Article

Directional evolution in temperature-responsive gene expression in yeast

Xueying C. Li[1] and Justin C. Fay[2,3,*]

[1]Molecular Genetics and Genomics Program, Washington University, St. Louis, MO

[2]Department of Genetics, Washington University, St. Louis, MO

[3]Center for Genome Sciences and System Biology, Washington University, St. Louis, MO

*Corresponding author: Justin Fay (jfay@genetics.wustl.edu)

Key words: *Saccharomyces*, allele-specific expression, thermotolerance, gene expression, *cis*-regulatory evolution, interspecific hybrid

Abstract

Gene regulation is a ubiquitous mechanism by which organisms respond to their environment. While organisms are often found to be adapted to the environments they experience, the role of gene regulation in environmental adaptation is not often known. In this study, we examine divergence in *cis*-regulatory effects between two *Saccharomyces* species, *S. cerevisiae* and *S. uvarum*, that have substantially diverged in their thermal growth profile. We measured allele specific expression (ASE) in the species' hybrid at three temperatures, the highest of which is lethal to *S. uvarum* but not the hybrid or *S. cerevisiae*. We find that *S. uvarum* alleles can be expressed at the same level as *S. cerevisiae* alleles at high temperature and most *cis*-acting differences in gene expression are not dependent on temperature. However, a set of 136 genes with temperature dependent ASE show a directional pattern of evolution with *S. cerevisiae* alleles exhibiting a stronger temperature response than *S. uvarum* alleles. Although we find binding sites enriched upstream of temperature responsive genes, there are only weak correlations between binding site and expression divergence. Our results indicate that temperature divergence between *S. cerevisiae* and *S. uvarum* has not caused widespread divergence in *cis*-regulatory activity, but that a subset of genes show evidence of directional evolution, mediated by positive selection or loss of constraint. The difficulty of explaining divergence in *cis*-regulatory sequences with models of transcription factor binding sites and nucleosome positioning highlights the importance of identifying mutations that underlie *cis*-regulatory divergence between species.

Introduction

Changes in gene regulation are thought to play an important role in evolution (Carroll 2000). Regulatory change may be of particular importance to morphological evolution where tissue specific changes and co-option of existing pathways can modulate essential and conserved developmental pathways without a cost imposed by more pleiotropic changes in protein structure. Indeed, many examples illustrate this view and

there is a strong tendency for *cis*-acting changes in gene expression to underlie morphological evolution between species (Stern and Orgogozo 2008).

However, gene regulation is also critical to responding to environmental changes and all organisms that have been examined exhibit diverse transcriptional responses that depend on the environmental alteration (López-maury et al. 2008). Environment-dependent gene regulation enables fine-tuning of metabolism depending on nutrient availability as well as avoiding the potential costs of constitutive expression of proteins that are beneficial in certain environments but deleterious in others. Despite the general importance of responding to changing environments, the role of gene regulation in modulating these responses between closely related species is not known and may involve structural changes in proteins whose expression is already environment-dependent.

Studies of genetic variation in gene expression within and between species have revealed an abundance of variation (reviewed in Whitehead and Crawford 2006, Zheng et al. 2011, Romero et al. 2012). When examined, a significant fraction of this variation is environment-dependent (Fay et al. 2004, Landry et al. 2006, Li et al. 2006, Smith and Kruglyak 2008, Tirosh et al. 2009, Fear et al. 2016, He et al. 2016; reviewed in Gibson 2008, Grishkevich and Yanai 2013). However, distinguishing between adaptive and neutral divergence in gene expression is challenging (Fay and Wittkopp 2008), since *trans*-acting changes can cause correlated changes in the expression of many genes and the rate of expression divergence depends on the mutation rate and effect size, which is likely gene-specific and not known for all but a few genes (Gruber et al. 2012, Yun et al. 2012, Metzger et al. 2015).

One potentially powerful means of identifying adaptive divergence in gene expression is through a sign test of directional *cis*-acting changes in gene expression measured by allele-specific expression (ASE) (Fraser 2011). By testing whether a group of functionally related or co-regulated group of genes have evolved consistently higher or lower expression levels, the test does not assume any distribution of effect sizes and more importantly is specifically targeted to identifying polygenic adaptation. Applications of this or related sign tests (Fraser et al. 2010, Naranjo et al. 2015) have revealed quite a few cases of adaptive evolution (Bullard et al. 2010, Fraser et al. 2010, Fraser et al. 2011, Fraser et al. 2012, Martin et

al. 2012, Chang et al. 2013, Naranjo et al. 2015, He et al. 2016, Roop et al. 2016), some of which have been linked to organismal phenotypes. However, in only two of these studies was condition-specific divergence in gene expression examined (He et al. 2016, Roop et al. 2016), leaving open the question of how often such changes exhibit evidence for adaptive evolution. Of potential relevance, the majority (44-89%) of environment-dependent differences in gene expression have been found to be caused by *trans*- rather than *cis*-acting changes in gene expression (Smith and Kruglyak 2008, Tirosh et al. 2009, Grundberg et al. 2011, Fear et al. 2016), suggesting that *trans*-acting changes in gene expression may be more important to modulating environment-dependent gene expression.

In this study, we examine allele-specific differences in expression between two *Saccharomyces* species that have diverged in their thermal growth profiles. Among the *Saccharomyces* species, the most prominent phenotypic difference is in their thermal growth profile (Gonçalves et al. 2011, Salvadó et al. 2011). The optimum growth temperature of *S. cerevisiae* and *S. paradoxus* is 29-35°C, whereas the optimum growth temperature for *S. uvarum* and *S. kudriavzevii* is 23-27°C (Salvadó et al. 2011). Furthermore, *S. cerevisiae* is able to grow at much higher temperatures (maximum 41-42°C) than *S. uvarum* (maximum 34-35°C, Gonçalves et al. 2011), while *S. uvarum* grows much better than *S. cerevisiae* at low temperature (4°C, Figure 1). Because *S. cerevisiae* × *S. uvarum* hybrids grow well at high temperature, we were able to measure *cis*-regulatory divergence in gene expression across a range of temperatures by measuring ASE in the hybrid. We use this approach to determine how ASE is influenced by temperature and specifically whether *S. uvarum* alleles are misregulated at temperatures not experienced in their native context. We find that while most ASE is independent of temperature, a small subset of genes with temperature dependent ASE show directional changes in gene expression with *S. cerevisiae* alleles exhibiting a much stronger response than *S. uvarum* alleles.

Material and methods

Strains and RNA sequencing

A hybrid strain YJF1484 was made by crossing an *S. cerevisiae* strain YJF153 (*MATa hoΔ*::*dsdAMX4*, derived from an oak tree isolate YPS163) and an *S. uvarum* strain YJF1450 (*MATα hoΔ*::*NatMX*, derived from CBS7001 and provided by C. Hittinger). The hybrid was typed by PCR (Albertin et al. 2013) and found to carry *S. cerevisiae* mitochondrial DNA. A diploid *S. cerevisae* strain YJF1463 was made by crossing YJF153 (*MATa hoΔ*::*dsdAMX4*) and YJF154 (*MATα hoΔ*::*dsdAMX4*, derived from YPS163). The diploid *S. uvarum* strain YJF2602 was made by crossing YJF1449 (*MATa*, *hoΔ*::*NatMX*, derived from CBS7001) and YJF1450 (*MATα hoΔ*::*NatMX*).

Three replicate overnight cultures of the diploid hybrid YJF1484 were used to inoculate 50 ml YPD cultures (1% yeast extract, 2% peptone, 2% glucose) and incubated at either 22°C, 33°C or 37°C at 300 rpm. Cells were harvested at mid-log phase and RNA was extracted with phenol/chloroform. The nine RNA samples were enriched for mRNA by poly A purification, reverse transcribed, fragmented, ligated to indexed adaptors and sequenced on a HiSeq (1×50 bp run) at Washington University's Genome Technology Access Center.

Allele-specific expression differences

Reads were mapped using Bowtie2 (Langmead and Salzberg 2012) to a combined *S. cerevisiae* and *S. uvarum* genome. The YJF153 genome was generated by converting the S288c (R64-1-1) reference to YJF153 using GATK (v3.3-0) and YJF153 variants. YJF153 variants were called using GATK and 5.3 million paired-end (2×101 bp) HiSeq reads (SRAXXXXX). Annotations for the YJF153 genome were obtained using S288c annotations and the UCSC LiftOver tool. The YJF1450 genome and annotation files were obtained from Scannell et al. (2011). We obtained an average of 5.5 million mapped reads per sample after removing duplicate reads and reads with low mapping quality (MQ < 2). Read counts for each gene were generated using HTSeq-count (Anders et al. 2015) and species-specific counts of 5,055 orthologs were generated using previously defined orthologs (Scannell et al. 2011). To quantify any systematic bias in read mapping we calculated the ratio of normalized *S. cerevisiae* to *S. uvarum* expression levels and found a median of 0.998, indicating no systematic read mapping bias. In our data,

expression differences did not correlate with GC content (p = 0.74, linear regression), which was a concern in a previous report (Bullard et al. 2010).

Significant differences in expression were tested using a generalized linear model with a negative binomial error model (Anders et al. 2010). Using normalized read counts we tested each gene for i) temperature effects, ii) allele effects, and iii) temperature-allele interactions by dropping terms from the full model: *counts ~ allele + temperature + allele*temperature*, where *allele* and *temperature* are terms indicating the species' allele and temperature effect and the star indicates an interaction. A false discovery rate (FDR) cutoff of 0.05 was used for significance.

Quantitative PCR

Quantitative PCR (qPCR) was used to quantify the expression of *HSP104* in the hybrid as well as both parental strains following temperature treatment. Overnight cultures were grown at 23°C, diluted to an optical density (OD600) of 0.1 in YPD for temperature treatment and grown at 10°C, 23°C and 37°C for two days, 6 hours and 5 hours respectively. The middle and high temperature cultures were shaken at 250 rpm whereas the low temperature cultures were grown without shaking. At the time of collection, the OD600 of the cultures were all within the range of 0.5 - 1.9. RNA was extracted as described above, DNase I treated (RQ1 RNase-free DNase, Promega) and cDNA was synthesized (Protoscript II Reverse Transcriptase, New England Biolabs). qPCR amplifications used the Power SYBR Green Master Mix (Thermo Fisher Scientific Inc.) and were quantified on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Each PCR reaction was run in triplicate and one sample was removed from analysis due to a high standard error of deltaCt values (>0.4) among the three technical replicates. For each sample, expression of *HSP104* was measured relative to *ACT1* expression. Because we used allele-specific primers to distinguish *S. cerevisiae* and *S. uvarum* alleles of *HSP104*, the expression levels were corrected using the PCR efficiency of each primer sets, determined by standard curves. Genomic DNA of YJF1484 was used as a calibrator and to remove any plate-to-plate differences.

Sign test for directional divergence

Pathways and groups of co-regulated genes were tested for directional divergence using a sign test as previously described (Bullard et al. 2010). Each gene was assigned a score 0 if the gene showed no ASE, 1 if the gene showed ASE and the *S. cerevisiae* allele was expressed higher than the *S. uvarum* allele and -1 if the gene showed ASE and the *S. cerevisiae* allele was expressed lower than the *S. uvarum* allele. Scores for all the genes in a co-regulated group (Gasch et al. 2004) were summed and tested for significant deviations from 0 by permutation resampling of scores across all 5055 genes. To correct for multiple comparisons, the false discovery rate was estimated from the permuted data across all groups. The analysis was independently applied to genes up- and down-regulated at 22°C.

Association with genomic features

Expression levels were associated with features of intergenic sequences, defined as sequences between annotated coding sequences. Intergenic sequences were obtained from http://www.SaccharomycesSensuStricto.org and pairwise alignments were generated using FSA (Bradley et al. 2009). Substitution rates were calculated using the HKY85 model of nucleotide substitution implemented in PAML (Yang 2007).

Transcription factor binding site scores were generated by Patser (Hertz and Stormo 1999) with 244 position weight matrix (PWM) models from YeTFasCo (expert-curated database, de Boer and Hughes 2012), using a pseudocount of 0.001. Binding site scores are the log-likelihood of observing the sequence under the motif model compared to a background model of nucleotide frequencies (G+C = 34.2% for *S. cerevisiae* and 36.3% for *S. uvarum*). For each gene we used the highest scoring binding site within its upstream intergenic region. Negative scores were set to zero. Binding sites associated with temperature effects were identified by linear regression with the average binding site score of the two species. Mann-Whitney tests were used to assess enrichment of binding sites in temperature-responsive genes compared to genes without a temperature response. Motif models that were significant for both linear regression and Mann-Whitney tests after Holm-Bonferroni correction were considered positive hits.

Predicted nucleosome occupancy was generated by NuPoP (Xi et al. 2010), using the yeast model for both species. The average nucleosome occupancy across each promoter was used. For each intergenic region, we calculated a weighted score: the average binding site score of the two species * (1- nucleosome occupancy of *S. cerevisiae* promoter). Linear regression and Mann-Whitney tests were used to predict temperature effects by the weighted scores.

Binding site divergence for each binding site model was calculated by the difference between the highest scoring site for each allele. To test for associations between expression and the combined divergence of all binding sites we use used the average of the absolute value of binding site divergence. For each motif model, linear regression was used to test association between binding site divergence and allele specific effects.

Results

Effects of temperature on allele-specific expression

To measure the effects of temperature on allele-specific expression (ASE) we generated RNA-seq data from an *S. cerevisiae* × *S. uvarum* hybrid during log phase growth at low (22°C), intermediate (33°C) and high (37°C) temperatures. We found 58% of 5,055 orthologs exhibited allele-specific expression (FDR < 0.05, Supplementary Data File 1). Allele differences were similar across temperatures with Pearson's correlation coefficients of 0.90, 0.92 and 0.96 for 22-37°C, 22-33°C and 33-37°C, respectively, and only 136 genes were found to have significant allele-by-temperature interactions. In addition, the proportion of genes with the *S. cerevisiae* allele expressed at higher levels than the *S. uvarum* allele was 49.9, 50.7, 49.8% at 22, 33 and 37°C, respectively. Thus, there is no tendency toward higher *S. cerevisiae* allele expression at high temperature. *S. uvarum* alleles can be expressed at the same level as their *S. cerevisiae* ortholog at 37°C despite the fact that these promoters don't experience high temperature in *S. uvarum* due to its temperature restriction.

Effects of temperature on hybrid expression

To characterize temperature-dependent changes in gene expression we examined 211 genes that showed both a significant temperature effect (FDR < 0.05) and a 2-fold or more difference between the low (22°C) and high (37°C) temperatures. Unexpectedly, genes expressed at higher levels at the low temperature were enriched for genes involved in protein folding (*AHA1*, *MDJ1*, *BTN2*, *SSA2*, *HSP104*, *HSC82*, *SIS1*, *STI1*, *HSP82*, *CUR1*, p = 0.00829, Table S1). Typically, protein chaperones are induced in response to heat stress or misfolded proteins (Verghese et al. 2012).

To confirm the higher expression of genes involved in protein folding at 22°C and test whether this expression is specific to the hybrid or also found in one of the parents, we examined *HSP104* expression by quantitative PCR (Figure 2). Similar to our RNA-seq data, in the hybrid *HSP104* is expressed at higher levels at low temperatures (10°C and 23°C) compared to high temperatures (37°C) (5-fold change, p = 0.0006, t-test). Consistent with prior work (Gasch et al. 2000), in both parental species *HSP104* is expressed at the same level across temperatures and any transient induction that might have occurred upon a shift to 37°C is no longer present (linear regression, p = 0.11 for *S. cerevisiae* and 0.13 for *S. uvarum*). However, in *S. uvarum HSP104* is expressed at higher levels than *S. cerevisiae* across all temperatures (t-test, p = 0.007, 0.013, 0.006 for 10°C, 23°C and 37°C, respectively). The atypical pattern of *HSP104* expression in the hybrid can be explained by a change in the dominant trans-acting environment. At low temperatures (10°C and 23°C) *S. uvarum* tends to dominate the trans-environment leading to high levels of *HSP104* expression whereas at 37°C *S. cerevisiae* completely dominates the *trans*-environment leading to low levels of *HSP104* expression.

Directional divergence in temperature dependent allele-specific expression

Genes that show species-specific differences in expression that depend on temperature may have evolved in response to temperature adaptation. We therefore examined the 136 genes with a significant temperature-by-allele interaction and found *S. cerevisiae* alleles tend to have stronger temperature responses than *S. uvarum* alleles (Figure 3). Furthermore, there was a positive correlation between the temperature (37°C over 22°C) and allele effect (*S. uvarum* over *S. cerevisiae*) at 22°C (Spearman's rho =

0.33, p = 8.1e-05) but not the allele effect at 33°C (Spearman's rho = -0.01, p = 0.91) or 37°C (Spearman's rho = -0.13, p = 0.12). The positive correlation indicates that genes induced at 22°C tend to have higher *S. cerevisiae* expression while genes repressed at 22°C tend to have lower *S. cerevisiae* expression. These correlations imply a consistent directional effect in the *cis*-regulatory evolution of temperature responsive genes such that the *S. cerevisiae* alleles response more strongly than the *S. uvarum* alleles.

Under a neutral model with no change in the selective constraints on gene expression, allele-specific differences in gene expression between species are expected to be symmetrically distributed. Parallel directional changes in gene expression among a group of functionally related or co-regulated genes might reflect selection (Bullard et al. 2010, Fraser 2011). Consistent with a neutral model, the fraction of genes where the *S. cerevisiae* is expressed at higher levels than the *S. uvarum* allele is 50.10% (1478/2950) among all genes exhibiting significant ASE. However, this fraction (F_{SC}) deviates from the neutral expectation for the 136 genes exhibiting temperature-by-allele interactions. F_{SC} is 63.3% among genes up-regulated at 22°C and 32.6% among those down-regulated at 22°C (Fisher's exact test, p = 0.014 and 0.025, respectively) when calculated using ASE at 22°C. This difference is specific to ASE at 22°C since F_{SC} calculated using ASE at 37°C is not associated with the genes' temperature response. The same trend is present for all genes with significant temperature and allele effects, where F_{SC} is 54.8% (319/582) in genes up-regulated at 22°C and 45.9% (250/545) in genes down-regulated at 22°C.

Divergence in *S. cerevisiae* and *S. uvarum* expression could be the result of positive selection for a stronger response in *S. cerevisiae*, or loss of constraint in *S. uvarum*. One prediction derived from positive selection for a stronger response is a higher substitution rate along the *S. cerevisiae* compared to the *S. uvarum* lineage. Using previously derived intergenic substitution rates across four *Saccharomyces* species (Engle and Fay, 2012) we failed to find evidence for this prediction. For 275 genes up-regulated at 22°C with higher *S. cerevisiae* expression, the intergenic substitution rates leading to the *S. cerevisiae* and *S. uvarum* lineages do not deviate from background substitution rates derived from all other genes (p = 0.85, Mann-Whitney test). For 253 genes down-regulated at 22°C with lower *S. cerevisiae* expression, the *S. uvarum* lineage has higher substitution rates than that of *S. cerevisiae*, *S. paradoxus* and *S. mikatae*,

and these differences are greater than background substitution rates (Mann-Whitney tests, p = 0.00053, 0.0026 and 0.0067, respectively).

To identify functionally related or groups of co-regulated genes that might underlie the directional changes in ASE we used a previously developed sign test for selection (Bullard et al. 2010, Fraser 2011). Because of the small number of genes (136) with significant allele-temperature interactions, we used all genes with significant temperature effects. Using previously defined groups of related genes (Gasch et al. 2004), we separately tested for asymmetric 22°C ASE among genes expressed at higher (617 genes) or lower (529 genes) levels at 22°C compared to 37°C (see Methods for details). We found two clusters (Lysine biosynthesis, Cluster_ASH1) exhibiting significant asymmetric divergence in ASE among the 22°C-up and one (Membrane localization) among the 22°C-down set, respectively ($p < 0.006$, Table 1, S2, S3). Of the significant groups, the lysine group was previously found in *S. cerevisiae* × *S. uvarum* hybrids (Bullard et al. 2010). As the allele effects in the lysine group and membrane associated group are in the same direction as the general trend (higher *S. cerevisiae* expression for 22°C-up genes, lower *S. cerevisiae* expression for 22°C-down genes), we consider them to be contributors to the directional divergence in temperature responses.

Promoter changes associated with expression divergence

To identify promoter features that could explain allele-specific differences in expression we examined intergenic substitution rate, transcription factor binding site scores and their interaction with nucleosome occupancy. Among ASE genes, intergenic substitution rates are weakly correlated with gene expression divergence (Spearman's rho = 0.064, p = 0.002). Given these differences we also calculated rates of binding site divergence using binding sites scores from 244 transcription factor binding site models (de Boer and Hughes 2012) and found a weak correlation between expression divergence and binding site divergence (Spearman's rho = 0.05, p = 0.0119).

To identify binding sites that could explain allele-specific expression we first tested each binding site model for its ability to predict temperature responsive genes (22°C vs 37°C). We identified 17 motifs

associated with genes induced at 22°C and 13 motifs associated with genes repressed at 22°C (Holm–Bonferroni corrected $p < 0.05$ for both linear regression and Mann-Whitney test, Figure S1). Many of the motifs (11/17) associated with up-regulated genes are similar to the stress response element (AGGGG), including the canonical stress response factors *MSN2* and *MSN4*. Other motifs known to be involved in the stress response include the heat shock factor *HSF1*, which is consistent with the observed up-regulation of heat shock genes at 22 °C (Table S1). Motifs enriched upstream of down-regulated genes are involved in glucose repression, e.g. *MIG1*, *MIG2*, *MIG3* and *ADR1*. *UME6* was also found, consistent with down-regulation of meiotic genes at 22°C revealed by GO analysis (Table S1). We also examined the correlations using a weighted score that accounts for both TF binding and nucleosome occupancy (see Methods), but the correlations were not greatly improved with the nucleosome weighted binding site scores.

Given the motifs associated with the temperature response, we tested each motif for an association between binding site divergence and ASE at 22°C. Within genes down-regulated at 22°C, divergence of 5 motifs was found to have a weak but significant association with expression divergence (*MIG1*, *MIG2*, *MIG3*, *TDA9* and *YGR067C*, linear regression, Holm-Bonferroni corrected $p < 0.05$, Figure S1). Only one motif (*ARO80*) was correlated with ASE in genes up-regulated at 22°C when genes that show allele-by-temperature effect were specifically examined ($p = 0.0028$, adjusted r-squared = 0.11). The weak correlations suggest that ASE is likely to be caused by *cis*-regulatory mutations outside of known binding sites.

Discussion

Environment-dependent gene expression is an important component of fitness. As a consequence, adaptive changes in gene expression may be found in environmental responsive genes. In this study we show that *S. cerevisiae* alleles respond more strongly to temperature than *S. uvarum* alleles. The directional divergence of *S. cerevisiae* alleles towards a stronger temperature response is consistent with

thermal adaptation, although loss of constraint in *S. uvarum* cannot be ruled out. Below, we discuss these results in relation to prior studies of variation in gene expression across environments and discuss the challenge of identifying changes in promoter sequences responsible for divergence in gene expression.

Environment-dependent cis-effects

Changes in gene regulation may be an important aspect of how species adapt to different environments. Although there is extensive variation in gene expression-by-environment interactions (Hodgins-Davis and Townsend 2009), the extent to which these differences are caused *cis*- or *trans*-acting factors is not as well characterized. We find that most *cis*-effects do not depend on temperature, only 136 of the 2,950 genes exhibiting ASE show temperature-dependent ASE. Thus, even though *S. uvarum* promoters have never been exposed to high temperatures, they can drive expression levels similar to those of *S. cerevisiae*. The consistent *cis*-effects across temperatures suggest that most *cis*-regulatory divergence is not associated with thermal divergence between the two species. Previous studies also found *cis*-effects tend to be constant across environments and only a small subset of them are environment-dependent (Smith and Kruglyak 2008, Tirosh et al. 2009, He et al. 2016, Fear et al. 2016). Although we did not examine *trans*-effects genome-wide, the shift in the *trans*-effect of *HSP104* with temperature is consistent with prior work showing that *trans*-effects play a more pronounced role in environment-dependent differences in gene expression (Smith and Kruglyak 2008, Tirosh et al. 2009).

Unexpected heat shock response at low temperatures

The non-canonical expression of heat shock genes at 22°C is mysterious. Because we measured expression at constant temperatures we did not expect to see induction of heat shock genes, which normally occurs within 30 minutes of treatment and then dissipates (Gasch et al. 2000). Given the high expression level of *HSP104* in *S. uvarum* across all temperatures, one potential explanation for the heat shock response is a *trans*-signal produced by the *S. uvarum* genome. The absence of the heat shock response in the hybrid at high temperature may be a consequence of loss of the *S. uvarum trans*-signal,

although this does not explain the high *HSP104* expression at high temperature in *S. uvarum*. Sample mix-up is unlikely as the *HSP104* experiment was done independently and is consistent with the original RNA-seq experiment.

The heat shock gene expression profile shows that the hybrid is under stress at 22°C but not 37°C. To better understand this counterintuitive phenomenon, we compared the hybrid expression profile to previously published *S. cerevisiae* (Gasch et al. 2000) and *S. uvarum* (Caudy et al. 2013) datasets. The hybrid temperature effect (37°C over 22°C) associates with 285 of 477 stress responses of either *S. cerevisiae* or *S. uvarum* (Spearman's correlation test, Holm-Bonferroni corrected $p < 0.05$). However, 232 of the 285 correlations are negative, implying that 22°C is more stressful than 37°C in the hybrid. Interestingly, the strongest positive correlation is between the hybrid's temperature response and *S. uvarum*'s 17°C to 30°C response at 60 min (Spearman's rho = 0.23, Holm-Bonferroni corrected p = 5.39E-48). In contrast, the correlations with *S. uvarum*'s 25°C to 37°C or 25°C to 42°C response are negative. Similar to the hybrid, heat shock genes are expressed higher at 17°C than 30°C in *S. uvarum*, but the pattern is not seen in the other two temperature shifts (Caudy et al. 2013). These differential correlations indicate *S. uvarum*'s heat shock response may be sensitive to specific temperatures used in the shifts. Taken together, the stress response induced in the hybrid at 22°C may reflect a contribution from the non-canonical temperature response in *S. uvarum.*

Directional evolution of the temperature response

The ratio of *S. cerevisiae* to *S. uvarum cis*-effects (F_{SC}) significantly deviate from a neutral model with no change in constraint. The pattern of direction evolution towards a strong *S. cerevisiae* response is only apparent by differentiating between up/down temperature changes, since F_{SC} is *S. cerevisiae* biased for up-regulated genes and *S. uvarum* biased for down-regulated genes. This signature of directional evolution could be the result of positive selection for stronger responses in *S. cerevisiae*, or loss of constraint in *S. uvarum*. In prior studies, Fraser et al. (2012) distinguished between these two possibilities using polymorphism within species. However, the thermal differentiation of *S. cerevisiae* and *S. uvarum*

is not recent and so would not be expected to affect linked polymorphism through hitchhiking. Bullard et al. (2010) used McDonald-Kreitman method to test for positive selection in coding sequences, but there was not enough statistical power for testing gene groups that we found (lysine biosynthesis, membrane associated groups, Bullard et al. 2010). Our analysis of intergenic substitution rates did not distinguish between adaptation or loss of constraint, since no difference was found for up-regulated genes and a higher rate in *S. uvarum* was found for down-regulated genes.

The cause and implications of a stronger or weaker temperature response remains to be determined. One indication that *S. cerevisiae* may have evolved a more robust response to environmental stress is that it generally exhibits better growth across diverse, often stressful environmental conditions compared to *S. uvarum* and a related cryophilic species *S. kudriavzevii* (Warringer et al. 2011). Alternatively, the cryophilic species might have evolved in low temperature environments where selection for a robust heat response was lacking. Studies with Antarctic species have shown the *HSP70* heat shock response can be lost in cold environments (La Terza et al. 2001, Place and Hofmann 2005). However, the *S. uvarum* response is not easily interpreted as we found higher *HSP104* expression compared to *S. cerevisiae* at multiple temperatures. Previous reports have shown that *S. uvarum* does exhibit a canonical stress response under acute heat shock (Caudy et al. 2013). For *S. uvarum*, we found weaker responses specifically for metabolic genes such as lysine or aromatic amino acid catabolic genes (Table S3). Given these observations, it is reasonable to also consider the possibility that *S. uvarum*'s weaker response is advantageous to its cryophilic lifestyle.

The lysine and membrane associated groups are the main co-regulated groups associated with directional evolution of the temperature response. The membrane group is enriched for genes in the plasma membrane (Table S1) and this cluster was previously found to be expressed at higher levels in *S. paradoxus* compared to *S. cerevisiae*, *S. kudriavzevii* and *S. uvarum* (Martin et al. 2012). However, of the 10 genes that contributed to the directional pattern we observed (*CTF13*, *GRX6*, *NGL3*, *OSW2*, *RIM101*, *RTA1*, *SNO1*, *SNZ1*, *SPS100*, *YNL208W*), only three are annotated as localizing in membranes (*GRX6*, *OSW2*, *RTA1*). Even so, temperature is known to influence membrane fluidity and function, and the

ability to maintain membrane integrity is critical for heat tolerance (Vigh et al. 1998). The fact that the *S. uvarum* alleles of these genes are not repressed as much as *S. cerevisiae* alleles at low temperature might reflect species' difference in membrane protein composition and potentially thermal tolerance.

The expression of genes in the lysine group is of particular interest as it was previously found to show evidence of directional evolution between *S. cerevisiae* and *S. uvarum* (Bullard et al. 2010). Seven genes in the cluster (*LYS1*, *LYS2*, *LYS4*, *LYS9*, *LYS12*, *LYS20*, *LYS21*) show both directional allele effects and temperature effects (Table 1) and *S. uvarum* alleles respond more weakly than those of *S. cerevisiae* for six of them. The lysine biosynthesis pathway is induced by mitochondrial retrograde signaling in response to compromised mitochondrial respiratory function (Liu and Butow, 2006). While it is unclear why *S. cerevisiae* alleles of the lysine pathway are more responsive than those of *S. uvarum*, respiratory deficient petite mutants are known to be thermosensitive in *S. cerevisiae* (Davidson and Schiestl 2001) and thus mitochondrial respiratory function may be related to *S. cerevisiae*'s thermotolerance.

Binding sites are only weakly related to expression divergence

Consistent with previous reports (Tirosh et al. 2008; Tirosh and Barkai 2008; Chen et al. 2010; Zeevi et al. 2014), we were only able to find weak correlations between binding site changes and allele-specific expression. Previous work has shown that binding sites in nucleosome depleted regions are more likely to cause changes in gene expression (Swamy et al. 2011). Yet, incorporation of nucleosome positioning did not improve our ability to predict gene expression, consistent with another study that found no relationship between divergence in nucleosome occupancy and gene expression in yeast (Tirosh et al. 2010). One explanation for the weak correlations is that ASE may often be caused by *cis*-regulatory mutations outside major binding sites, e.g. Levo et al. (2015). Genes in the lysine biosynthesis pathway provide a good example of conserved binding sites: seven genes in the pathway show higher *S. cerevisiae* expression, yet binding sites for *LYS14*, the major transcription factor that regulates these genes (Becker et al. 1998), are conserved in all of them. Furthermore, the lysine genes are also not enriched for

divergence in other motifs present upstream of these genes (e.g. *MOT2*, *XBP2*, *RTG1*, *RTG3*, $p > 0.05$, Mann-Whitney test).

Despite binding site divergence being only weakly related to ASE, we found a few significant associations with specific binding sites. One of these, *ARO80* sites, correlated with temperature-dependent expression differences largely due to two genes *ARO9* and *ARO10* (Figure S1, S2). In both cases, the *S. uvarum* promoters have lower binding scores and lower expression of the *S. uvarum* allele (Figure S2). Interestingly, the number of monomers in the *ARO80* binding sites also differs between *S. cerevisiae* and *S. uvarum*. In both genes, *S. cerevisiae* sites are tetrameric and *S. uvarum* sites are trimeric (Figure S2). The example of *ARO80* suggests expression divergence might associate with changes in the number of binding sites, instead of gain/loss of binding sites.

Signatures of selection on cis-acting divergence in gene expression

The sign test of allele imbalance across functionally related genes has been a powerful approach to detecting polygenic adaptation. The test has been used in a variety of configurations to identify adaptive changes in gene expression levels (Fraser et al. 2010, Fraser 2011, Naranjo et al. 2015, He et al. 2016). However, previous configurations of the test were applied to expression levels under standard growth conditions. By applying a modified configuration of the sign test to condition-specific ASE we avoid the need to *a priori* specify groups of functionally related or co-regulated genes. By grouping genes based on their environmental response the test does not rely on external information or arbitrary classification of gene groups, but only on gene expression data generated in the same experiments. Further applications of the sign test to environment-dependent changes in gene expression may therefore yield new insights into adaptive changes in gene expression.

Acknowledgements

We thank Chris Hittinger for sharing the *S. uvarum* strain, Ching-Hua Shih for help with the RNA-seq analysis pipeline and other members of Fay lab for constructive comments. This work was supported by an NIH grant (GM080669) to JCF.

References

Albertin W, et al. 2013. The mitochondrial genome impacts respiration but not fermentation in interspecific *Saccharomyces* hybrids. *PLoS One* 8:e75121.

Anders S, et al. 2010. Differential expression analysis for sequence count data. *Genome Biol*. 11:R106.

Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.

Becker B, Feller A, El Alami M, Dubois E, Piérard A. 1998. A nonameric core sequence is required upstream of the *LYS* genes of *Saccharomyces cerevisiae* for Lys14p-mediated activation and apparent repression by lysine. *Mol. Microbiol*. 29:151–163.

de Boer CG, Hughes TR. 2012. YeTFaSCo: A database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res*. 40:169–179.

Bradley RK, et al. 2009. Fast statistical alignment. *PLoS Comput. Biol.* 5:e1000392

Bullard JH, Mostovoy Y, Dudoit S, Brem RB. 2010. Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc. Natl. Acad. Sci. U. S. A.* 107:5058–63.

Carroll SB. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101:577–80.

Caudy AA, et al. 2013. A new system for comparative functional genomics of *Saccharomyces* yeasts. *Genetics* 195:275–287.

Chang J, et al. 2013. The molecular mechanism of a *cis*-regulatory adaptation in yeast. *PLoS Genet.* 9:1–8.

Chen K, Van Nimwegen E, Rajewsky N, Siegal ML. 2010. Correlating gene expression variation with *cis*-regulatory polymorphism in *Saccharomyces cerevisiae*. *Genome Biol. Evol.* 2:697–707.

Davidson JF, Schiestl RH. 2001. Mitochondrial respiratory electron carriers are involved in oxidative stress during heat Stress in Saccharomyces cerevisiae. Mol. Cell. Biol. 21:8483–8489.

Engle EK, Fay JC. 2012. Divergence of the yeast transcription factor FZF1 affects sulfite resistance. *PLoS Genet*. 8.

Fay JC, et al. 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol*. 5:R26.

Fay JC, Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* 100:191–9.

Fear JM, et al. 2016. Buffering of genetic regulatory networks in *Drosophila melanogaster*. *Genetics* 203:1177-1190.

Fraser HB. 2011. Genome-wide approaches to the study of adaptive gene expression evolution: systematic studies of evolutionary adaptations involving gene expression will allow many fundamental questions in evolutionary biology to be addressed. *BioEssays* 33:469–477.

Fraser HB, et al. 2011. Systematic detection of polygenic *cis*-regulatory evolution. *PLoS Genet*. 7:e1002023.

Fraser HB, et al. 2012. Polygenic *cis*-regulatory adaptation in the evolution of yeast pathogenicity. *Genome Res.* 22:1930–1939.

Fraser HB, Moses AM, Schadt EE. 2010. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* 107:2977–82.

Gasch AP, et al. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11:4241–4257.

Gasch AP, et al. 2004. Conservation and evolution of *cis*-regulatory systems in ascomycete fungi. *PLoS Biol*. 2:e398.

Gibson G. 2008. The environmental contribution to gene expression profiles. *Nat. Rev. Genet.* 9:575–81.

Gonçalves P, Valério E, Correia C, de Almeida JMGCF, Sampaio JP. 2011. Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces* species. *PLoS One*

6:e20739.

Grishkevich V, Yanai I. 2013. The genomic determinants of genotype × environment interactions in gene expression. *Trends Genet.* 29:479–487.

Gruber JD, Vogel K, Kalay G, Wittkopp PJ. 2012. Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: Frequency, effects, and dominance. *PLoS Genet.* 8:e1002497.

Grundberg E, et al. 2011. Global analysis of the impact of environmental perturbation on *cis*-regulation of gene expression. *PLoS Genet.* 7:e1001279.

He F, et al. 2016. The footprint of polygenic adaptation on stress-responsive *cis*-regulatory divergence in the *Arabidopsis* genus. *Mol. Biol. Evol*. 33:1–34.

Hertz GZ, Stormo GD. 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* 15:563–577.

Hodgins-Davis A, Townsend JP. 2009. Evolving gene expression: from G to E to G×E. *Trends Ecol. Evol*. 24:649–658.

Landry CR, Oh J, Hartl DL, Cavalieri D. 2006. Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene* 366:343–351.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.

Levo M, et al. 2015. Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.*:gr.185033.114.

Li Y, et al. 2006. Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet*. 2:2155–2161.

Liu Z, Butow RA. 2006. Mitochondrial retrograde signaling. *Annu. Rev. Genet*. 40:159–185.

López-maury L, Marguerat S, Bähler J. 2008. Tuning gene expression to changing to evolutionary adaptation. *Nat. Rev. Genet*. 9:583–593.

Martin HC, Roop JI, Schraiber JG, Hsu TY, Brem RB. 2012. Evolution of a membrane protein regulon in *Saccharomyces*. *Mol. Biol. Evol*. 29:1747–56.

Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. 2015. Selection on noise constrains variation in a eukaryotic promoter. *Nature* 521:344–347.

Naranjo S, et al. 2015. Dissecting the genetic basis of a complex *cis*-regulatory adaptation. *PLoS Genet*. 11:1–19.

Place SP, Hofmann GE. 2005. Constitutive expression of a stress-inducible heat shock protein gene, *hsp70*, in phylogenetically distant Antarctic fish. *Polar Biol*. 28:261–267.

Romero I, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet*. 13:505–516.

Roop JI, Chang KC, Brem RB. 2016. Polygenic evolution of a sugar specialization trade-off in yeast. *Nature*:1–14.

Salvadó Z, et al. 2011. Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Appl. Environ. Microbiol.* 77:2292–302.

Scannell DR, et al. 2011. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 (Bethesda)*. 1:11–25.

Smith EN, Kruglyak L. 2008. Gene-environment interaction in yeast gene expression. *PLoS Biol*. 6:e83.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62:2155–77.

Swamy KBS, Chu W-Y, Wang C-Y, Tsai H-K, Wang D. 2011. Evidence of association between nucleosome occupancy and the evolution of transcription factor binding sites in yeast. *BMC Evol. Biol*. 11:150.

La Terza A, Papa G, Miceli C, Luporini P. 2001. Divergence between two Antarctic species of the ciliate *Euplotes*, *E. focardii* and *E. nobilii*, in the expression of heat-shock protein 70 genes. *Mol. Ecol. 10*:1061–1067.

Tirosh I, Barkai N. 2008. Evolution of gene sequence and gene expression are not correlated in yeast. *Trends Genet.* 24:109–113.

Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324:659–662.

Tirosh I, Sigal N, Barkai N. 2010. Divergence of nucleosome positioning between two closely related yeast species: genetic basis and functional consequences. *Mol. Syst. Biol.* 6:365.

Tirosh I, Weinberger A, Bezalel D, Kaganovich M, Barkai N. 2008. On the relation between promoter divergence and gene expression evolution. *Mol. Syst. Biol.* 4:159.

Verghese J, Abrams J, Wang Y, Morano KA. 2012. Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. *Microbiol. Mol. Biol. Rev.* 76:115–58.

Vigh L, Maresca B, Harwood JL. 1998. Does the membrane's physical state control the expression of heat shock and other genes? *Trends Biochem. Sci*. 23:369–374.

Warringer J, et al. 2011. Trait variation in yeast is defined by population history. *PLoS Genet.* 7:e1002111.

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

Xi L, et al. 2010. Predicting nucleosome positioning using a duration Hidden Markov Model. *BMC Bioinformatics* 11:346.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol*. 24:1586–1591.

Yun Y, Adesanya TMA, Mitra RD. 2012. A systematic study of gene expression variation at single nucleotide resolution reveals widespread regulatory roles for uAUGs. *Genome Res.* 22:1089–1097.

Zeevi D, et al. 2014. Molecular dissection of the genetic mechanisms that underlie expression

conservation in orthologous yeast ribosomal promoters. *Genome Res.* 24:1991–1999.

Zheng W, Gianoulis TA, Karczewski KJ, Zhao H, Snyder M. 2011. Regulatory variation within and between species. *Annu. Rev. Genomics Hum. Genet.* 12:327–346.

Tables

Table 1. Groups of genes showing directional evolution in response to temperature

Class	Group[1]	Total number of genes in the group[2]	Sum of scores[3]	p-value[4]
22°C > 37°C	Lysine biosynthesis	9 (7)	7	4×10^{-4}
22°C > 37°C	Cluster_ASH1	34 (6)	-5	0.0047
22°C < 37°C	Membrane localization	71 (10)	-10	$<10^{-4}$

[1] Groups are named by GO term analysis of genes in the cluster, defined by Gasch et al. (2004), Lysine biosynthesis (Cluster_Lysine), Membrane localization (Cluster_adata-CalciumSpecific), except for Cluster_ASH1 for which no enrichment was found. Results for all groups with or without temperature effects are given in Table S2, S3 and S4.

[2] Number of genes in the temperature-response class is shown in parentheses.

[3] Positive scores indicate *S. cerevisiae* alleles are expressed higher than *S. uvarum* alleles; negative scores indicate *S. cerevisiae* alleles are expressed lower than *S. uvarum* alleles.

[4] FDR at $p < 0.006$ is 0.28 for 22°C-up genes and 0.17 for 22°C-down genes.

Figure legends

Fig. 1. Temperature dependent growth difference between *S. cerevisiae* and *S. uvarum*. Growth is after 17 days at 4°C, 3 days at 20°C and 2 days at 33°C and 37°C, with platings on YPD at 1:3 serial dilutions.

Fig. 2. Temperature dependent *HSP104* expression in *S. cerevisiae*, *S. uvarum* and their hybrid. Expression is based on qPCR with points showing the mean and bars the standard errors. Hybrid expression is the sum of the two alleles.

Fig. 3. *S. cerevisiae* alleles response more strongly to temperature than *S. uvarum* alleles. 136 genes are ranked by the temperature effect of *S. cerevisiae* allele. For each gene, red points represent *S. cerevisiae* alleles and blue points represent *S. uvarum* alleles. *S. uvarum* alleles that show weaker or opposite responses than *S. cerevisiae* alleles (96 genes) are shown in grey and stronger responses (40 genes) are shown in orange. Sc, *S. cerevisiae*. Su, *S. uvarum*.

Figure 1.

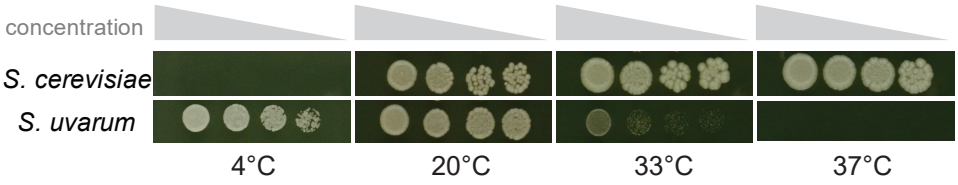

Figure 2.

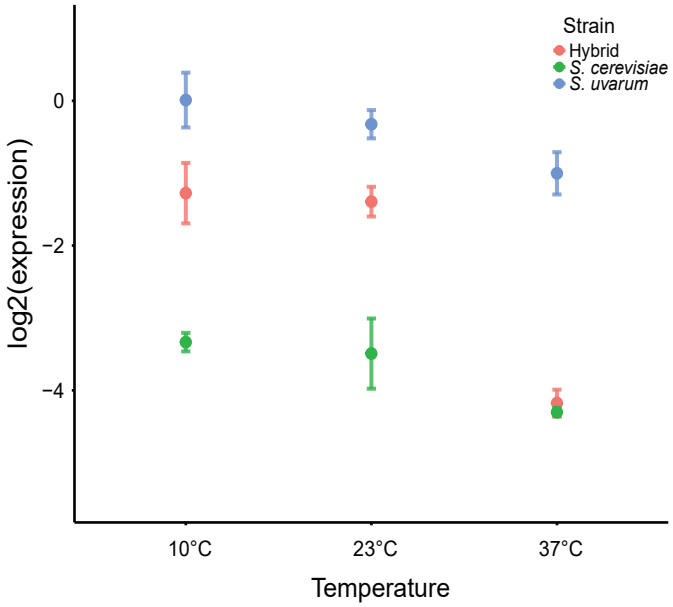

Figure 3.