Roles of Trans and Cis Variation in Yeast Intraspecies Evolution of Gene Expression

Huang-Mo Sung,†* Tzi-Yuan Wang,*† Daryi Wang,‡ Yu-Shan Huang,§ Jen-Pey Wu,§
Huai-Kuang Tsai,‖ Jengnan Tzeng,* Chih-Jen Huang,§ Yi-Chen Lee,* Peggy Yang,* Joyce Hsu,*
Tiffany Chang,* Chung-Yi Cho,‖ Li-Chuan Weng,‡ Tso-Ching Lee,§ Tien-Hsien Chang,*¶
Wen-Hsiung Li,*‡‡ and Ming-Che Shih**††

*Genomics Research Center, Academia Sinica, Taipei, Taiwan; †Department of Life Sciences, National Cheng Kung University, Tainan, Taiwan; ‡Biodiversity Research Center, Academia Sinica, Taipei, Taiwan; §Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan; ‖Institute of Information Science, Academia Sinica, Taipei, Taiwan; **Department of Molecular Genetics, The Ohio State University; #Department of Ecology and Evolution, University of Chicago; §§Agricultural Biotechnology Research Center, Academia Sinica, Taipei, Taiwan; and ††Department of Biology, University of Iowa

Both cis and trans mutations contribute to gene expression divergence within and between species. We used Saccharomyces cerevisiae as a model organism to estimate the relative contributions of cis and trans variations to the expression divergence between a laboratory (BY) and a wild (RM) strain of yeast. We examined whether genes regulated by a single transcription factor (TF; single input module, SIM genes) or genes regulated by multiple TFs (multiple input module, MIM genes) are more susceptible to trans variation. Because a SIM gene is regulated by a single immediate upstream TF, the chance for a change to occur in its trans-acting factors would, on average, be smaller than that for a MIM gene. We chose 232 genes that exhibited expression divergence between BY and RM to test this hypothesis. We examined the expression patterns of these genes in a BY–RM coculture system and in a BY–RM diploid hybrid. We found that trans variation is far more important than cis variation for expression divergence between the two strains.

However, because in 75% of the genes studied, cis variation has significantly contributed to expression divergence, cis change also plays a significant role in intraspecific expression evolution. Interestingly, we found that the proportion of genes with diverged expression between BY and RM is larger for MIM genes than for SIM genes; in fact, the proportion tends to increase with the number of transcription factors that regulate the gene. Moreover, MIM genes are, on average, subject to stronger trans effects than SIM genes, though the difference between the two types of genes is not conspicuous.

Introduction

The expression of a gene is initiated via binding of specific transcription factors (TFs) to cis-regulatory sites in the promoter of the gene. Thus, expressional changes in a gene can arise from cis changes, or trans changes, or both. Cis changes are allele-specific, affecting only the expression of the allele linked to the changes but not the expression of the other allele in a diploid individual. Cis changes refer mainly to changes in the promoter region of the gene, especially cis-regulatory elements, though a change in the transcribed region, including 5′ and 3′ untranslated regions, can affect the RNA stability and thus also the expression level of the allele. In contrast, trans changes are not allele-specific and can affect the expression of both alleles in a diploid. Trans effect can be due to changes that affect the timing, level, or activity of the TFs or other regulators that control the expression of TF genes or the target gene.

Several recent studies reported that both intra and interspecies diversifications in gene expression during development in Drosophila melanogaster, Drosophila Yakuba, and Drosophila simulans are mainly due to cis variation (Rifkin et al. 2003; Osada et al. 2006). Using pyrosequencing to estimate the relative allelic expression levels between a D. melanogaster and D. simulans hybrid, Wittkopp et al. (2004) found that in 29 genes that differed in expression between the two species, every one of them exhibited cis variation. On the other hand, only 16 genes (55%) showed evidence for trans effect. In their more recent study (Wittkopp et al. 2008), the authors compared intraspecific and interspecific variation and found that the variation between D. melanogaster and D. simulans was dominated by cis variation to a larger degree than within species variation, leading to the conclusion that differences between species are more likely to be driven by mutations in cis elements than in trans factors.

In Saccharomyces cerevisiae, Yvert et al. (2003) used microarray and linkage analysis to map the genetic changes responsible for the expression differences of 2,294 genes between the BY (a laboratory strain) and RM (a wild strain) of yeast, and found that the expression divergence of 1,716 genes (~75%) between these two strains did not show self-linkage, indicating that the genetic variation in expression in these genes is mainly due to changes in trans-acting factors. Moreover, their data suggested that the differences in trans-acting factors do not reside in TFs themselves, although TFs are natural candidates for harboring the trans-acting regulatory variation (Yvert et al. 2003).

Following Yvert et al. (2003), we also used yeast as the model organism. We considered two main factors that can influence the chance of a mutation affecting the expression level of a gene: 1) the number of trans-acting factors that influence the expression of the target gene and 2) the number of cis-regulatory elements that control the expression of the target gene. A very simple situation is that a gene is regulated by a single TF; such genes are called SIM (single input module) genes. Because a SIM gene is regulated by a single immediate upstream TF, the chance for a change...
to occur in its trans-acting factors would, on average, be smaller than that for a MIM (multiple input module) gene, because a MIM gene is regulated by more TFs. There are other types of regulatory modules (see Harbison et al. 2004). Our purpose is to understand the relative contributions of cis- and trans-regulatory variations to expression evolution in yeast for genes with different types of regulatory motifs. We used pyrosequencing to measure the relative expression levels of the BY and RM alleles in a BY–RM hybrid and in a coculture system in which BY and RM cells were grown together to eliminate the environmental effect on gene expression. As will be explained below, a comparison of the coculture and hybrid data on the relative expression levels can help us evaluate the relative contributions of cis variation and trans variation (Wittkopp et al. 2004, 2008). The data can also be used to address whether there is a significant difference between MIM and SIM genes.

Materials and Methods

Yeast Strains and Growth Conditions

The laboratory strain BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) is a descendant of S288C. The wild strain RM11-1Δz (MATa lys2Δ0 ura3Δ0 ho::KAN) was a gift from Dr Lee Hartwell (Fred Hutchinson Cancer Research Center); it was a haploid strain derived from Bb32(3), a natural isolate collected by Mortimer et al. (1994). WL201 is a hybrid strain of BY4741 × RM11-1Δz constructed in our laboratory. These three strains will be called BY, RM, and the BY–RM hybrid, respectively.

Yeast strains were grown in yeast extract, peptone, adenine and dextrose media (YPAD) and harvested at the mid-log phase. Overnight, yeast cultures were used to prepare the starting cultures with OD600 = 0.1 and were grown in YPAD media at 30 °C with 250 rpm shaking.

Allele Quantification with Pyrosequencing

The yeast cells were harvested at OD600 = 1.0 and the total RNAs were extracted by the hot acid phenol–chloroform method. An aliquot of 5 μg total RNA from each strain was used for cDNA synthesis. The reverse transcription was carried out with oligo-dT primers and the Super-script II kit (Invitrogen) following the manufacturer’s instructions. After identification of strain-specific nucleotide differences in the coding region of a gene, a 150–200 base-pair (bp) fragment of the coding region in the hybrid strain, WL201, and in the coculture of BY and RM, was amplified and sequenced. Pyrosequencing reactions were used to measure the relative abundances of the two alleles in genomic DNA and in cDNA samples from both coculture and hybrid pools and were performed according to the manufacturer’s instructions (http://www.pyrosequencing.com/). The pyrosequencing software reports a peak height directly proportional to the number of molecules incorporated into the growing DNA chain. The ratio of allele-specific frequencies (RMc/ycoc/BYc/ycoc, RMhybr/ycoc/BYhybr/ycoc), which corresponds to the relative abundances of the BY and RM alleles in the starting sample, was also reported by the pyrosequencing software, PSQ 96MA.

2.1.1. The cDNA ratios were then normalized with genomic DNA measurements as described in Wittkopp et al. (2004). Because both alleles are extracted and measured in a single sample, this method is insensitive to differences in extraction efficiency and eliminates the need for controlling quantification of total RNA recovery. The relative expression ratio of BY and RM alleles of each gene was estimated using at least three biological replicates.

Let BYhybr and RMhybr be the expression levels of the BY and RM alleles in the hybrid diploid and BYc and RMc be the expression levels of the BY and RM alleles when the two strains are grown in the same culture (coculture). Let $R_1 = RM_{hybr}/BY_{hybr}$ and $R_2 = RM_{hybr}/BY_{hybr}$. If $R_1$ is different from 1 (i.e., $R_1 \neq 1$), the difference in the expression levels of the two alleles can be due to the cis effect or the trans effect or both. Note that in a hybrid, the genetic background for the two alleles is the same, so the expression difference in the expression level between the two alleles in “hybrid” (i.e., $R_2 \neq 1$) is completely due to the cis effect, whereas if $R_2 = 1$, then there is no cis effect (fig. 1). On the other hand, if $R_2 = R_1$, then the expression difference between the two alleles in the “ko-culture” is completely due to the cis effect, because homogenization (hybridization) of the genetic background does not reduce the expression differences between the two alleles. Thus, the trans- and cis-effects on the expression differences between BYc and RMc can be judged by the following guidelines (see fig. 1):

1. If $R_1 \neq 1$ and $R_2 = R_1$, then the expression difference is due to the cis effect alone.
Table 1

Expression Divergence of SIM, MIM, and TF-Unknown Genes between the BY and RM Strains Inferred from Published Microarray Data

<table>
<thead>
<tr>
<th>Category</th>
<th>TF Unknown</th>
<th>SIM</th>
<th>2–4 TFs</th>
<th>5–7 TFs</th>
<th>&gt;7 TFs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of genes</td>
<td>2,272</td>
<td>1,537</td>
<td>1,773</td>
<td>445</td>
<td>246</td>
<td>6,273</td>
</tr>
<tr>
<td>Expression divergence</td>
<td>771</td>
<td>578</td>
<td>704</td>
<td>219</td>
<td>147</td>
<td>2,419</td>
</tr>
<tr>
<td>Proportion of genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>divergence (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a* TF-binding sites were taken from MYBS (Tsai et al. 2007; Wu and Li 2008).
*b* Including 250 genes that are not listed in MYBS or Wu and Li (2008).
*c* Including 11 genes that are not listed in MYBS or Wu and Li (2008).

Identification of SIM and MIM Genes

We are interested in determining whether differences in regulatory complexities between SIM and MIM genes can affect the relative contributions of cis and trans variations to expression divergence. To classify genes into SIM and MIM genes, we used a collection of *S. cerevisiae* TFs and their putative target genes in the mining yeast binding sites (MYBS) database http://cg1.iis.sinica.edu.tw/~mybs/ (Tsai et al. 2007) and the inferred TF–gene relationships from Wu and Li (2008) (called the Wu–Li database below), which incorporated ChIP-chip data (Harbison et al. 2004), the known binding site information and the mutation analysis data. This combined database contains 6,273 yeast genes and the putative targets of 132 known TFs. A gene is classified as a SIM gene if both databases indicate that it is the target of a (same) single TF or if it is indicated as the target of a single TF in one database but as a TF-unknown gene in the other database (i.e., no TF is known or predicted to regulate that gene). From this procedure, we inferred 1,537 putative SIM genes, among which 695 genes were classified as TF-unknown genes in the MYBS database, 580 genes were TF-unknown in the Wu-Li database, and 245 genes were identified to be regulated by the same TF in both databases. On the other hand, a gene is classified as a MIM gene if the sum of the numbers of the TFs that putatively regulate that gene in the two databases is 2 or more.

2. If \( R_1 > 1, R_2 \neq R_1, \) and \( R_2 = 1 \), then the expression difference is due to the trans effect alone.

3. If \( R_1 > 1 \) and \( R_2 > 1 \)

   (a) if \( (R_2 - 1)(R_1 - 1) \leq 0.25 \), then the expression difference is mainly due to the trans effect—we say it is a “major trans effect.”

   (b) if \( 0.25 < (R_2 - 1)(R_1 - 1) < 0.75 \), we say the expression difference is due to both cis and trans effects.

   (c) if \( (R_2 - 1)(R_1 - 1) > 0.75 \) (but \( \neq 1 \)), then the expression difference is mainly due to the cis effect—we say it is a “major cis effect.”

4. If \( R_1 > 1 \) but \( R_2 < 1 \), to be on the same scale as \( R_1 \), we consider \( 1/R_2 \).

   (a) If \( R_1 > 1/R_2 \), it is a major trans effect, because it can be shown that the cis effect \((1/R_2 - 1)\) is smaller than half of the trans effect.

   (b) If \( R_1 \leq 1/R_2 \), then it is due to both cis and trans effects because the cis effect \((1/R_2 - 1)\) is greater than half of the trans effect.

5. If \( R_1 \approx 1 \) but \( R_2 \neq 1 \), then the expression difference is due to both cis and trans effects, that is, the cis and trans effects in the coculture are equal so that \( R_1 = 1 \), but the cis effect \((R_2 \neq 1)\) is seen in the hybrid.

All the above equalities and inequalities are tested by the student t-test. We calculate the standard errors (SEs) of \( R_1 \) and \( R_2 \) from at least three biological repeats for each data point and test the null hypothesis of \( R_1 = 1 \) or \( R_2 = 1 \). We use the two-tailed t-test to test \( R_1 = R_2 \).

In the above, we considered the case of \( \text{RM}_{\text{cocul}}/\text{BY}_{\text{cocul}} \geq 1 \). If \( \text{RM}_{\text{cocul}}/\text{BY}_{\text{cocul}} < 1 \), then \( R_1 = \text{BY}_{\text{cocul}}/\text{RM}_{\text{cocul}} \) instead of \( R_1 = \text{RM}_{\text{cocul}}/\text{BY}_{\text{cocul}} \) and \( R_2 = \text{BY}_{\text{hybrid}}/\text{RM}_{\text{hybrid}} \) instead of \( R_2 = \text{RM}_{\text{hybrid}}/\text{BY}_{\text{hybrid}} \) should be used in conditions (2)–(4).

Identification of SIM and MIM Genes

We are interested in determining whether differences in regulatory complexities between SIM and MIM genes can affect the relative contributions of cis and trans variations to expression divergence. To classify genes into SIM and MIM genes, we used a collection of *S. cerevisiae* TFs and their putative target genes in the mining yeast binding sites (MYBS) database http://cg1.iis.sinica.edu.tw/~mybs/ (Tsai et al. 2007) and the inferred TF–gene relationships from Wu and Li (2008) (called the Wu–Li database below), which incorporated ChIP-chip data (Harbison et al. 2004), the known binding site information and the mutation analysis data. This combined database contains 6,273 yeast genes and the putative targets of 132 known TFs. A gene is classified as a SIM gene if both databases indicate that it is the target of a (same) single TF or if it is indicated as the target of a single TF in one database but as a TF-unknown gene in the other database (i.e., no TF is known or predicted to regulate that gene). From this procedure, we inferred 1,537 putative SIM genes, among which 695 genes were classified as TF-unknown genes in the MYBS database, 580 genes were TF-unknown in the Wu-Li database, and 245 genes were identified to be regulated by the same TF in both databases. On the other hand, a gene is classified as a MIM gene if the sum of the numbers of the TFs that putatively regulate that gene in the two databases is 2 or more.
larger. For example, if a gene is indicated to be regulated by two TFs in one database, then it is classified as a MIM gene regardless of what the other database says. As another example, if a gene is said to be the target of TF \( x \) in one database and the target of TF \( y \) in the other database, then it is also classified as a MIM gene (i.e., regulated by two TFs). From this procedure, we inferred a total of 2,464 MIM genes. Among these, 210 (23%) were classified as a single-TF regulated genes in both databases but the TFs in the two databases were different. Note that some of the putative MIM genes may be actually SIM genes, whereas some of the SIM genes may be actually MIM genes; thus, the actual difference between the SIM and MIM genes is likely to be larger than to be inferred below. This problem will be discussed in Results and Discussion below.

Results and Discussion

Identification of Genes with Expression Divergence between BY and RM

We used microarray data from Brem et al. (2002) and Wang et al. (2007) to examine the gene expression profiles of BY and RM. Among the 6,228 genes included in the microarray studies, 1,528 (Brem et al. 2002) and 1,360 genes showed significant \( (P < 0.5 \times 10^{-4}) \) expression differences between the two strains. We considered the expression of a specific gene different between the two strains if it showed a significantly different expression level in either one of the two microarray data sets. In total, we identified 2,419 genes with different expression levels for further analysis (table 1 and supplementary table S1, Supplementary Material online).

To confirm the expression difference between BY and RM observed by microarray analysis and to rule out possible environmental effects on the expression divergence because the BY and RM strains were grown separately, we used pyrosequencing to measure the relative expression levels of BY and RM alleles in a coculture system in which BY and RM cells were grown together. We randomly selected 194 genes for pyrosequencing analysis (supplementary table S2, Supplementary Material online). We found nine genes that showed no significant expression difference between BY and RM when these two strains of cells were cultured together, suggesting that the expression difference was mainly due to environmental effects or was a false positive in the microarray analysis. These results also indicated that at least 90% of these genes identified by microarray analysis were indeed not false positives. The remaining 185 genes were used in further pyrosequencing analysis.

SIM Genes versus MIM Genes

We were interested in knowing whether MIM genes are in general more susceptible to expression divergence than SIM genes. Because a MIM gene is regulated by several immediate upstream TFs, the chance for a change to occur in its trans-acting factors or on the TF-binding motifs would, on average, be greater than that for a SIM gene, which is regulated by only one TF and one TF-binding motif. Our analyses showed that the proportion of MIM genes (43%) that showed expression divergence was significantly higher than that of SIM genes (38%) \( (P < 0.05, \text{Chi square test}) \). For genes whose upstream regulatory TFs have not been identified (i.e., TF-unknown genes), the proportion of genes with expression divergence is even somewhat lower (34%). Note that these TF-unknown genes are likely to be, on average, regulated by fewer TFs than the SIM genes because none of the known TFs were inferred to regulate these genes, whereas a SIM is inferred to be regulated by at least one of the 132 known TFs. Together, our results suggested that the number of upstream regulators (TFs) is a factor that affects the expression divergence between BY and RM (table 1).

Importance of Trans-Variation for Intraspecies Expression Divergence

For the 223 genes, 185 genes from this study plus 38 genes from our previous study (Wang et al. 2007) that showed expression divergence in the BY–RM coculture or in the BY–RM hybrid, we found that 25.1% (56/223) of the cases can be attributed to differences in trans variations alone, 29.6% (66/223) mainly due to trans variations,
Table 2
Molecular Function of Genes Characterized

<table>
<thead>
<tr>
<th>GO ID</th>
<th>GO Term</th>
<th>Trans Alone</th>
<th>Major trans</th>
<th>Both cis &amp; trans</th>
<th>Major cis</th>
<th>Cis Alone</th>
<th>GO Term/All Yeast Genes in SGD</th>
<th>GO Term/Differently Expressed Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>16787</td>
<td>Hydrolase activity</td>
<td>16.1% (9)</td>
<td>22.7% (15)</td>
<td>20.3% (12)</td>
<td>13.3% (2)</td>
<td>12.9% (8)</td>
<td>14.1% (341)</td>
<td>13.1% (341)</td>
</tr>
<tr>
<td>16491</td>
<td>Oxidoreductase activity</td>
<td>14.3% (8)</td>
<td>12.1% (8)</td>
<td>10.2% (6)</td>
<td>13.3% (2)</td>
<td>4.3% (268)</td>
<td>7.2% (174)</td>
<td></td>
</tr>
<tr>
<td>16740</td>
<td>Transferase activity</td>
<td>8.9% (5)</td>
<td>12.1% (8)</td>
<td>18.6% (11)</td>
<td>6.7% (1)</td>
<td>11.2% (706)</td>
<td>13.1% (318)</td>
<td></td>
</tr>
<tr>
<td>5215</td>
<td>Transporter activity</td>
<td>10.7% (6)</td>
<td>12.1% (8)</td>
<td>10.2% (6)</td>
<td>6.7% (1)</td>
<td>5.9% (369)</td>
<td>7.2% (175)</td>
<td></td>
</tr>
<tr>
<td>5515</td>
<td>Protein binding</td>
<td>5.4% (3)</td>
<td>12.1% (8)</td>
<td>6.8% (4)</td>
<td>13.3% (2)</td>
<td>9.6% (602)</td>
<td>10.6% (256)</td>
<td></td>
</tr>
<tr>
<td>30528</td>
<td>Transcription</td>
<td>5.4% (3)</td>
<td>6.1% (4)</td>
<td>3.4% (2)</td>
<td>0%</td>
<td>5.1% (319)</td>
<td>5.3% (127)</td>
<td></td>
</tr>
<tr>
<td>3677</td>
<td>Structural molecule</td>
<td>3.6% (2)</td>
<td>3.0% (2)</td>
<td>8.5% (5)</td>
<td>3.7% (1)</td>
<td>0%</td>
<td>5.1% (319)</td>
<td>4.6% (111)</td>
</tr>
<tr>
<td>16829</td>
<td>Lyase activity</td>
<td>3.6% (2)</td>
<td>1.5% (1)</td>
<td>1.7% (1)</td>
<td>0%</td>
<td>1.4% (85)</td>
<td>1.8% (43)</td>
<td></td>
</tr>
<tr>
<td>3723</td>
<td>RNA binding</td>
<td>3.6% (2)</td>
<td>4.5% (3)</td>
<td>0%</td>
<td>0%</td>
<td>4.4% (278)</td>
<td>4.9% (119)</td>
<td></td>
</tr>
<tr>
<td>16874</td>
<td>Ligase activity</td>
<td>3.6% (2)</td>
<td>0%</td>
<td>5.1% (3)</td>
<td>0%</td>
<td>2.6% (162)</td>
<td>2.1% (51)</td>
<td></td>
</tr>
<tr>
<td>3774</td>
<td>Motor activity</td>
<td>1.8% (1)</td>
<td>0%</td>
<td>1.7% (1)</td>
<td>0%</td>
<td>0.3% (16)</td>
<td>0.2% (5)</td>
<td></td>
</tr>
<tr>
<td>8233</td>
<td>Peptidase activity</td>
<td>1.8% (1)</td>
<td>3.0% (2)</td>
<td>1.7% (1)</td>
<td>7.4% (2)</td>
<td>2.3% (147)</td>
<td>2.8% (67)</td>
<td></td>
</tr>
<tr>
<td>4386</td>
<td>Helicase activity</td>
<td>0%</td>
<td>3.0% (2)</td>
<td>0%</td>
<td>0%</td>
<td>1.3% (83)</td>
<td>1.4% (33)</td>
<td></td>
</tr>
<tr>
<td>30234</td>
<td>Enzyme regulator activity</td>
<td>1.8% (1)</td>
<td>1.5% (1)</td>
<td>0%</td>
<td>0%</td>
<td>3.3% (208)</td>
<td>3.4% (83)</td>
<td></td>
</tr>
<tr>
<td>8289</td>
<td>Lipid binding</td>
<td>1.8% (1)</td>
<td>0%</td>
<td>1.7% (1)</td>
<td>0%</td>
<td>1.0% (64)</td>
<td>0.9% (21)</td>
<td></td>
</tr>
<tr>
<td>16779</td>
<td>Nucleotidyl transferase activity</td>
<td>1.8% (1)</td>
<td>0%</td>
<td>3.7% (1)</td>
<td>0%</td>
<td>1.7% (110)</td>
<td>2.3% (55)</td>
<td></td>
</tr>
<tr>
<td>5198</td>
<td>Structural molecule activity</td>
<td>0%</td>
<td>1.5% (1)</td>
<td>0%</td>
<td>7.4% (2)</td>
<td>5.3% (331)</td>
<td>4.6% (111)</td>
<td></td>
</tr>
<tr>
<td>4871</td>
<td>Signal transducer activity</td>
<td>0%</td>
<td>1.5% (1)</td>
<td>1.7% (1)</td>
<td>0%</td>
<td>0.7% (44)</td>
<td>0.9% (21)</td>
<td></td>
</tr>
<tr>
<td>4721</td>
<td>Phosphoprotein phosphatase activity</td>
<td>1.8% (1)</td>
<td>1.5% (1)</td>
<td>1.7% (1)</td>
<td>3.7% (1)</td>
<td>0.7% (47)</td>
<td>0.6% (14)</td>
<td></td>
</tr>
<tr>
<td>16853</td>
<td>Isomerase activity</td>
<td>0%</td>
<td>1.5% (1)</td>
<td>1.7% (1)</td>
<td>0%</td>
<td>1.3% (2)</td>
<td>0.9% (59)</td>
<td>1% (23)</td>
</tr>
<tr>
<td>4672</td>
<td>Protein kinase activity</td>
<td>0%</td>
<td>1.5% (1)</td>
<td>0%</td>
<td>6.7% (1)</td>
<td>2.1% (129)</td>
<td>2.1% (51)</td>
<td></td>
</tr>
<tr>
<td>30533</td>
<td>Triplet codon-amino acid adaptor activity</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>45182</td>
<td>Translation regulator activity</td>
<td>0%</td>
<td>0%</td>
<td>1.7% (1)</td>
<td>6.7% (1)</td>
<td>0.8% (52)</td>
<td>0.8% (19)</td>
<td></td>
</tr>
<tr>
<td>3674</td>
<td>Molecular function unknown Other</td>
<td>32.1% (18)</td>
<td>24.2% (16)</td>
<td>23.7% (14)</td>
<td>33.3% (9)</td>
<td>38.3% (2407)</td>
<td>33.5% (810)</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
<td>1.8% (1)</td>
<td></td>
<td>3.7% (1)</td>
<td>0%</td>
<td>2% (125)</td>
<td>2.3% (55)</td>
<td></td>
</tr>
</tbody>
</table>

* The number in parentheses is the number of genes.

Transversus Cis Variation in Yeast Expression Evolution 2537

26.5% (59/223) due to both cis- and trans variations, 12.1% (27/223) mainly due to cis variations, and 6.7% (15/223) due to cis variations alone (fig. 2). In total, 54.7% of genes were mainly affected by trans variations (trans effect alone or major trans effect) between the BY and RM strains. Only 18.8% of genes were mainly affected by cis variations (cis effect alone or major cis effect). Our results suggest that trans-acting factors play a more important role in intraspecies expression divergence than cis elements, largely in agreement with the observations of Brem et al. (2002) and Yvert et al. (2003). However, cis variation contributed significantly to the expression divergence even in the case of a “trans major effect” because R2 was significantly different from 1. Indeed, in 74.9% of the 223 genes examined, cis variation contributed significantly to the expression divergence.

Role of Trans-Variation in Expression Evolution in SIM and MIM Genes

We next determined whether a SIM gene is less susceptible to trans-variation than a MIM gene. Among the 223 genes examined by pyrosequencing, 57 are SIM genes, 123 are MIM genes, and 43 are TF-unknown genes. Our results showed that MIM genes were more affected by trans effect alone than SIM genes, that is, 29% versus 25% (fig. 3). To determine whether the number of upstream TFs affects the pattern of expression divergence, we compared the contribution of cis and trans variations in genes with different numbers of upstream TFs. We found a positive correlation between the number of TFs and the proportion of genes affected by the trans variation (fig. 4). For genes with 1 one TF, 2 TFs, and >2 TFs, 48%, 42%, and 68% were, respectively, affected by trans variations (trans effect alone or major trans effect). Unexpectedly, genes regulated by two TFs were more affected by cis variations (cis effect alone or major cis effect) than SIM genes and genes regulated by more than two TFs (30%, 21%, and 11%, respectively). However, as mentioned above, some of the genes predicted to be regulated by two TFs may actually be SIM genes and some SIM may actually be MIM genes, so the difference between SIM genes and genes regulated by two TFs may actually be smaller than seen in figure 4. By the same reasoning, the difference in the contribution of trans variation to expression between MIM and SIM
may actually be larger than that shown in figure 4. Thus, trans variation seems to play a more important role in expression divergence in MIM genes than in SIM genes and the number of TFs that regulate a gene seems to be a significant factor in determining the contribution of trans variation to expression divergence.

To examine whether divergence in TF expression between BY and RM can contribute to the expression divergence of its target genes, we examined the expression difference of a TF between BY and RM. Our results showed that half of the TFs showed different expression levels between BY and RM. Figure 5 shows that genes regulated by the TFs that showed significant expression divergence between BY and RM tend to be more affected by trans variation effect (61%). These results support our hypothesis that MIM genes, compared with SIM genes, tend to diverge faster in expression and are more susceptible to trans variation effect.

Molecular Function of the Genes Showing Different Cis or Trans Variation Effects

To examine whether regulatory evolution of a gene is related to its molecular function, we grouped the genes according to their molecular function in the Gene Ontology (GO) annotation of Saccharomyces Genome Database (SGD) http://www.yeastgenome.org/. Our results showed that hydrolase activity genes tend to be affected by both cis and trans variations (major trans effect, both cis and trans effect, or major cis effect) but not by trans effect alone or by cis effect alone (table 2). On the other hand, transferase genes showed a low level of trans alone variation effect or cis alone variation effect. Because the molecular functions of genes that showed expression divergence between the BY and RM strains are still not well defined, the conclusion here should be taken with caution.

Supplementary Material

Supplementary tables S1 and S2 and Supplementary figures S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank J.J. Emerson for valuable suggestions on the paper. This study was supported by Academia Sinica, Taiwan and by NIH grant (GM081724), USA to W.H.L.

Literature Cited


Takashi Gojobori, Associate Editor

Accepted July 22, 2009