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†, 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*‡, Maitreya J. Dunham***

* Co-corresponding authors

‡Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, M5S 3E1 Canada

†Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, 08544
‡Department of Computer Science, Princeton University, Princeton, NJ, 08540
**Department of Genome Sciences, University of Washington, Seattle, WA 98195
†Current Address: Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI 48109

GEO: GSE16544 and GSE47613

Running title: S. bayanus, a model for comparative genomics

Keywords: comparative genomics, yeast, gene expression

Please address correspondence to:
Maitreya Dunham
Department of Genome Sciences
Foege Building S403B, Box 355065
3720 15th Ave NE
Seattle, WA 98195-5065
206-543-2338
maitreya@uw.edu
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 dataset 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.
INTRODUCTION

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 last decade (reviewed in (DJON 2010)), and comparative analyses have revealed a history of gene duplication (DIETRICH et al. 2004; KELLIS et al. 2004), conservation at DNA binding sites (CLIFEN 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 *S. 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 (BORNMAN 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 NIEDUSZYNISKI 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 non-model 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 (HIBBS et al. 2007; HUTTENHOWER et al. 2007; STUART et al. 2003; VAN NOORT et al. 2003).

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 (BERGMANN et al. 2004; STUART et al. 2003), 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 (AIROLDI et al. 2009; HESS et al. 2006; ZILL and RINE 2008).

To examine the conservation and divergence of gene function, we selected the yeast *S. bayanus* var. *uvarum* (henceforth referred to as *S. bayanus* for simplicity) for comparison with *S. cerevisiae*. The two species diverged approximately 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 (Gallagher et al. 2009; Jones et al. 2008; Serra et al. 2003; Talarek et al. 2004; Zill and Rine 2008; Zill et al. 2010), little experimental work has been performed in S. bayanus, and even less at genome-scale (Bullard et al. 2010; Busby et al. 2011; Guan et al. 2011; Muller and Nieduszynski 2012; Tsankov et al. 2010).

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 have 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 co-expression can be used to predict the functional roles of genes (SHARAN et al. 2007).

Our analysis of these datasets reveals both regulatory change and evolution of gene function amid overall 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.

**METHODS**

*Methods summary*

The strains used in this study are described in 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](http://bayanusfunction.princeton.edu).

Complete methods information is included as a supplement.

### 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 1B,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 *S. cerevisiae* (Figure 1D, E). This heat sensitivity precludes efficient lithium acetate transformation using heat shock at 42°, so
we modified our procedure to use a milder $37^\circ$ heat shock for *S. bayanus* (see supplemental methods for details).

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 2 micron 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 approximately 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 *MATa* 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 *MATa* 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 dataset*

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 2-fold or greater change in at least one treatment.
Hierarchical clustering of this *S. bayanus* gene expression compendium revealed a number of groups of genes co-expressed under a variety of conditions (Figure 2, numerical data in Table S3). Although clustering was performed solely 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 showed 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 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 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 only induced 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 carbon sources including ethanol, raffinose, sucrose, and glycerol (Figure 3A). The derepression of galactose metabolism genes on non-glucose 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 non-glucose 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 non-glucose 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 over 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 is 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 *MATa* *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 *Shayime1* 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 datasets in both species. These limits can be overcome using computational interpretation of expression data, which accurately predicts gene function over much larger datasets 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 AMNI. 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 MATa insertion mutant of Sbayard1, we observed repression of MATa 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 4A,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
The functional predictions for \textit{SbayOPT1} include cation homeostasis, the GO parent term that includes copper ion homeostasis (our functional predictions did not include GO terms with few members). \textit{ScerOPT1} (also named \textit{HGT1}) has been characterized as a high affinity glutathione transporter induced by sulfur starvation (BOURBOULOUX \textit{et al.} 2000; SRIKANTH \textit{et al.} 2005). Copper resistance had not been investigated in this mutant, although sensitivity to cadmium had been noted (SERERO \textit{et al.} 2008). The \textit{OPT1} mutant in \textit{S. cerevisiae} also showed increased resistance to copper (Figure 4E). Of note, \textit{S. bayanus} is more sensitive to copper than the laboratory strain of \textit{S. cerevisiae}; our screen in the sensitized background of \textit{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 \textit{S. bayanus} provide information about conserved gene function in \textit{S. cerevisiae}.

\textit{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 \textit{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 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 16–32 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 (Kitanovic et al. 2009; Winter et al. 2008), so we propose 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< 7x10^{-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 (Causton et al. 2001; Gasch et al. 2000). Five of the S. cerevisiae orthologs of this cluster have been assigned the GO biological process of response to toxin (GO enrichment, p<4.07x10^{-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 twelve proteins in this cluster that have S. cerevisiae homologs, five 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 7 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 (Cliften et al. 2006; Gordon et al. 2009; Hittinger et al. 2010; Hittinger et al. 2004; 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 non-glucose 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 only present in *S. bayanus*, forming a species-specific network.

**DISCUSSION**

Though the genomes of many non-model 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 dataset 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 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 in 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 non-model 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.

ACKNOWLEDGEMENTS

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. OGT is supported by the NSF CAREER award DBI-0546275, and by NIH R01 grants GM071966 and HG005998. MJD is supported in part by a grant from the National Institute of General Medical Sciences (8 P41 GM103533-17) from the National Institutes of Health. AAC is supported in part by grants from the Canadian Institutes for Health Research. All authors were supported by the 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. MJD is a Rita Allen Scholar and a CIFAR Fellow.

AUTHOR CONTRIBUTIONS

All authors participated in the design and execution of experiments through the educational activities of the NIGMS Center for Quantitative Biology at Princeton University. AAC, YG, MJD, and OGT analyzed the data and wrote the manuscript.
REFERENCES

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. Molecular Biology and Evolution 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. DELROIT and A. K. BACHHAWAT, 2000 Hgt1p, a high affinity glutathione transporter from the yeast Saccharomyces cerevisiae. The Journal of biological chemistry 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. Proceedings of the National Academy of Sciences 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. Molecular biology of the cell 12: 323-337.

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

CLIFEN, 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. Molecular Biology and Evolution 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. Nature Reviews Genetics 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 genetics 2: e212.

GALLAGHER, J. E. G., J. E. BABIARZ, L. TETTELMAN, 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., 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.

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.

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., 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., M. J. DUNHAM, O. TROYANSKAYA and A. CAUDY, 2013 Comparative gene expression between two yeast species. BMC Genomics 14: 33.

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., V. YAO, K. TSUI, M. GEBBIA, M. J. DUNHAM et al., 2011 Nucleosome-coupled expression differences in closely-related species. BMC Genomics 12: 466.

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 (Oxford, England) 23: 2692-2699.
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.

Hittinger, C. T., A. Rokas and S. B. Carroll, 2004 Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. Proceedings of the National Academy of Sciences of the United States of America 101: 14144-14149.

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.

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

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. Molecular systems biology 8: 602.

Kitanovic, A., T. Walther, M. O. Lore, 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 Research.

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 Research 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. Proceedings of the National Academy of Sciences of the United States of America 106: 5755-5760.

Libkind, D., C. T. Hittinger, E. Valério, C. Goncalves, J. Dover et al., 2011 Microbe domestication and the identification of the wild genetic stock of lager-
brewing yeast. Proceedings of the National Academy of Sciences 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. Molecular and cellular biology 1: 83-93.

Müller, C. A., and C. A. Nieduszynski, 2012 Conservation of replication timing reveals global and local regulation of replication origin activity. Genome Research 22: 1953-1962.

Osborn, M. J., and J. R. Miller, 2007 Rescuing yeast mutants with human genes. Briefings in Functional Genomics and 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 & Development 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. Applied and environmental microbiology 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 (Bethesda, Md.) 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 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. Current genetics 47: 345-358.
St John, T. P., and R. W. Davis, 1981 The organization and transcription of the galactose gene cluster of Saccharomyces. Journal of Molecular Biology 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 Biology 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. Snell and M. A. Huygen, 2003 Predicting gene function by conserved co-expression. Trends in genetics : TIG 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. Proceedings. Biological sciences / The Royal Society 273: 1507-1515.

Wu, A., J. A. Wemmie, N. P. Edgington, M. Goebi, J. L. Guevara et al., 1993 Yeast bZip proteins mediate pleiotropic drug and metal resistance. The Journal of biological chemistry 268: 18850-18858.

Yocom, 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. Molecular and cellular biology 4: 1985-1998.

Yocom, 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.

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 & Development 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 Biology 8: e1000550.
FIGURE LEGENDS

**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 minute 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 mM, except for Cadmium (10x µM), Ammonium (M), and Ethanol 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.

**Figure 2.** *S. bayanus* gene expression megacluster. 46 *S. bayanus* gene expression datasets are shown as indicated by color-coded experiment labels with genes hierarchically clustered along the other dimension. Each dataset was either zero-transformed or mean-centered to remove the reference. By assuming 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. Datasets are ordered to group similar conditions. The data as plotted are available in Table S3.
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 in 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.

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 5 mM CuSO₄. 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 CuSO₄. 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.

**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. 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 log₂-ratios of expression data from microarrays, zero-transformed to the initial timepoint in wild type cells.
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.
A. Heat shock, osmotic stress, ethanol toxicity, ammonium toxicity, sulfate toxicity, cadmium toxicity, copper toxicity, lead toxicity, nickel toxicity, hydrogen peroxide, bleach, MMS, zeocin, hydroxyurea, tunicamycin, MG-132, Lovastatin, 2-deoxyglucose, Rapamycin.

| S. bayanus gene | S. bayanus inferred Biological Process | S. cerevisiae ortholog | S. cerevisiae Biological process Annotations |
|-----------------|--------------------------------------|------------------------|---------------------------------------------|
| 658.54          | vitamin metabolism, sulfur metabolism | OYE3                   | response to toxin                            |
| 380.3           | 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           | mitotic cell cycle, M phase |                       |                                             |
| 601.1           | cofactor metabolism, vitamin metabolism | GTT2                   | sulfur metabolic process, cofactor metabolic process, response to toxin |
| 453.5           | response to toxins, proteolysis and peptidolysis |                       |                                             |
| 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 |                       |                                             |
| 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, mRNA splicing |
| 591.11          | drug transport, transcription from RNA polymerase II promoter | YDR132C                 |                                             |
| 565.14          | cofactor metabolism, proteolysis and peptidolysis | YLR346C                 | response to toxin                            |

B. 640.6 S. bayanus unique, 602.1 S. bayanus unique, 621.2 S. bayanus unique, 602.5 S. bayanus unique.