Inheritance of Gene Expression Level and Selective Constraints on *Trans-* and *Cis-*Regulatory Changes in Yeast

Bernhard Schaefke,[†,1,2,3] J.J. Emerson,[†,4,5] Tzi-Yuan Wang,[2] Mei-Yeh Jade Lu,[2] Li-Ching Hsieh,[6,7] and Wen-Hsiung Li*[,2,8]

[1]National Yang-Ming University, Taipei, Taiwan

[2]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

[3]Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan

[4]Department of Ecology & Evolutionary Biology, University of California, Irvine

[5]Center for Complex Biological Systems, University of California, Irvine

[6]Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung, Taiwan

[7]Biotechnology Center, National Chung Hsing University, Taichung, Taiwan

[8]Department of Ecology and Evolution, University of Chicago

[†]These authors contributed equally to this work.

Corresponding author: E-mail: whli@uchicago.edu.

Associate Editor: Jianzhi Zhang

Abstract

Gene expression evolution can be caused by changes in *cis-* or *trans*-regulatory elements or both. As *cis* and *trans* regulation operate through different molecular mechanisms, *cis* and *trans* mutations may show different inheritance patterns and may be subjected to different selective constraints. To investigate these issues, we obtained and analyzed gene expression data from two *Saccharomyces cerevisiae* strains and their hybrid, using high-throughput sequencing. Our data indicate that compared with other types of genes, those with antagonistic *cis–trans* interactions are more likely to exhibit over- or underdominant inheritance of expression level. Moreover, in accordance with previous studies, genes with *trans* variants tend to have a dominant inheritance pattern, whereas *cis* variants are enriched for additive inheritance. In addition, *cis* regulatory differences contribute more to expression differences between species than within species, whereas *trans* regulatory differences show a stronger association between divergence and polymorphism. Our data indicate that in the *trans* component of gene expression differences genes subjected to weaker selective constraints tend to have an excess of polymorphism over divergence compared with those subjected to stronger selective constraints. In contrast, in the *cis* component, this difference between genes under stronger and weaker selective constraint is mostly absent. To explain these observations, we propose that purifying selection more strongly shapes *trans* changes than *cis* changes and that positive selection may have significantly contributed to *cis* regulatory divergence.

Key words: **gene regulation, expression evolution, *cis* effect, *trans* effect, functional constraint, natural selection.**

Article

Introduction

Phenotypic variation within or between species can be caused by differences either in protein sequences or in the abundance level or timing of gene expression. Divergence in gene expression has been proposed to be a major factor in the evolution of phenotypic differences between closely related species (Ohno 1972; King and Wilson 1975). Identifying the genetic changes underlying such expression differences is of great importance for understanding the evolution of gene regulation and its role in phenotypic evolution and speciation.

The genetic causes of gene expression changes can be classified into two categories: changes in *cis*-acting elements (e.g., promoters and enhancers), which are on the same chromosome of the gene they affect, and changes in *trans*-acting factors (e.g., transcription factors and chromatin modifiers), which are diffusible and can influence the expression of genes on other chromosomes. The way in which gene expression is changed can affect its inheritance pattern and evolution (Ronald and Akey 2007). Thus, it is important to distinguish between these two types of change to understand the causes of intraspecific variation and interspecific divergence in gene expression.

Two complementary methods exist for uncovering the regulatory basis of gene expression differences: expression quantitative trait loci (eQTL) mapping and hybrid experiments. Although eQTL mapping can be used to locate the elements responsible for expression variation and to differentiate between local and distant regulators, it cannot reliably distinguish between *cis-* and *trans*-acting elements, as a *trans* regulator also can be located on the same chromosome in close proximity to the target gene, and a *cis* regulatory element (e.g., an enhancer) can be distantly located from it (Rockman and Kruglyak 2006; Emerson and Li 2010). The first

eQTL study on a genomic scale was conducted in *Saccharomyces cerevisiae* (Brem et al. 2002), and the method was also successfully applied to investigate the roles of distant-acting variation (Yvert et al. 2003) and of epistatic interactions (Brem and Kruglyak 2005; Brem et al. 2005) in determining different gene expression phenotypes. In contrast, hybrid experiments can be used to distinguish between *cis*- and *trans*-regulatory changes contributing to differences in gene expression. An additional advantage of the hybrid approach used in this study is that only the two parental strains and their F1 hybrid have to be assayed. Hybrid experiments require the measurement of mRNA levels in homo- or hemizygous parental strains and of allele-specific expression (ASE) in their cross. As the same set of diffusible elements acts on both parental alleles in the hybrid, *trans* variation produces no differential effect on the two alleles, so that ASE differences in the F1 hybrid can be interpreted as a direct representation of *cis*-regulatory variation (Cowles et al. 2002; Wittkopp et al. 2004). Several studies have used this approach to investigate the relative importance of *cis* and *trans* variation for the expression response in yeast to different environmental conditions (Tirosh et al. 2009; Li et al. 2012) or in the evolution of specific pathways (Chang et al. 2008). Furthermore, it has been successfully applied to investigate the role of *cis* and *trans* effects in nucleosome positioning (Tirosh et al. 2010), in protein expression (Khan et al. 2012), and in the regulation of DNA replication timing in yeast (Muller and Nieduszynski 2012).

Past research has unraveled the genetic causes of gene expression differences between various *Saccharomyces* strains and species and their relevance for phenotypic evolution. In a recent study, several transcription factors were identified whose expression profiles differ between wine strains, resulting in the production of volatile aroma compounds in a predictable manner (Rossouw et al. 2012). Another study identified a subset of genes for which the expression divergence between two yeast strains correlates well with changes in predicted transcription factor binding sites (TFBSs) (Chen et al. 2010). The Barkai lab compared the expression of genes induced during mating among three different *Saccharomyces* species and found that divergence in TFBSs of the transcription factor STE12 could explain about half of the expression differences, although they found no general correlation between promoter sequence divergence and gene expression evolution in yeasts and mammals (Tirosh et al. 2008). The same group found a positive correlation between divergence in promoter sequence and the *cis*-component of gene expression differences in a comparison of two *Saccharomyces* species and their hybrid (Tirosh et al. 2009).

Several properties of the promoter region play an important role in the evolution of gene regulation, especially in *trans*: Genes with a significant *trans* effect in a comparison between *S. cerevisiae* and *S. paradoxus* lack a pronounced nucleosome-free region, tend to contain a TATA box in their promoter region, and consistently display larger expression differences between different yeast strains or species (Tirosh et al. 2009). In addition, the sensitivity of expression levels to genetic perturbations in mutation accumulation lines is enhanced for TATA box-containing genes (Landry et al. 2007). The relationship between nucleosome occupancy of the promoter region and gene expression evolution is complex. Yeast genes can be roughly classified into two groups: those with a promoter containing a well-defined nucleosome-free region close to the transcription start site, referred to as DPN (depleted proximal-nucleosome) genes and those with a promoter lacking such a region, referred to as OPN (occupied proximal-nucleosome) genes (Tirosh and Barkai 2008). The latter group exhibits more expression plasticity between different environmental conditions and more cell-to-cell variability (Tirosh and Barkai 2008). Although some studies found that changes in nucleosome occupancy are related to divergence in gene expression (Field et al. 2009; Tsankov et al. 2010), a comparison of *S. cerevisiae* and its closest relative, *S. paradoxus*, found no relationship between divergence of gene expression and divergence of nucleosome positioning (Tirosh et al. 2010). In contrast, a recent experimental evolution study selecting yeast strains for overexpression of a target gene found different evolutionary mechanisms for the two classes: DPN genes have a stronger tendency to be duplicated (which could be considered as a *cis* acting mutation) than OPN genes, which have predominantly undergone *trans* regulatory changes (Rosin et al. 2012).

Several studies have shown a relationship between the mode of inheritance of gene expression levels and the molecular mechanisms of gene regulatory differences. Alleles conferring *cis* regulatory variation tend to have an additive influence on gene expression level, with the expression level in the hybrid being intermediate between those of the two parents (Lemos et al. 2008; McManus et al. 2010). This is hypothesized to contribute to positive selection on *cis*-regulatory elements over long evolutionary time (Lemos et al. 2008), because the expression levels of single genes can be "fine-tuned," so that a gradual adaptation to changing selective pressures can take place. On the other hand, genes with antagonistic *cis-trans* interactions have been found to be enriched for over- or underdominant inheritance of gene expression level (with the mRNA level in the hybrid being either higher or lower than those in both parents) and could play a role in the development of hybrid incompatibilities (Landry et al. 2005; McManus et al. 2010). It has been proposed that even when the expression level of a gene is under stabilizing selection, its regulatory elements may undergo divergent evolution between species if mutations in *cis* (*trans*) are balanced by compensatory mutations in *trans* (*cis*) (Landry et al. 2005).

Different inheritance patterns of changes in *cis* and *trans* and the potentially pleiotropic nature of *trans* mutations are likely to result in different evolutionary constraints. Changes in *trans* regulators can impact the expression of multiple downstream genes and can thus be expected to affect multiple phenotypic traits more often than *cis* regulatory changes.

Indeed, a study of mutation accumulation and natural isolate lines in *Caenorhabditis elegans* found that most *trans*-acting mutations that resulted in expression changes of multiple genes were quickly removed by selection in

natural populations (Denver et al. 2005). Furthermore, differences in *cis* regulatory elements appear to play a larger role in expression differences between species than within species (Wittkopp et al. 2008a; Emerson et al. 2010). Additionally, genes which show a significant gene expression difference in *trans* between two different *S. cerevisiae* strains also tend to exhibit more gene expression divergence in *trans* between *S. cerevisiae* and *S. paradoxus*, while in *cis* this trend is weak or absent (Emerson et al. 2010).

These findings could be explained by stronger positive selection on *cis* divergence and stronger selective constraint on *trans*-acting factors. As selective constraint is expected to affect essential genes more strongly than nonessential genes, its impact on gene regulatory evolution in *cis* and *trans* can be evaluated by comparing the *cis* and *trans* components of within-species and between-species gene expression differences for genes of higher and lower importance.

In this study, we investigate the relationship between gene expression inheritance patterns and regulatory differences in *cis* and *trans* between two strains of *Saccharomyces cerevisiae*, RM11-1 (RM) and BY4741 (BY). Our results indicate that genes with antagonistic *cis*–*trans* interactions are more likely to show an under- or overdominant inheritance pattern in our within-species hybrids, whereas essential genes are less likely to exhibit an underdominant inheritance pattern.

In addition, we integrate the data from an interspecies comparison (Tirosh et al. 2009) with our data and evaluate the role of selective constraint on changes in *cis* and *trans* factors. We show that *trans* regulatory mutations indeed tend to be under stronger selective constraint than *cis* regulatory mutations and that this observation may explain the relative contributions of *cis* and *trans* changes to intra- and interspecific gene expression differences.

Results

Transcriptome Sequencing and Expression Level Estimation

We selected 4,442 genes for our study (Methods; Emerson et al. 2010) and estimated their expression levels using Illumina paired-end (PE) sequencing with a read length of 151 base pairs (bp). In the hybrid sample, 4,558,258 reads were mapped specifically to the BY genome and 4,564,464 reads to the RM genome. In the coculture sample, 3,745,275 reads were mapped as BY specific and 3,776,660 reads as RM specific. This new data set enabled us to analyze ASE differences with greater power than the expression data obtained in a previous study (Emerson et al. 2010). In this study, 4,237 out of the 4,442 genes under study have more than 10 sequence reads for both alleles in both experiments (coculture and hybrid) and were used for further analyses. Among the 4,237 genes under study, 2,268 genes (53.5%) show a significant expression polymorphism in coculture and 1,207 (28.5%) show a significant ASE difference in the hybrid (binomial exact test, false discovery rate [FDR] <5%, see Materials and Methods). These are two times higher than the corresponding numbers in Emerson et al. (2010), in which the numbers of genes with significant ASE differences are 1,294 (35.1%) and 488 (13.2%) out of the 3,685 genes under study, for coculture and hybrid, respectively, using the same criteria for statistical significance. Thus, the new data set allows us to do more rigorous statistical analyses.

Classifying Gene Expression Differences in Terms of *Cis* and *Trans* Effects

As discussed earlier, a *cis*-regulatory factor influences the expression level of only the allele on the same chromosome, whereas a *trans*-regulatory factor can affect the expression of both alleles in a cell. Therefore, it is possible to estimate the relative contributions of *cis*- and *trans*-regulatory changes to differences in gene expression between the RM and the BY strain by comparing the ASE in the hybrid to expression differences between the two parental strains (Wittkopp et al. 2004). We assume that there is no allele-specific preferential binding of the maternal or paternal transcription factor (Takahasi et al. 2011) and that the expression of an allele is independent of the other, that is, there is no transvection. Therefore, the expression differences between the two parental alleles in the hybrid are interpreted as a direct representation of *cis*-regulatory differences (Cowles et al. 2002) (see examples in supplementary figs. S1 and S2, Supplementary Material online), because in the same cell the *trans*-regulatory milieu is identical for the two alleles. The expression difference between the two parental strains in coculture is thus interpreted as a combination of *cis*- and *trans*-effect (see examples in supplementary figs. S1 and S2, Supplementary Material online). In agreement with previous studies (Wittkopp et al. 2008b; Emerson et al. 2010), we found that *trans* effects dominate in our within-species comparison: 1,577 (37.3%) of the 4,237 genes under study show a significant *trans* effect, whereas only 1,267 (30%) show a significant *cis* effect (significance was determined using the likelihood ratio test, FDR < 5%, see Materials and Methods). The median absolute *trans* effect (0.301) is significantly higher than the median *cis* component (0.166) (Wilcoxon rank sum test, P value $< 2.2 \times 10^{-16}$). Although genes with significant expression differences in coculture or in the hybrid showed a higher single nucleotide polymorphism (SNP) density than nondifferentially expressed genes (Wilcoxon rank sum test: P value $< 2.2 \times 10^{-16}$), we found no difference between *cis* or *trans* regulatory changes related to gene SNP density (supplementary table S1, Supplementary Material online). Differentially expressed genes also showed a significantly higher sequence divergence in the promoter region (defined as 500 bp upstream of the transcription start site) (supplementary table S2, Supplementary Material online). This trend is significant for both the *cis* and the *trans* effect but stronger for the *cis* component of expression differences, as can be expected (Wilcoxon rank sum test: *cis*: P value $= 7.79 \times 10^{-12}$, *trans*: P value $= 1.497 \times 10^{-5}$). Similarly, genes whose promoter region contains a TATA box (Basehoar et al. 2004) are more likely to be differentially expressed than those without a TATA box (supplementary table S3, Supplementary Material online). Genes without a well-defined nucleosome-free region close to the

transcription start site (i.e., OPN genes) are more likely to be differentially expressed than those with such a region (i.e., DPN genes) (the gene sets were defined in Tirosh et al. 2008). This relationship is observed in *trans* and in *cis* (supplementary table S4, Supplementary Material online) but is stronger for the *trans* effect (Fisher's exact test: *trans*: P value = 3.993×10^{-5}, *cis*: P value = 0.0027). We tested whether genes with significant and consistent *cis* or *trans* effects (in the old and the new data set) were enriched in specific biological processes or cellular components in the Gene Ontology (GO) annotation, using the FunSpec analysis tool (Robinson et al. 2002). Genes with a significant *trans* component were enriched in mitochondrial electron transport (GO term "mitochondrial electron transport, ubiquinol to cytochrome *c*" [GO identifier: 0006122], P value: 1.09×10^{-7}, and overlapping terms) and in the biosynthesis of ergosterol (GO term "ergosterol biosynthetic process" [GO identifier: 0006696], P value: 1.8×10^{-7}, and overlapping terms), a major component of the fungal cell membrane. Genes with a significant *cis* effect were enriched in oxidation/reduction among biological processes (GO term "oxidation–reduction process," GO identifier: 0055114, P value: 9.45×10^{-7}, and overlapping terms) and in the cell wall among cellular components (GO term: "cell wall," GO identifier: 0005618, P value: 2.19×10^{-9}, and related terms). This is consistent with the previous finding of an enrichment for cell wall related genes among those with local regulatory differences between the RM and BY strains (Chen et al. 2010). To estimate the importance of transcription factors in *trans* regulatory evolution versus changes in sensory and signaling molecules or chromatin modifiers, we compared gene pairs which either share a common regulator (Teixeira et al. 2006) but belong to different expression modules (Ihmels et al. 2002), or which belong to the same module(s), but are not known to be regulated by an identical transcription factor. We did not find any significant difference between these two sets of genes regarding the probability of both genes in a pair having expression differences in *trans* in the same direction, that is, favoring the allele from the same strain (either BY or RM) for both genes (supplementary table S5, Supplementary Material online).

The genes under study were classified into five categories, as in McManus et al. (2010), but with some different category names as follows:

1) Nondifferential: no significant expression difference between the RM and the BY allele in coculture or hybrid. It is the same as the "conserved" category in McManus et al. (2010).
2) *Cis* only: A significant *cis*- component but no significant *trans* difference.
3) *Trans* only: A significant *trans*-component but no significant *cis* difference.
4) *Cis* + *Trans*: The *cis* and *trans* components are both significant and work in the same direction (supplementary fig. S1, Supplementary Material online).
5) *Cis* — *Trans*: The *cis* and *trans* components are both significant but have opposite effects. It can be divided into three subcategories according to the relative magnitudes of the *cis* and *trans* components (supplementary fig. S2, Supplementary Material online):
   a) "*Cis* — *Trans* (t > c)" (i.e., "*cis* — *trans*" with a greater absolute *trans* effect): The $\log_2$ expression ratios in coculture and in the hybrid have different signs (the allele which is more highly expressed in the hybrid has lower expression levels in the parental comparison); it is equivalent to the "*cis* × *trans*" category in McManus et al. (2010).
   b) "*Cis* — *Trans* (c = t)": The *cis* and the *trans* component work in opposite directions and have approximately the same absolute value; no significant expression difference between the two alleles in the parental strains; it is equivalent to the "compensatory" category in McManus et al. (2010).
   c) "*Cis* — *Trans* (c > t)" (i.e., "*cis* — *trans*" with a greater absolute *cis* effect): The *cis* and *trans* components have opposite signs, but the $\log_2$ expression ratios in hybrid and in coculture have the same sign (i.e., the same allele is favored in coculture and hybrid, but the absolute expression ratio in hybrid is greater than that in coculture); it was assigned to the "*cis* + *trans*" category by McManus et al. (2010).

Among the 4,237 genes under study, 2,077 (49%) showed no significant expression difference between the RM and BY alleles in hybrid or in coculture and were classified as nondifferential. Among the 2,160 (51%) "differentially expressed" genes, 583 genes (13.8%) were classified as "*cis* only," whereas 893 (21.1%) as "*trans* only" (fig. 1 and table 1). The group "*cis* + *trans*" comprises only 172 genes (4.1%). The total number of genes falling into the "*cis* – *trans*" category is 512 (12.1%). Among these, 234 genes (5.5%) have *cis* and *trans* effects of approximately equal magnitude and were classified as "*cis* — *trans* (c = t)". Only 71 genes (1.7%) in the "*cis* — *trans*" category have a larger *cis* component and were classified as "*cis* — *trans* (c > t)." In contrast, 207 genes (4.9%) fall into the "*cis* — *trans* (t > c)" category, having a stronger *trans* effect than a *cis* effect. These observations show the overall prevalence of *trans* regulatory changes in our within-species comparison.

Inheritance Mode of Gene Expression Level Versus ASE Differences in *Cis* and *Trans*

To study the mode of inheritance, the expression levels of the hybrid and the parental strains were compared for each gene in three comparisons: 1) the expression of the gene in the parental BY strain ("BY") versus in the parental RM strain ("RM") in coculture; 2) the expression of the gene in BY versus the total expression level in the hybrid; and 3) the expression of the gene in RM versus the total expression level in the hybrid. A gene was classified as conserved if the expression difference in each of the three comparisons was not statistically significant or was less than 25%. This category comprised 43.7% of the genes (1,852/4,237). The other 2,385 genes (56.3%) were nonconserved and assigned to one of the

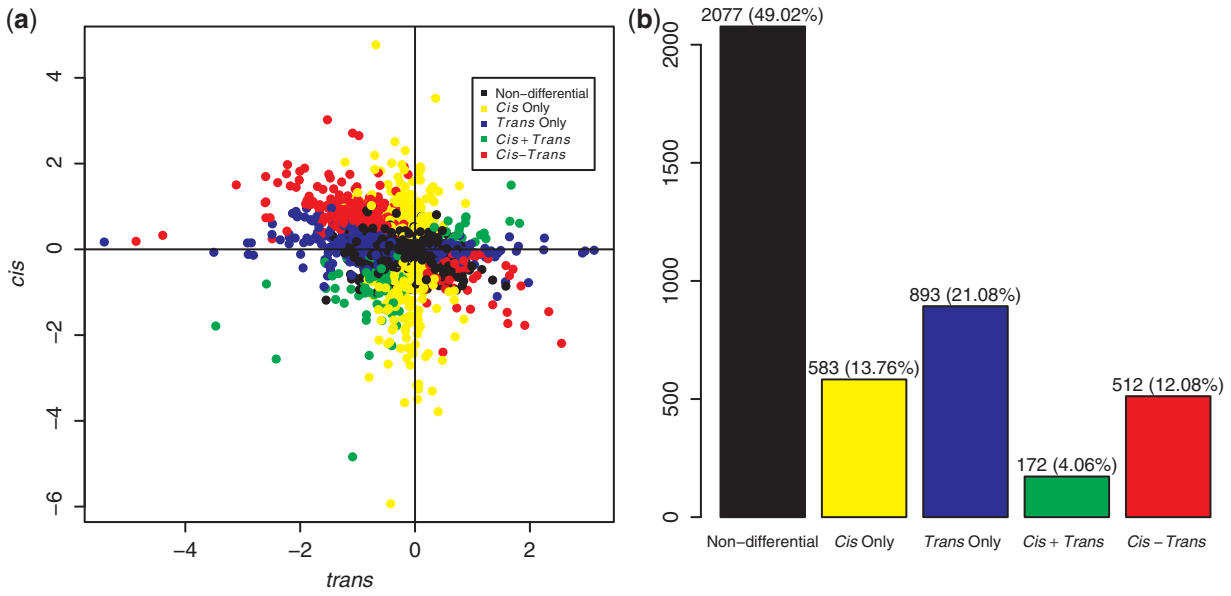

Fig. 1. Classification of genes according to *cis* or *trans* effects. (*a*) Scatterplot: *Y* axis: the *cis* component [the $\log_2$-ratio of reads in the hybrid sample mapped to the RM and BY genomes: $\log_2(e_{cis}) = \log_2(e_{Hy}) = \log_2(\varepsilon_{RM,Hy}/\varepsilon_{BY,Hy})$]. *X* axis: the *trans* component [difference between parental and hybrid $\log_2$-transformed ASE ratios: $\log_2(e_{trans}) = \log_2(e_{Co}/e_{cis}) = \log_2(\varepsilon_{RM,Co}/\varepsilon_{BY,Co}) - \log_2(\varepsilon_{RM,Hy}/\varepsilon_{BY,Hy})$]. Notations: $\varepsilon_{RM,Hy}$, expression level of the RM allele in the hybrid; $\varepsilon_{BY,Hy}$, expression level of the BY allele in the hybrid; e_{Hy}, ASE ratio in the hybrid; $\varepsilon_{RM,Co}$, expression level of the RM allele in the coculture; $\varepsilon_{BY,Co}$, expression level of the BY allele in the coculture; and e_{Co}, ASE ratio in the coculture. (*b*) The bar graph shows the number of genes in each *cis*/*trans* category.

Table 1. Number of Genes Falling into Different Combinations of Inheritance and *cis*/*trans* Categories.

Inheritance mode	Regulatory Effect					
	Nondifferential	*Trans* Only	*Cis* Only	*Cis* + *Trans*	*Cis* − *Trans*	Sum
Conserved	1,265	199	165	6	217	1,852
RM dominant	371	376	149	66	86	1,048
BY dominant	147	116	85	24	73	445
Additive	53	134	141	67	39	434
Overdominant	69	11	23	1	60	164
Underdominant	172	57	20	8	37	294
Sum	2,077	893	583	172	512	4,237

categories: "additive," "BY dominant," "RM dominant," "overdominant," and "underdominant" (fig. 2). The "additive" category comprised 434 genes (10.2%). Interestingly, 1,048 genes (24.7%) were classified as "RM dominant," but only approximately half as many (445 genes, 10.5%) were classified as "BY dominant." In total, 458 genes (10.8%) were misexpressed (overdominant or underdominant) in the hybrid; the underdominant expression pattern was found in 294 genes (6.9%) and the overdominant pattern in 164 genes (3.9%).

To investigate how the molecular mechanism of gene expression differences influences the inheritance mode of the expression level, we examined whether an inheritance mode is enriched for genes belonging to a specific expression divergence pattern (table 1). Consistent with previous studies (Lemos et al. 2008; McManus et al. 2010), we found a weak but significant relationship between *cis* regulation and additive inheritance. The median percent *cis* for genes with additive inheritance (39.84%) was significantly higher than for those with the other inheritance modes (37.46%) (Wilcoxon rank sum test: *P* value = 0.0014).

Additionally, in agreement with previous findings (Lemos et al. 2008), genes with dominant inheritance (either RM or BY dominant) showed a strong enrichment for *trans* regulatory variation. The median percent *trans* was significantly higher for genes with dominant inheritance (68.83%) than for the other genes (59.34%) (Wilcoxon rank sum test: *P* value $< 1.4 \times 10^{-14}$).

Furthermore, we investigated whether genes in the "*cis* — *trans*" category disproportionately contributed to misexpression in our within-species cross, as previously described for between-species hybrids (Landry et al. 2005; McManus et al. 2010). Indeed, we found an enrichment for misexpressed genes in the "*cis* — *trans*" category (Fisher's exact test, *P* value $< 5 \times 10^{-9}$, table 2). This relationship remains significant even when both "conserved" and "nondifferential" genes are removed from the analysis (table 2).

Different Constraints on *Cis* and *Trans* Regulatory Components

We divided genes into different classes expected to be under relatively weak or strong selective constraint, using three criteria: 1) the ratio of the rate of nonsynonymous substitution to the rate of synonymous substitution (ω): genes with an ω higher than the median value (~0.09) were classified as less conserved and those with a lower ω as more conserved; 2) connectivity in protein–protein interaction (PPI) networks (Stark et al. 2006; Collins et al. 2007): genes with more than

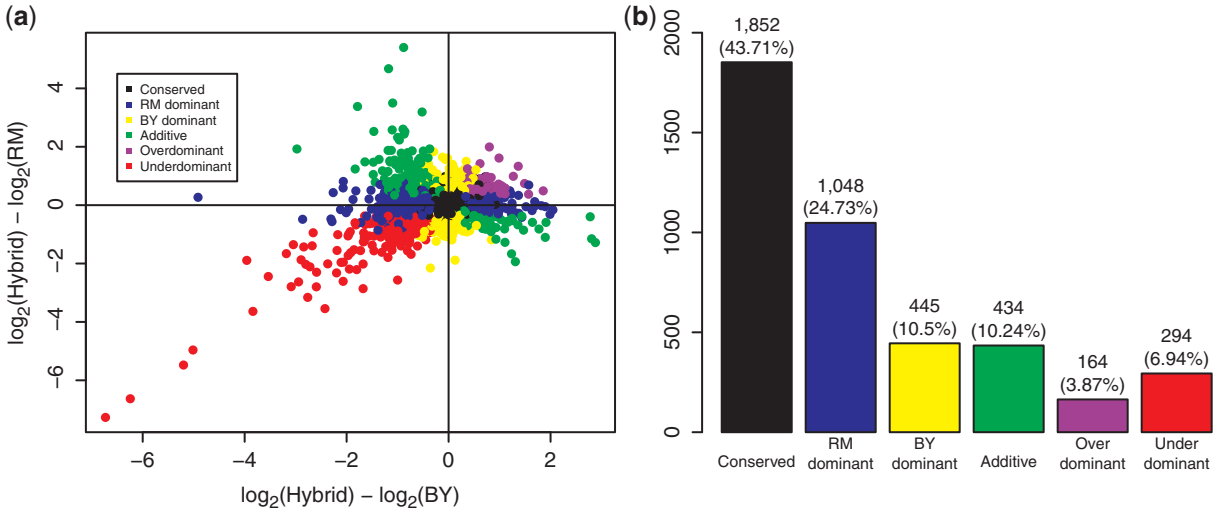

FIG. 2. Inheritance modes. (*a*) The scatterplot compares the differences in expression level between the F1 hybrid and each of the parental strains (BY on the *X* axis and RM on the *Y* axis). (*b*) The bar graph shows the number of genes in each inheritance category.

Table 2. Enrichment of Genes with Under- or Overdominant Inheritance in the "*Cis* — *Trans*" Category.

Regulatory Effect	Inheritance mode	
	Misexpressed	Other
***Cis* — *trans* vs. other categories**[a]	**97**	**415**
	361	**3,364**
***Cis* — *trans*[b] vs. other categories**[b]	**97**	**198**
	120	**1,158**

[a]Misexpressed genes are enriched for genes in the "*cis* — *trans*" category, Fisher's exact test: *P* value = 4.99×10^{-9}.

[b]Indicates that both the "conserved" and "nondifferential" genes are removed from the analysis, Fisher's exact test: *P*-value < 2.2×10^{-16}.

the median (four known interaction partners) were classified as more constrained, whereas those with no known interaction partners as less constrained; and 3) essentiality: essential genes versus genes with a fitness >0.85 in knock-out experiments (Deutschbauer et al. 2005). We used the ratio $p/(p + d)$ [polymorphism/(polymorphism + divergence)] as a measure to detect shifts toward polymorphism or divergence, where p represents the absolute value of either the *cis* or the *trans* component in gene expression differences between different strains of the same species, whereas d represents the respective value for the interspecies comparison (figs. 3 and 4) (the divergence data were obtained from Tirosh et al. 2009). The ratios $p_{cis}/(p_{cis} + d_{cis})$ and $p_{trans}/(p_{trans} + d_{trans})$ were then each compared between categories with expected strong selective constraints or with expected weak selective constraints for each of the three criteria of selective constraints (figs. 3 and 4). All three comparisons showed a significant relative abundance of polymorphism in *trans* for the less constrained category when compared with the more constrained category (Wilcoxon rank-sum test: *P* values < 0.01 in all three comparisons). In contrast, in *cis* none of the three comparisons showed a significant difference between the more and the less constrained category (*P* values > 0.2). We tested for the equality of the distributions of $p/(p + d)$ in the more constrained and less constrained categories and found significant differences for all three comparisons in *trans* (bootstrapped Kolmogorov–Smirnov [KS] tests: *P* values $\leq$ 0.002 in all three comparisons), but not in *cis* (bootstrapped KS tests: *P* values > 0.2 in all three comparisons). In agreement with these observations, we find that essential genes are significantly less likely than nonessential genes to have a significant *trans* effect (supplementary table S6, Supplementary Material online, Fisher's exact test: *P* value = 0.018), whereas there is no significant difference in *cis* (supplementary table S6, Supplementary Material online, Fisher's exact test: *P* value = 0.37).

Misexpressed inheritance modes are slightly under-represented among essential genes for our within-species data; this difference is not statistically significant in comparison with all other inheritance categories (supplementary table S7, Supplementary Material online, Fisher's exact test: *P* value = 0.35). However, misexpressed genes are significantly less likely to be essential if the comparison is restricted only to genes with a conserved total expression level (supplementary table S7, Supplementary Material online, Fisher's exact test: *P* value = 0.022). This tendency for misexpressed genes to be less essential appears to be largely due to an enrichment of nonessential genes among those with underdominant inheritance. Underdominant genes are significantly less likely to be essential, in comparison with all other inheritance categories (table 3; Fisher's exact test: *P* value < 9.763×10^{-9}) and also in comparison with overdominant genes only (table 3; Fisher's exact test: *P* value = 1.14×10^{-13}).

As genes in the "*cis* — *trans*" category are more likely to be misexpressed (table 2), we test whether this category exhibits a similar enrichment for nonessential genes. Indeed, nonessential genes are more likely to fall into the "*cis* — *trans*" category when compared with all other categories taken together. This is true not only for our within-species comparison (table 4, Fisher's exact test: *P* value = 0.078) but also for the between-species data of Tirosh et al. (table 4, *P* value = 0.0003). In contrast, compared with "*cis* + *trans*" genes only, this enrichment for nonessential genes is not

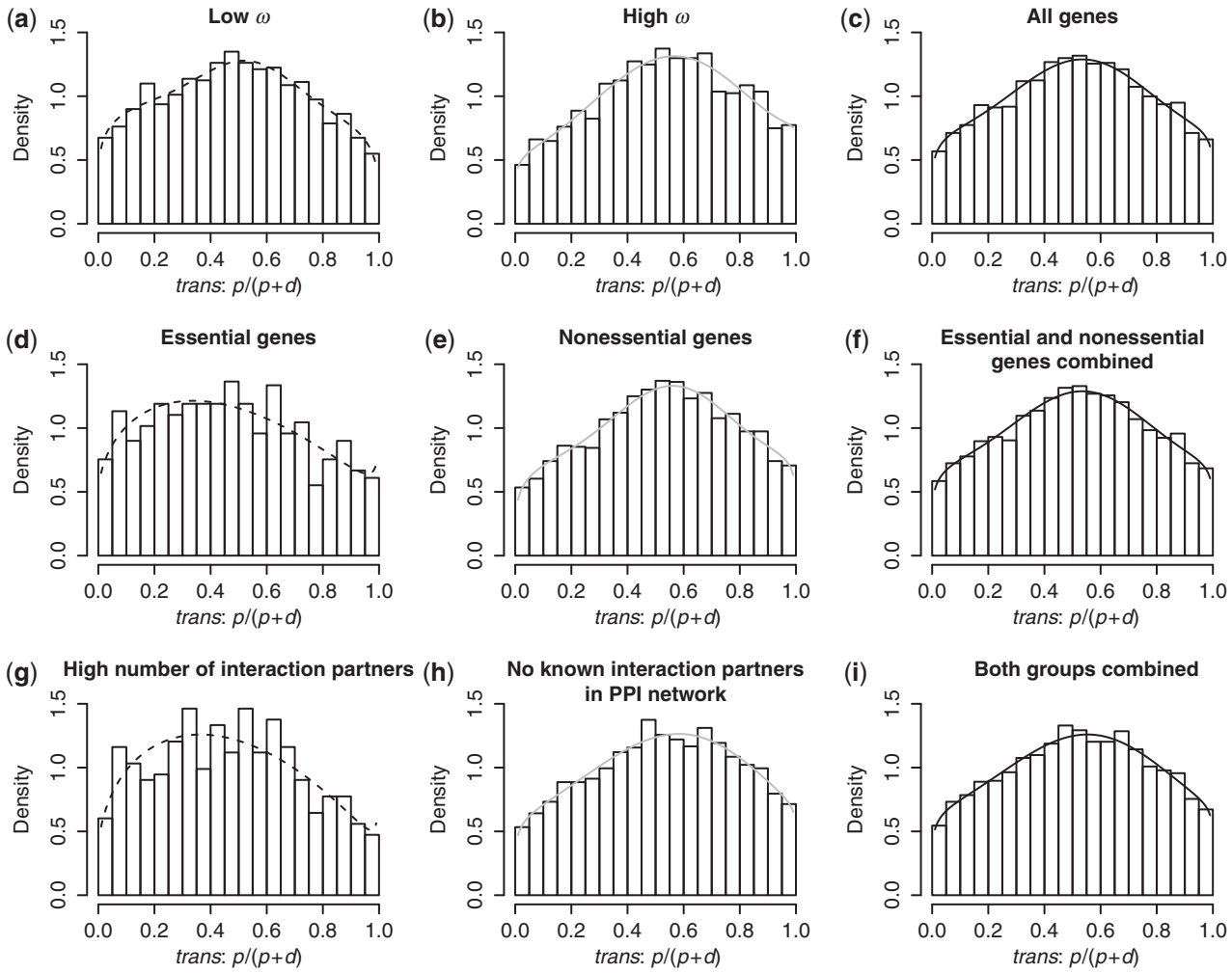

Fig. 3. The distribution of the ratio $p/(p + d)$ is significantly different between constrained and less constrained categories for all three classification systems in *trans* (*a–c*: the ratio of the rate of nonsynonymous substitution to the rate of synonymous substitution [ω]; *d–f*: essentiality; *g–i*: connectivity in PPI networks). Notations: *p*, *trans* polymorphism; *d*, *trans* divergence.

significant (P value = 0.49 within-species, and P value = 0.51 between species, table 4).

Discussion

Cis and *trans* regulatory factors differ in the way in which they influence gene expression levels and their inheritance patterns. Thus, they may be subjected to different selection pressures, and this should be reflected in the way gene regulatory networks evolve. It remains difficult to tease apart these different gene regulatory mechanisms and their evolutionary pathways. However, an elegant approach is the use of hybrid experiments and the comparison of *cis* and *trans* effects in crosses between and within species (Wittkopp et al. 2004), although some of the underlying assumptions might lead to an overestimation of the relative contribution of *cis* changes to gene expression differences (Takahasi et al. 2011).

Our new data set provides more power to detect gene expression differences than our previous data (Emerson et al. 2010). Indeed, only 48.7% (1,011/2,078) of the genes with a significant expression difference in coculture in the new data set were also significantly different in the old data set, but 78.2% (1,011/1,293) of the genes with a significant expression difference in the old data set are also significantly different in the new data set (supplementary table S8, Supplementary Material online). For ASE differences in the hybrid, the respective numbers are 27.6% (305/1,106) and 62.8% (305/486) (supplementary table S8, Supplementary Material online). The relatively small number of genes which were differentially expressed in the old data set but not in the new one (282 in coculture and 181 in the hybrid) could be due to variation between biological or technical replicates and stochastic fluctuations (Busby et al. 2011).

Our new analysis examines the different effects of functional constraints on the *cis* and *trans* components of gene expression differences. In general, highly deleterious mutations would be quickly removed from the population and are unlikely to be observed either as differences between strains or between species, whereas slightly deleterious mutations may be found in polymorphisms. The chance for slightly deleterious mutations to be observed as *trans* polymorphisms could be high, because of frequent *trans* changes (Wittkopp 2005; Landry et al. 2007) and the insufficient evolutionary time for selection to remove the mutations from the population. However, they are unlikely to contribute significantly

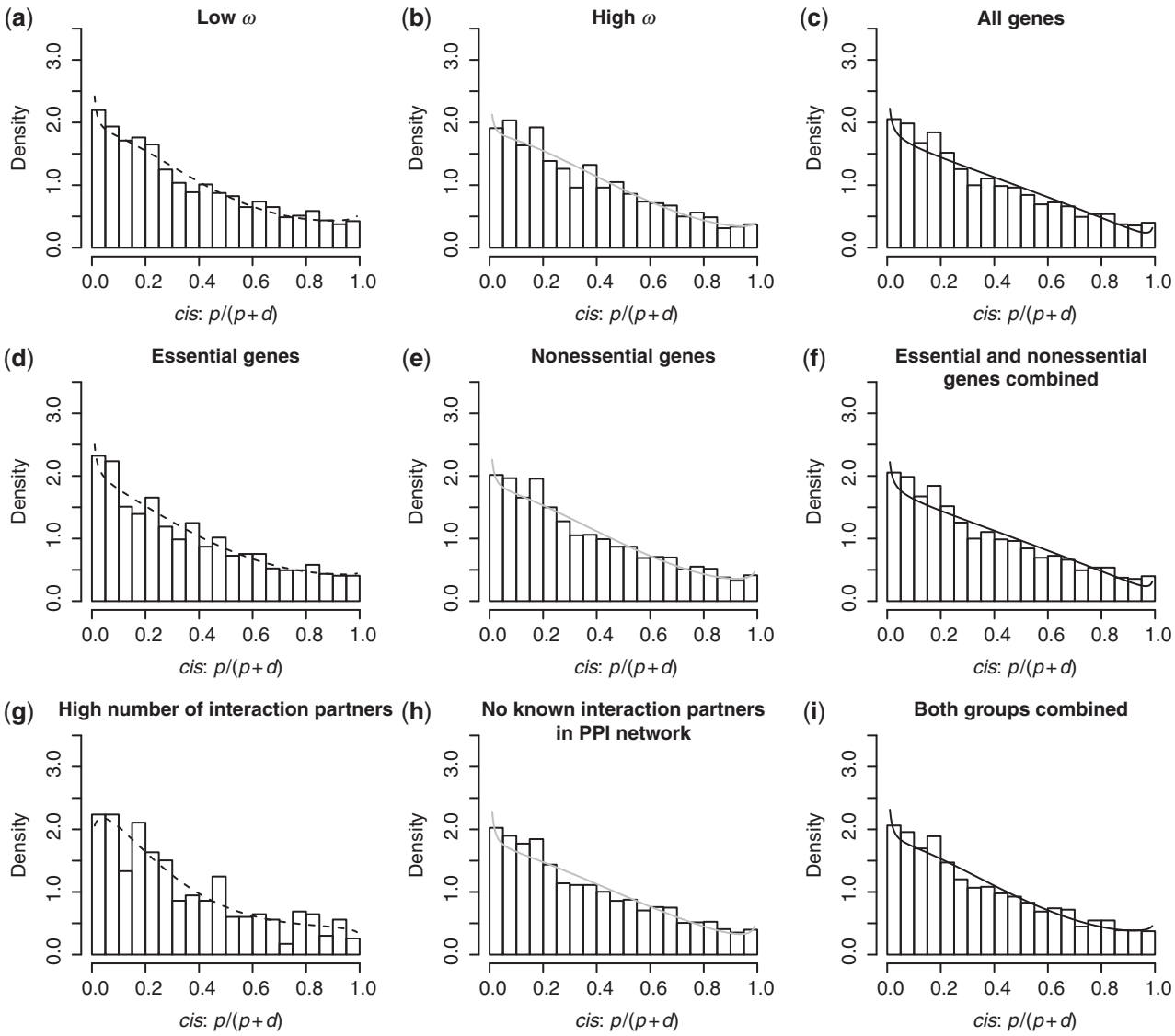

FIG. 4. In *cis*, the distribution of the ratio $p/(p + d)$ is not significantly different between constrained and less constrained categories (*a–c*: the ratio of the rate of nonsynonymous substitution to the rate of synonymous substitution [ω]; *d–f*: essentiality; *g–i*: connectivity in PPI networks). Notations: *p*, *cis* polymorphism; *d*, *cis* divergence.

Table 3. Proportions of Essential Genes in the Underdominant Inheritance Category and in the Other Inheritance Categories.

Essentiality	Inheritance Mode			
	Underdominant Only	All Other Inheritance Categories		
		Total	Overdominant	Nonmisexpressed
Essential	24[a,b]	831[a]	59[b]	772
Nonessential (fitness > 0.85)	243[a,b]	2,783[a]	86[b]	2,697

[a]Underdominant genes are significantly less likely to be essential in comparison with all other inheritance categories, Fisher's exact test: P value $< 9.763 \times 10^{-9}$.
[b]Essential genes are significantly less likely to be underdominant than overdominant, Fisher's exact test: P value $= 1.14 \times 10^{-13}$.

Table 4. Enrichment of the "*Cis* − *Trans*" Category in Nonessential Genes.

	Regulatory Effect within Species (RM vs. BY)		Regulatory Effect between Species (*S. cerevisiae* vs. *S. paradoxus*)	
Essentiality	*Cis* − *Trans*	Other Categories (*Cis* + *Trans*)	*Cis* − *Trans*	Other Categories (*Cis* + *Trans*)
Essential	91	764 (34)	166	689 (182)
Nonessential (fitness > 0.85)	392	2,634 (123)	766	2,260 (776)

NOTE.—*Cis–trans* genes are enriched for nonessential genes compared with genes in all other categories (within species: Fisher's exact test: P value = 0.078; between species: Fisher's exact test: P value = 0.0003). But there is no significant difference between the "*cis* − *trans*" and "*cis* + *trans*" (numbers in brackets) categories in the proportion of essential genes (within species: Fisher's exact test: P value = 0.49; between species: Fisher's exact test: P value = 0.51).

to fixed expression differences between species. Thus, a trend toward polymorphism in less constrained categories (representing slightly deleterious mutations) can be expected in *trans*, if different selective constraints are important in the evolution of *trans* regulation. Conversely, if gene expression evolution in *trans* is primarily neutral and mutation-driven, no such trend is expected. Our data show a clear shift toward polymorphism for less constrained categories compared to highly constrained categories in *trans* (fig. 3) and a significant difference between essential and nonessential genes in the probability of having a significant *trans* component (supplementary table S6, Supplementary Material online). The positive association of divergence with polymorphism in *trans* may imply that some of the *trans* mutations that contribute to within-species differences are neutral or nearly neutral.

In *cis*, the earlier-mentioned trends are largely absent. The association between polymorphism and divergence is much weaker in *cis* than in *trans* (supplementary table S9, Supplementary Material online). As in previous studies (Wittkopp et al. 2008a; Emerson et al. 2010), we found a stronger impact of *cis* regulatory divergence on gene expression differences between species than on those within species in comparison with the *trans* effect. Indeed, for the *cis* effect, 56% (530) of the 940 genes showing significant polymorphism show significant divergence and 52% (1,174) of the 2,263 genes showing nonsignificant polymorphism show significant divergence, whereas for the *trans* effect, the corresponding proportions are only 29% and 21%. These observations suggest positive selection has contributed to *cis* expression divergence.

Thus, our data are compatible with the view that *trans* regulatory factors are subjected to stronger selective constraint than *cis* regulatory factors. As the mutational target size in *trans* is larger than that in *cis*, *trans* differences contribute relatively more to gene expression differences within species. However, as *trans* changes are subjected to stronger selective constraint, they contribute less to between species divergence than *cis* changes (supplementary table S9, Supplementary Material online).

The fact that changes in *cis* and *trans* regulators impact gene expression in different ways is reflected in different inheritance patterns. In accordance to previous studies, genes with *cis* regulatory variants tend to show an additive inheritance pattern, while those with *trans* regulatory differences are enriched for dominant inheritance of expression level (Lemos et al. 2008; McManus et al. 2010). The lower number of "BY dominant" genes might be due to fixation of rare recessive alleles in the BY laboratory strain. Genes with antagonistic *cis*–*trans* interactions are more likely to be misexpressed in hybrids, in agreement with previous findings (Lemos et al. 2008; McManus et al. 2010). The percentage of misexpressed genes (10.8%) in our study is higher than that found in an interspecies-hybrid between *S. cerevisiae* and *S. paradoxus* (2–8%) (Tirosh et al. 2009). These values are difficult to compare because of the differing sensitivity of the different experimental tools used: detecting subtle differences in gene expression is easier with next generation sequencing (our within-species data) than with microarrays (Tirosh et al.'s between-species data). Otherwise, this finding might be surprising, as misexpression is expected to contribute to hybrid incompatibilities and speciation. Essential genes have a significantly lower probability to be segregating for mutations exhibiting underdominant inheritance (table 3). If we assume that the allelic differences between the two strains in the majority of their polymorphisms are present in "natural" *S. cerevisiae* populations, especially in human-associated environments (e.g., vineyards, as for RM), and could thus be found in heterozygotes (Magwene et al. 2011), this result may be expected, as genes which are required for reproduction and survival of the organism tend to be under stabilizing selection. If mutations in two or more regulatory loci occur which in combination lead to significantly lower expression levels of these essential genes, they will be removed from the population quickly.

Our data showed that *cis*–*trans* genes are enriched for misexpressed genes. Although more of these misexpressed genes are overdominant (60) than underdominant (37; table 1), we might also expect essential genes to be underrepresented in the "*cis* − *trans*" category. This is true not only for our within-species comparison but also for the between-species data. A possible explanation is that most of the regulatory differences in the "*cis* − *trans*" category are due to the two mutations having occurred in the same lineage, with the first mutation being slightly deleterious and the second (partially) compensating. For the first mutation to become fixed in the population, it cannot have a strongly deleterious effect. Alternatively, this observation might just reflect that essential genes are less likely to accumulate several regulatory mutations over time, which is a necessary condition for a gene to be classified as either "*cis* − *trans*" or "*cis* + *trans*." The fact that the "*cis* + *trans*" category is not significantly different from the "*cis* − *trans*" category (table 4) and is equally enriched for nonessential genes in the interspecific comparison supports this simpler hypothesis.

Materials and Methods

Yeast Strains and Culturing

Two yeast strains were used: one, designated as "BY," is a haploid laboratory strain, officially named BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) and is a descendant of S288C (Brachmann et al. 1998), and the other, designated as "RM", is formally named either RM11-1a (MATa lys2Δ0 ura3Δ0 ho::KAN) or RM11-1α (MATα lys2Δ0 ura3Δ0 ho::KAN), both of which are haploid and derived from Bb32(3), a natural isolate described previously (Mortimer et al. 1994; Brem et al. 2002). The hybrid of BY (MATa) and RM (MATα) constructed in our laboratory is named WL201.

Two culture types were prepared: coculture and hybrid. The coculture is a mixture of approximately equal numbers of BY (MATa) cells and RM (MATa) cells, whereas the hybrid strain was derived from a cross between BY (MATa) and RM (MATα). All strains were grown on the standard YPAD medium at 30 °C with 250 rpm shaking as described previously (Emerson et al. 2010).

Transcriptome Sequencing

To sequence the transcriptomes of hybrid and cocultured yeast strains, RNA was extracted using the Hot Acidic Phenol method (Kohrer and Domdey 1991) and subjected to mRNA-seq library preparation using the Illumina TruSeq mRNA-seq Sample Prep kit with some modifications. Briefly, 5 μg of total RNA from each sample was used to purify for polyA-RNA, and the mRNA was fragmented by heat at 94 °C for 8 min. Double-stranded cDNA was synthesized by random priming, end repaired, and ligated to the Y-shaped TruSeq adaptors. Samples were cleaned by AMPure beads (Agencourt) and split into two halves, which were independently assembled into two polymerase chain reaction (PCR) reactions, one using the PCR reagents provided in the Illumina kit, and the other using the KAPA PCR reagent (KAPA HiFi HotStart ReadyMix). All reactions went through 12 cycles of PCR amplifications and were cleaned by AMPure beads to remove primer dimers. The purified products were quantified by Qubit (Invitrogen) and BioAnalyzer 2100 High Sensitivity DNA Assay (Agilent). The library profiles showed a wide spectrum of fragment sizes ranging from 220 to 500 bp with a peak at approximately 285 bp. Each of the four libraries (two libraries amplified from the coculture and the hybrid experiments) were put in one lane of PE sequencing on Illumina GA IIx (IGA-IIx) in the High Throughput Sequencing Core Facility of Academia Sinica, Taiwan. The sequencing data were processed by CASAVA 1.8.2 to generate pass-filtered reads for downstream analyses.

Whole-Genome Sequencing

To quantify the relative cell numbers of the two yeast strains in coculture, genomic DNA (gDNA) from the cocultured yeast sample was sequenced (Emerson et al. 2010). The gDNA was extracted using the Qiagen Q100 Genomic Purification Kit (Qiagen), and 1 μg of the gDNA was sonicated to fragments of approximately 300–400 bp. Using the Illumina Paired-End (PE) DNA Sample Prep Kit, the fragments were end-repaired, A-tailed, and ligated to the PE adaptors. To control the precise fragment size for downstream mapping, the ligation product was fractionated on agarose gel, and fragments ranging from 400 to 500 bp in length were gel purified and amplified using the KAPA PCR kit (same as mentioned earlier). The library was cleaned using Ampure beads and assayed on Qubit and BioAnalyzer 2100. The library showed a narrow distribution with a peak approximately 437 bp and was sequenced in the same manner as the mRNA-seq libraries.

Mapping Reads to the Reference Genomes

To map sequencing reads to the reference genomes and obtain genome-specific reads, the software tool ASAP ("Allele-Specific Alignment Pipeline") was downloaded from the Bioinformatics Group at the Babraham Institute (http://www.bioinformatics.bbsrc.ac.uk/projects/, last accessed September 2012). To determine whether a given sequence matches one of the two reference genomes specifically, it performs alignments against both sequences in parallel using the Bowtie program (Langmead et al. 2009). In our analysis, the seed length was set to 40 and the maximum number of mismatches permitted in the seed was set to 2.

The BY reference genome was downloaded from the SGD project, "*Saccharomyces* Genome Database" (http://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/, last accessed April 2008). The RM reference genome was downloaded from the *Saccharomyces cerevisiae* RM11-1a Sequencing Project, Broad Institute (http://www.broadinstitute.org/annotation/genome/saccharomyces_cerevisiae, last accessed April 2008). In addition, 893 error sites against our strains, 309 from the BY strain and 584 from the RM strain, detected previously (Emerson et al. 2010) were corrected in our updated reference genomes.

From the two channels of cDNA IGA-IIx sequencing for each of the two samples, we obtained in total 60,909,895 and 52,151,477 raw reads from the hybrid sample and the coculture sample, respectively. Genes that are known to be mating-type specific or have been found to be differentially expressed between mating-types or between haploids and diploids were excluded from all further analyses (Galitski et al. 1999; Tirosh et al. 2009).

Assigning Gene Expression Differences to *Cis*- or *Trans*-Regulatory Changes

The ASE ratios and their *cis*/*trans* contributions were estimated as previously described (Emerson et al. 2010). To account for the difference between cell numbers of BY and RM in coculture, we estimated a normalization parameter based on the gDNA ratio of the two strains in the coculture experiment.

We calculated the *cis*-regulatory component (e_{cis}) of gene expression differences as the ratio of the reads mapped to the RM genome and those mapped to the BY genome in the hybrid sample: $e_{cis} = e_{Hy} = \varepsilon_{RM,Hy}/\varepsilon_{BY,Hy}$, where $\varepsilon_{RM,Hy}$ is the expression level of the RM allele in the hybrid, $\varepsilon_{BY,Hy}$ is the expression level of the BY allele in the hybrid, and e_{Hy} is the ASE ratio in the hybrid. The expression difference between the two parental strains can be attributed to both *cis* and *trans* effects (supplementary fig. S1, Supplementary Material online). We assume that *cis* and *trans* effects are multiplicative (additive, if the logarithm of the ASE ratios is considered). Thus, $e_{Co} = e_{trans}e_{cis}$ and $\log_2(e_{Co}) = \log_2(e_{trans}) + \log_2(e_{cis})$, or $e_{trans} = e_{Co}/e_{cis} = (\varepsilon_{RM,Co}/\varepsilon_{BY,Co})/(\varepsilon_{RM,Hy}/\varepsilon_{BY,Hy})$, where $\varepsilon_{RM,Co}$ = expression level of the RM allele in the coculture, $\varepsilon_{BY,Co}$ = expression level of the BY allele in the coculture, and e_{Co} = ASE ratio in the coculture. We obtain the *P* value of a hypothesis test using the likelihood ratio test with one degree of freedom under the null hypotheses of $e_{cis} = 1$ and $e_{trans} = 1$. The FDR cutoff for each test was set to 2.5% to give a combined FDR of approximately 5% (Benjamini and Hochberg 1995). Under the above formulation, genes are sorted into five categories, using R (v 2.14.1, CRAN) with the methodology described by McManus et al. (2010).

Sequence Analysis and GO Term Enrichment

The SNP density of each gene was calculated as the number of its SNPs between RM and BY divided by the length of the transcribed region in bp. The promoter region was defined as 500 bp upstream of the transcription start site. Promoter regions with a putative loss or gain of a transcription factor binding site (D. Wang, unpublished data) were excluded from this analysis. Gene classification into those with TATA box containing promoter regions and TATA-less genes was taken from Basehoar et al. (2004). Gene sets of DPN genes (genes with a well-defined nucleosome-free region close to the transcription start site) and OPN genes (genes without such a region) were defined as in Tirosh et al. (2008). Enrichment in specific biological processes or cellular components in the GO annotation was analyzed using the FunSpec analysis tool with Bonferroni correction for multiple testing (Robinson et al. 2002). It uses the hypergeometric distribution to calculate the probability (*P* value) that the intersection of a given gene list with any given functional category occurs by chance.

Transcription Factors and Gene Expression Modules

We obtained regulatory associations between transcription factors and their target genes from the YEASTRACT database (Teixeira et al. 2006). Gene expression modules as groups of coexpressed genes under several conditions (Ihmels et al. 2002) were downloaded from the Barkai lab website (http://barkai-serv.weizmann.ac.il/Modules/page/details.html, level 10, last accessed May 2013). We only considered genes with significant *trans* effects, present in one of the co-expression modules and with at least one known transcription factor regulating the gene. Genes were also required to have consistent expression differences between our old and our new data set for hybrid and coculture (either the BY or the RM allele must be higher in both data sets). Among these genes, we compared the values for $\log_2(e_{trans})$ of all possible gene pairs which either share a common regulator, but belong to different modules, or which belong to the same module(s), but are not known to be regulated by any identical transcription factor.

Inheritance Mode Classification

To account for the unequal total amounts of mRNA in different samples, we calculate the RNA ratios for BY and RM by dividing the total number of mapped RNA reads for each of the two strains by that of the hybrid, after removing outliers with extreme ratios (i.e., values below the 2.5% quantile or above the 97.5% quantile). The total expression for each gene is normalized by dividing the number of mapped mRNA reads by the respective RNA ratio; for example, the mRNA reads of each gene for RM are divided by the RNA ratio RM/Hy. The parental and hybrid data sets are analyzed for evidence of differential expression using the exact binomial test. We set the FDR cutoff to 1.696%, and only the *P* values below this threshold are considered significant, so that the probability of a false positive in one of three comparisons (discussed later) is $\leq$5% (Benjamini and Hochberg 1995). We determine the mode of expression level inheritance for a gene by comparing the following three expression levels: 1) the total expression level in the hybrid (referred to as "hybrid"); 2) the expression level of the gene in the parental RM strain (measured as the expression level of the RM allele in coculture; referred to as "RM"); and 3) the expression level of this gene in the parental BY strain (measured as the expression level of the BY allele in coculture; referred to as "BY"). In each of these comparisons, expression levels are considered as "similar" if their difference is not statistically significant or less than 1.25-fold. Genes are categorized into six different inheritance modes using R (v 2.14.1, CRAN) according to the classification of McManus et al. (2010): conserved, additive, BY dominant, RM dominant, overdominant, and underdominant (supplementary fig. S3, Supplementary Material online). Genes with similar total expression levels in hybrid and the parental strains are classified as "conserved." Genes with more than 25% expression difference and with a significant exact binomial test in at least one of the three comparisons (BY-RM, BY-hybrid, and RM-hybrid) are classified as nonconserved and further assigned to one of the five nonconserved categories:

- "Additive": The expression level in the hybrid lies in between the levels of the parental strains.
- "BY dominant": The expression level of the hybrid is similar to the parental BY strain, but significantly different from RM.
- "RM dominant": The expression level of the hybrid is similar to the parental RM strain, but significantly different from BY.
- "Underdominant": The expression level is significantly lower in the hybrid than in both parent strains.
- "Overdominant": The expression level is significantly higher in the hybrid than in both parent strains.

Selective Constraint Analysis

The *cis* and *trans* components in within-species differences (labeled here as p_{cis} and p_{trans} to represent polymorphism) were derived from our own data set, as previously described (Emerson et al. 2010). The *cis* and *trans* components for the divergence between *S. cerevisiae* and *S. paradoxus* (labeled here as d_{cis} and d_{trans} to represent divergence) were obtained from Tirosh et al. (2009). To analyze different constraints in *trans* and *cis* changes, we compared the ratio $p/(p + d)$ for the genes in two different categories of expected weak or strong selective constraint. Three such comparisons were executed with gene groupings according to three different characteristics: 1) the connectivity in PPI networks (Stark et al. 2006; Collins et al. 2007), contrasting "high connectivity" (more than four known interaction partners, which is the data set's median) against no known interaction partner; 2) sequence divergence: the ratio of the rate of nonsynonymous substitution to the rate of synonymous substitution (ω) between *S. cerevisiae* and *S. paradoxus* (lower than the data set's median of 0.0925 vs. $\geq$ 0.0925); and 3) essentiality (essential genes, which are lethal when knocked out, vs. nonessential genes, defined as those genes exhibiting a fitness of more than

0.85 in knock-out experiments) (Deutschbauer et al. 2005). The ratio $p_{cis}/(p_{cis} + d_{cis})$ (or $p_{trans}/[p_{trans} + d_{trans}]$) was then compared between categories with strong selective constraint proxies (low ω, high connectivity, and high essentiality) and weak expected selective constraint (high ω, no known interaction partner, and low essentiality) using the Wilcoxon rank-sum test. Distributions of the values for strongly constrained and weakly constrained categories were tested for equality using the bootstrap version of the Kolmogorov–Smirnov test in R (Sekhon 2011).

Supplementary Material

Supplementary tables S1–S9 and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by Academia Sinica, Taiwan and National Science Council (NSC) grants (99-2628-B-001-009-MY3 and 99-2321-B-001-041-MY2). The authors thank Fu-Jung Yu for experimental assistance, Daryi Wang for providing unpublished data, and Krishna B. Swamy and Nathan Barham for thoughtful discussion of the manuscript. They also thank two anonymous reviewers for constructive criticisms and valuable suggestions.

References

Basehoar AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116(5):699–709.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Stat Methodol.* 57(1):289–300.

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD. 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14(2):115–132.

Brem RB, Kruglyak L. 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A.* 102(5):1572–1577.

Brem RB, Storey JD, Whittle J, Kruglyak L. 2005. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 436(7051):701–703.

Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* 296(5568):752–755.

Busby MA, Gray JM, Costa AM, Stewart C, Stromberg MP, Barnett D, Chuang JH, Springer M, Marth GT. 2011. Expression divergence measured by transcriptome sequencing of four yeast species. *BMC Genomics* 12:635.

Chang YW, Robert Liu FG, Yu N, Sung HM, Yang P, Wang D, Huang CJ, Shih MC, Li WH. 2008. Roles of cis- and trans-changes in the regulatory evolution of genes in the gluconeogenic pathway in yeast. *Mol Biol Evol.* 25(9):1863–1875.

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.

Collins SR, Kemmeren P, Zhao XC, Greenblatt JF, Spencer F, Holstege FC, Weissman JS, Krogan NJ. 2007. Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol Cell Proteomics.* 6(3):439–450.

Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. 2002. Detection of regulatory variation in mouse genes. *Nat Genet.* 32(3):432–437.

Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. 2005. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat Genet.* 37(5):544–548.

Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, Nislow C, Giaever G. 2005. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169(4):1915–1925.

Emerson JJ, Hsieh LC, Sung HM, Wang TY, Huang CJ, Lu HH, Lu MY, Wu SH, Li WH. 2010. Natural selection on cis and trans regulation in yeasts. *Genome Res.* 20(6):826–836.

Emerson JJ, Li WH. 2010. The genetic basis of evolutionary change in gene expression levels. *Philos Trans R Soc Lond B Biol Sci.* 365(1552):2581–2590.

Field Y, Fondufe-Mittendorf Y, Moore IK, Mieczkowski P, Kaplan N, Lubling Y, Lieb JD, Widom J, Segal E. 2009. Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nat Genet.* 41(4):438–445.

Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR. 1999. Ploidy regulation of gene expression. *Science* 285(5425):251–254.

Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N. 2002. Revealing modular organization in the yeast transcriptional network. *Nat Genet.* 31(4):370–377.

Khan Z, Bloom JS, Amini S, Singh M, Perlman DH, Caudy AA, Kruglyak L. 2012. Quantitative measurement of allele-specific protein expression in a diploid yeast hybrid by LC-MS. *Mol Syst Biol.* 8:602.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188(4184):107–116.

Kohrer K, Domdey H. 1991. Preparation of high molecular weight RNA. *Methods Enzymol.* 194:398–405.

Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. 2007. Genetic properties influencing the evolvability of gene expression. *Science* 317(5834):118–121.

Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. 2005. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics* 171(4):1813–1822.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3):R25.

Lemos B, Araripe LO, Fontanillas P, Hartl DL. 2008. Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc Natl Acad Sci U S A.* 105(38):14471–14476.

Li CM, Tzeng JN, Sung HM. 2012. Effects of cis and trans regulatory variations on the expression divergence of heat shock response genes between yeast strains. *Gene* 506(1):93–97.

Magwene PM, Kayikci O, Granek JA, Reininga JM, Scholl Z, Murray D. 2011. Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A.* 108(5):1987–1992.

McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* 20(6):816–825.

Mortimer RK, Romano P, Suzzi G, Polsinelli M. 1994. Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae* derived from natural fermentation of grape musts. *Yeast* 10(12):1543–1552.

Muller CA, Nieduszynski CA. 2012. Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Res.* 22(10):1953–1962.

Ohno S. 1972. Argument for genetic simplicity of man and other mammals. *J Hum Evol.* 1(6):651–662.

Robinson MD, Grigull J, Mohammad N, Hughes TR. 2002. FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinformatics* 3:35.

Rockman MV, Kruglyak L. 2006. Genetics of global gene expression. *Nat Rev Genet.* 7(11):862–872.

Ronald J, Akey JM. 2007. The evolution of gene expression QTL in Saccharomyces cerevisiae. *PLoS One* 2(7):e678.

Rosin D, Hornung G, Tirosh I, Gispan A, Barkai N. 2012. Promoter nucleosome organization shapes the evolution of gene expression. *PLoS Genet.* 8(3):e1002579.

Rossouw D, Jacobson D, Bauer FF. 2012. Transcriptional regulation and the diversification of metabolism in wine yeast strains. *Genetics* 190(1):251–261.

Sekhon JS. 2011. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J Stat Softw*. 42(7):1–52.

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34(Database issue):D535–D539.

Takahasi KR, Matsuo T, Takano-Shimizu-Kouno T. 2011. Two types of cis-trans compensation in the evolution of transcriptional regulation. *Proc Natl Acad Sci U S A.* 108(37): 15276–15281.

Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sa-Correia I. 2006. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 34(Database issue):D446–D451.

Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* 18(7):1084–1091.

Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324(5927):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.

Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. 2010. The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biol.* 8(7):e1000414.

Wittkopp PJ. 2005. Genomic sources of regulatory variation in cis and in trans. *Cell Mol Life Sci.* 62(16):1779–1783.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. *Nature* 430(6995):85–88.

Wittkopp PJ, Haerum BK, Clark AG. 2008a. Independent effects of cis- and trans-regulatory variation on gene expression in *Drosophila melanogaster*. *Genetics* 178(3):1831–1835.

Wittkopp PJ, Haerum BK, Clark AG. 2008b. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet.* 40(3):346–350.

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L. 2003. Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet.* 35(1):57–64.