Young Genes are More Responsive to Environmental Stress than Ancient Genes in Budding Yeasts

Tyler W. Doughty[1], Iván Domenzain[1], Aaron Millan-Oropeza[2], Noemi Montini[3], Philip A. de Groot[4], Rui Pereira[1], Jens Nielsen[1], Céline Henry[2], Jean-Marc G. Daran[4], Verena Siewers[1]* & John P. Morrissey[3]*

[1]Department of Biology and Biological Engineering and Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, SE-41296, Gothenburg, Sweden
[2]Plateforme d'Analyse Protéomique Paris Sud-Ouest (PAPPSO), INRA, MICALIS Institute, Université Paris-Saclay, 78350 Jouy-en-Josas, France.
[3]School of Microbiology, University College Cork, T12YN60, Ireland and Environmental Research Institute, Cork, Ireland.
[4]Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands
*Corresponding Authors

Abstract:
Members of the subphylum Saccharomycotina (budding yeasts) are found in many of Earth's biomes, with some species, including *Saccharomyces cerevisiae, Kluyveromyces marxianus*, and *Yarrowia lipolytica*, exhibiting potential for industrial production of biochemicals. One complication for production is that ideal growth conditions (e.g. temperature, pH, salinity) for these species may not match reaction parameters for industrial processes, resulting in cellular stress and lower product yields. This work exposed budding yeasts to common industrial stressors to assess commonalities amongst stress-adaptation responses. The primary finding was that each stress-response, for each yeast species, was enriched for expression changes amongst genus and species-exclusive (young) genes compared to broadly conserved (ancient) genes. Further, young genes exhibited lower expression, less influence on viability, and accumulated nonsynonymous substitutions more frequently than ancient genes. These findings demonstrate that young genes are enriched for stress-responsiveness, may adapt rapidly, and therefore, may be important targets for improving industrial stress tolerances.

Introduction

Yeasts in the Saccharomycotina subphylum, “budding yeasts”, have proven to be useful platforms for the production of ethanol, flavors, nutritional supplements, biopharmaceuticals, and other valuable chemicals[1-3]. At present, industrial production using budding yeasts as a platform is dominated by *Saccharomyces cerevisiae,* an extensively characterized organism that is readily engineered by homologous recombination. *S. cerevisiae* exhibits common budding yeast phenotypes (fast growth and diverse substrate utilization) as well as a less common adaptation amongst budding yeasts, high ethanol tolerance[4]. Together, these traits enable cost-effective production of 100 billion liters of ethanol annually using *S. cerevisiae* as a platform[1]. Other yeast species adaptations that make them well-suited for production of specific biomolecules, but have been difficult to engineer until the recent emergence of CRISPR/Cas9[5,6]. One example is *Yarrowia lipolytica,* a lipid-rich or oleaginous budding yeast, which facilitates high-yield production of fatty acid- and acetyl-CoA-derived molecules including aromas and precursors to bioplastics[5]. Yet another promising budding yeast, *Kluyveromyces marxianus* is capable of growing on lactose and plant fructans, and exhibits high temperature tolerance, which is a beneficial feature for industrial processes[6,7]. Despite progress in characterizing these yeasts, the underlying causes of many of their unique phenotypes remain enigmatic.

Amongst these enigmatic phenotypes are tolerances to various forms of environmental stress (robustness). Understanding how robustness is achieved in diverse budding yeasts, especially those with exceptional tolerance to a particular stress, might enable the engineering of improved tolerance in industrial strains, thereby reducing process costs and increasing yields[8,9]. Previous works have characterized stress tolerances in *S. cerevisiae* via systems biology, rational design experiments, adaptive laboratory evolution, and gene deletion or overexpression arrays[8,10,11]. However, while these analyses have elucidated mechanisms that influence robustness in *S. cerevisiae*, understanding the intricacies of the stress response well enough to engineer more robust *S. cerevisiae* strains without physiological trade-offs remains challenging[9]. In addition, the mechanisms by which certain budding yeasts achieve genus or species-specific tolerances, such as the uniquely high temperature tolerance exhibited by *K. marxianus*[12]*,* remain unclear.

Here, *S. cerevisiae, Y. lipolytica,* and *K. marxianus* were exposed to industrial stress conditions to assess commonalities and differences in gene expression and protein abundance changes after stress adaptation. Analysis of 2,903 orthologous protein-coding genes found that few significant changes were shared between these organisms after adaptation to the same stress. The differences between the stress responses was further illustrated by the finding that genes that were shared broadly amongst eukaryotes (ancient genes) were less responsive to stress than genus or species-specific (young) genes for each of the yeasts in response to each stress. This work presents a novel approach to understanding stress gene expression and finds that efforts to engineer stress tolerance could benefit from consideration of the evolutionary age and adaptation rate of target genes.

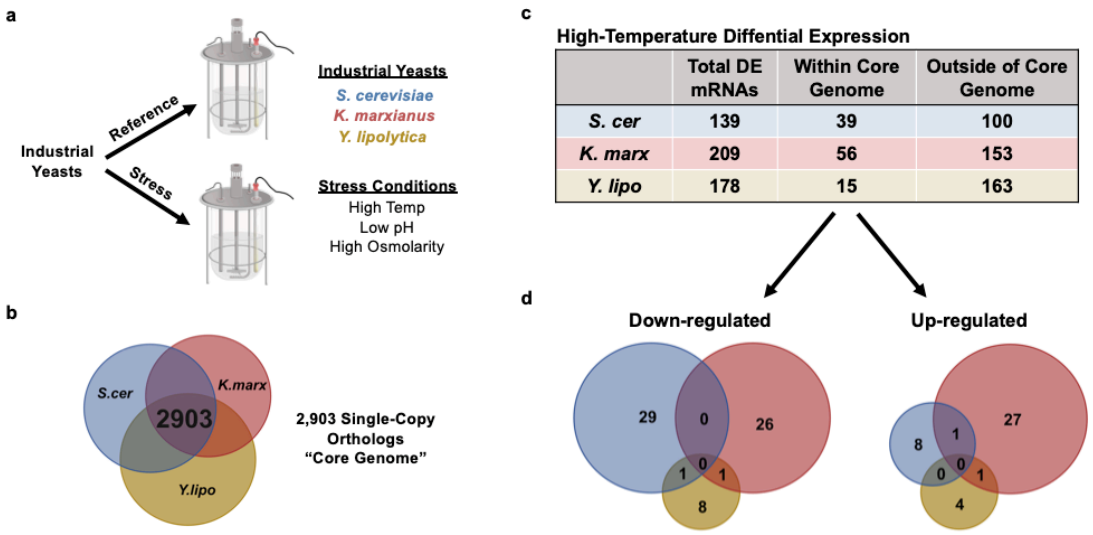
Results

Stress-Adaptation in Industrially-Relevant Species

In this work, *S. cerevisiae, K. marxianus,* and *Y. lipolytica* were exposed to conditions designed to mimic typical stresses caused by feedstocks (high osmotic pressure and low pH) or process conditions (elevated temperatures) during industrial fermentations[9]. These organisms were chosen due to their high potential for industrial production and their phylogenetic diversity, which spans much of the Saccharomycotina subphylum[13], which is referred to as the budding yeast subphylum hereafter. Organisms were cultivated in steady-state chemostats under standard or stress conditions for 50 hours to assess the transcriptomic and proteomic changes that occur in response to long-term stress (stress adaptation) (Supplemental Figure 1a). This experimental design differs from studies focused on the stress shock response, which is often analyzed 10 to 60 minutes after a change in environment[14,15]. Low overlap was observed between significant expression changes when comparing short-term shock to longer-term adaptation to high osmotic pressure (Supplemental Figure 1b and Supplemental Figure 1c)[16,15].

Stress-Induced Gene Expression Suggests Few Responses are Conserved at Gene-Level Amongst Budding Yeasts During Stress Adaptation

To explore potential conservation of the stress response between budding yeasts at the level of individual genes, we sought to compare differential gene expression (DE) amongst orthologs shared between the three yeasts in this study (Figure 1a and 1b). Orthologs were inferred using OrthoFinder, which enables proteome-wide matching based on amino-acid sequence and chain length similarity in order to predict proteins that descend from a common ancestor[17]. This analysis predicted 2,903 proteins as 1:1:1 orthologs between the three yeasts, these proteins are referred to as “core” proteins in the following text. To assess the fidelity of ortholog predictions, protein complexes and enzymatic processes that are expected to be conserved amongst budding yeasts were searched for amongst the core proteins[18]. This analysis found that orthology inference identified the majority of the expected complex members and enzymes as core proteins (Supplemental Figure 2a), which supports the high fidelity of OrthoFinder predictions that was observed previously[17]. Core proteins from this analysis were matched to their corresponding genes, which we refer to as core genes (Figure 1b). The 1:1:1 relationship of these core genes allowed for comparison of gene expression changes at gene level between the three yeasts. Figure 1c shows the number of genes within the 2,903 core genes that were differentially expressed in response to high temperature adaptation (log2FC>1, FDR<0.01). Differentially expressed core genes were compared between each species and few expression changes were found to be shared between the predicted 1:1:1 orthologs (Figure 1d, Supplemental Figures 2b and 2c). Core protein analysis also showed few core proteins with significantly altered abundance for multiple yeasts in response to the same stress (Supplemental Table 2). These data suggest that stress adaptation responses are not well conserved at the level of differential expression/abundance of specific orthologous transcripts or proteins in *S. cerevisiae, K. marxianus,* and *Y. lipolytica*.



	Total DE mRNAs	Within Core Genome	Outside of Core Genome
S. cer	139	39	100
K. marx	209	56	153
Y. lipo	178	15	163

Figure 1) Budding Yeasts Exhibit few Orthologous Gene Expression Changes in Response to High Temperature.
a *S. cerevisiae* (blue), *K. marxianus* (red), and *Y. lipolytica* (gold) were cultivated in chemostats at reference growth conditions or in the presence of stress. **b** Orthologous protein coding genes were inferred between *S. cerevisiae*, *K. marxianus*, and *Y. lipolytica* using OrthoFinder, 2,903 core genes were predicted to be shared as 1:1:1 orthologs. **c** The number of differentially expressed genes in the core genome and non-core genome is shown for each yeast species. **d** Differentially expressed core genes are shown, downregulated genes (left) and upregulated genes (right). Overlapping regions indicate differential expression of an orthologous gene in response to elevated temperature.

Core Genes and Proteins are Unenriched for Stress-Responsive Expression

The expression changes in Figure 1 indicated that stress-induced expression was observed more often in non-core genes compared to core genes (Figure 1c). This phenomenon was intriguing as core genes are predicted to have descended from a last common ancestor between the three yeast species that diverged approximately 325 million years ago[13]. Could the core gene underrepresentation in Figure 1c indicate that genes that originated earlier in evolutionary time were less stress responsive than their younger (non-core) counterparts? To assess this, each of the differentially expressed transcripts were grouped by their presence in (core) or absence from the list of orthologs (non-core) (Figure 2a). This grouping showed that, for each organism, there was a significantly higher percentage of differentially expressed genes in the non-core gene group for each of the stress conditions (Figure 2b). Examination of significant changes in protein abundance also showed a higher frequency amongst the non-core proteins for many of the stress conditions, with two of the conditions exhibiting significant bias (Figure 2d). Together, these findings suggest an inverse correlation between the likelihood of differential expression in response to stress and gene/protein conservation amongst these diverse budding yeasts.

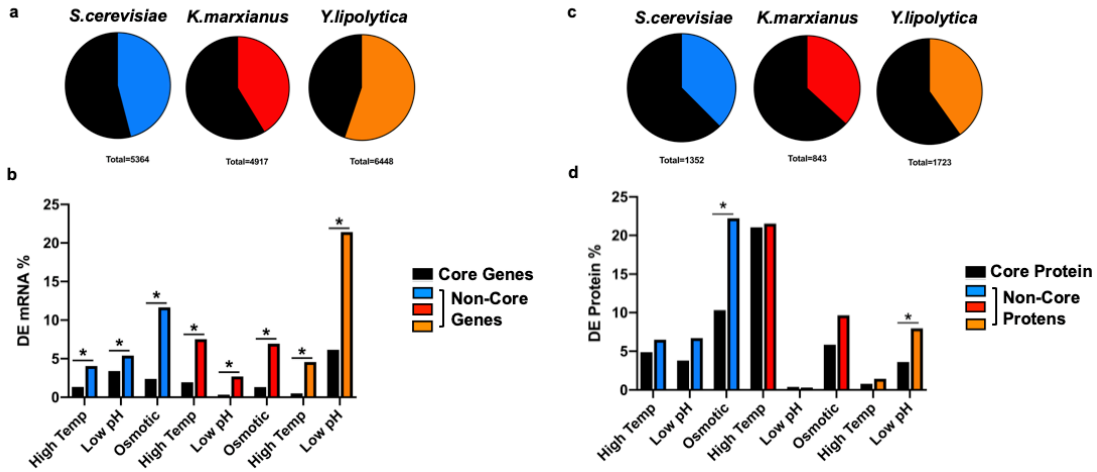


Figure 2) Differential Expression in Response to Stress Occurs Preferentially in Non-Core Genes and Proteins.
The transcripts (**a**) and proteins (**c**) that were measured in this study were classified into core genes/proteins (black) or non-core genes/proteins (color). The number of differentially expressed ($log_2FC>1$, FDR<0.01) mRNAs (**b**) and proteins (**d**) were normalized to the total number of measured genes inside of their respective group; orthologs (black) or non-orthologs (color). Fisher's exact test was used to determine significance between core and non-core for each condition, * indicates p<0.01.

The Stress Adaptation Response in Budding Yeast is Enriched for Young Genes

The results in Figure 2 suggested a relationship between gene/protein age and differential expression in response to stress. To test this relationship more stringently, the protein-coding genes of each yeast were divided into five groups representing a broad swath of eukaryotic evolution. The goal of this analysis was not to compare expression changes of the same ancestral gene between the three yeasts (as in Figure 1), but to compare expression patterns amongst groups of genes stratified by their predicted age. Gene age was determined by ortholog presence in common ancestors that date from approximately one billion years ago to approximately 20 million years ago[13]. A similar approach called phylostratigraphy, was used to stratify genes from *Drosophila melanogaster* by predicted age and found that genes associated with specific adaptations correlate with specific periods in evolution[19]. Similarly, we sought to understand if the yeast stress responses from this study were enriched for genes that emerged more recently in evolution.

Gene grouping based on gene age was assessed using OrthoFinder[17]. A separate gene grouping analysis was carried out for each yeast to include relatives that had similar estimated times of common ancestor divergence. Orthologous gene predictions were used to sort each protein, and its coding gene into a single bin based on the most distant ancestor with a predicted orthologous protein (Supplemental Figure 3a and Supplemental Figure 3b). This analysis matched 2,440 protein coding genes from *K. marxianus* to a distant eukaryotic common ancestor shared with *H. sapiens*, these genes are referred to as ancient herein. Of these 2,440 protein coding genes, 96.4% to 99.2% were found in other budding yeasts queried in this study, suggesting that these genes are broadly conserved amongst budding yeasts (Supplemental Figure 3c). Conversely, genes that were predicted to have an ortholog only amongst members of the genus (*Kluyveromyces*) and/or the species

(*Kluyveromyces marxianus*) are referred to as young genes in this work. We assayed the ancient gene predictions from our analysis for the presence of previously described ancient eukaryotic genes[18], and found that orthology inference was able to detect the expected ancient genes with high fidelity (Supplemental Figure 3d).

The gene groupings in Supplemental Figure 3 stratified protein-coding genes from *K. marxianus* based on their traceability to eukaryotes of various phylogenetic distances (Figure 3a). The resulting groups (Figure 3b) were matched to DE genes to determine the percentage of each gene group that was significantly differentially expressed in response to stress. For *K. marxianus*, genes that were conserved to *H. sapiens* (ancient genes from group I) were 7 to 10-fold less likely to be differentially expressed after stress adaptation compared to *K. marxianus* specific genes (group V) (Figure 3c). In addition, few group IV and V (young) genes were found to respond to all stresses, suggesting that the majority of expression changes amongst young genes were unique to a given stress (Figure 3d). Relative proteomics was not integrated into this analysis due to low detection frequency (<10%) of proteins that correspond to young genes (Supplemental Figure 4).

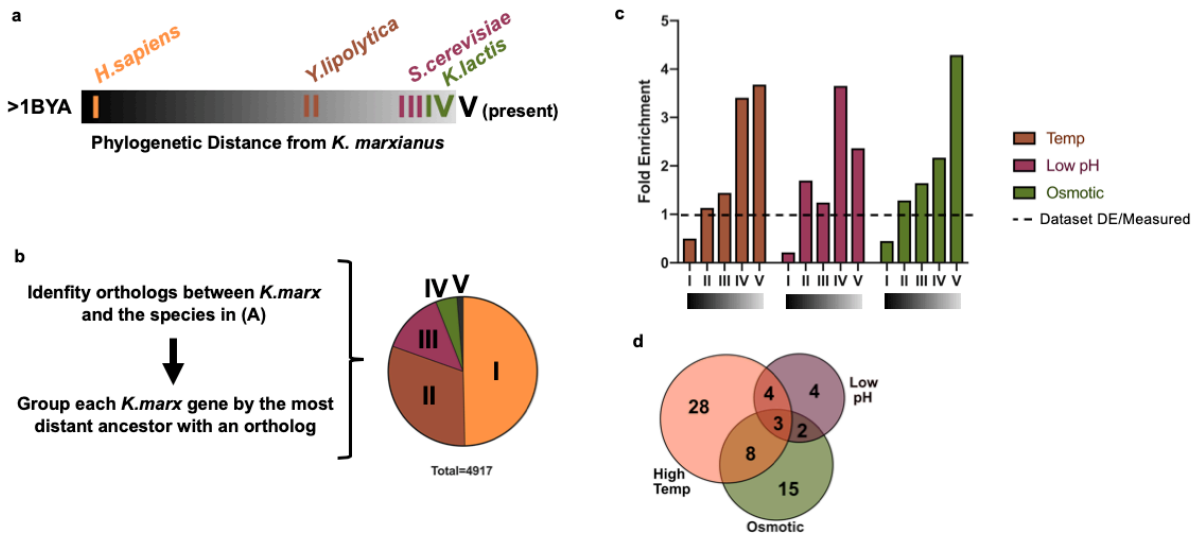


Figure 3) The Stress Adaptation Response in *K. marxianus* is Enriched for Young Genes. **a** The approximate phylogenetic distance between *K. marxianus* and the organisms listed is shown. **b** The measured transcripts of *K. marxianus* were put into groups based on workflow listed (left), the resulting gene groupings are shown (right). Gene group colors indicate the presence of an ortholog shared between *K. marxianus* and the species in a. **c** Differentially expressed genes were normalized to the bin size and the total measured DE % (dashed line). Fold change between group I and group V was 7.33 (High Temp), 10.89 (Low pH), and 9.53 (Osmotic). **d** Transcripts in group IV and V were assessed for shared differentially expressed genes. Colors represent stress conditions in c, overlapping regions indicate differential expression of the same genes in multiple stress responses.

Gene groupings based on gene age were created for the protein-coding genes from *S. cerevisiae* and *Y. lipolytica* using the same hierarchal sorting concept described above. Analysis of the *Y. lipolytica* and *S. cerevisiae* gene groups in relation to each stress adaptation condition showed similar patterns to *K. marxianus*, with ancient genes exhibiting a low percentage of significant differential expression compared to young gene groups (Supplemental Figure 5 and Figure 4). Further, these findings were not exclusive to high temperature, low pH, and high osmotic pressure, as additional stress adaptation experiments with *S. cerevisiae* exposed to ethanol[16] or anaerobic stress also showed similar enrichment for young genes (Supplemental Figure 6). Together, these findings show that the young genes were enriched for stress-responsiveness compared ancient genes for each yeast species in this study after adaptation to each stress condition. As the species chosen for this analysis span much of the diversity of the budding yeast subphylum[13], these results may be indicative of a conserved stress adaptation mechanism amongst budding yeasts.

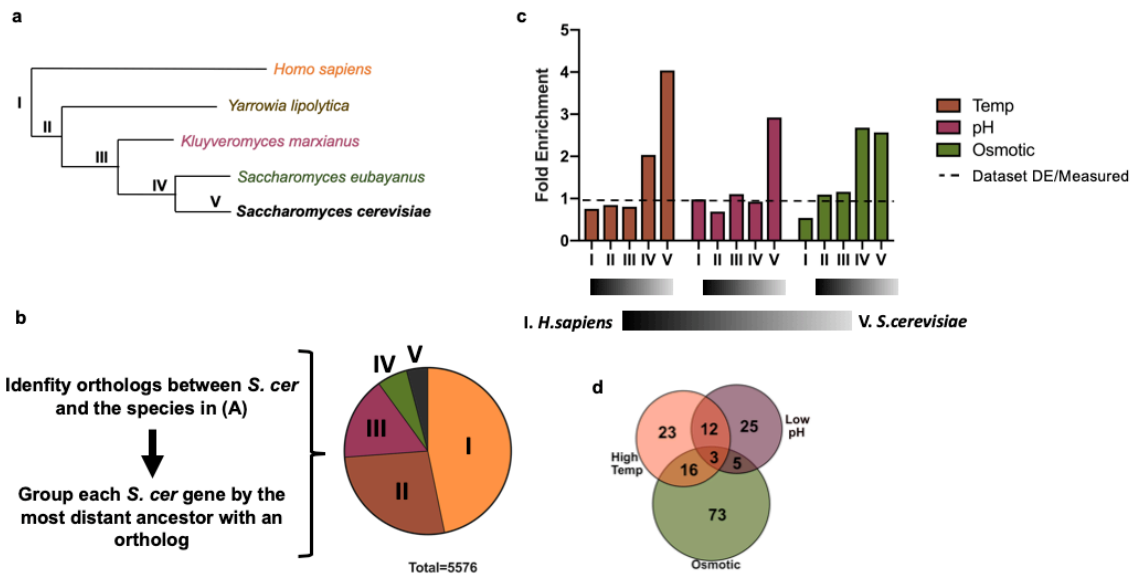


Figure 4) The Stress Adaptation Response in *S. cerevisiae* is Enriched for Young Genes
a A simplified phylogenetic tree showing the relationship between the species used for orthology grouping in this figure. **b** The measured transcripts of *S. cerevisiae* were put into groups based on workflow listed (left), the resulting gene groupings are shown (right). Gene group colors indicate the presence of an ortholog shared between *S. cerevisiae* and the species in a. **c** Differentially expressed genes for *S. cerevisiae* were parsed by their grouping shown in b, then normalized to the group size and the total measured DE % (dashed line). **d** Transcripts in group IV and V were assessed for shared differentially expressed genes. Colors represent stress conditions in c, overlapping regions indicate differential expression of the same genes in multiple stress responses.

Stress Adaptation and Stress Shock Show Different Trends in Relation to Gene Age in *S. cerevisiae*

An alternative approach to understanding the stress response is stress shock analysis, where the reaction to stress is often assessed within the first hour of stress exposure. Stress shock and stress adaptation exhibit different gene expression changes (Supplemental Figure 1b and 1c), which may be indicative of distinct response mechanisms. To assess whether the shock response showed similar biases towards young genes for *S. cerevisiae*, the gene groups from Figure 4b were used to classify the gene expression changes in response to osmotic and temperature shock (Supplemental Figure 7). Analysis of osmotic shock (400 mM NaCl, 15 min) experiments from Silva 2017 found that ancient genes were enriched amongst significantly downregulated genes and that young genes were weakly enriched amongst upregulated genes (Supplemental Figure 7b). Similarly, analysis of heat shock (20 min at 42°C) experiments from Hughes Hallett 2014 showed an enrichment for ancient genes amongst significantly downregulated transcripts and a weak enrichment of young genes amongst upregulated transcripts (Supplemental 7d and e). In contrast, the stress adaptation experiments from this report showed young genes were enriched amongst both upregulated and downregulated genes in response to osmotic and temperature stress (Supplemental 7c and 7f). These results suggest that stress shock and stress adaptation gene expression differ in relation to gene age, especially amongst downregulated genes.

Young Genes are Less Expressed, Shorter, Less Likely to be Essential, and Accumulate More Sequences Changes than Ancient Genes.

To understand the functions associated with the gene groups in Figure 5, we assessed Gene Ontology Enrichment for each gene group from *S. cerevisiae*, which was chosen due to its extensive characterization. This analysis employed YeastMine and showed that young genes (groups IV and V) were enriched amongst specialized processes, including sporulation and anatomical structure morphogenesis, as well as specialized components like fungal-type vacuole and the cell wall. In contrast, ancient genes were associated with fundamental biological processes amongst eukaryotes, including primary metabolism, tRNA aminoacylation, and DNA strand elongation, as well as cellular components like ribosomal subunits, mitochondrial proteins, and RNA polymerases. The fundamental nature of ancient gene functional associations was reflected by their high likelihood to be essential or required for optimal growth compared to young genes (Figure 5c)[20].

Additional characterization of young protein-coding genes found that they exhibited lower median gene expression and shorter protein length compared to ancient protein-coding genes amongst the yeasts in this study (Figures 5a and 5b), which might explain the low frequency of detection of young proteins via mass spectrometry (Supplemental Figure 4). Furthermore, previous works have shown that low expression and non-essentiality correlate with increased adaptation rates[21,22], suggesting that young genes could adapt more rapidly compared to ancient genes (Figures 5a and 5c). To test this, amino acid sequence homology was compared between orthologous proteins from closely related members of the same genus using BLAST+. Since the query species were closely related, protein sequences from groups I through IV could be analyzed for sequence identity change over the same

span of evolutionary time to assess adaptation rates. This analysis was adjusted to reflect the estimated evolutionary time elapsed[13] between each pair of species and showed that the average frequency of amino acid identity changes was more than two-fold higher for young protein groups compared to ancient protein groups (Figure 5d).

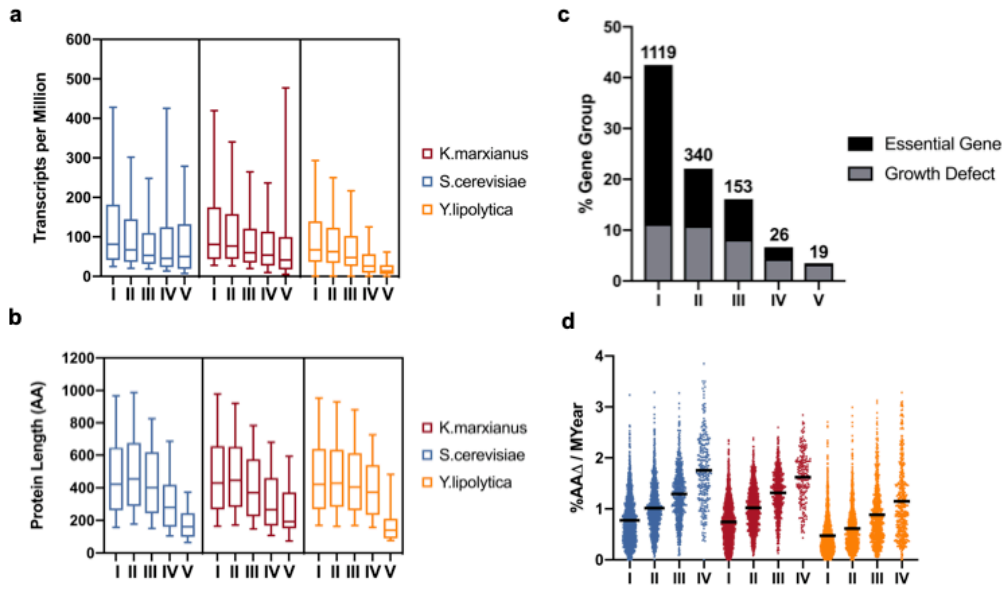


Figure 5) Young Genes are Less Expressed, Shorter, Less Likely to be Essential, and Accumulate More Sequence Changes than Ancient Genes.
a Transcripts measured in standard growth conditions were normalized to the read depth and gene length. Median Transcripts per Million (TPM) are shown for each gene group, whiskers mark the upper 90% and lower 10% of each distribution. **b** Median protein length is shown for each group, whiskers mark the upper 90% and lower 10% of protein sizes. **c** The percentage of essential genes (black) and non-essential genes associated with slow growth (gray) is shown for each gene group in *S. cerevisiae*. The total number of genes is listed above each bar. Essential and slow growth ORFs were obtained from Giaever 2002[20]. **d** The percentage of amino acid identity changes for each protein in comparison to its respective ortholog from a member of the same genus. Results were adjusted to the percent amino acid change per million years (%AAΔ/MYear) using the estimated divergence time between pairs of organisms[13]. The mean for each group is denoted by a black line. Queries were performed between *S. cerevisiae/S. eubayanus* (blue), *K. marxianus/K. lactis* (red), or *Y. lipolytica/Y. bubula* (orange).

Discussion

Budding yeasts are attractive for industrial production of biomolecules, since they grow rapidly, utilize inexpensive substrates, and are readily engineered to produce heterologous gene products[1-3]. However, stresses that result from feedstocks, products, and reaction temperatures can lower the cost-effectiveness of some industrial processes by diminishing yields[9]. This study and previous stress tolerance investigations have identified several dozen to several hundred significant gene expression changes during stress adaptation[16] or stress shock[11,14,15,23]. Despite identification of stress-responsive genes in several species and conditions, rational engineering of more stress-tolerant strains remains difficult, possibly due to an incomplete understanding of how these responses evolved. In this report, we exposed industrially-relevant yeasts to common industrial stresses to assess the conservation of the transcriptomic and proteomic response to stress. Core gene/protein analysis found that conserved genes and proteins shared between diverse budding yeasts were less responsive to stress than non-conserved genes and proteins. Further analysis showed that expression changes upon stress adaptation were enriched for genus or species-specific (young) genes compared to broadly conserved (ancient) genes for each organism in response to each stress condition.

The observation that ancient genes were unenriched in multiple stress adaptation responses, amongst diverse budding yeasts, suggests that this may be a common theme amongst the budding yeast subphylum. Further, these findings could indicate that many ancient genes are off-limits for differential expression during long-term stress responses. This may be due to the association between ancient genes and fundamental cellular processes and/or the high percentage of ancient genes that influence viability. In contrast to adaptation, stress shock responses were enriched for ancient gene downregulation in *S. cerevisiae*. These ancient gene expression results could help to explain the differences between the observed stress shock and stress adaptation expression patterns. For example, downregulation of ancient genes upon stress shock could facilitate robust upregulation of stress-tolerance genes, however, these expression patterns might not be sustainable in the long-term. Thus, a subsequent response (stress adaptation), in which most ancient genes are not downregulated, might be necessary. While further work is needed to characterize the relationship between stress shock and adaptation, we propose that consideration of both responses can inform the design of strains that achieve increased stress tolerance without altering fundamental cellular processes.

Each stress adaption response in this report showed enrichment for differential expression of young genes (group IV and V genes), which represented 6%, 10%, and 21% of protein-coding genes in *K. marxianus, S. cerevisiae,* and *Y. lipolytica*, respectively. These group sizes are similar to the 7-19% of genes in *C. elegans, D. melanogaster,* and *H. sapiens* that lack recognizable homologs in other organisms[19,24]. While many species and genus-specific genes (young genes) are poorly characterized, previous works have linked some of these genes to species and genus-specific adaptations. Examples of adaptations involving young genes include movement on the surface of fast water in *Rhagovelia* water striders, HIV-1 resistance in owl monkeys, and cold tolerance in artic fishes[25-30]. These examples suggest that young genes can confer novel adaptations in response new or changing

evolutionary pressures. In this report, we found that young genes were unlikely to be essential or influence growth, were on average lowly expressed, and exhibited more than double the rate nonsynonymous substitutions in recent evolutionary time compared to ancient genes. These attributes indicate that young genes are more tolerant to sequence change, which could increase their likelihood to acquire and hone niche-specific functional adaptations. One intriguing hypothesis derived from this work is that some of the young genes that responded to industrial stress may influence tolerance to analogous stressors in natural environments. For example, young genes from *K. marxianus* that responded to temperature stress in this work might influence it's increased temperature tolerance (45°C)[7] compared to the closely related species *K. lactis* (37°C)[12]. We propose that further analysis of young gene functions, especially amongst organisms that exhibit exceptional tolerance to a stress condition, could help to identify adaptations that can advance the field of industrial stress tolerance engineering.

Methods

Strains and Cultivation Conditions

Y. lipolytica (W29), *K. marxianus* (CBS6556), and *S. cerevisiae* (CEN.PK113-7D) were grown in 30mL synthetic media for 24hours in shake flasks, followed by inoculation of bioreactors and an initial batch growth phase. After the completion of the batch phase, chemostat cultivation was started with a dilution rate of 0.1/h and a working volume of 500mL (*S. cerevisiae*) or 1L (*K. marxianus* and *Y. lipolytica*). Stress conditions were achieved by altering either temperature, pH, or osmotic pressure (KCl) for the duration of the cultivation, specific conditions are listed in Supplemental 1A. Standard growth temperature was adjusted to reflect organism specific tolerances. Cultivations for were performed in synthetic medium (SM)[31] containing 5 $g \cdot L^{-1}$ $(NH_4)_2SO_4$, 3 $g \cdot L^{-1}$ KH_2PO_4, 0.5 $g \cdot L^{-1}$ $MgSO_4 \cdot 7H_2O$, 7.5 $g.L^{-1}$ glucose, trace elements and vitamins with 1 g L^{-1} pluronic PE6100 to reduce foaming. Stress conditions were enforced throughout the batch and chemostat phases, specific conditions are listed in Supplemental 1A. Sample collection was carried out after at least five volume changes (50hours) in steady state growth conditions. Steady state growth was defined as less than 5% deviation in biomass dry weight.

Ortholog Prediction with OrthoFinder

For Figures 1 and 2, proteome-wide homology matching was executed using OrthoFinder[17]. Proteins were excluded from the core genome if the analysis predicted zero or more than 1 copy of a sequence for any of the yeasts. This allowed each of the measured 2903 genes/proteins to be compared between the fungi. These 1:1:1 genes and proteins were matched to significantly altered differential expression datasets to probe for equivalent changes shared amongst the organisms. Core orthologs can be found in Supplemental Table 1.

For Figure 3, Figure 4, and Supplemental Figure 5, OrthoFinder was used to identify orthologs between each yeast and a set of eukaryotic organisms. Orthologs predicted between, for example, *K. marxianus* and each query organism were identified, followed by binning each protein into a single group. When a protein sequence matched to multiple of the query species, the protein was placed in the group of the most distant common ancestor (Supplemental Figure 3b)[13,32]. For this analysis, orthologous proteins predicted between two organisms were counted as a match, regardless of the gene copy number present in each organism. This allowed matching to include the entire proteome. FASTA proteome files were obtained from uniprot.org, except for *Saccharomyces eubayanus, Yarrowia bubula,* and *Candida hispaniensis,* which were obtained from Shen 2018[13]. Protein groups were matched to their underlying genes for gene expression analyses.

RNAseq Preparation and Mapping

RNA extractions were performed on samples that were mechanically lysed with 0.5mm acid washed beads using an MP-Biomedicals™ FastPrep-24 for three one-minute cycles. Further extraction was performed using an RNeasy® Kit from Qiagen. Libraries were prepared using the TrueSeq mRNA Stranded HT kit. Sequencing was carried out using an Illumina NextSeq 500 High Output Kit v2 (75 cycles), with a minimum of 8 million paired-end reads per replicate. The Novo Nordisk Foundation Centre for Biosustainability (Technical University of Denmark), performed the RNA

sequencing and library preparation. RNAseq datasets will be made public after manuscript publication under SRA accession PRJNA531619. RNAseq read mapping was performed after analysis in FASTQC, which identified one sample from *K. marxianus* as having overrepresented sequences. This sample was excluded from the analysis herein. Analysis for TPM in Figure 4a was performed using Hisat2[33] and StringTie[34]. RNAseq mapping for differential expression was mapped with STAR[35] and reads were assigned with featureCounts[36]. Differential expression results can be found in Supplemental Table 3.

Mass Spectrometry Measurements for Proteomics

Sample preparation details can be found in supplementary data. Peptides were analyzed in an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo Fisher Scientific) in data-dependent acquisition mode using High Collision Dissociation (details in supplementary data). Protein identification was performed using the search engine X!Tandempipeline 3.4.4[37]. Data filtering was set to peptide E-value < 0.05, protein log(E-value) < –2 and to a minimum of two identified peptides per protein. Relative quantification of protein abundance was carried out using two complementary methods, one based on spectral counting and other considering the area of eXtracted Ion Chromatograms (XIC) of a given peptide at specific retention time as described in Millan-Oropeza 2017[38]. MS data will become available after manuscript publication via the PRIDE[39] repository with the dataset identifier PXD011426 and PROTICdb[37,40] or via the following URLs:
http://moulon.inra.fr/protic/chassy_kluyveromyces
http://moulon.inra.fr/protic/chassy_yarrowia
http://moulon.inra.fr/protic/chassy_saccharomyces

Differential Expression Analysis

Differential expression data was generated using limma[41] and edgeR[42] R packages, with filtering to remove lowly expressed genes/proteins. In addition, each dataset was filtered to remove genes/proteins for which the relative standard deviation was greater than 1 (RSD>1) for replicates for a given condition and organism. Differential expression was defined by a significance cutoff of absolute $\log_2$FC>1, False Discovery Rate<0.01 for a stress condition compared to control.

References

1 Mohd Azhar, S. H. *et al.* Yeasts in sustainable bioethanol production: A review. *Biochem Biophys Rep* **10**, 52-61 (2017).

2 Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185-1197 (2016).

3 Sanchez-Garcia, L. *et al.* Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microbial Cell Factories* **15**, 33 (2016).

4 Ma, Y. J., Lin, L. L., Chien, H. R. & Hsu, W. H. Efficient utilization of starch by a recombinant strain of *Saccharomyces cerevisiae* producing glucoamylase and isoamylase. *Biotechnol Appl Biochem* **31 (Pt 1)**, 55-59 (2000).

5 Ledesma-Amaro, R. & Nicaud, J. M. *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. *Prog Lipid Res* **61**, 40-50 (2016).

6 Varela, J. A., Gethins, L., Stanton, C., Ross, P. & Morrissey, J. P. Applications of *Kluyveromyces marxianus* in Biotechnology. In: *Yeast Diversity in Human Welfare*. 439-453 (Springer, Singapore, 2017).

7 Lane, M. M. & Morrissey, J. P. *Kluyveromyces marxianus*: A yeast emerging from its sister's shadow. *Fungal Biology Reviews* **24**, 17-26 (2010).

8 Mans, R., Daran, J. G. & Pronk, J. T. Under pressure: evolutionary engineering of yeast strains for improved performance in fuels and chemicals production. *Curr Opin Biotechnol* **50**, 47-56 (2018).

9 Deparis, Q., Claes, A., Foulquie-Moreno, M. R. & Thevelein, J. M. Engineering tolerance to industrially relevant stress factors in yeast cell factories. *FEMS Yeast Res* **17** (2017).

10 Caspeta, L. C., Y.; Ghiaci, P.; Feizi, A.; Buskov, S.; Hallström, B.; Petranovic, D.; Nielsen, J. Altered Sterol Composition Renders Yeast Thermotolerant. *Science* **346**, 75-78 (2014).

11 Gibney, P. A., Lu, C., Caudy, A. A., Hess, D. C. & Botstein, D. Yeast metabolic and signaling genes are required for heat-shock survival and have little overlap with the heat-induced genes. *Proc Natl Acad Sci U S A* **110**, E4393-4402 (2013).

12 Steensma, H. Y., de Jongh, F. C. M. & Linnekamp, M. The use of electrophoretic karyotypes in the classification of yeasts: *Kluyveromyces marxianus* and *K. lactis Current Genetics* **14**, 311-317 (1988).

13 Shen, X. X. *et al.* Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* **175**, 1533-1545 (2018).

14 Hughes Hallett, J. E., Luo, X. & Capaldi, A. P. State transitions in the TORC1 signaling pathway and information processing in *Saccharomyces cerevisiae*. *Genetics* **198**, 773-786 (2014).

15 Silva, A. *et al.* Regulation of transcription elongation in response to osmostress. *PLoS Genet* **13**, e1007090 (2017).

16 Lahtvee, P. J. *et al.* Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell Syst* **4**, 495-504 e495 (2017).

17 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157, doi:10.1186/s13059-015-0721-2 (2015).

18 Prachumwat, A. & Li, W.-H. Protein Function, Connectivity, and Duplicability in Yeast. *Molecular Biology and Evolution* **23**, 30-39 (2005).

19 Domazet-Lošo, T. B., J.; Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends in Genetics* **23**, 533-539 (2007).
20 Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391 (2002).
21 Pál, C., Papp, B. & Hurst, L. D. Highly expressed genes in yeast evolve slowly. *Genetics* **158**, 927-931 (2001).
22 Mata, J., Bahler, J. Correlations Between Gene Expression and Gene Conservation in Fission Yeast. *Genome Research* **13(12)**, 2686–2690 (2003).
23 Kasavi, C., Eraslan, S., Oner, E. T. & Kirdar, B. An integrative analysis of transcriptomic response of ethanol tolerant strains to ethanol in *Saccharomyces cerevisiae*. *Molecular BioSystems* **12**, 464-476 (2016).
24 Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. & Bosch, T. C. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet* **25**, 404-413 (2009).
25 Santos, M. E., Le Bouquin, A., Crumière, A. J. J. & Khila, A. Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. *Science* **358**, 386-390 (2017).
26 Sayah, D. M., Sokolskaja, E., Berthoux, L. & Luban, J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**, 569-573 (2004).
27 Stremlau, M. *et al.* The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-853 (2004).
28 Zhuang, X., Yang, C., Murphy, K. R. & Cheng, C.-H. C. Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids. *Proceedings of the National Academy of Sciences* **116**, 4400-4405 (2019).
29 Chen, L., DeVries, A. L. & Cheng, C. H. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc Natl Acad Sci U S A* **94**, 3817-3822 (1997).
30 Chen, S., Krinsky, B. H. & Long, M. New genes as drivers of phenotypic evolution. *Nat Rev Genet* **14**, 645-660 (2013).
31 Verduyn, C., Postma, E., Scheffers, W. A. & Van Dijken, J. P. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **8**, 501-517 (1992).
32 Hedges, S. B. The origin and evolution of model organisms. *Nature Reviews Genetics* **3**, 838 (2002).
33 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015).
34 Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).
35 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
36 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
37 Langella, O. *et al.* X!TandemPipeline: A Tool to Manage Sequence Redundancy for Protein Inference and Phosphosite Identification. *J Proteome Res* **16**, 494-503 (2017).

38 Millan-Oropeza, A. *et al.* Quantitative Proteomics Analysis Confirmed Oxidative Metabolism Predominates in *Streptomyces coelicolor* versus Glycolytic Metabolism in Streptomyces lividans. *J Proteome Res* **16**, 2597-2613 (2017).
39 Vizcaíno, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic acids research* **44**, 11033 (2016).
40 Ferry-Dumazet, H. *et al.* PROTICdb: A web-based application to store, track, query, and compare plant proteome data. *PROTEOMICS* **5**, 2069-2081 (2005).
41 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
42 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297 (2012).

Acknowledgements
This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation - Grant Agreement No. 720824. This work was furthermore supported by the Knut and Alice Wallenberg Foundation and the Novo Nordisk Foundation (Grant no NNF10CC1016517). We appreciate Erik de Hulster and Fredrik Schubert for technical support and training, as well as Gang Li, Lucy Chao, Benjamín Sánchez, and colleagues in the CHASSY consortium for scientific discussions.

Contributions
T.W.D., N.M., P.A.G., and R.P. performed chemostat cultivations. A.M-O. performed protein extraction/identification/quantification. I.D. created the analysis pipeline in R and performed differential expression data analysis. T.W.D. prepared/mapped RNA-seq, performed gene grouping analyses, and wrote the manuscript. J.N., C.H., J-M.D., V.S., and J.M. conceived and supervised the project.

Competing Interests
The authors declare no competing interests.