

# Supporting Information

Tirosh et al. 10.1073/pnas.1113718108

## SI Materials and Methods

**Strains and Growth Conditions.** Strains used in this study are listed in Table S1. Yeasts were grown at 28 °C in rich media (YP: 1% Yeast extract and 2% Peptone) supplemented with 2% dextrose as the carbon source (YPD) until midlog phase (OD<sub>600</sub> ~ 0.6), and were then subjected to three different stresses separately. For osmotic shock, yeasts were treated with 0.4 M NaCl at 30 °C for 5 min. For heat shock, cells were transferred to a prewarmed flask at 39 °C for 10 min. For carbon starvation, cells were harvested by centrifugation, washed twice with and continued to grow in prewarmed YP media for 6 h. For each condition, 10 mL of cells were collected on ice and immediately harvested by centrifugation at 4 °C and the cell pellet was kept frozen at –80 °C before RNA extraction.

**RNA Isolation, First-Strand cDNA Synthesis and Library Construction for Sequencing Using the HeliScope Sequencer.** Total crude RNAs were isolated using the hot acid phenol method. Briefly, cell pellet was resuspended in 600 µL of TES buffer (10 mM Tris, pH 7.5/10 mM EDTA/0.5% SDS) and mixed with 500 µL of acid phenol (Ambion). The mixture was incubated at 65 °C for at least 1 h with vortexing every 10 min. The suspension was then cooled on ice for 5 min and followed by centrifugation at room temperature for 5 min. Approximately 500 µL of aqueous phase was mixed with 400 µL of fresh acid phenol, vortexed, and centrifuged for 5 min at room temperature. This step was repeated once, and followed by ethanol precipitation at –80 °C for at least 1 h. RNA was pelleted by centrifugation for 20 min at 4 °C and washed with cold 70% ethanol, and the air-dried pellet was resuspended in 100–200 µL of TE. poly(A)+ RNA was purified from total crude RNA using the Oligotex mRNA mini kit (Qiagen) following the manufacturer's instruction. For one of the two biological repeats, two rounds of poly(A)+ RNA purification were performed to further reduce rRNA contamination. The poly(A)+ RNA was quantified using Agilent BioAnalyzer RNA pico assay at the Biopolymers facility of Harvard Medical School.

To allow for comparisons among different yeast samples without the concern of sample-to-sample variation from reactions of first strand cDNA synthesis and sequencing library construction, equal mass of poly(A)+ RNA of different yeasts from the same growth conditions was pooled together and then subjected to the following treatments. First strand cDNA synthesis was performed using the Invitrogen SuperScript III RT system. Approximately 100 µg of the mixed polyA+ RNA was concentrated, when needed, by speedvac to 100 µL volume and was fragmented by heating at 95 °C for 5 min. Subsequently, the fragmented poly(A)+ RNA was mixed with 2 µg of random hexamers and 10 µL of 10 mM dNTP, incubated at 65 °C for 5 min, and then chilled on ice while the reaction mix (40 µL of Invitrogen SuperScript III 5× first-strand buffer, 10 µL of 0.1 M DTT) was added. Actinomycin D (Sigma) was included in the reverse transcription reaction at a final concentration of 9 µg/µL to suppress the synthesis of second-strand cDNA by reverse transcriptase (RT). The mixture was incubated at 15 °C for 15 min before the addition of 10 µL of RNaseOUT (Invitrogen) and 2,000 units of Invitrogen SuperScript III RT. The reaction was incubated at 25 °C for 10 min, 40 °C for 40 min, 50 °C for 50 min, and 70 °C for 10 min. Upon completion of the reaction, samples were treated with 1 µL of RNase mixture (Ambion) and RNase H (Invitrogen) at 37 °C for 30 min. First-strand cDNA was cleaned up twice using Performa Spin Columns (EdgeBio) with each sample loaded onto four separate columns in each round of purification. The purified cDNA was concentrated by speedvac to a volume of ~50 µL and quantified by Nanodrop.

For library preparation for sequencing using the HeliScope sequencer, 100 ng of cDNA was mixed with 5 µL of Helicos poly (A) Control Oligonucleotides A, heated to 95 °C for 5 min and quickly cooled on ice for 2 min. The following reaction mix [5 µL each of CoCl<sub>2</sub>, 10× TdT buffer, and poly(A) tailing dATP (Helicos) and 2 µL of terminal transferase (NEB)] was then added before incubating at 42 °C for 1 h, 70 °C for 10 min and 4 °C until the next step. A final 3' blocking step was performed on the poly (A)-tailed cDNA sample by heating at 95 °C for 5 min, then cooling on ice for 2 min before the addition of 0.4 µL of biotin-ddATP (PerkinElmer) and 2 µL of Terminal transferase (NEB). The blocking reaction was carried out at 37 °C for 1 h and Terminal transferase was inactivated at 70 °C for 10 min. The resultant library was kept at 4 °C in dark and was quantified and sequenced using the HeliScope sequencer at the Molecular Biology Core Facility of Dana Farber Cancer Institute.

**Sequence Mapping.** Raw data were mapped using bioinformatics tools available from the HeliSphere bioinformatics software package released by the Helicos BioSciences Corporation (<http://open.helicosbio.com>). The genome sequences of *S. cerevisiae*, *C. albicans*, *K. lactis*, and *Y. lipolytica* were obtained from SGD (<http://www.yeastgenome.org/>), CGD (<http://www.candidagenome.org/>), and Genolevures (<http://www.genolevures.org/>), and were combined to one reference file, which was put through the preprocessDB LINUX script with an optional seed size setting at 18. Raw sequencing data were mapped against the processed combined reference genome (NCBI accession number SRA029166.1, SRP005689) together with an assignment containing annotation information of the four yeasts using the DGE pipeline. To avoid ambiguity, reads were mapped only to regions with the best alignment scores, and reads that can be aligned equally to more than one region within the combined reference genome would be excluded from the analysis. The following parameters were used for the DGE pipeline: Global.minLength 10; Global.minScore 4; Global.bestOnly True; Global.globalAmbig none.

**Comparison of Expression Response Between Species.** Expression levels at each condition were averaged over two biological repeats for *S. cerevisiae* and *K. lactis*. For *C. albicans* and *Y. lipolytica* we noticed that one replicate is of lower quality than the other: one *Y. lipolytica* replicate had very low coverage, and one *C. albicans* replicate had very low similarity ( $r = 0.15–0.3$ ) with previous microarray stress data (1). For these species we decided to include only the apparent higher-quality dataset in most of the analysis, although averaging of the two replicates does not change our main results (Fig. S11). In Fig. S4 we used one replicate to estimate basal (YPD) mRNA levels and the other replicate to estimate stress-response (stress levels divided by YPD levels), to avoid spurious correlations between these measures.

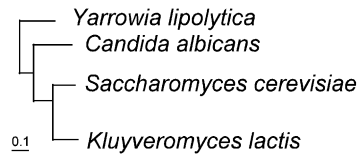
The expression levels in each stress were divided by the expression at YPD and the resulting log<sub>2</sub> ratios were centered to mean of zero. Response to a particular stress was defined as any ratio above 1.75-fold (increase or decrease) and *P* value (binomial test) below 0.05, and ESR genes were defined as those responding to all three stresses in a given species. We chose to focus on a 1.75-fold threshold, rather than the more common twofold threshold, both to examine more expression differences and because much of our analysis requires a margin of error around that threshold to identify species-specific responses. These were defined as genes with response >2-fold (and  $P < 0.05$ ) in one species and <1.4-fold in others. Exclusion of genes with low overall read counts (for

which the estimation of fold change is less accurate) did not significantly alter our results. In each pairwise species comparison we analyzed one-to-one orthologs as defined by Wapinski (2), and

comparisons among more than two species were restricted to genes with one-to-one orthologs among all relevant pairwise comparisons.

1. Enjalbert B, Nantel A, Whiteway M (2003) Stress-induced gene expression in *Candida albicans*: Absence of a general stress response. *Mol Biol Cell* 14:1460–1467.

2. Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449:54–61.



**Fig. S1.** Phylogenetic tree of the four yeast species, based on evolutionary distances from Fitzpatrick et al. (1). Although *S. cerevisiae* has been widely used as a model yeast to study many biological phenomena including responses to stresses, the other species are relatively less well studied. *C. albicans* is an opportunistic pathogen, which can often be found on skin and in intestinal tract of human. On the other hand, *K. lactis* are frequently found in dairy products. Being regarded as safe for use in food industries, *K. lactis* is widely used as a source of lactase for production of lactose-free products. In addition, *K. lactis* is also widely used for heterologous protein expression both in laboratories as well as in industries. Similarly, *Y. lipolytica* is regarded as safe and has been used for protein expression in industries. *Y. lipolytica* can be found in oil-contaminated environment, and due to its ability to degrade and grow on hydrocarbons, it has also emerged as a useful species in bioremediation (2).

1. Fitzpatrick DA, Logue ME, Stajich JE, Butler G (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol Biol* 6:99.  
2. Bankar AV, Kumar AR, Zinjarde SS (2009) Environmental and industrial applications of *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 84:847–865.





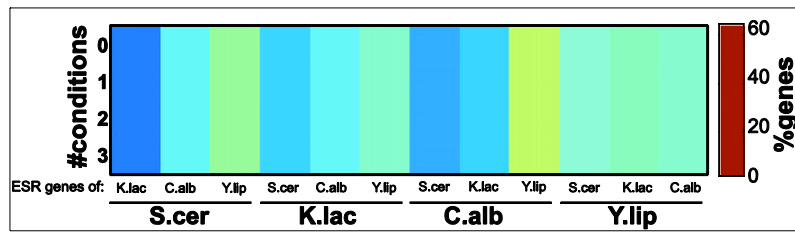


Fig. S5. Down-regulated genes. Number of conditions in which down-regulated genes from one species (Upper) are down-regulated in another species (Lower).

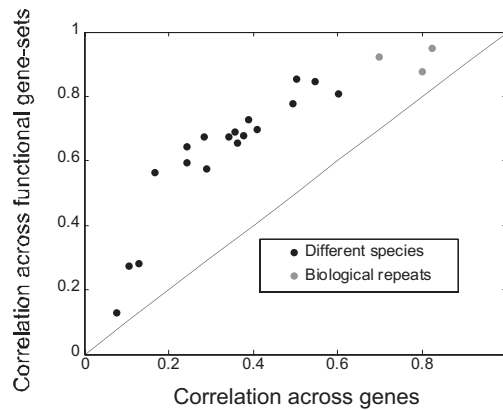


Fig. S6. Analysis of functional gene categories. Correlations between the responses of each species-pair to each stress over all one-to-one orthologs (x axis) and over the averages of functional gene-sets (GO terms) with at least 20 and at most 100 annotated genes (y axis). Gray dots reflect a similar comparison of biological repeats of *S. cerevisiae* (instead of different species).

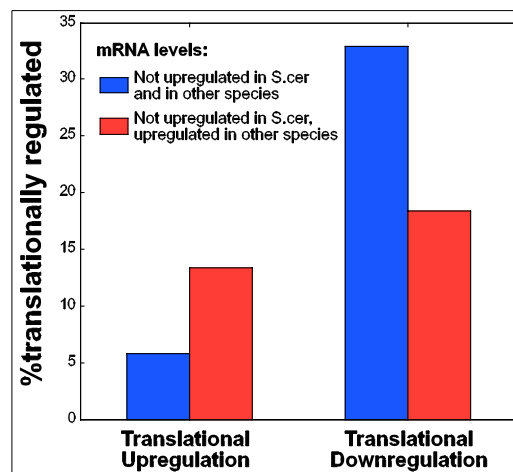
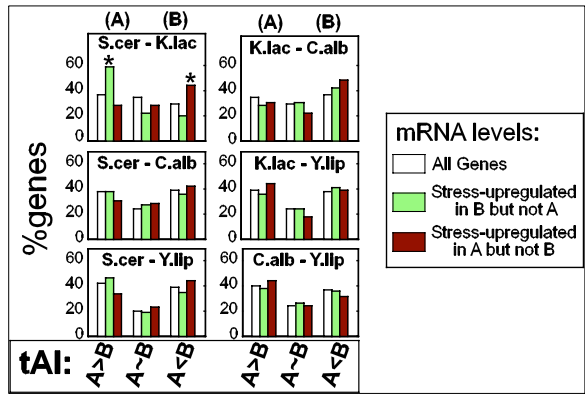


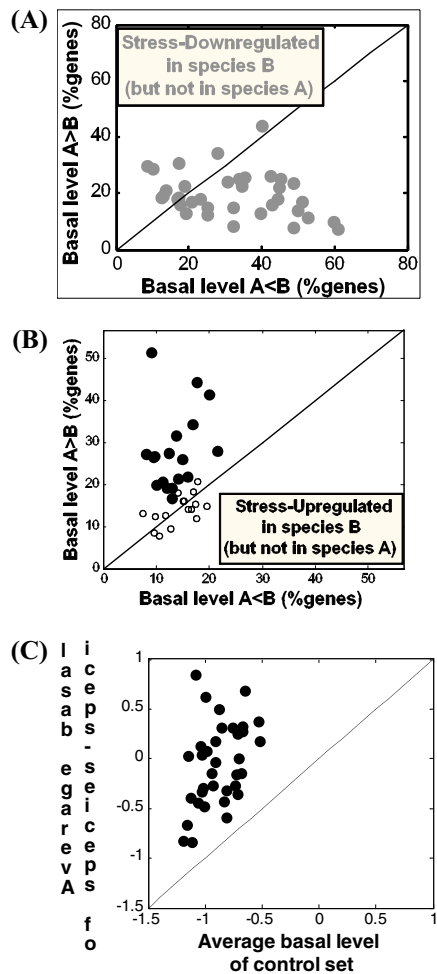
Fig. S7. Compensation by translational regulation. The percentage of genes whose ribosome association increases (translational up-regulation) or decreases (translational down-regulation) in osmotic shock (1 M sorbitol for 10 min) (1), among genes whose mRNA levels are not up-regulated (less than 1.4-fold increase) in any of the four species (blue), and among genes whose mRNA levels are up-regulated (more than 1.75-fold) in at least one species but not in *S. cerevisiae* (red). Genes up-regulated in non-*S. cerevisiae* species (red) are significantly enriched with translational up-regulation and significantly depleted with translational down-regulation ( $P < 0.01$ ), suggesting that some evolutionary changes in mRNA up-regulation are compensated by translational up-regulation. Such effects were not observed in analysis of ribosome association upon heat shock (2), in which only 158 genes were detected as being translationally up-regulated and these did not include any of the genes whose mRNA levels are up-regulated only in the non-*S. cerevisiae* species.

- Halbeisen RE, Gerber AP (2009) Stress-Dependent Coordination of Transcriptome and Translatome in Yeast. *PLoS Biol* 7:e105.
- Preiss T, Baron-Benhamou J, Ansong W, Hentze MW (2003) Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock. *Nat Struct Biol* 10:1039–1047.



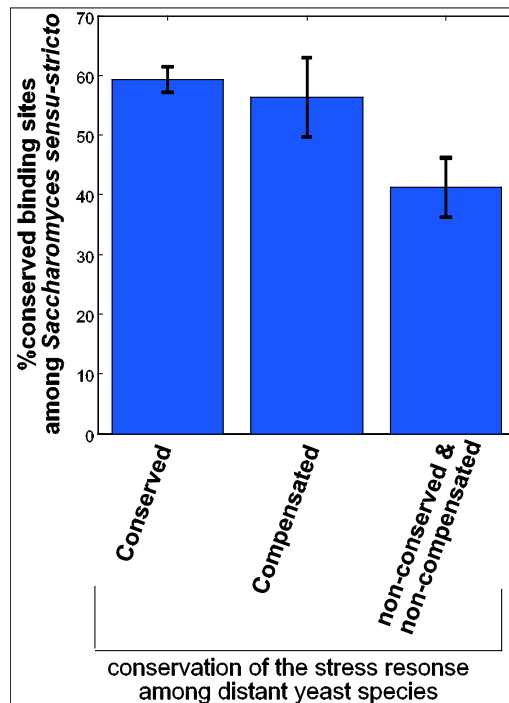
**Fig. 58.** Compensation by putative translation efficiency. For each pair of species (A and B), we identified genes whose mRNA levels are up-regulated in at least two stresses in one species but not up-regulated in any of the stresses at the other species (see color legend), and classified them according to changes in translation Adaptation Index (tAI) which presumably reflects translation efficiency (1). Genes were defined as having differential tAI by a threshold of 1.5-fold. Overall, genes that respond to stress in one species tend to have higher tAI in the other species (on average of all pairwise comparisons,  $P < 0.05$ ). This effect is significant ( $P < 0.05$ ) for the average of all species comparisons (37% of all genes have  $A < B$  tAI compared with 41% of the genes up-regulated only in A) and for the *S. cerevisiae* vs. *K. lactis* comparison, but is not statistically significant in all other species comparisons.

1. Man O, Pilpel Y (2007) Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species. *Nat Genet* 39:415–421.



**Fig. S9.** Additional analysis of the transitions between induced and constitutive activation. (A and B) Similar analysis as in Fig. 4 for down-regulation (instead of up-regulation; A) and when basal mRNA levels are taken from Tsankov et al. (1) (B). Empty circles in B reflect analysis of shuffled gene-pairs (i.e., comparison of nonorthologs). (C) Genes that respond to stress tend to have lower levels (before the stress) than other genes. To verify that the lower-than-average basal levels of the species-specific stress genes is not enough to account for their increased basal levels in other species, we performed the following analysis: For each pair of species and each stress we identified the genes responding to the stress in A but not B and examined the average basal level of these genes in A. We selected a subset of A genes that do not respond to the respective stress and that have the same average basal level as the genes in *i*. We compared the average basal levels of the B orthologs of the gene-sets defined in *i* and *ii*. The average  $\log_2$  of the normalized basal levels (zero reflects the genome-wide-average) is shown for the orthologs (in B) of the control gene-sets defined in *ii* and the species-specific stress genes defined in *i*. In all 36 analyses (12 species pairs times 3 stresses), the basal levels of the B orthologs of the species-specific stress genes defined in *i* were higher than those of the control genes defined in *ii*. In 35 of the 36 comparisons, the difference was significant ( $P < 0.05$ , *t* test).

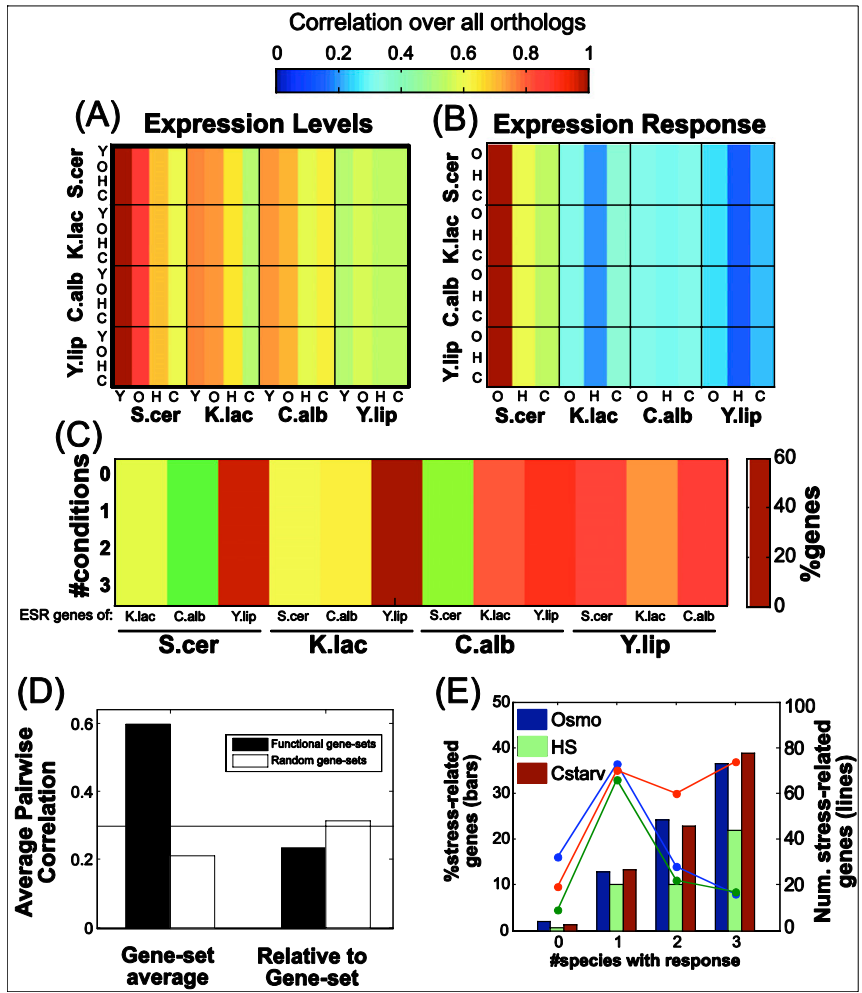
1. Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ (2010) The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol* 8:e1000414.



**Fig. S10.** *S. cerevisiae* stress responses, which are not conserved or compensated among the other three species, are associated with promoter transcription factor binding sites, which are preferentially not conserved among closely related yeast species. Genes that respond to any of the three stresses in *S. cerevisiae* were classified based on conservation of the response in the three other species (*K. lactis*, *C. albicans* and *Y. lipolytica*) and the evidence for compensation in these species (increased basal mRNA level by at least 1.75-fold, or annotation to a GO term whose genes respond to stress on average). "Conserved" represent genes whose response is conserved among at least three species, "Compensated" represent genes whose response is not conserved in any of the three species but shows evidence for compensation in at least one species, and "non-conserved & non-compensated" represent genes whose response is not conserved or compensated in any of the three species. For each class, we calculated the percentage of promoter transcription factor binding sites that are conserved among three closely related species of the *Saccharomyces sensu-stricto* yeasts, as previously defined (1). The depletion of conserved sites in the third class is highly significant.

1. Maclsaac KD, et al. (2006) An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* 7:113.





**Fig. S11.** Replicate analysis. Analyses from Fig. 1 A and B, Fig. 2C, and Fig. 3 D and E were repeated after averaging the two replicate experiments for each species.

**Table S1.** Yeast strains used in this study

Species	Genotypes	Strain
<i>S. cerevisiae</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	BY4741
<i>K. lactis</i>	<i>MATa met- k1 k2</i>	CLIB210
<i>C. albicans</i>	Wild type isolate	SC5314
<i>Y. lipolytica</i>	<i>MATa</i>	CLIB89