Transcriptional Network Structure Has Little Effect on the Rate of Regulatory Evolution in Yeast

Artyom Kopp* and Lauren M. McIntyre*

1Department of Evolution and Ecology, University of California, Davis
2Department of Molecular Genetics and Microbiology, University of Florida

*Corresponding author: E-mail: akopp@ucdavis.edu.
Associate editor: Matthew Hahn

Introduction

Recent analyses suggest that the structure of genetic pathways may bias the fixation of natural variation toward particular nodes in these pathways. For example, parallel evolution of similar phenotypes in different lineages is sometimes caused by independent regulatory changes in the same genes (Sucena et al. 2003; Protas et al. 2006; Chan et al. 2010). A possible explanation is that these genes represent the optimally pleiotropic nodes in the genetic pathways that control the evolving traits. In other words, mutations in these genes have large effects on the trait under selection and little or no effect on other traits (Gompel et al. 2009; Kopp 2009; Stern and Orgogozo 2009; Streisfeld and Rausher 2009). A broad interpretation of this hypothesis is that a gene’s pleiotropy in the context of intersecting regulatory pathways is a key predictor of its role in phenotypic evolution.

Evidence for this hypothesis comes from case studies in which the genetic basis of phenotypic variation was identified by forward genetic analysis. These studies are still few in number and may suffer from ascertainment biases because phenotypic changes that are controlled by one or a few genes of large effect are more likely to be chosen for detailed molecular dissection than more complex quantitative variation. A useful complementary approach may be to examine the effects of gene network topology on genome-wide patterns of regulatory evolution. Regulatory mutations play a major role in the evolution of many phenotypic traits (Wray 2007; Carroll 2008; Streisfeld and Rausher 2011). Thus, if pleiotropy exerts a strong and universal influence on the genetic basis of evolution, we expect a systematic relationship between the number of regulatory interactions in which a gene participates and the rate at which its expression diverges between species.

Here, we test this hypothesis by examining gene expression divergence between sibling species of yeast, Saccharomyces cerevisiae and Saccharomyces paradoxus. These species are closely related and share the same set of transcription factors (TFs) (Kellis et al. 2003; Liti et al. 2006), suggesting that their gene regulatory networks are similar. Despite strong reproductive isolation, S. cerevisiae and S. paradoxus can be hybridized in the laboratory (Naumov 1996; Scannell et al. 2007). This ability can be exploited to decompose gene expression divergence between these species into the cis- and trans-regulatory components (Tirosh et al. 2009). The cis-regulatory component provides a direct estimate of the cumulative effect of regulatory mutations that have accumulated at each locus since species divergence, while ignoring the indirect effects of changes in the more upstream levels of gene networks (the trans-component). At the same time, because yeast are single-celled organisms, it is possible to reconstruct their genome-wide transcriptional networks by identifying the direct downstream targets of each TF and the direct transcriptional regulators of each target gene (Lee et al. 2002; Harbison et al. 2004; Luscombe et al. 2004; Balaji, Babu, et al. 2006, 2008; Farkas et al. 2006; Maclsaac et al. 2006; Jothi et al. 2009). By integrating publicly available data sets on...
transcriptional network architecture and *cis*-regulatory expression divergence, we test the hypothesis that the position of genes in the gene network is a systemic predictor of their regulatory evolution.

Materials and Methods

Genome-wide estimates of *cis*-regulatory expression divergence between *S. cerevisiae* and *S. paradoxus* were obtained by Tirosh et al. (2009) based on allele-specific gene expression in *F*1 interspecific hybrids. In these hybrids, both alleles at each locus are present in the same *trans*-regulatory background, so that any differences in expression levels between the two alleles must be due to *cis*-regulatory divergence between the parental genotypes. The expression level of each allele was quantified separately by hybridizing *F*1 mRNA samples extracted under four different environmental conditions to microarrays that carried species-specific probes. Neither copy number variation nor epigenetic effects made a major contribution to allele-specific expression, and *cis*-regulatory effects were strongly correlated among environmental treatments (*r* ~ 0.85) (Tirosh et al. 2009). We therefore used the average absolute difference between log2-transformed expression levels of the *S. cerevisiae* and *S. paradoxus* alleles across all conditions as the estimate of *cis*-regulatory divergence between these species (calculated from supplementary table S1 in Tirosh et al.). Using sum of squared differences across all conditions as the measure of *cis*-regulatory divergence produced similar results (supplementary tables 1 and 2, Supplementary Material online, compare with tables 1 and 2). An important caveat is that regulatory divergence was measured in a single *F*1 genotype produced by crossing isogenic parental strains (Tirosh et al. 2009). Divergence estimates may therefore confound fixed interspecific differences with standing variation segregating within either or both species.

Transcriptional network topologies were inferred by several groups based mainly on genome-wide chromatin immunoprecipitation (ChIP) experiments. This approach uses in vivo binding between candidate TFs and regions of genomic DNA, quantified by hybridizing immunoprecipitated samples to tiling microarrays (ChIP-chip), as evidence of direct transcriptional regulation. Gene expression analysis shows that not all TF–DNA binding is biologically relevant (Gao et al. 2004; Nuzhdin et al. 2010). At the same time, many TF–DNA interactions may occur only under specific environmental conditions. Thus, ChIP-chip data contain a significant number of false positives and false negatives. Despite these caveats, TF–DNA binding provides the core evidence that, in combination with gene expression and genetic interaction studies, allows the

### Table 1. Relationship between the Number of Outgoing Connections and Mean *cis*-Regulatory Expression Divergence.

<table>
<thead>
<tr>
<th>Network</th>
<th>Spearman’s <em>p</em> Value</th>
<th>Spearman <em>P</em> Value</th>
<th>Kruskal–Wallis <em>P</em> Value</th>
<th>Wilcoxon <em>P</em> Value</th>
<th>Number of Genes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole genome</td>
<td>0.062</td>
<td>0.484</td>
<td>0.405</td>
<td>0.983</td>
<td>123</td>
<td>Balaji et al. (2006a)</td>
</tr>
<tr>
<td>Whole genome</td>
<td>0.067</td>
<td>0.453</td>
<td>0.425</td>
<td>0.957</td>
<td>125</td>
<td>Jothi et al. (2009)</td>
</tr>
<tr>
<td>Whole genome</td>
<td>0.128</td>
<td>0.187</td>
<td>0.625</td>
<td>0.451</td>
<td>107</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Transcription factors</td>
<td>0.104</td>
<td>0.262</td>
<td>0.759</td>
<td>0.588</td>
<td>115</td>
<td>Balaji et al. (2006a)</td>
</tr>
<tr>
<td>Transcription factors</td>
<td>0.125</td>
<td>0.246</td>
<td>0.066</td>
<td>0.367</td>
<td>86</td>
<td>MacIsaac et al. (2006)</td>
</tr>
<tr>
<td>Diauxic shift</td>
<td>0.320</td>
<td>0.023*</td>
<td>0.081</td>
<td>0.058</td>
<td>51</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>DNA damage response</td>
<td>0.084</td>
<td>0.555</td>
<td>0.846</td>
<td>0.926</td>
<td>49</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>−0.070</td>
<td>0.624</td>
<td>0.475</td>
<td>0.780</td>
<td>49</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Sporulation</td>
<td>0.321</td>
<td>0.020*</td>
<td>0.213</td>
<td>0.075</td>
<td>53</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Stress response</td>
<td>0.430</td>
<td>0.005**</td>
<td>0.008**</td>
<td>0.003*</td>
<td>43</td>
<td>Luscombe et al. (2004)</td>
</tr>
</tbody>
</table>

*a* Rank correlation between the number of connections and expression divergence.  
*b* Genes were split into quartiles according to the number of connections; the test shows whether *cis*-regulatory divergence is significantly different among all four quartiles.  
*c* Test shows whether *cis*-regulatory divergence differs significantly between the top and bottom quartiles (those with the most and fewest connections).  
*d* Corresponds to nominal *P* < 0.05, ** corresponds to *P* < 0.01, **** corresponds to *P* < 0.0001.

### Table 2. Relationship between the Number of Incoming Connections and Mean *cis*-Regulatory Expression Divergence.

<table>
<thead>
<tr>
<th>Network</th>
<th>Spearman’s <em>p</em> Value</th>
<th>Spearman <em>P</em> Value</th>
<th>Kruskal–Wallis <em>P</em> Value</th>
<th>Wilcoxon <em>P</em> Value</th>
<th>Number of Genes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole genome</td>
<td>0.094</td>
<td>&lt;0.0001***</td>
<td>&lt;0.0001***</td>
<td>&lt;0.0001***</td>
<td>3,038</td>
<td>Balaji et al. (2006a)</td>
</tr>
<tr>
<td>Whole genome</td>
<td>0.101</td>
<td>&lt;0.0001***</td>
<td>&lt;0.0001***</td>
<td>&lt;0.0001***</td>
<td>3,098</td>
<td>Jothi et al. (2009)</td>
</tr>
<tr>
<td>Whole genome</td>
<td>0.110</td>
<td>&lt;0.0001***</td>
<td>0.302</td>
<td>0.041*</td>
<td>2,408</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Diauxic shift</td>
<td>0.158</td>
<td>&lt;0.0001***</td>
<td>0.771</td>
<td>0.285</td>
<td>558</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Transcription factors</td>
<td>0.149</td>
<td>0.110</td>
<td>0.675</td>
<td>0.424</td>
<td>115</td>
<td>Balaji et al. (2006a)</td>
</tr>
<tr>
<td>DNA damage response</td>
<td>0.216</td>
<td>&lt;0.0001***</td>
<td>0.384</td>
<td>0.635</td>
<td>486</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>0.117</td>
<td>0.008**</td>
<td>0.033*</td>
<td>0.800</td>
<td>203</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Sporulation</td>
<td>0.159</td>
<td>0.031*</td>
<td>0.553</td>
<td>0.669</td>
<td>185</td>
<td>Luscombe et al. (2004)</td>
</tr>
<tr>
<td>Stress response</td>
<td>0.233</td>
<td>&lt;0.0001***</td>
<td>0.697</td>
<td>0.379</td>
<td>278</td>
<td>Luscombe et al. (2004)</td>
</tr>
</tbody>
</table>

*a* Rank correlation between the number of connections and expression divergence.  
*b* Genes were split into quartiles according to the number of connections; the test shows whether *cis*-regulatory divergence is significantly different among all four quartiles.  
*c* Test shows whether *cis*-regulatory divergence differs significantly between the top and bottom quartiles (those with the most and fewest connections).  
*d* Corresponds to nominal *P* < 0.05, ** corresponds to *P* < 0.01, **** corresponds to *P* < 0.0001.
identification of all or most direct transcriptional regulators of each target gene and most direct transcriptional targets of each TF. We used several versions of the S. cerevisiae transcriptional network (listed below), which were inferred by integrating ChIP-chip data with other data sets and have somewhat different topologies as a result. At present, no comparable data are available for S. paradoxus. Analyses presented here are predicated on the assumption that transcriptional networks are similar in S. cerevisiae, S. paradoxus, and their hybrids. This assumption seems reasonable in light of the close relationship between these species, but it remains to be verified. In each network, we quantified the number of “incoming” connections as the number of TFs inferred to regulate each gene and the number of “outgoing” connections as the number of inferred target genes of each TF.

The following network estimates were examined (tables 1 and 2):

1. Whole-genome networks of Luscombe et al. (2004), Balaji, Babu, et al. (2006), Balaji, Iyer, et al. (2006), and Jothi et al. (2009). The network of Luscombe et al. (obtained from http://sandy.topnet.gersteinlab.org/index2.html) includes 142 TFs, 3,420 target genes, and 7,074 inferred interactions. This network was constructed by integrating the original genome-wide ChIP-chip analysis performed under fixed environmental conditions (Lee et al. 2002) with gene expression data from other studies. The other two networks are based on expanded ChIP-chip analysis (Harbison et al. 2004), which used several different environmental conditions and identified DNA motifs that were bound by candidate TFs at high confidence (P < 0.001) and were conserved among multiple Saccharomyces species. The network of Balaji, Babu, et al. (2006) and Balaji, Iyer, et al. (2006) integrate data from the Lee et al. (2002) Harbison et al. (2004), Luscombe et al. (2004), and several smaller-scale studies and consists of 157 TFs, 4,410 target genes, and 12,873 regulatory interactions. The Jothi et al. network (from http://www.nature.com/msb/journal/v5/n1/supplinfo/msb200952_S1.html) is updated from the version by Balaji et al. and includes 158 TFs, 4,369 target genes, and 13,385 interactions.

2. Most target genes in whole-genome networks have only a small number of incoming connections (≤3), whereas TFs have on average a larger number of incoming connections. For this reason, analyses were repeated for two subnetworks that consisted solely of TFs. The network of Maclasac et al. (2006) was updated from the version by Harbison et al. using a refined algorithm to measure the evolutionary conservation of binding sites. The TF-only subnetwork was obtained from http://fraenkel.mit.edu/improved_map/ and corresponds to figure 5 in Maclasac et al. The TF-only subnetwork of Balaji et al. (2006a) is described in the supplementary table S5 from that publication.

3. Gene expression analysis shows that different transcriptional interactions are realized under different conditions, that is, the same TFs regulate different targets depending on the circumstances (Luscombe et al. 2004). We therefore examined five condition-specific subnetworks constructed by Luscombe et al.: cell cycle, sporulation, diauxic shift, DNA damage, and stress response (fig. 1 in Luscombe et al. 2004). These networks, consisting of 482–1,218 interactions each, were obtained from http://sandy.topnet.gersteinlab.org/index2.html.

Transcript abundance measurements are described by Holstege et al. (1998) and were obtained from http://web.mit.edu/youdng/expression/transcriptome.html. Genome-wide D_N and D_S values for the S. cerevisiae–S. paradoxus pairwise comparison are described by Drummond et al. (2005).

The main question of interest is whether there is a relationship between the number of network connections and cis-regulatory divergence of the regulator genes and their targets. The distribution of the number of connections deviates from normality in all networks and for both incoming and outgoing connections (Kolmogorov–Smirnov test, P < 0.01 in all data sets), making Pearson correlation analysis inappropriate. Assuming that the suspected relationship would be linear, Spearman rank correlations were calculated for each network and tested for difference from zero (tables 1 and 2). The number of connections may however exhibit a nonlinear relationship with cis-regulatory divergence. In order to examine whether there were nonlinear trends in the data, genes in each data set were separated into four quartiles according to the number of outgoing or incoming connections and the median divergence was compared among quartiles using a Kruskal–Wallis test. To test more specifically whether genes with few connections differed from those with many connections, the median cis-regulatory divergence was compared between the first and fourth quartiles using a Wilcoxon test (tables 1 and 2). The number of network connections, cis-regulatory divergence, protein divergence, and transcript abundance data were not available for all genes, leading to the different numbers of genes and gene pairs in different comparisons.

**Results**

In whole-genome transcriptional networks, the number of outgoing connections, and cis-regulatory divergence are not significantly related (table 1 and fig. 1A), and there is no significant difference between the TFs with the lowest and highest numbers of downstream targets (table 1 and fig. 1B). There is a significant relationship between cis-regulatory divergence and the number of incoming connections, although the magnitude of the correlation is small (table 2 and fig. 1C). Genes regulated by many TFs have higher median cis-regulatory divergence than genes regulated by few TFs (table 2 and fig. 1D). In TFs-only subnetworks, where the numbers of incoming and outgoing connections are more evenly balanced, cis-regulatory divergence and the number of outgoing, incoming, and total connections are not significantly associated (tables 1 and 2 and data not shown).

The rate of protein evolution is negatively correlated with expression levels, and this effect often confounds other evolutionary rate comparisons (Koonin 2005; Pal et al. 2006). However, the relationship between cis-regulatory expression divergence and transcript abundance
(Holstege et al. 1998) is not significant ($\rho = -0.041, P = 0.104$), whereas the relationship between transcript abundance and the number of network connections is positive ($\rho = 0.349, P = <0.001$), suggesting that transcript abundance is unlikely to explain the relationship between connectivity and cis-regulatory divergence.

Whole-genome networks represent the average behavior of transcriptional regulators and their targets across multiple conditions. If a TF functions separately in two or more genetic pathways that regulate different traits, its connectivity in the whole-genome network may not be a correct reflection of the influence exerted by any of these intersecting pathways on cis-regulatory evolution. In such cases, smaller gene networks that mediate specific cellular conditions and responses may have greater predictive value. To test this proposition, we examined the relationship between network connectivity and regulatory divergence in five condition-specific subnetworks (Luscombe et al. 2004). Only one of these subnetworks (mediating stress response) reveals a significant positive correlation between the number of outgoing connections and expression divergence (table 1 and fig. 2A) and a significant difference between TFs with the lowest and highest numbers of target genes (table 1 and fig. 2B). This relationship appears to be due to a small number of TFs with the most targets (fig. 2A and B). Most of the remaining subnetworks show no significant evidence of a relationship between outgoing connectivity and cis-regulatory divergence (table 1). A significant positive correlation is seen between cis-regulatory divergence and the number of incoming connections, but there is no significant difference between the quartiles of genes with the highest and lowest numbers of upstream regulators in any of the subnetworks (table 2).

Jothi et al. (2009) have classified TFs as “top,” “core,” or “bottom,” according to their position in the regulatory hierarchy. The core TFs have the highest average number of connections (Jothi et al. 2009). Although the mean cis-regulatory divergence is higher for the core TFs compared with the top, bottom, and unconnected TFs (0.588 vs. 0.431–0.466), these differences are not significant (Kruskal–Wallis test, $P = 0.561$). Jothi et al. (2009) also distinguish between 32 “hubs,” which together account for >20% of all regulatory connections in the whole-genome network, and the remaining TFs. The mean cis-regulatory divergence is 0.666 for the hubs, compared with 0.462 for the non-hub TFs, but this difference is also not significant (Wilcoxon test, $P = 0.124$). Thus, there is little evidence that the most highly connected TFs undergo faster regulatory evolution than other TFs.

**Fig. 1.** Relationship between the number of connections and cis-regulatory expression divergence in the whole-genome transcriptional network (Balaji et al. 2006a). (A). Scatter plot of cis-regulatory divergence versus the number of outgoing connections. (B). Box plot of cis-regulatory divergence for genes with different numbers of outgoing connections. On the x axis, genes were divided into four quartiles according to the number of outgoing connections. Each box shows the second and third quartiles of cis-regulatory divergence values, with the dividing line indicating the median and the “whiskers” showing the first and fourth quartiles. (C). Scatter plot of cis-regulatory divergence versus the number of incoming connections. (D). Box plot of cis-regulatory divergence for genes with different numbers of incoming connections. The first and fourth quartiles of cis-regulatory divergence values are not shown due to their wide range.
It has been suggested that evolutionary changes may often be concentrated at regulatory genes that occupy nexus positions in genetic pathways (i.e., receive multiple inputs and generate multiple outputs) (Gompel and Prud’homme 2009; Kopp 2009; Stern and Orgogozo 2009). We classified TFs in the five condition-specific subnetworks of Luscombe et al. (2004) as “nexus” if they had at least two incoming and at least two outgoing connections and as “non-nexus” otherwise. The nexus TFs showed higher cis-regulatory divergence than non-nexus TFs in the sporulation subnetwork (0.746 vs. 0.434; Wilcoxon test, \( P < 0.014 \)) but not in the other four subnetworks or in the combined data from all five subnetworks. This result suggests that faster regulatory evolution of nexus regulators may be limited to relatively few genetic pathways.

Some TFs control largely the same sets of target genes under all conditions, whereas others regulate largely non-overlapping sets of targets depending on the cell state. Luscombe et al. (2004) quantified these differences in the form of interchange indices, such that \( I = 0 \) means that a TF regulates the same targets under all conditions and \( I = 1 \) means that a TF changes its targets completely between conditions. This is analogous to the “party hubs” and “date hubs” in protein–protein interaction networks, respectively (Han et al. 2004). There is no correlation between interchange index and cis-regulatory divergence (Spearman \( \rho = 0.017, P = 0.734 \)) and no systematic difference in the rate of divergence between TFs that retain all interactions across multiple states (\( I < 0.1 \)) and those that replace all interactions in switching conditions (\( I > 0.9 \)) (Wilcoxon \( P = 0.877 \)). This differs from protein interaction networks, where party hubs show significantly slower evolution than date hubs (Fraser 2005).

**Discussion**

In the whole-genome transcriptional network, there is no evidence that TFs with a greater number of direct targets evolve faster at the cis-regulatory level. More restricted, condition-specific networks show only limited evidence for a relationship between regulatory divergence and the number of outgoing connections (table 1). The overall conclusion emerging from this analysis is that although the more highly connected TFs tend to show faster cis-regulatory evolution, this difference is slight and explains only a small fraction of variation in evolutionary rates.

There is a significant relationship between cis-regulatory divergence and the number of incoming connections, but the magnitude of this effect is small. This relationship might be explained by differences in mutational target size. Genes that are regulated by a greater number of direct...
activators and repressors are likely to have more TF binding sites, increasing the probability of cis-regulatory mutations that alter their expression. The slight positive correlation between expression divergence and the number of incoming connections is therefore consistent with neutral or nearly neutral evolution of gene expression.

Although these results do not refute the pleiotropy conjecture, they do not provide support for it either. There may be several, not mutually exclusive, explanations for this finding. First, the concentration of evolutionary changes at 1 or a few nexus loci may represent relatively rare exceptions, affecting only the pathways with unusual topological features. Most pathways may not have such nexus points, leading to a more diffuse genetic basis of evolutionary change. Second, pleiotropy in yeast may be fundamentally different from the type seen in multicellular organisms. In animals and plants, modular organization of cis-regulatory sequences allows each gene to participate in several distinct regulatory pathways that operate in different cell types or at different stages of development. This modularity allows gene function is each of the cell type-specific pathways to evolve independently of the others (Wray 2007; Carroll 2008). Thus, unicellularity and simple cis-regulatory regions—the very features that make it meaningful to speak of a whole-genome transcriptional network in yeast—may make them poor models for understanding the evolution of genetic pathways in multicellular organisms. Finally, the low correlation between connectivity and regulatory divergence may reflect the limitations of transcriptional network data based on genome-wide ChIP analysis. By pooling or omitting information about condition-specific regulatory interactions, whole-genome networks can obscure biologically meaningful trends. On the other hand, experimental dissection of evolutionary changes in precisely defined regulatory pathways can pinpoint the loci of evolutionary change with high confidence, yet provides only anecdotal, case-by-case evidence that is difficult to generalize in statistically sound ways. Indeed, a recent meta-analysis shows that reproducible biases in the genetic basis of phenotypic evolution that are evident in trait-by-trait analyses can be obscured by the pooling of data from multiple traits (Streisfeld and Rausher 2011). Perhaps the greatest promise lies in the integration of these approaches. Detailed experimental analysis will be necessary to decompose whole-genome transcriptional networks into modular genetic pathways that control specific phenotypic traits. A systematic analysis of evolutionary changes in these pathways, ideally in multiple independent lineages (Kopp 2009; Streisfeld and Rausher 2011), will show whether pathway topology exercises a consistent influence on the genetic basis of phenotypic evolution.

Supplementary Material

Supplementary tables 1 and 2 are available at Molecular Biology and Evolution online (http://www.mbe. oxfordjournals.org/).

Acknowledgments

We are grateful to Tim Connallon, Illes Farkas, Richard Jovelin, and Matt Hahn for sharing data sets and to Matt Hahn, Ondrej Podlaha, and two anonymous reviewers for comments on the manuscript. This work was supported by National Institutes of Health (grant 5-R01-GM-082843 to A.K.).

References


