The Impact of Trans-Regulation on the Evolutionary Rates of Metazoan Proteins

Yi-Ching Chen1, Jen-Hao Cheng1, Zing Tsung-Yeh Tsai1,2, Huai-Kuang Tsai1,* and Trees-Juen Chuang3,*

1Institute of Information Science, Academia Sinica, Taipei 115, Taiwan, 2Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei 115, Taiwan and 3Genomic Research Center, Academia Sinica, Taipei 115, Taiwan

Received February 7, 2013; Revised April 10, 2013; Accepted April 14, 2013

ABSTRACT

Transcription factor (TF) and microRNA (miRNA) are two crucial trans-regulatory factors that coordinate control gene expression. Understanding the impacts of these two factors on the rate of protein sequence evolution is of great importance in evolutionary biology. While many biological factors associated with evolutionary rate variations have been studied, evolutionary analysis of simultaneously accounting for TF and miRNA regulations across metazoans is still uninvestigated. Here, we provide a series of statistical analyses to assess the influences of TF and miRNA regulations on evolutionary rates across metazoans (human, mouse and fruit fly). Our results reveal that the negative correlations between trans-regulation and evolutionary rates hold well across metazoans, but the strength of TF regulation as a rate indicator becomes weak when the other confounding factors that may affect evolutionary rates are controlled. We show that miRNA regulation tends to be more important than TF regulation, and the combination of TF and miRNA regulations has a significant dependent effect on protein evolutionary rates. We also show that trans-regulation (especially miRNA regulation) is much more important in human/mouse than in fruit fly in determining protein evolutionary rates, suggesting a considerable variation in rate determinants between vertebrates and invertebrates.

INTRODUCTION

Gene expression is largely controlled by actions of various trans-regulatory factors. Undoubtedly, transcription factor (TF) and microRNA (miRNA) are the most conspicuous classes of trans-regulatory factors and are regarded as primary gene regulators in metazoans. TFs are proteins that facilitate or repress the transcription of their target genes through binding to specific DNA sequences, the so-called TF-binding sites (TFBSs), in the gene promoter regions (1). On the other hand, miRNAs are ~22 nucleotide noncoding RNAs, which target mRNAs and reduce stability and/or translation activity of mRNA to regulate gene expression at the posttranscriptional level (2). TFs and miRNAs may work together and form a complex regulatory network that generally consists of intricate feedback and feed-forward loops (3–5). The coordinated regulation of TFs and miRNAs may play important roles in a wider diversity of biological processes (6). A previous study reported that genes with more TFBSs tend to be targeted by miRNAs and have more miRNA-binding sites, suggesting a positive correlation between these two trans-regulatory factors (7). Although the mechanism of how miRNAs cooperate with TFs in the regulatory network remains largely unknown (8), accumulating evidence indicates its biological significance. Thus, it is of interest to investigate the relationship between these two trans-regulatory factors.

In terms of molecular evolution, it was shown that genes regulated by more TFs (N\textsubscript{TF}) tend to evolve more slowly in yeast (9,10). Similarly, genes targeted by more miRNAs (N\textsubscript{miR}) were suggested to experience more functional constraints and thereby evolve more slowly in human and mouse (11,12). These observations revealed that trans-regulation complexity is an important indicator of evolutionary rates, regardless of TF regulation at the transcriptional level or miRNA regulation at the posttranscriptional level. However, comparative studies of trans-regulatory factors have been hampered by the paucity or incompleteness of TF and miRNA information. To our knowledge, there is currently no systematic evolutionary analysis available that simultaneously accounts for these two trans-regulatory factors across metazoans. Whether the negative correlation between the number of trans-regulators that regulate a

*To whom correspondence should be addressed. Tel: +886 2 27871244; Fax: +886 2 27898757; Email: trees@gate.sinica.edu.tw
Correspondence may also be addressed to Huai-Kuang Tsai. Tel: +886 2 2788 3799 (Ext. 1718); Fax: +886 2 2782 4814; Email: hktai@iis.sinica.edu.tw

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
gene (i.e. \( N_{\text{TF}} \) and \( N_{\text{miR}} \)) and evolutionary rates is maintained across metazoans, whether \( N_{\text{TF}} \) and \( N_{\text{miR}} \) have a dependent effect on evolutionary rates, and which of these two factors has a greater effect on metazoan protein evolution still await investigation.

In addition to \textit{trans}-regulation, many other biological factors associated with and potentially underlying evolutionary rates of proteins have been reported. These factors include protein connectivity in protein–protein interaction (PPI) networks (9,10,13–19), expression level (or expression abundance) (9,11,13,14,17–30), tissue specificity (or expression breadth) (13,21,23,25,26,31–34), length of untranslated regions (UTRs) (12,21,26), intron length (13,21,26,35), intron number (23,26), solvent accessibility (36–39) and disorder content (11,40–42). Some of these factors were also shown to be correlated with \( N_{\text{TF}} \) or \( N_{\text{miR}} \) (9,11,12,43,44). We classified these 10 factors into five categories: \textit{trans}-regulation (\( N_{\text{TF}} \) and \( N_{\text{miR}} \)), protein connectivity, gene expression (expression level and tissue specificity), gene compactness (UTR length, intron length and intron number) and protein structure (solvent accessibility and disorder content). It is worth exploring whether the last four categories of confounding factors contribute to the strength of \( N_{\text{TF}} \) and \( N_{\text{miR}} \) as indicators of evolutionary rates.

To address these issues, we widely collect TF- and miRNA-binding data from human (\textit{Homo sapiens}), mouse (\textit{Mus musculus}) and fruit fly (\textit{Drosophila melanogaster}) and then systematically examine the correlations between these two \textit{trans}-regulatory factors (\( N_{\text{TF}} \) and \( N_{\text{miR}} \)) and evolutionary rates: nonsynonymous substitution rate (\( d_S \)), synonymous substitution rate (\( d_K \)) and \( d_S/d_K \) ratio. We show that genes regulated by more different TFs/miRNAs evolve more slowly is generally maintained in human, mouse and fruit fly. By controlling for the other confounding factors (i.e. protein connectivity, gene expression, gene compactness and protein structure), the partial correlations between \( N_{\text{miR}} \) and evolutionary rates still hold well, whereas the strength of \( N_{\text{TF}} \) as a rate indicator is greatly decreased in human/mouse and even disappears in fruit fly. We further find two trends: miRNA regulation tends to be much stronger than TF regulation in determining the rate of protein sequence evolution, and TF and miRNA regulations have a dependent effect on evolutionary rates, both of which are generally maintained across metazoans (human, mouse and fruit fly). We also observe that \textit{trans}-regulation seems to play a much greater role in human/mouse than in fruit fly in causing variation in protein evolutionary rates. This result reveals that the relative impact of \textit{trans}-regulation on the evolutionary rates appears to be different between vertebrates and invertebrates.

**MATERIALS AND METHODS**

**Data retrieval and extraction**

The protein-coding genes in human and mouse, orthology assignments and human–mouse and mouse–human evolutionary rates (\( d_S \), \( d_K \) and \( d_S/d_K \)) were downloaded from the Ensembl genome browser at http://www.ensembl.org/ (release 69) (45). For an alternatively spliced gene, only its longest isoform was selected. To avoid the confounding factor of gene duplication, only 1:1 orthologs between human and mouse genes were considered. Meanwhile, fruit fly protein-coding genes were downloaded from Flybase (release 4.3) (46). Fruit fly genes with single-copy orthologs across five other \textit{Drosophila} species (i.e. \textit{D.melanogaster}, \textit{Drosophila simulans}, \textit{Drosophila yakuba}, \textit{Drosophila erecta}, \textit{Drosophila sechellia} and \textit{Drosophila ananassae}) and the evolutionary rates were obtained from Larracuente et al.’s study (23). Here, \( d_S \) values = 0 and \( d_S/d_K \) values \( \geq 2 \) were not considered. Additionally, the human and mouse genes on chromosome Y were excluded for reducing the possibility of irregular evolutionary rates in the short single-copy sex chromosome.

**Protein connectivity**

The connectivity of a protein was defined by the total number of distinct proteins interacting with the protein. The PPI datasets of human, mouse and fruit fly were downloaded from STRING 9.0 (57), which retrieved known and predicted PPIs from literature.
Gene expression

Normalized expression datasets of 78 nonpathogenic human tissues and 77 nonpathogenic mouse tissues were downloaded from BioGPS (58), and a normalized expression dataset of 27 fruit fly nonpathogenic tissues was downloaded from FlyAtlas (59). If multiple probe sets refer to the same gene, the signals from different probe sets of the same gene were averaged. Here, expression was analyzed in terms of expression level and tissue specificity. A larger RCVE indicates a more important contribution of the factor of interest to the regression model (65).

RESULTS AND DISCUSSION

miRNA regulation is much more important than TF regulation in determining the evolutionary rates of metazoans

Previous studies have shown that the number of regulatory TFs that regulate a gene (N_{TF}) is negatively correlated with d_{N}/d_{S} in a yeast transcriptional regulatory network (9,10), leading to that genes with more regulatory TFs tend to evolve more slowly. We are then interested to know whether the trend is maintained in multicellular organisms. We first extract experimentally determined TFBS data (i.e. TF ChIP-binding datasets) from human, mouse and fruit fly (‘Materials and Methods’ section; Table 1) and estimate the Spearman’s rank correlation (ρ) between N_{TF} and evolutionary rates (i.e. d_{N}, d_{S} and d_{S}/d_{N}). In general, we find that evolutionary rates are negatively correlated with N_{TF} in the three species examined (Table 2). In terms of miRNA regulation, we extract human, mouse and fruit fly miRNA target data (Table 1) and also show negative correlations between N_{miR} and evolutionary rates in the three species examined (Table 2). These observations reveal a common trend that genes regulated by more TFs or miRNAs evolve more slowly at both the protein and RNA levels in metazoans.

The above results, however, should be treated carefully because many confounding factors that may affect evolutionary rates of protein-coding genes have not been controlled. As stated above (see ‘Introduction’ section), the confounding factors include protein connectivity, gene expression [expression level and tissue specificity (or expression breadth)], gene compactness (UTR length, intron length and intron number), protein structure (solvent accessibility and disorder content) and so on. Some of these confounding factors have also been reported to be correlated with N_{TF} or N_{miR}. For example, N_{TF} was reported to be positively correlated with mRNA expression (9) and UTR length (12). Meanwhile, N_{miR} was shown to be positively correlated with protein connectivity (11,43), expression breadth (43), 3'UTR length (12) and disorder content (11). Thus, we reevaluate the correlations between trans-regulation (N_{TF} and N_{miR}) and evolutionary rates by using partial correlation analyses (66) to simultaneously control for these confounding factors. As shown in Table 2, N_{TF} is still negatively correlated with evolutionary rates in human and mouse after controlling for N_{miR} and the other eight potential confounding factors. However, the partial correlations between N_{TF} and evolutionary rates are substantially reduced in human/mouse and even disappear in fruit fly (Table 2). This result suggests that the evolutionary effect of N_{TF} is considerably affected by these confounding factors. On the other hand, the negative correlations between N_{miR} and evolutionary rates remains strong in all three species examined after controlling for N_{TF} and the other confounding factors (Table 2). These observations reveal that N_{TF} and N_{miR} tend to have different effects on evolutionary rates. We find that the partial correlation between N_{miR} and evolutionary rates is remarkably stronger than that between N_{TF} and evolutionary rates.
when the other confounding factors are controlled, suggesting that \(N_{\text{miR}}\) is much more important than \(N_{\text{TF}}\) in affecting \(d_N, d_S\) and \(d_{N/S}\) (Table 2). This trend is maintained across metazoans.

### Dependent effects of TF and miRNA regulations on protein evolutionary rates in metazoans

It is known that TF and miRNA would cooperate with each other in gene regulation (3–5). In addition, genes with more TFBSs have a higher probability to be targeted by miRNAs and tend to have more miRNA-binding sites in human (7). Also, highly connected TFs in human regulatory network tend to regulate more miRNAs and tend to have more miRNA-targeted sites in human (7). Also, highly connected TFs tend to evolve more slowly than those regulated by only one type of \(G_{\text{trans}}\), suggesting that combination of TF and miRNA regulations has a dependent effect on protein evolutionary rates in metazoans.

We further conduct a stepwise multiple regression analysis including \(N_{\text{TF}}, N_{\text{miR}}\) and the other eight confounding factors to explore the interaction effects on \(d_N/d_S\) between any two of these 10 factors. According to the stepwise model selection, the trend that the coefficients of the \(N_{\text{TF}}-N_{\text{miR}}\) interaction term \((\beta_{1,2})\) significantly deviate from zero holds in all three species examined (Supplementary Table S1), further supporting the dependence between \(N_{\text{TF}}\) and \(N_{\text{miR}}\) in affecting \(d_N/d_S\).

**Trans-regulation is much more important in mammals than in fruit fly in determining protein evolutionary rates**

We have shown that \(G_{\text{trans}}\) is an important indicator of evolutionary rates in metazoans (Table 2). Considering the other biological factors

### Table 1. The numbers of TFs and miRNAs used in this study

<table>
<thead>
<tr>
<th>Species</th>
<th>TF Data source</th>
<th>Number of TFs</th>
<th>miRNA Data source</th>
<th>Number of miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>ENCODE (hg19)</td>
<td>162</td>
<td>TargetScanHuman (release 6.2)</td>
<td>1267</td>
</tr>
<tr>
<td>Mouse</td>
<td>hmChIP (mm8)</td>
<td>59</td>
<td>TargetScanMouse (release 6.2)</td>
<td>663</td>
</tr>
<tr>
<td>Fruit fly</td>
<td>DroID (May 2011)</td>
<td>149</td>
<td>TargetScanFly (release 6.2)</td>
<td>121</td>
</tr>
</tbody>
</table>

### Table 2. Spearman’s rank coefficient of correlation \((\rho)\) between evolutionary rates \((d_N, d_S\) and \(d_{N/S}\)) and experimentally determined \(N_{\text{TF}}\) (or \(N_{\text{miR}}\)) before and after controlling for \(N_{\text{miR}}\) (or experimentally determined \(N_{\text{TF}}\)) and the other eight confounding factors: protein connectivity, expression level, tissue specificity \((\tau)\), UTR length, intron length, intron number, solvent accessibility and disorder content

<table>
<thead>
<tr>
<th>Indicator and species</th>
<th>Before control</th>
<th>After control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(d_N)</td>
<td>(d_S)</td>
</tr>
<tr>
<td>Evolutionary rate versus (N_{\text{TF}})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human(^a)</td>
<td>−0.1852**</td>
<td>−0.1641***</td>
</tr>
<tr>
<td>Mouse(^b)</td>
<td>−0.1499**</td>
<td>−0.1087***</td>
</tr>
<tr>
<td>Fruit fly(^c)</td>
<td>−0.0759***</td>
<td>−0.1081***</td>
</tr>
<tr>
<td>Evolutionary rate versus (N_{\text{miR}})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human(^d)</td>
<td>−0.3760***</td>
<td>−0.2941***</td>
</tr>
<tr>
<td>Mouse(^e)</td>
<td>−0.3165***</td>
<td>−0.2710***</td>
</tr>
<tr>
<td>Fruit fly(^f)</td>
<td>−0.1172***</td>
<td>−0.2027***</td>
</tr>
</tbody>
</table>

\(^a\)The analysis was based on 6870 human genes and their mouse orthologs.
\(^b\)The analysis was based on 4903 mouse genes and their human orthologs.
\(^c\)The analysis was based on 1768 fruit fly genes. The \(d_N, d_S\) and \(d_{N/S}\) values were estimated in single-copy orthologs within the six *Drosophila* group species (23).

Significance: * \(P < 0.05\), ** \(P < 0.01\), and *** \(P < 0.001\).
associated with evolutionary rates of proteins [protein connectivity, gene expression (expression level and tissue specificity), gene compactness (UTR length, intron length and intron number) and protein structure (solvent accessibility and disorder content)], we then ask which biological factor(s) is/are the dominant determinant(s) of evolutionary rates. To this end, we measure the relative effect of each individual factor in determining the evolutionary rates by calculating the RCVE (see ‘Materials and Methods’ section). As shown in Figure 3 and Supplementary Figure S1A, the most dominant determinants of \( d_S \) and \( d_S/d_S \) common to human and mouse are trans-regulation (\( N_{TF} \) and \( N_{miR} \)) and protein structure (solvent accessibility and disorder content), whereas only protein structure is shown as a dominant determinant in fruit fly. Regarding \( d_S \), trans-regulation also exhibits influential determinants in human and mouse; however, the trend is not observed in fruit fly (Supplementary Figure S1B). Our results suggest that the effect of trans-regulation (especially miRNA regulation) on protein evolutionary rates is much stronger in mammals than in insects, in consistent with our above finding that the correlations between trans-regulation and evolutionary rates are relatively less significant in fruit fly than in human/mouse (Table 2). The results reveal that trans-regulation seems to be much more important in human/mouse than in fly in determining the rate of protein sequence evolution.

The above results thus suggest that the relative impacts of trans-regulation on evolutionary rates are different between vertebrates and invertebrates. In view of the relationship between regulatory complexity and organismal complexity, there are two possible reasons. First, for TF regulations, previous studies have indicated that organismal complexity might arise from progressively more elaborate gene regulation and the number of TFs per gene is positively correlated with the size of the genome (68–70). Second, in terms of miRNAs, a recent study showed an exponential correlation between the 3’UTR length and morphological complexity (71). The median 3’UTR length is much longer in human than in fruit fly, leading to the conclusion that human genes generally have longer potential miRNA-targeted regions and more complex miRNA regulations (71). Several studies also demonstrated that miRNAs regulate 20–30% of vertebrate genes (72–75) but only 15% of Drosophila genes (75). These notions imply that regulatory complexity might increase with the increase of organismal complexity and trans-regulations tend to play a much greater role in mammals than in insects, leading to a higher correlation between trans-regulation and evolutionary rates in mammals.

Potential caveats

Although the trends that miRNA regulation is much stronger than TF regulation in determining the rate of protein sequence evolution, and TF and miRNA regulations have a dependent effect on evolutionary rates generally hold in metazoans (human, mouse and fruit fly), the limited experimental data (i.e. ChIP-supported TFs and TFBSs) probably cause bias in our results. To address this possibility, we retrieve 843 TRANSFAC human TFs with known PWMs, filter out potentially false-positive TFBSs using the GBP scores (see ‘Materials and Methods’ section) and then conduct the same analyses. Obviously, the number of the TRANSFAC human TFs is much larger than that of the ChIP-supported TFs used above (843 versus 162; Table 1). On the basis of TRANSFAC-based \( N_{TF} \) (or predicted \( N_{TF} \)), we find that the abovementioned trends still hold well (Table 3, Supplementary Figures S2A and S3A and Supplementary Table S1). Although highly accurate TFBS predictions (which are currently more comprehensive than experimental data) remain challenging (76–78), the predicted TFBS data used here were generated by integrating multiple evidence sources (including sequence conservation, cis-feature, transcriptional information, epigenetic information, and so on) with motif information (54), which were shown to be highly predictive of true locations of TF binding (54,79–81). It is worthwhile to apply our evolutionary analyses to other species (or newly generated data) as the dramatic increase of publicly available trans-regulation data. Because the probability of observing the same trends from two biased datasets appears to be small, our results are likely unbiased.

Moreover, because rodents have a faster molecular clock than primates (82,83), it is possible to yield different tendencies between comparison of human–mouse orthologs and that of two species with similar molecular clocks. We therefore ask whether our results may be biased toward different molecular clocks. To address this question, we conduct the same statistical analyses for mouse–rat orthologs, which have similar molecular clocks, and show the same tendencies as above (Table 4, Supplementary Figures S2B and S3B and Supplementary Table S1). These results indicate that these observed trends are not affected by species selection or different molecular clocks. Therefore, our results can be regarded, in a broad view, as exploring the impacts of trans-regulation on evolutionary rates.
Concluding remarks

This study analyzes the impacts of two trans-regulatory factors (\(N_{TF}\) and \(N_{miR}\)) on the evolutionary rates in the metazoan protein-coding genes. Our results indicate that (i) both \(N_{TF}\) and \(N_{miR}\) are negatively correlated with evolutionary rates (\(d_N\), \(d_S\) and \(d_N/d_S\) in metazoans, but the strength of \(N_{TF}\) becomes weak in human/mouse and even disappears in fruit fly if the other confounding factors are controlled for; (ii) evolutionary rates tend to more strongly correlated with \(N_{miR}\) than with \(N_{TF}\); (iii) genes simultaneously regulated by TFs and miRNAs are subject to stronger selection pressure than those regulated by only TFs or miRNAs, and the stepwise multiple regression analysis also reveals that the coefficients of the \(N_{TF}-N_{miR}\) interaction term (\(\beta_{1,2}\)) significantly deviate from zero, both of which suggest the dependence between \(N_{TF}\) and \(N_{miR}\) in affecting \(d_N/d_S\); and (iv) compared with other biological factors, trans-regulation exhibits an influential...
Figure 3. The RCVE of the ten factors: $N_{TF}$, $N_{miR}$, protein connectivity (PPI), expression level (ExpLvl), tissue specificity (t), UTR length (UTRLeng), intron length (InLeng), intron number (InNum), solvent accessibility (SolAcc) and disorder content (DisCont) on $d_{s}/d_{S}$ in human, mouse and fruit fly. The analyses were based on 6870 human genes, 4903 mouse genes and 1768 fruit fly genes.
determinants in determining $d_S$ and $d_S/d_S$ in vertebrates, whereas the effect of trans-regulation on protein evolutionary rates is relatively weaker in invertebrates. The first and fourth trends show a great variation in rate determinants between vertebrates and invertebrates, also echoing the previous notion that the rules governing evolutionary rates may not be the same for all species (21). Because the currently available trans-regulatory data may only partially represent the reality, we compare the impacts of TFs and miRNAs across species and evaluated the impacts of them by controlling for potential confounding factors. It is found that the second and third trends hold well in diverse species including vertebrates and invertebrate. We therefore suggest that these two observations should be generally maintained in metazoans, although the roles of various rate determinants might be different between species (21) (also see the first and fourth trends). In addition, our result shows remarkable dependent effects of TF and miRNA regulations on protein evolutionary rates. We thus demonstrate the intricate relationships between gene regulations and the actions of natural selection in metazoan protein evolution.

## Supplementary Data

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–3.

## Acknowledgements

We especially thank Ben-Yang Liao and Chia-Ying Chen for their valuable suggestions, and Ting-Wei Hsu for a part of the data preprocessing.

## Funding

Funding for open access charge: Genomics Research Center and Institute of Information Science of Academia Sinica; National Science Council of Taiwan [NSC99-2628-B-001-008-MY3 to T.-J.C., and NSC100-2628-E-001-006-MY3 to H.-K.T.].

Conflict of interest statement. None declared.

## References


