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*

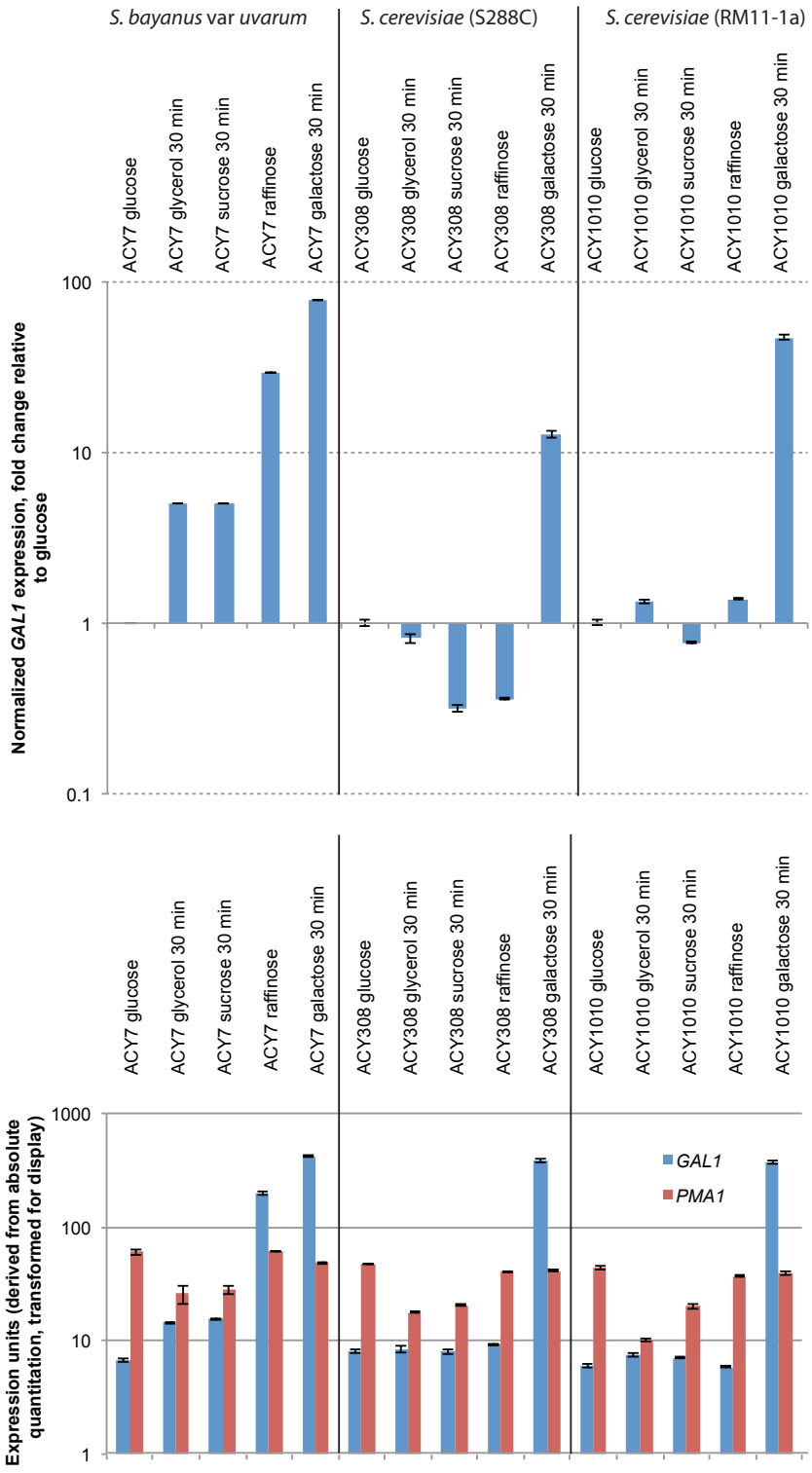


Figure S1 *GAL1* expression in *S. bayanus* is increased during carbon switches. A. Expression of *GAL1* normalized to *PMA1* expression. The indicated *S. bayanus* and *S. cerevisiae* strains were grown in minimal medium containing glucose and switched to minimal medium containing sucrose or glycerol, 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*.

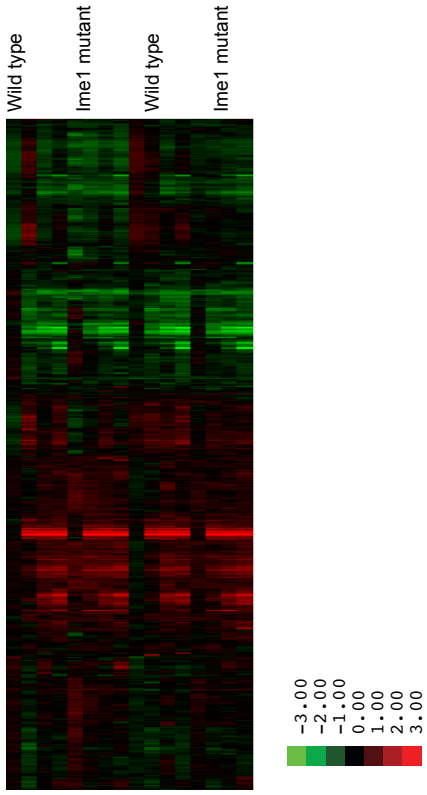


Figure S2 Pheromone response in *670.55* knockout mutant (ortholog of *IME1*) is similar to wild type. The indicated mutant and wild type strains were grown in YPD and exposed to alpha factor pheromone. Samples were harvested for expression analysis. A heat map of the genome-wide expression data is presented,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

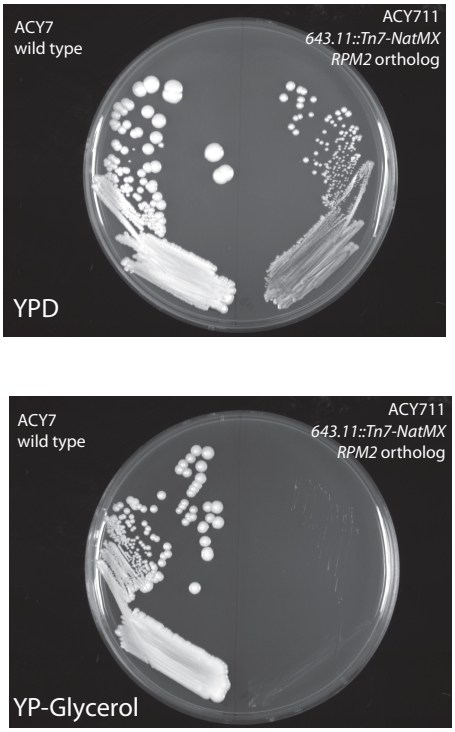
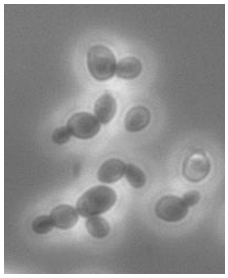
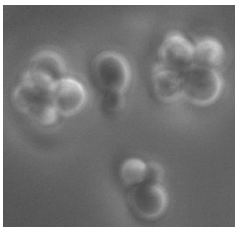


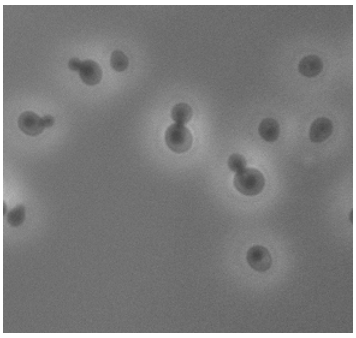
Figure S3 The *643.11* knockout mutant (ortholog of *RPM2*) is respiratory deficient. The *643.11* knockout strain ACY711 and a wild type strain were struck for growth on SC glycerol and on YNB glucose medium.



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.

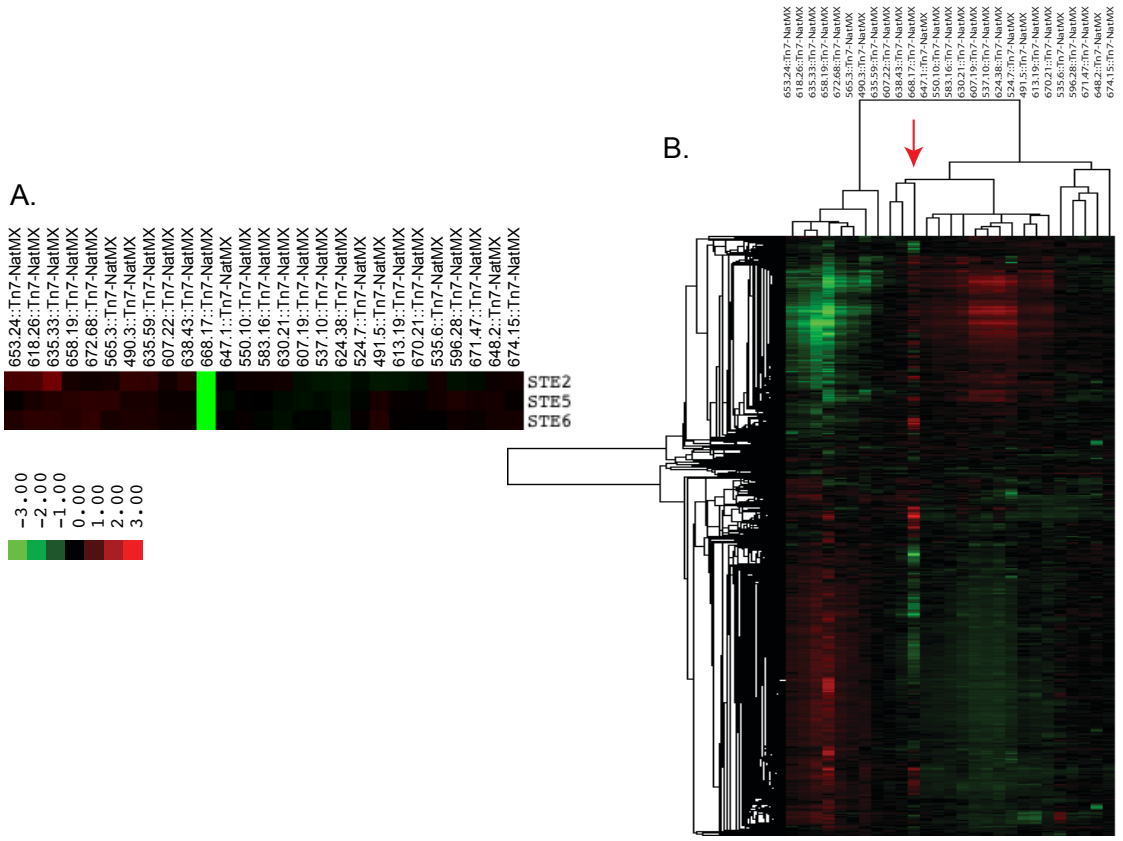


Figure S5 The *668.17* knockout mutant (ortholog of *ARD1)* shows changes in gene expression of mating-type specific genes, as well as global changes in expression. A. The data for the Tn7 insertion strains were selected from the megacluster (data in Table S3), and expression levels of *STE2, STE5,* 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.

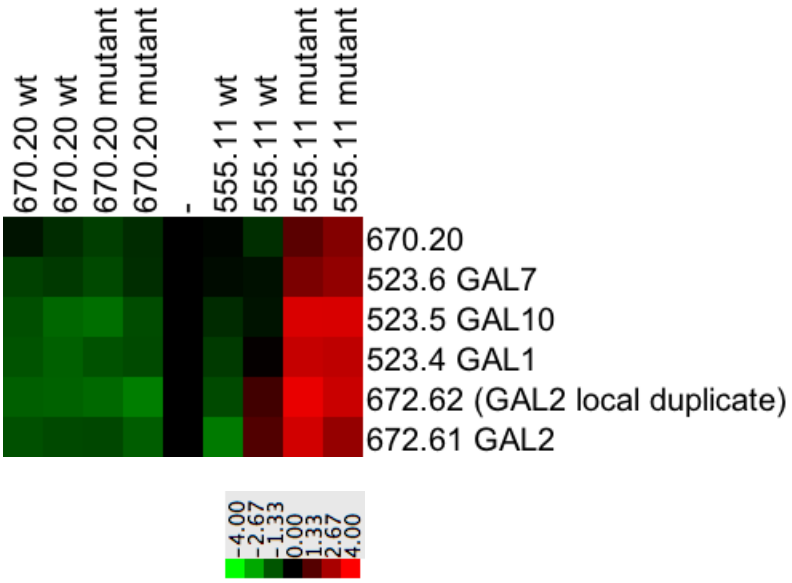
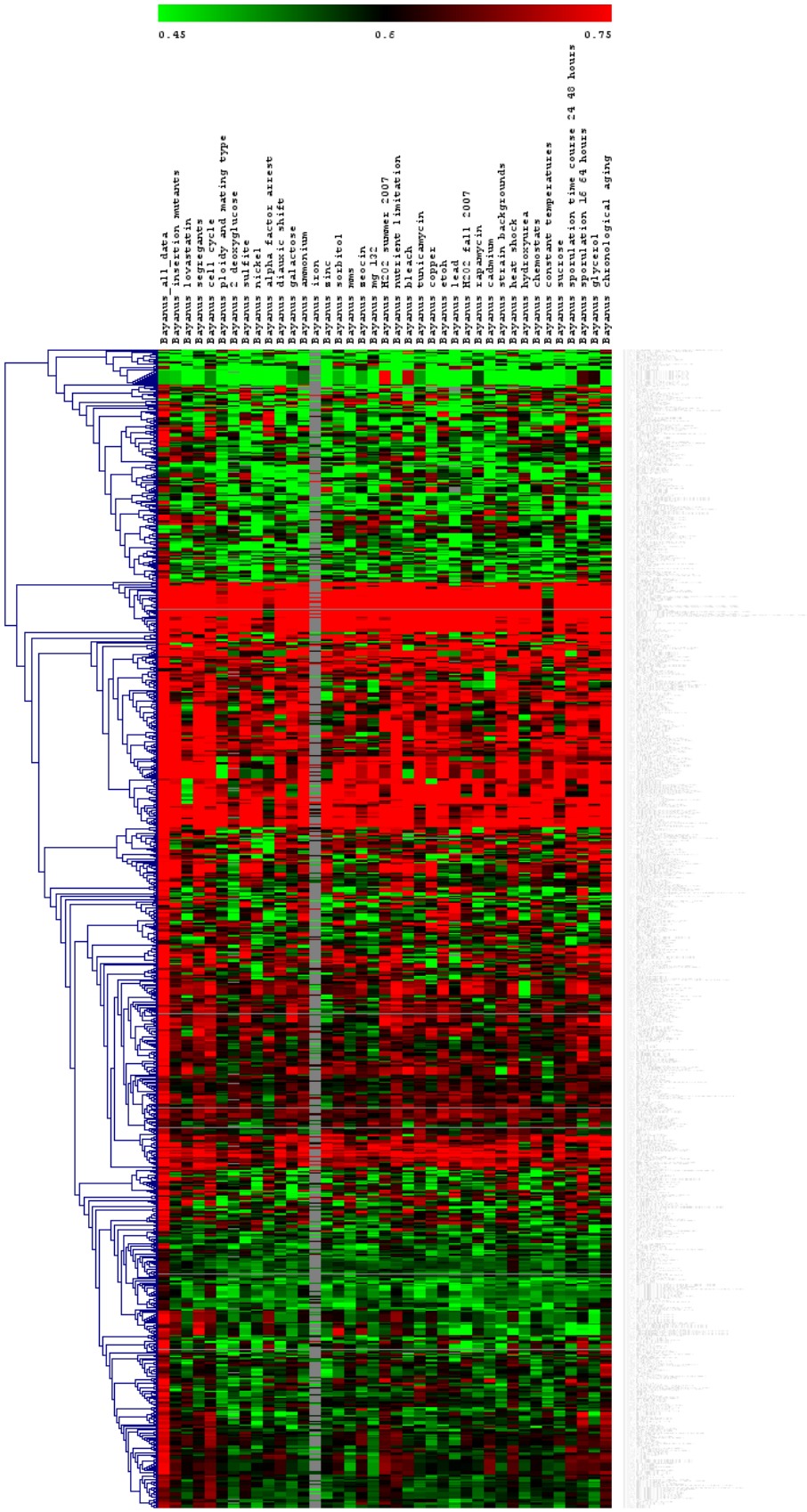


Figure S6 The *S. bayanus GAL80* ortholog *555.11* represses the expression of galactose structural genes when glucose is present. The indicated strains were grown in YPD to mid log phase and harvested for RNA. The expression of a set of genes whose orthologs in *S. cerevisiae* are involved in galactose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

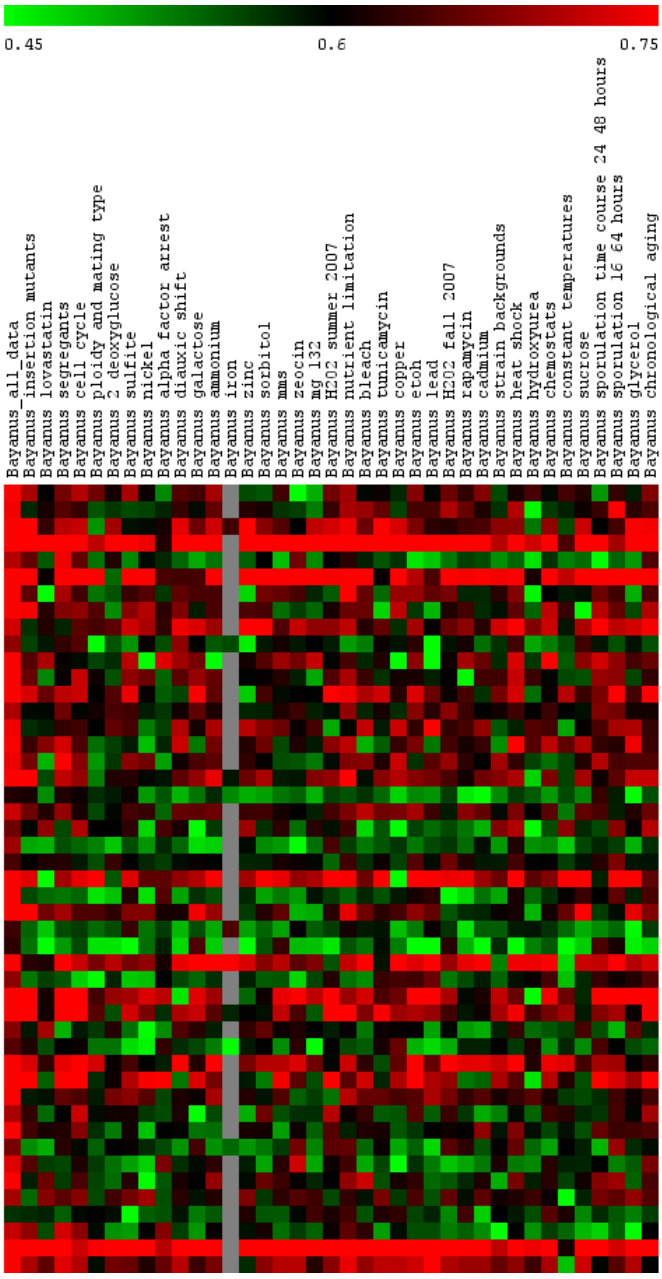


Figure S7 The AUC (area under the precision/recall curve) for various biological functions predicted from *S. bayanus* gene expression data, shown for A. all GO terms and B. GO SLIM terms.

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.