* and 6423 *S. cerevisiae* 70mer oligos were purchased from Illumina.

For array printing, 70mer oligonucleotides were resuspended at 40 μM in 3X SSC and printed using a pin-style arraying robot onto aminosilane slides in a controlled-humidity environment. Slides were UV crosslinked at 70

mJ. On the day of hybridization, the slides were blocked by agitating for 35 minutes at 65ºC with 1% Roche blocking agent in 5X SSC and 0.1% SDS. Slides were then rinsed with water for 5 minutes and spun dry.

RNA Preparation and Labeling: Cells were harvested by fast-filtration followed by snap freezing in liquid nitrogen. Total RNA was prepared from the cells by acid phenol extraction followed by ethanol precipitation. Crude total RNA was further purified using Qiagen RNeasy kits, and labeled by direct incorporation of Cy3-dUTP or Cy5-dUTP (Enzo Biosciences) into cDNA primed using oligo-dT. Labeling yield and efficiency were measured using a nanodrop spectrophotometer. Reverse transcription reactions were purified on Zymo DNA Clean and Concentrator kits and hybridized using Agilent hybridization buffer and blocking reagent as described below. All microarray manipulations were performed in an ozone-free environment.

Nearly all samples were hybridized versus a common reference prepared from a mixture of RNA from *MAT***a** (ACY14), *MAT*α (ACY12), and *MAT***a**/α cells (ACY9). These three strains were sampled in both exponential and stationary phase. Additionally, RNA from stress conditions was included: ACY12 treated with hydrogen peroxide and sampled at 10, 30 and 45 minutes, and ACY12 cells treated with heat shock from 25ºC to 37ºC and sampled at 10 and 30 minutes. Total RNA was prepared from all these samples, mixed, and aliquoted.

Samples from the following datasets were not hybridized versus the common reference (reference used in parenthesis): cell cycle (asynchronous ACY7), constant temperatures (log phase ACY1354), mating type and ploidy (log phase ACY1354), diauxic shift (log phase ACY1354), aging (mixture of all timepoints), sporulation (asynchronous cells in YPD, or mixture of timepoints, depending on experiment series), and strain backgrounds (log phase CY1).

Array processing: Experiments testing cross-hybridization vs. the *S. cerevisiae* probes allowed us to optimize hybridization and wash parameters for high performance (data not shown).

Labeled cDNAs were mixed with Agilent blocking reagent and 2X hybridization buffer in a total volume of 400 μl, heated at 95°C for 5 minutes, and hybridized to a crosslinked and blocked microarray using an Agilent gasket slide. Hybridizations were performed overnight at 65ºC in a rotating hybridization oven (Agilent). Gasket slides were removed in 1X SSC and 0.1 % SDS solution. Arrays were agitated for 10 minutes in a 65ºC bath of the same wash buffer, then washed on an orbital shaker for 10 minutes in a new rack in 1X SSC, ending with 5 minutes in 0.1 X SSC. Arrays were then spun dry and scanned in an Agilent or Axon scanner. The resulting images were analyzed using

Axon Genepix software version 5. Complete microarray data can be downloaded from the Princeton Microarray Database and GEO (accession number GSE16544 and GSE47613).

Data corresponding to *S. bayanus* probes were linearly normalized and filtered for spots with intensity of at least 2 times over background in at least one channel. Manually flagged spots were also excluded. Processed microarray data as presented in Figure 2 are available in Table S3.

Construction of Tn7 insertion library: Briefly, this mutagenesis approach uses a library of Tn7 transposon insertions in a *S. bayanus* genomic DNA library. The construct has a selectable marker for transformation into yeast, allowing the selection of disruption alleles. In order to remove contamination of mtDNA, we created a ρ^0 *S. bayanus* strain in a cir^0 background (Figure 1I) by treatment with ethidium bromide. Petite colonies were identified for lack of growth on glycerol and lack of visible mtDNA nucleoids by DAPI staining. Genomic DNA was isolated and fragmented by sonication to an average length of 3 kb. The ends of the DNA were blunted and cloned into the pZero-Blunt vector (Invitrogen). Approximately 50,000 colonies were recovered from the transformation into DH5α bacteria; colonies were scraped from the plates for plasmid purification.

We constructed a version of the Tn7 transposon by amplifying the promoter from the Tet-on pCM224 (BELLÍ *et al.* 1998). The cassette of the Tet-on promoter and the ClonNAT resistance gene was amplified using PCR primers containing lox and *Bam*HI sites and cloned into the *Bam*HI site of the NEB vector pGPS3, producing the vector pAC13. The ClonNAT resistance gene confers ClonNAT resistance in both yeast and bacteria. The tet-on promoter should permit inducible expression when placed near genes in strains carrying the appropriate tetracycline responsive transcription factor, but this has not been tested. This transposon construct was hopped into the *S. bayanus* genomic DNA library in vitro using the transposon kit from NEB. Background from pAC13 was removed by digesting with I-*Sce*I, which cuts the parent vector. Initial selection (50,000 colonies) was on ClonNAT/Zeo; subsequent replating of the library for DNA preps was on ClonNAT/Kan, reducing the background of transposon insertions into the kanamycin gene on the plasmid backbone. The transposed library was cut by *Hin*dIII and *Xba*I to release the genomic DNA from the pZero backbone, linearizing the genomic DNA for efficient recombination. The library was then transformed into *S. bayanus* and selected on YPD-ClonNAT.

The library is available upon request transformed into ACY12 (300,000 colonies), and also in DH5α *E. coli*.

Transposons were mapped using the method described in Gabriel et al. (2006), using probes gaacataaacaaccatgggt and ggcggggataacttcgtata. When whole genomic DNA was used, the extracted material was compared to sonicated control DNA, and the highest point of the peak was estimated to be the insertion site. When genomic DNA was digested using enzymes that cut within the transposon, the separate 5' and 3' extractions were labeled with opposite dyes. The insertion position of the transposon is apparent as the site between which the dyes change color.

Growth and treatments for expression analysis: Samples for gene expression analysis were grown, harvested, and hybridized as follows.

Alpha factor arrest. S. bayanus ACY7 (*MAT***a** prototroph, note that the *S. bayanus* strain used, a derivative of the sequenced strain, is a *bar1* mutant via a naturally occurring frameshift) culture at 1 x 10^7 cells was exposed to 50 ng/mL alpha factor in YPD. The indicated time points were harvested. The reference was a mixed reference of all time points from the experiment.

Diauxic shift. ACY1354 was grown in overnight YPD cultures and diluted back to an OD600 of 0.025 into fresh glucose-limited chemostat media with 0.24% glucose. Samples were harvested at intervals indicated. Glucose and ethanol assays were performed on filtrate samples to ensure coverage over the diauxic shift using enzymatic/spectrophotometric assays (R-Biopharm).

Drug treatments. ACY12 growing in YPD was used for all experiments. Drugs were purchased from Sigma-Aldrich except where noted. The final drug concentrations used were 50 mM hydroxyurea, 2 mM 2-deoxy-D-glucose, 0.002% tunicamycin, 43 μg/mL lovastatin, 1 μg/mL zeocin (Invitrogen), 0.1% MMS, 100 μM MG132, and 100 nM rapamycin. The reference used was the universal mixed reference.

Hyperosmotic shock. Exponentially growing *S. bayanus* ACY12 in YPD at 2 x 10^7 cells/mL were treated with hyperosmotic shock of 1 M sorbitol and timepoints were taken as indicated. The reference used was the universal mixed reference.

Carbon sources. ACY12 was used for all experiments. For sucrose and glycerol, cells were grown overnight in YNB (glucose), freshly diluted in YNB (glucose) to 5 x 10^6 cells/mL, grown 5 hours (two doubling times), recovered by filtration, and switched to YNB + 2% sucrose or YNB + 2% glycerol. For galactose, cells were grown overnight in YNB (raffinose), diluted to 5 x 10^6 cells/mL, grown 5 hours (two doubling times), and galactose was added to 2% final concentration. Cells were harvested at indicated time points. Reference was the universal mixed reference.

Cell cycle data. ACY7 at 1 x 10^7 cells/mL was synchronized with 50 ng/mL alpha factor in YPD for 5 hours at 25°. The culture was spun down and released into YPD plus 50 μg/mL pronase. The volume of the culture was adjusted to return the cell density to 1 x 10^7 cells/mL. Samples were taken at the indicated time points and measured by FACS to monitor cell cycle progression (data not shown). The reference for the arrays was a population of asynchronous ACY7 cells at 1 x 10^7 cells/mL.

Stresses. The critical concentrations of each treatment were derived from growth assays (Figure 1G), and were 5 μM $CdCl_2$, 2 mM $CuSO_4$, 1 mM $MnCl_2$, 1.5 mM $NiCl_2$, 2 mM $PbCl_2$, 1.5 mM $ZnCl_2$, 0.25% bleach (dilution from a stock of 2% NaOCl), 5% Ethanol, 0.15 mM and 1.5 mM H_2O_2, 10 mM $K_2S_2O_5$, or 0.5 M NH_4Cl. Cells growing exponentially in YPD were treated as indicated and sampled at the indicated time points.

Tn7 Transposon insertion strains. Library construction is described above. 28 plasmids from the Tn7 insertion library were identified by sequencing to have Tn7 insertions within genes (see Table S1 for identities). These plasmids were digested and transformed into ACY332; the Tn7 insert was selected with ClonNAT and colony purified. Selected strains were grown in YPD overnight cultures, diluted 1:50 in fresh media, grown to midlog phase (klett colorimeter 65-80), and harvested. Upon inspection of the data, it was noted that insertion 7 is aneuploid for chromosome 16.

Aging. The indicated strains were grown to mid-exponential phase in YPD at room temperature. The cells were pelleted by centrifugation and resuspended in sterile water at a concentration of 7.5 x 10^5 cells/mL. Incubation continued with shaking, and cells were harvested for array analysis at the indicated timepoints. The reference was a mixture of all samples.

Strain backgrounds. Strains obtained from a variety of sources and identified as *S. bayanus* were grown in YPD overnight cultures, diluted 1:50 in fresh YNB media, grown to midlog phase (klett colorimeter 65-80), and harvested.

Segregants. Data from the "strain backgrounds" experiment described above identified ACY1356 as having divergent gene expression vs the other strains. ACY1356 was sporulated as below and resulting tetrads dissected on rich media. Individual ACY3 cells were placed immediately adjacent to each spore. Clones from the resulting colonies were colony-purified and sporulated. Tetrads were tested for segregation of *lys2*, drug resistance, and mating type to identify diploids that were the product of mating between the strains. One segregant from each tetrad was grown in YPD overnight culture, diluted 1:50 in fresh media, grown to midlog phase (klett colorimeter 65-80), and

harvested. Upon further investigation of the origins of ACY1356 and ACY3, it was discovered that both are actually derivatives of MCYC623. Therefore, this experiment is not a true outcross.

Sporulation. ACY1354 and ACY1356 were grown in overnight YPD cultures and diluted back to 5 x 10^6 cells/mL in YPD, allowed to grow 5 hours at room temperature with shaking, and pelleted in a centrifuge. The YPD was discarded and cells were resuspended in a 10-fold larger volume of SPO++ media (2.5 g/L yeast extract, 15 g/L potassium acetate, 0.25% dextrose, 40 mg/L adenine, 40 mg/L uracil, 40 mg/L tyrosine, 20 mg/L histidine, 20 mg/L leucine, 20 mg/L lysine, 20 mg/L tryptophan, 20 mg/L methionine, 20 mg/L arginine, 100 mg/L phenylalanine, 350 mg/L threonine). Samples were collected as indicated for the gene expression timecourse. For sporulation for strain construction, cultures were allowed to sporulate at room temperature for 1-7 days before dissection.

Ploidy and mating type. ACY8, ACY331, and ACY1354 were grown in overnight YPD cultures and diluted back to an OD600 of 0.125 into fresh glucose-limited chemostat media with 0.24% glucose. Samples were harvested at midlog (klett of 100). Residual glucose and ethanol concentrations in the filtrates were measured using enzymatic/spectrophotometric assays (R-Biopharm) to ensure cultures had not reached the diauxic shift (data not shown).

Constant temperature. ACY1354 was grown in overnight YPD cultures at 25º and diluted back to an OD600 of 0.125 into fresh glucose-limited chemostat media with 0.24% glucose. Samples were grown at 20º, 25º, or 30º and harvested at midlog (klett of 100). Residual glucose and ethanol concentrations in the filtrates were measured using enzymatic/spectrophotometric assays (R-Biopharm) to ensure cultures had not reached the diauxic shift.

Nutrient starvation. Auxotrophs for uracil, lysine, and tryptophan were grown overnight in YNB glucose media supplemented with a limiting concentration of the appropriate additive (4 mg/L, 6 mg/L, and 2 mg/L respectively). Cultures were diluted 1:50 into new flasks of limiting media and harvested at intervals indicated.

Chemostat cultivation. ACY1354 was grown to steady state in ATR Sixfors modified to run as chemostats. Cultures were grown in 300 mL nutrient-limited minimal medium at a dilution rate of 0.17 hr^{-1} at 25º or 30º as indicated. Chemostats were well-mixed and sparged with sterile humidified air.

Galactose metabolism. Deletion mutants of the *GAL80* ortholog *555.11* and the *GAL80* ohnolog *670.20* were prepared by standard knockout methods. To examine the repression of *GAL* genes, the four spores from a full tetrad from each knockout (ACY677, ACY678, ACY1357, ACY1358, ACY1361, ACY1362, ACY1363, ACY1364 were grown to exponential

phase in YPD. To examine the response to galactose, cells were grown overnight in YNB raffinose. When cells reached 1 x 10^7 cells/mL, galactose was added to a final concentration of 2% and timepoints were taken as indicated.

Quantitative PCR: Yeast were grown exponentially in 2% glucose (for switch to sucrose or glycerol) or 2% raffinose (for switch to galactose). Cells were harvested by filtration and snap frozen for acid phenol RNA preparation. RNA was converted into cDNA using Superscript III with a T20VN primer, and analyzed by quantitative PCR using the absolute quantitation method using a dilution series of galactose cDNA as a calibrator. cDNAs were amplified using TaqComplete (AllianceBio) amended to contain a 1:50,000 dilution of SybrGreen (Life Technologies). The PCR primers for *S.bayGAL1* were 5'-TGGTTGCACTGTTCACTTGG-3' and 5'-ATTGCGTCTTCCAACTCAGC-3'; for *S.bayPMA1 5'*-GAACCGACAGCCAACACAAT-3' and 5'-TCATTGCCATTTTCGCCGAT-3'; for *ScerGAL1 5'*-ACATTTCCACACCCTGGAAC-3' and 5'-GATTGTGCGACATCGTCAAC-3'; and for *ScerPMA1 5'*-TCTCCAAAGCCCGTTAAATG-3' and 5'-CCGTTCATAGCACCGAAGTT-3'.

Motif analysis of oxidative stress cluster: The sequence 1000 bp upstream of the start site of the genes in the oxidative stress cluster (Figure 5) was submitted to the SCOPE server (CARLSON *et al.* 2007). The motifs are available in Table S6. These were compared to the motifs derived from comparative sequence analysis of the *sensu stricto* yeasts (KELLIS *et al.* 2003) using STAMP (MAHONY *et al.* 2007).

Gene function prediction: Gene function prediction including the following steps.

Preprocessing of microarray data. We searched the literature to collect data from 2569 arrays covering 125 datasets in *S. cerevisiae* (GUAN *et al.* 2010). The following preprocessing steps were carried out on both the *S. bayanus* and *S. cerevisiae* datasets to allow later learning using support vector machines. For each dataset, genes that are represented in less than half of the arrays were removed, and missing values were inserted using KNNimpute (TROYANSKAYA *et al.* 2001) with K = 10, Euclidean distance. Technical replicates were averaged, resulting in datasets with each gene followed by a vector representing its expression values in a series of arrays.

Visualization of gene-gene correlation. Within a species, for each pair of genes j and k, we calculated the correlation coefficient of their expression pattern:

$$\rho(j_i, k_i) = \frac{\mathrm{cov}(j_i, k_i)}{\sigma_{j_i}\sigma_{k_i}} \qquad (1)$$

We then averaged these correlations over all datasets. These gene-gene correlations are available in a network view that presents the top five connected genes for each species(HU *et al.* 2009).

Bootstrap SVM in predicting function and evaluating the reliability of each dataset. Because *S. bayanus* does not have an annotated set of genes that could be used as gold standard in function prediction, a gold standard was constructed by transferring the biological process annotations from *S. cerevisiae* by orthology. The orthologs used were those determined by the joint assembly effort (CLIFTEN *et al.* 2003). To avoid errors caused by genes that changed function between the two species, we applied bootstrap SVM to predict functions. In each bootstrap, only the held-out values were used to estimate the probability that a gene is annotated to a certain function.

For each GO term, the positive examples were taken as genes annotated directly to certain biological process or to a descendent of this term. Negative examples were assumed to be all other genes. The basis of our approach is a support vector machine (SVM) classifier. Our previous work has shown that a single linear-kernel SVM often out-performs most of the more complicated machine learning methods in gene function prediction (GUAN *et al.* 2008). Therefore, we trained a linear-kernel SVM on each biological process. We used the SVM^{light} software to implement the SVM classifiers (JOACHIMS 1999). We have experimented with several parameters and alternative kernels and found only cost factor (*j*) plays an important role in the scenario of gene function prediction (data not shown). We set it as the ratio of negative examples to positive examples.

We applied 0.632 bootstrap aggregation in predicting gene functions. Intuitively, this method trains the SVM on a subset of genes and tests it on a different subset of genes repeatedly, thus minimizing the possibility of over-fitting or the effect of potentially mis-annotated genes. Specifically, examples (genes) were randomly sampled with replacement (0.632 bootstrap). For each bootstrap sample, a model was learned based on the selected examples, and the resulting classifier was used to give an output on non-selected (out-of-bag) examples. The final classifier outputs were taken as the median of out-of-bag values across 25 independent bootstraps, and the ROC curves were derived from these median values. Because only the results of the out-of-bag values were recorded, this approach minimizes contamination of mis-annotated genes.

We carried out this function prediction process for individual datasets and a concatenate of all datasets. The accuracy of each dataset in capturing the biological processes was evaluated using AUC (Figure S7).

Probability estimation. The value of the SVM output does not directly imply the probability of a gene to be annotated to the term. We therefore estimated this probability by fitting the SVM output distribution of positive and negative examples with two normal distributions. According to Bayesian theory,

$$p(y \mid X) = \frac{p(X \mid y)p(y)}{p(X)} = \frac{p(X \mid y)p(y)}{p(X \mid y)p(y) + p(X \mid n)p(n)}, \quad (2)$$

where

$$p(X \mid y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(x-\mu_y)^2}{2\sigma_y^2}} \quad (3)$$

$$p(X \mid n) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(x-\mu_n)^2}{2\sigma_n^2}} \quad (4)$$

Where X is the SVM output value, y represents positive examples, and n represents negative examples. σ_y and σ_n are the standard deviation of the SVM output for positive and negative examples respectively. μ_y and μ_n are the mean of the SVM output for positive and negative examples respectively. Based on these we could estimate the probability of a gene annotated to a term given its observation value X, i.e. p(y|X). SVM output with a value lower than the average of negative examples was assigned as zero. We provide the complete list of predictions in terms of probability in Table S4.

Calculating changes in gene function. The distribution of SVM output or the consequent probability scores differed by GO term and species due to the different reliability of the predictions. To quantify functional changes between species, for each function, we subtracted the normalized ranks (by the total number of genes on an array) of SVM output for a gene in *S. cerevisiae* from that of *S. bayanus*. This resulted in rank differences ranging from -1 to 1, where a higher value means that the ortholog is more biased towards this function in *S. bayanus* than in *S. cerevisiae*, and vice versa. Functional changes are most relevant when at least one member of the ortholog ranks high and therefore potentially is associated with the function under study. Taking into account that the GO terms vary in size, we examined the fraction of orthologs with at least one member ranking in the top as counted by the total number of genes annotated to the term in *S. cerevisiae*. The data are shown in Table S5.

Data visualization. The interactive network view of the expression data and searchable prediction results are available at *http://bayanusfunction.princeton.edu*

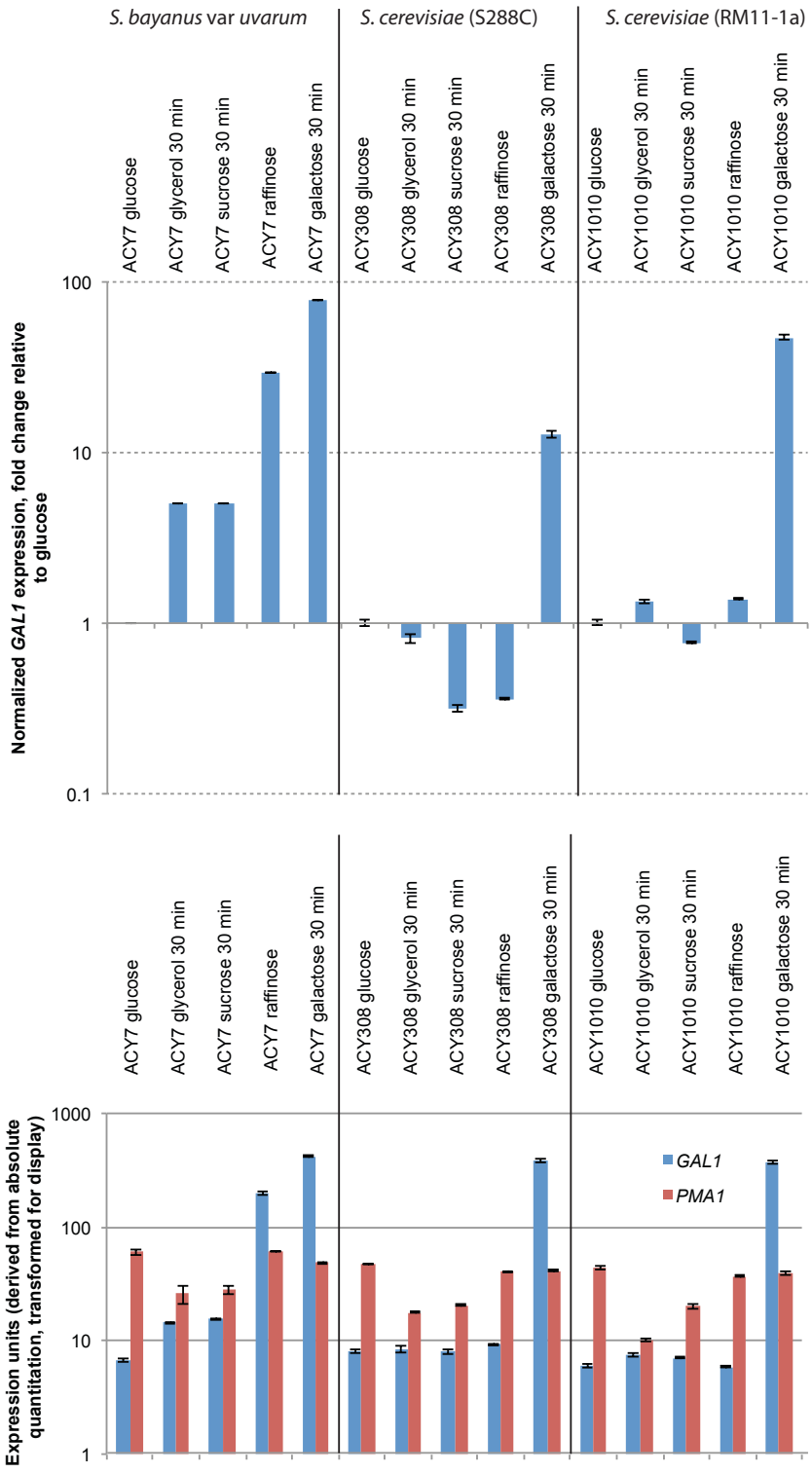


Figure S1 *GAL1* expression in *S. bayanus* is increased during carbon switches. A. Expression of *GAL1* normalized to *PMA1* expression. The indicated *S. bayanus* and *S. cerevisiae* strains were grown in minimal medium containing glucose and switched to minimal medium containing sucrose or glycerol, or in minimal medium containing raffinose and switched to minimal medium containing galactose. Cells were harvested, RNA was extracted and converted into cDNA, and expression of *GAL1* and the calibrator *PMA1* were measured by real time PCR in technical triplicate; quantities of mRNA were determined by using a calibration curve of RNA from the galactose induced sample. B. Levels of *PMA1* and *GAL1* transcripts shown in A are scaled for comparison but presented without normalization. Expression of the vacuolar protease *PMA1* decreases when glucose is removed, consistent with previous observations (Rao *et al.* 1993). Although this does affect the ratios presented in (A), the induction of *GAL1* by sucrose and glycerol are a greater magnitude in *S. bayanus* than in *S. cerevisiae*.

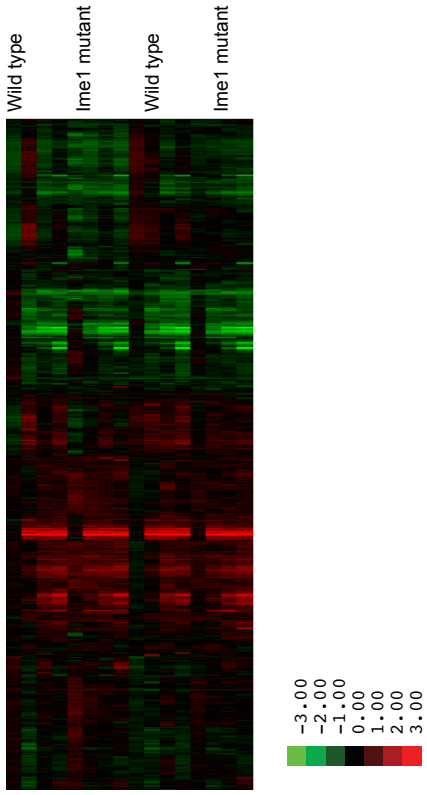


Figure S2 Pheromone response in *670.55* knockout mutant (ortholog of *IME1*) is similar to wild type. The indicated mutant and wild type strains were grown in YPD and exposed to alpha factor pheromone. Samples were harvested for expression analysis. A heat map of the genome-wide expression data is presented, with the data transformed to the average expression of wild type samples at the moment of alpha factor addition.

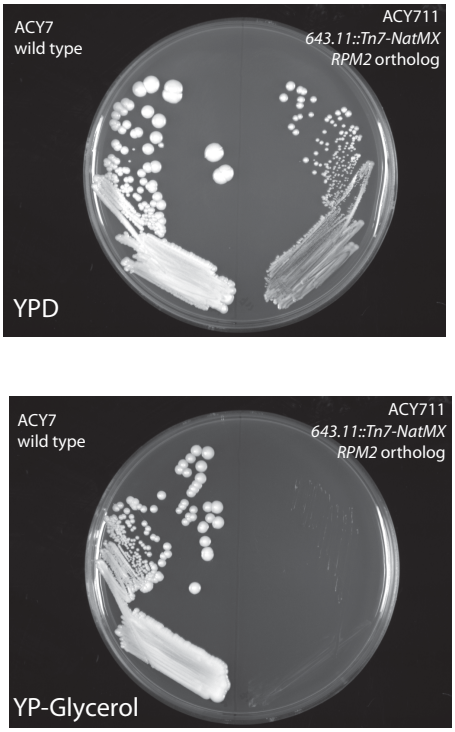
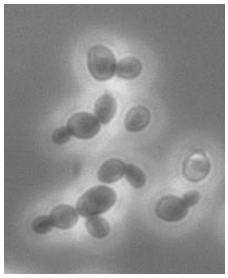
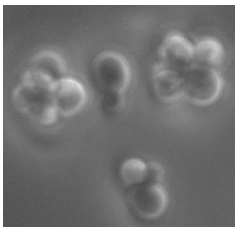


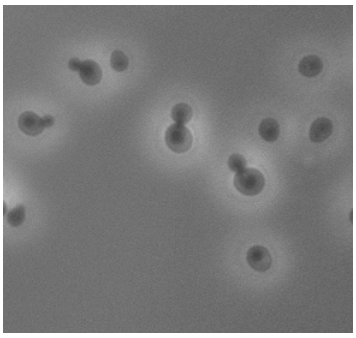
Figure S3 The *643.11* knockout mutant (ortholog of *RPM2*) is respiratory deficient. The *643.11* knockout strain ACY711 and a wild type strain were struck for growth on SC glycerol and on YNB glucose medium.



Diploid *S. bayanus*
Wild type



Haploid *S. bayanus*
Wild type



Haploid *S. bayanus*
678.66 KO
Sbayamn1

Figure S4 Haploid-specific cell clumping is disrupted in a mutant of *678.66* (ortholog of *AMN1*). Phase contrast micrographs are shown of diploid cells and of *678.66* and wild type haploid cells.

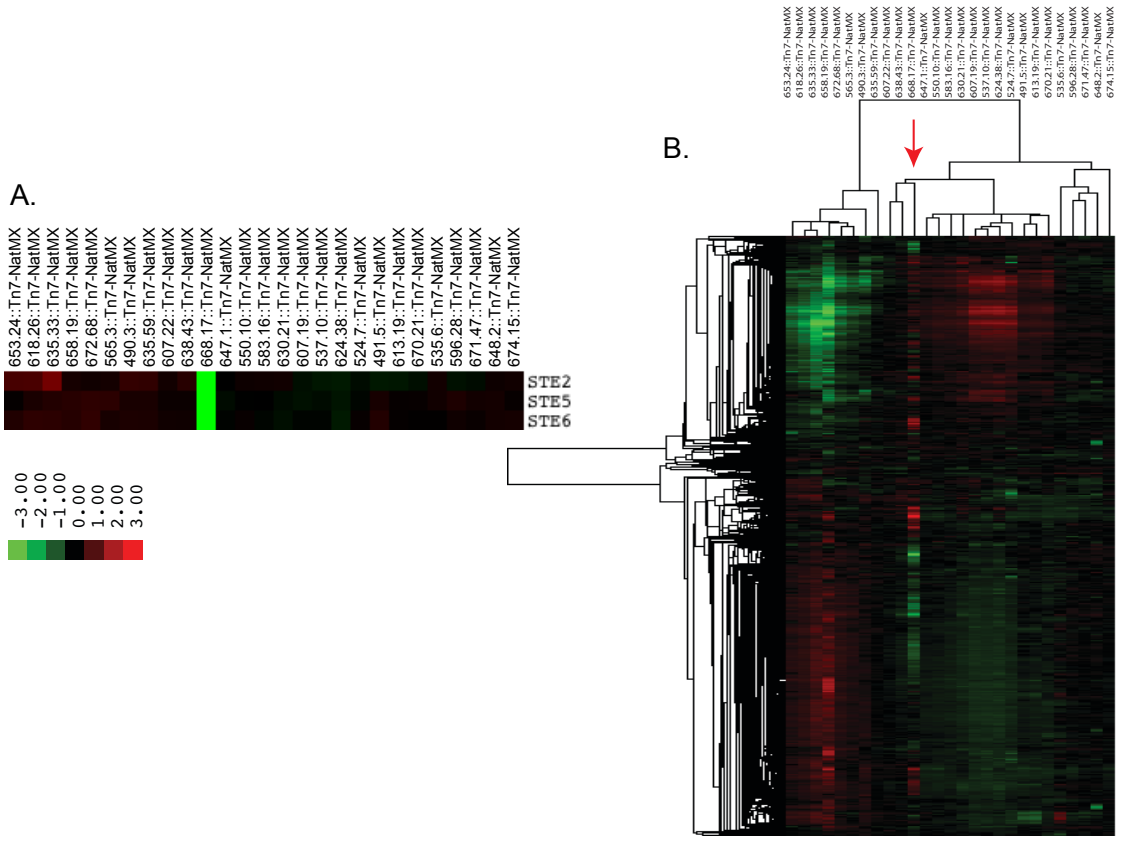


Figure S5 The *668.17* knockout mutant (ortholog of *ARD1)* shows changes in gene expression of mating-type specific genes, as well as global changes in expression. A. The data for the Tn7 insertion strains were selected from the megacluster (data in Table S3), and expression levels of *STE2, STE5,* and *STE6* is shown. B. The genome wide expression data Tn7 insertion strains was hierarchically clustered by both gene and experiment. The column corresponding to the *668.17::Tn7-NatMX* mutant is indicated with a red arrow.

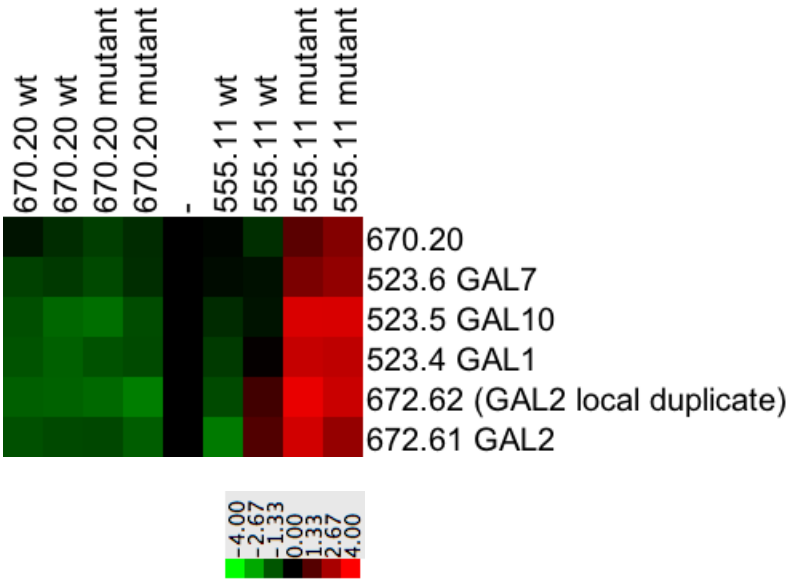
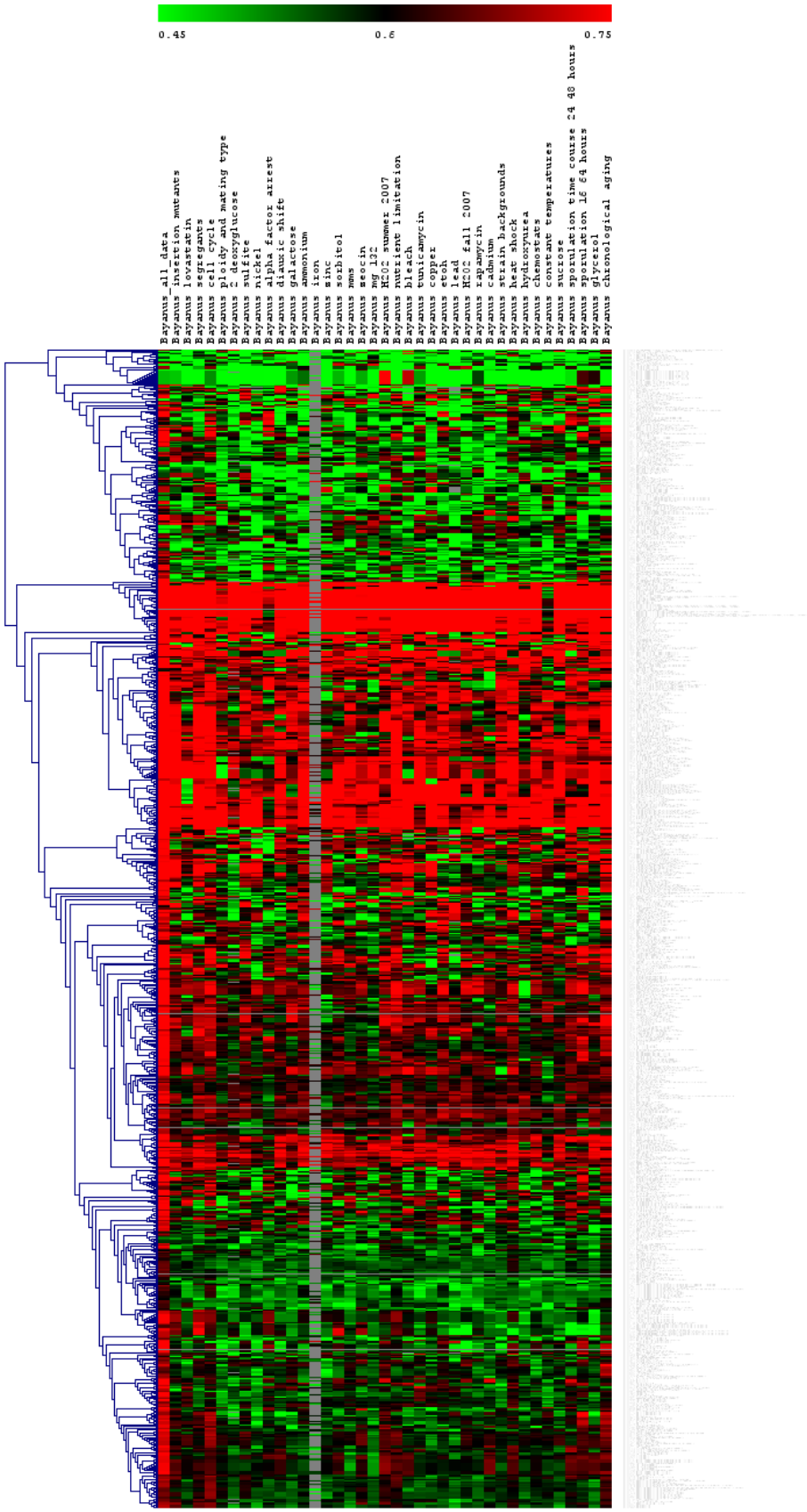


Figure S6 The *S. bayanus GAL80* ortholog *555.11* represses the expression of galactose structural genes when glucose is present. The indicated strains were grown in YPD to mid log phase and harvested for RNA. The expression of a set of genes whose orthologs in *S. cerevisiae* are involved in galactose metabolism are shown.

A


0.45
0.6
0.75
Bayanus_all_data
Bayanus insertion mutants
Bayanus lovastatin
Bayanus segregants
Bayanus cell cycle
Bayanus ploidy and mating type
Bayanus 2-deoxyglucose
Bayanus sulfite
Bayanus nickel
Bayanus alpha factor arrest
Bayanus diauxic shift
Bayanus galactose
Bayanus ammonium
Bayanus iron
Bayanus zinc
Bayanus sorbitol
Bayanus mms
Bayanus mg_ion
Bayanus HC0c summer 2007
Bayanus nutrient limitation
Bayanus starch
Bayanus tunicamycin
Bayanus copper
Bayanus etoh
Bayanus lead
Bayanus HC0c fall 2007
Bayanus rapamycin
Bayanus cadmium
Bayanus strain backgrounds
Bayanus heat shock
Bayanus hydroxyurea
Bayanus chemostats
Bayanus constant temperatures
Bayanus sucrose
Bayanus sporulation time course 24 48 hours
Bayanus sporulation 16 64 hours
Bayanus glycerol
Bayanus chronological aging

B

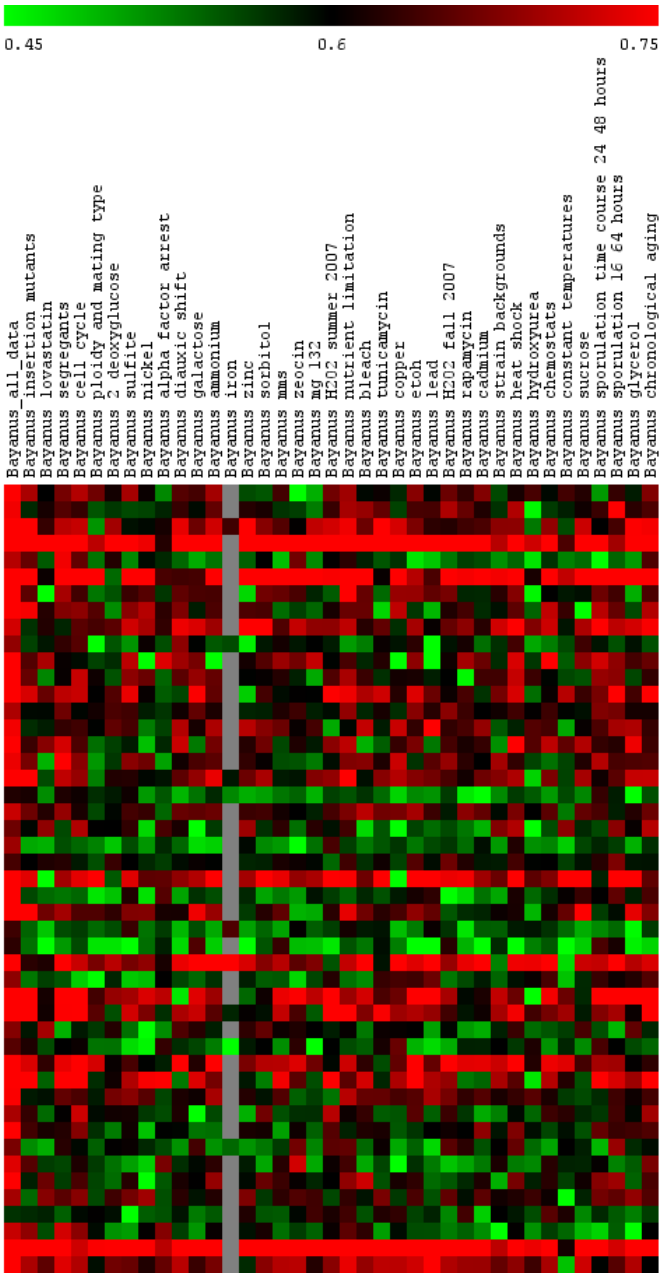


Figure S7 The AUC (area under the precision/recall curve) for various biological functions predicted from *S. bayanus* gene expression data, shown for A. all GO terms and B. GO SLIM terms.

Table S1 Description of *S. bayanus* strains used in this study.

Strain	Species	*Genotype*	Source
ACY3	*S. bayanus var. uvarum*	*MAT*α *hoΔ::KanMX lys2-1*	Douglas Koshland, derived from type strain (MCYC623) from Duncan Greig
ACY4	*S. bayanus var. uvarum*	*MAT*α *hoΔ::HygMX lys2-1 ura3Δ::NatMX*	Douglas Koshland, derived from type strain (MCYC623) from Duncan Greig
ACY7	*S. bayanus var. uvarum*	*MAT***a** *ho::KanMX*	This study. Haploid segregant from ACY1354.
ACY8	*S. bayanus var. uvarum*	*MAT*α *hoΔ::KanMX*	This study. Haploid segregant from ACY1354.
ACY9	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *hoΔ::KanMX/hoΔ::HygMX lys2-1/LYS2 ura3Δ::NatMX/URA3*	This study
ACY11	*S. bayanus var. uvarum*	*MAT*α *ura3Δ::NatMX*	This study. Haploid segregant from ACY1354.
ACY12	*S. bayanus var. uvarum*	*MAT*α *hoΔ::HygMX*	This study
ACY14	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HygMX*	This study
ACY331	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::KanMX*	This study. Haploid segregant from ACY1354.
ACY333	*S. bayanus var. uvarum*	*MAT*α *hoΔ::HisG*	This study. 480
ACY382	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG his3*Δ1	This study.
ACY468	*S. cerevisiae* X *S. bayanus var. uvarum*	*MAT***a***/MAT*α *Scerho/SbayhoΔ::HisG Scerabf1Δ::KanMX/SbayABF1*	This study. *S. cerevisiae* parent was ACY308, *S. bayanus* parent was ACY333.
ACY480	*S. cerevisiae*	*MAT***a** *ura3Δ0 ras1::rgs2::KanMX*	This study.
ACY488	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 678.66Δ::HygMX (Sbayamn1)*	This study.
ACY545	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 596.28::Tn7-NatMX*	This study.
ACY546	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 565.3::Tn7-NatMX*	This study.
ACY547	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 670.21::Tn7-NatMX*	This study.
ACY548	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 491.5::Tn7-NatMX*	This study.
ACY549	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 658.19::Tn7-NatMX*	This study.
ACY550	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 672.68::Tn7-NatMX*	This study.
ACY551	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 671.47::Tn7-NatMX*	This study.
ACY552	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 648.2::Tn7-NatMX*	This study.
ACY553	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 535.6::Tn7-NatMX*	This study.
ACY554	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 613.19::Tn7-NatMX*	This study.
ACY555	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 645.10::Tn7-NatMX*	This study.
ACY556	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 653.24::Tn7-NatMX*	This study.
ACY557	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 607.19::Tn7-NatMX*	This study.
ACY558	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 537.10::Tn7-NatMX*	This study.
ACY559	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 658.34::Tn7-NatMX*	This study.
ACY560	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 524.7::Tn7-NatMX*	This study.
ACY561	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 618.26::Tn7-NatMX*	This study.
ACY562	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 624.38::Tn7-NatMX*	This study.
ACY563	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 635.33::Tn7-NatMX*	This study.
ACY564	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 635.59::Tn7-NatMX*	This study.
ACY565	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 557.12::Tn7-NatMX*	This study.
ACY566	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 607.22::Tn7-NatMX*	This study.
ACY567	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 638.43::Tn7-NatMX*	This study.
ACY568	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 630.21::Tn7-NatMX*	This study.
ACY569	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 647.1::Tn7-NatMX*	This study.
ACY570	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 649.2::Tn7-NatMX*	This study.
ACY571	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 469.4::Tn7-NatMX*	This study.

ACY572	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 490.3::Tn7-NatMX*	This study.
ACY573	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 668.17::Tn7-NatMX*	This study.
ACY574	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 674.15::Tn7-NatMX*	This study.
ACY575	*S. bayanus var. uvarum*	*S. bayanus var. uvarum MAT***a** *hoΔ::HisG 583.16::Tn7-NatMX*	This study.
ACY624	*S. bayanus var. uvarum*	*MATα hoΔ::HisG 610.13::Tn7-NatMX*	This study. Insertion confers copper resistance, and is between ORFs *610.12 (SbayPEX2)* and *610.13 (SbayOPT1)* but functionally disrupts *610.13*.
ACY639	*S. cerevisiae*	*MAT***a** *can11Δ::MFA1prHIS3 lyp1Δ leu2Δ0 his3Δ1 ura3Δ0 met15Δ0 opt1Δ::KanMX*	This study. Retrieved from random spore analysis of backcross to "Magic Marker" deletion set (Pan et al. 2004, PMID 15525520)
ACY671	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 610.13distalΔ:KanMX*	This study. Deletion of terminal 500 nt of *Sbayopt1* to check for effects on divergently transcribed *SbayPEX2*.
ACY681	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 610.13Δ:KanMX*	This study. *Sbayopt1* deletion.
ACY677	*S. bayanus var. uvarum*	*MATα 670.20Δ::KanMX hoΔ::HisG lys2*	This study. Haploid segregant from heterozygous diploid knockout of *670.20*.
ACY678	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 670.20Δ::KanMX*	This study. Haploid segregant from heterozygous diploid knockout of *670.20*.
ACY703	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 555.11Δ::KanMX (SbayGAL80)*	This study.
ACY711	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG 643.11::Tn7-NatMX (SbayRPM2)*	This study.
ACY732	*S. bayanus var. uvarum*	*MATα hoΔ::HisG 670.20Δ::KanMX*	This study. Haploid segregant from heterozygous diploid knockout of *670.20*.
ACY888	*S. bayanus var. uvarum*	*MATα Tn7:trp3*	This study. Haploid segregant from heterozygous diploid knockout.
ACY957	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG lys2-1 670.55Δ::KanMX*	This study.
ACY958	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG lys2-1*	This study. Sister spore of ACY958.
ACY959	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG his3Δ1 670.55Δ::KanMX*	
ACY1354	*S. bayanus var. uvarum*	*MAT***a***/MATα hoΔ::KANMX/hoΔ::KANMX lys2-1/+ ura3Δ::cloNAT/+*	Douglas Koshland, derived from type strain (MCYC623) from Duncan Greig
ACY1355	*S. bayanus var. uvarum*	*MAT***a***/MATα HO/hoΔ::NatMX*	Jasper Rine, isogenic to strains in Zill 2008
ACY1356	*S. bayanus var. uvarum*	*MAT***a***/MATα*	This study.
ACY1356	*S. bayanus var. uvarum*	*MAT***a***/MATα HO/HO*	Cletus Kurtzman, NRRL-Y11845 (aka MCYC623 or CBS 7001)
ACY1357	*S. bayanus var. uvarum*	*MATα hoΔ::HisG his3*	This study. Haploid segregant from heterozygous diploid knockout of 670.20.
ACY1358	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG lys2 his3*	This study. Haploid segregant from heterozygous diploid knockout of 670.20.
ACY1361	*S. bayanus var. uvarum*	*MAT***a** *hoΔ::HisG*	This study. Haploid segregant *S. bayanus var. uvarum* A1 from heterozygous diploid knockout of

			555.11.
ACY1362	*S. bayanus var. uvarum*	*MAT*α *555.11*Δ*::KanMX ho*Δ*::HisG lys2 his3*	This study. Haploid segregant A2 from heterozygous diploid knockout of *555.11*.
ACY1363	*S. bayanus var. uvarum*	*MAT*α *ho*Δ*::HisG*	This study. Haploid segregant A3 from heterozygous diploid knockout of *555.11*.
ACY1364	*S. bayanus var. uvarum*	*MAT***a** *555.11*Δ*::KanMX ho*Δ*::HisG lys2 his3*	This study. Haploid segregant A4 from *555.11* heterozygous diploid knockout.
ACY1365	*S. bayanus var. uvarum*	*MAT*α *ho*Δ*::KanMX lys2-1*	This study. Segregant 6B-A1 from ACY3 crossed to ACY1356.
ACY1366	*S. bayanus var. uvarum*	*MAT*α *ho*Δ*::KanMX lys2-1*	This study. Segregant 6B-A2 from ACY3 crossed to ACY1356.
ACY1367	*S. bayanus var. uvarum*	*MAT***a** *ho*Δ*::KanMX*	This study. Segregant 6B-A3 from ACY3 crossed to ACY1356.
ACY1368	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO lys2-1/lys2-1*	This study. Segregant 6B-A4 from ACY3 crossed to ACY1356.
ACY1369	*S. bayanus var. uvarum*	*MAT*α *ho*Δ*::KanMX lys2-1*	This study. Segregant 6B-A5 from ACY3 crossed to ACY1356.
ACY1370	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO*	This study. Segregant 6B-A6 from ACY3 crossed to ACY1356.
ACY1371	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO*	This study. Segregant 6B-A7 from ACY3 crossed to ACY1356.
ACY1372	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO*	This study. Segregant 6B-A8 from ACY3 crossed to ACY1356.
ACY1373	*S. bayanus var. uvarum*	*MAT*α *ho*Δ*::KanMX lys2-1*	This study. Segregant 6B-A9 from ACY3 crossed to ACY1356.
ACY1374	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO*	This study. Segregant 6B-A10 from ACY3 crossed to ACY1356.
ACY1375	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO lys2-1/lys2-1*	This study. Segregant 8D-A1 from ACY3 crossed to ACY1356.
ACY1376	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO*	This study. Segregant 8D-A2 from ACY3 crossed to ACY1356.
ACY1377	*S. bayanus var. uvarum*	*MAT*α *ho*Δ*::KanMX lys2-1*	This study. Segregant 8D-A3 from ACY3 crossed to ACY1356.
ACY1378	*S. bayanus var. uvarum*	*MAT***a** *ho*Δ*::KanMX lys2-1*	This study. Segregant 8D-A4 from ACY3 crossed to ACY1356.
ACY1379	*S. bayanus var. uvarum*	*MAT***a** *ho*Δ*::KanMX lys2-1*	This study. Segregant 8D-A5 from ACY3 crossed to ACY1356.
ACY1380	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO*	This study. Segregant 8D-A6 from ACY3 crossed to ACY1356.
ACY1381	*S. bayanus var. uvarum*	*MAT***a** *ho*Δ*::KanMX lys2-1*	This study. Segregant 8D-A7 from ACY3 crossed to ACY1356.
ACY1382	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO*	This study. Segregant 8D-A8 from ACY3 crossed to ACY1356.
ACY1383	*S. bayanus var. uvarum*	*MAT***a***/MAT*α *HO/HO lys2-1/lys2-1*	This study. Segregant 8D-A9 from ACY3 crossed to ACY1356.
ACY1384	*S. bayanus var. uvarum*	*MAT***a** *ho*Δ*::KanMX lys2-1*	This study. Segregant 8D-A10 from ACY3 crossed to ACY1356.
ACY1493	*S. bayanus var. uvarum*	*MAT***a** *ho*Δ*::HisG 610.12*Δ*:KanMX*	This study. Deletion of *Sbaypex2.*

Table S2 List of experimental treatments in this *S. bayanus* gene expression compendium.

Condition	number of datasets (arrays)
growth at different temperatures	1 (3)
heat shock	4 (18)
ammonium	1 (6)
cadmium	1 (4)
copper	1 (6)
lead	1 (6)
nickel	1 (5)
sulfite toxicity	1 (6)
zinc	1 (6)
ethanol toxicity	1 (6)
sorbitol	1 (6)
bleach	1 (6)
hydrogen peroxide	3 (18)
2-deoxyglucose	1 (6)
hydroxyurea	1 (6)
lovastatin	1 (4)
MG-132	1 (5)
MMS	1 (6)
rapamycin	1 (6)
tunicamycin	1 (6)
zeocin	1 (6)
chronological aging	3 (13)
diauxic shift	1 (6)
galactose	1 (5)
glycerol	1 (4)
sucrose	1 (4)
auxotroph starvation	1 (11)
nutrient limited chemostat growth	3 (7)
mating type and ploidy	1 (3)
alpha factor	1 (8)
cell cycle	1 (30)
sporulation	3 (18)
strain backgrounds	1 (4)
cross progeny	1 (22)
Tn7 insertions	1 (27)
Total	46 (303)

Table S3 Filtered and transformed expression data from Figure 1.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152918/-/DC1.

Table S4 Gene function predictions for *S. bayanus*. We report the probability for each gene of belonging to each GO term, based on the SVM analysis and probability estimation.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152918/-/DC1.

Table S5 Changes in predicted gene function between *S. bayanus* and *S. cerevisiae*, filtered for those in the top probability score in either species.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152918/-/DC1.

Table S6 Position weight matrices of the motifs enriched in the oxidative stress cluster. These motifs were determined using the SCOPE server (http://genie.dartmouth.edu/scope/).

PWM: Similar to K31 from Kellis, M. et al., in Nature (2003), with p value 1.61×10^{-5} as determined by STAMP.

	m	t	k	a	s	t	a	a
a	29	0	0	54	0	0	54	54
c	25	0	0	0	32	0	0	0
g	0	0	9	0	22	0	0	0
t	0	54	45	0	0	54	0	0

PWM: Similar to K25 with 7×10^{-5} from Kellis, M. et al., in Nature (2003), with p value 1.61×10^{-5} as determined by STAMP.

	c	n	c	g	g	m
a	0	12	0	0	0	25
c	49	8	49	0	0	24
g	0	17	0	49	49	0
t	0	12	0	0	0	0

Supporting References

BELLÍ, G., E. GARÍ, M. ALDEA and E. HERRERO, 1998 Functional analysis of yeast essential genes using a promoter-substitution cassette and the tetracycline-regulatable dual expression system. Yeast (Chichester, England) **14:** 1127-1138.

BOZDECH, Z., J. ZHU, M. P. JOACHIMIAK, F. E. COHEN, B. PULLIAM *et al.*, 2003 Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. Genome Biol **4:** R9.

CARLSON, J. M., A. CHAKRAVARTY, C. E. DEZIEL and R. H. GROSS, 2007 SCOPE: a web server for practical de novo motif discovery. Nucleic Acids Research **35:** W259-264.

CLIFTEN, P., P. SUDARSANAM, A. DESIKAN, L. FULTON, B. FULTON *et al.*, 2003 Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. Science **301:** 71-76.

CLIFTEN, P. F., R. S. FULTON, R. K. WILSON and M. JOHNSTON, 2006 After the duplication: gene loss and adaptation in *Saccharomyces* genomes. Genetics **172:** 863-872.

GRESHAM, D., M. M. DESAI, C. M. TUCKER, H. T. JENQ, D. A. PAI *et al.*, 2008 The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet **4:** e1000303.

GUAN, Y., M. DUNHAM, A. CAUDY and O. TROYANSKAYA, 2010 Systematic planning of genome-scale experiments in poorly studied species. PLoS Computational Biology **6:** e1000698.

GUAN, Y., C. MYERS, D. HESS, Z. BARUTCUOGLU, A. CAUDY *et al.*, 2008 Predicting gene function in a hierarchical context with an ensemble of classifiers. Genome Biol **9 Suppl 1:** S3.

HESS, D., W. LU, J. RABINOWITZ and D. BOTSTEIN, 2006 Ammonium toxicity and potassium limitation in yeast. Plos Biol **4:** e351.

HU, Z., J. H. HUNG, Y. WANG, Y. C. CHANG and C. L. HUANG …, 2009 VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene …. Nucleic Acids Research.

JOACHIMS, T., 1999 Making large scale SVM learning practical. eldorado.uni-dortmund.de.

KELLIS, M., N. PATTERSON, M. ENDRIZZI, B. BIRREN and E. S. LANDER, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423:** 241-254.

MAHONY, S., P. AURON and P. BENOS, 2007 DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. PLoS Comput Biol **3:** e61.

RAINIERI, S., Y. KODAMA, Y. KANEKO, K. MIKATA, Y. NAKAO *et al.*, 2006 Pure and mixed genetic lines of *Saccharomyces bayanus* and *Saccharomyces pastorianus* and their contribution to the lager brewing strain genome. Appl Environ Microbiol **72:** 3968-3974.

RAO, R., D. DRUMMOND-BARBOSA and C. W. SLAYMAN, 1993 Transcriptional regulation by glucose of the yeast *PMA1* gene encoding the plasma membrane H(+)-ATPase. Yeast **9:** 1075-1084.

ROBERTS, C. J., B. NELSON, M. J. MARTON, R. STOUGHTON, M. R. MEYER *et al.*, 2000 Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science **287:** 873-880.

SALDANHA, A. J., M. J. BRAUER and D. BOTSTEIN, 2004 Nutritional homeostasis in batch and steady-state culture of yeast. Mol Biol Cell **15:** 4089-4104.

TEIXEIRA, M. C., P. MONTEIRO, P. JAIN, S. TENREIRO, A. R. FERNANDES *et al.*, 2006 The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. Nucleic Acids Research **34:** D446-451.

TROYANSKAYA, O., M. CANTOR, G. SHERLOCK, P. BROWN, T. HASTIE *et al.*, 2001 Missing value estimation methods for DNA microarrays. Bioinformatics **17:** 520-525.