# Evolution of gene regulation between two yeast species

## Introduction

Determining the causes of phenotypic divergence remains to this day a key research topic in evolutionary biology. While modifications of the coding sequence of proteins can bring about phenotypic changes, their occurrence is usually too rare to account for observed differences in phenotype (King and Wilson 1975). Instead, phenotypic diversity appears to be largely affected by changes in the regulation of gene expression (Carroll 2005; Rockman and Kruglyak 2006). Since transcription and mRNA translation are central to gene expression, investigating how these processes vary both intra- and interspecifically is crucial for clarifying which types of evolutionary forces are at play at different stages of gene regulation, and how they contribute to gene expression divergence and, ultimately, speciation.

The genome-wide analysis of both transcription and translation has been made possible by recent advances in the field of comparative genomics (e.g., Ingolia et al. 2009). The transcription of genes of interest can be examined via messenger RNA sequencing (mRNA-seq), which is used to measure mRNA abundance. The more a gene is transcribed, the higher its mRNA abundance will be. Comparing mRNA abundance levels between two species can thus allow us to assess interspecific differences in transcription. Translation can also be measured via ribosomal profiling, which provides estimates of ribosomal occupancy (i.e., number of active ribosomes for a specific gene). Both mRNA abundance and ribosomal occupancy estimates can then be used to measure translational efficiency, or the rate of mRNA translation into proteins (i.e., ribosomal occupancy divided by mRNA abundance).

Studies of variation in gene expression have shown that mRNA abundance levels differ greatly both within and across species (e.g., Tirosh et al. 2011; Rifkin et al. 2003; Schadt et al. 2003). Indeed, 25% or more of genes shared by conspecifics are characterized by divergent mRNA abundance levels, and this percentage increases when comparing closely related species (McManus et al. 2014). However, protein abundance appear to be quite conserved across species despite the divergence observed in mRNA abundance (Schrimpf et al. 2009; Laurent et al. 2010). Consequently, McManus and associates (2014) have proposed that post-transcriptional processes, such as mRNA translation, may act as buffers to gene expression divergence by mitigating the effects of variation in mRNA levels. To explore how the evolution of gene regulation occurs at both the transcriptional and translational levels, and affects species divergence, we first analyzed and compared mRNA-seq and ribosome profiling data from two species of yeast, *Saccharomyces cerevisiae*, the familiar lab yeast, and *S. paradoxus*, a wild species. We also examined whether coordinated (i.e., changes at transcriptional and translational levels act in concert) or compensatory (i.e., changes at the transcriptional level are offset by changes at the translational level, or vice-versa) regulatory evolution was acting upon each gene.

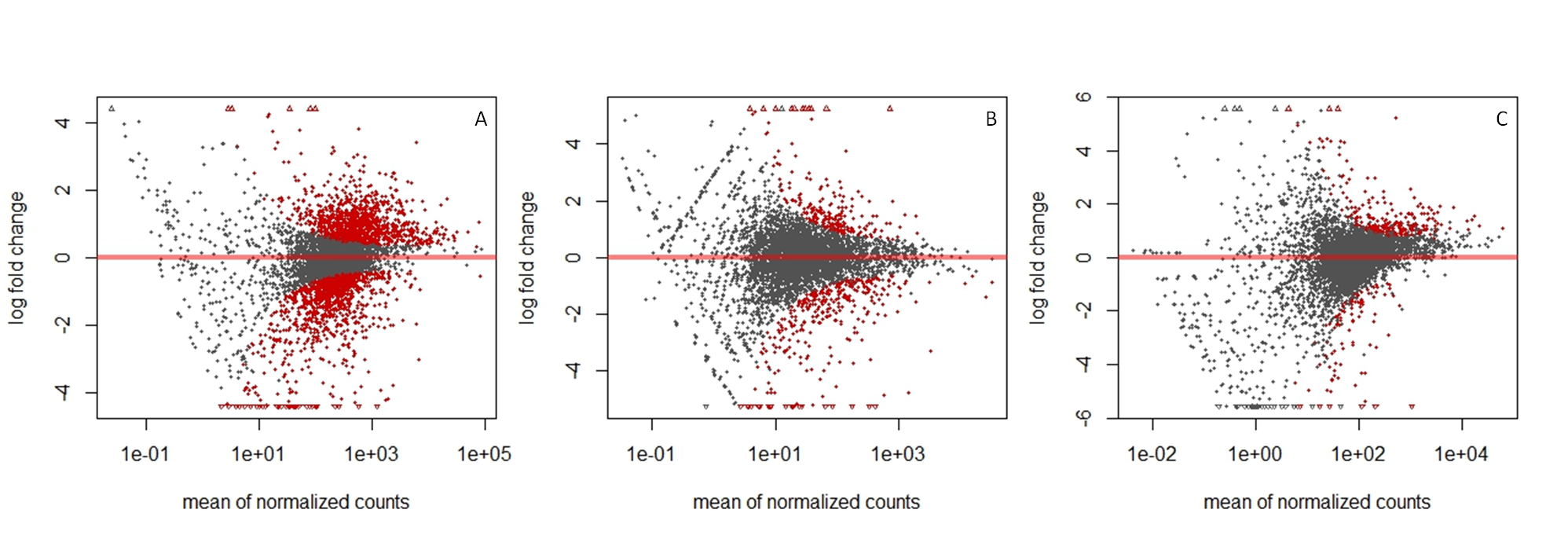
## Methods

We used part of the mRNA-seq and ribosome profiling data provided by McManus and associates (2014), from two currently sympatric yeast species, *S. cerevisiae* and *S. paradoxus*, which diverged about 5 million years ago (Sniegowski et al. 2002). The bed files "S\_cerevisiae\_genes.bed" and "S\_paradocus\_genes.bed", which contained gene locations for *S. cerevisiae* and *S. paradoxus*, respectively, were used to build transcriptomes for each strain ("Scer\_transcriptome.fa" and "Spar\_transcriptome.fa"). This was done using custom python scripts ("final\_Scer.py" and "final\_Spar.py"). We then used kallisto to index our transcriptome files ("Scer\_transcriptome.idx" and " Spar\_transcriptome.idx"), and then quantify gene expression in eight different replicates of RNA-seq and ribosome profiling data from the two species ("Scer\_ribo\_seq\_1.fastq.gz", "Scer\_ribo\_seq\_2.fastq.gz", "Scer\_RNA\_seq\_1.fastq.gz", "Scer\_RNA\_seq\_2.fastq.gz", "Spar\_ribo\_seq\_1.fastq.gz", "Spar\_ribo\_seq\_2.fastq.gz", "Spar\_RNA\_seq\_1.fastq.gz", "Spar\_RNA\_seq\_2.fastq.gz").

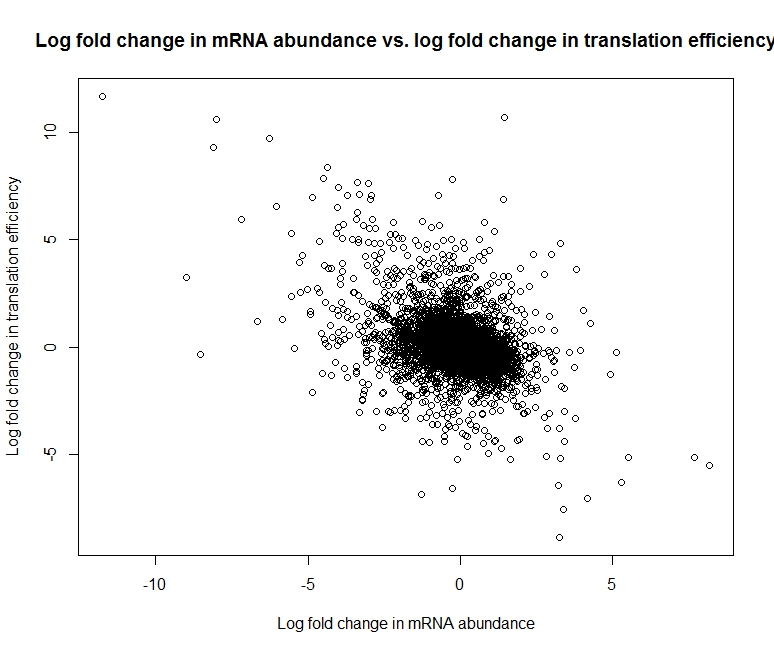
DESeq2 was used in R (version 3.4.0; "Final\_R\_code.R") to quantify differential gene expression, ribosomal occupancy, and translation efficiency (with likelihood ratio test) between species. For all three measures, an adjusted false discovery rate (FDR) of 10% was used to determine significant differences. The number of genes undergoing compensatory (i.e., mRNA abundance going up and translation efficiency going down, or mRNA abundance going down and translation efficiency going up) or coordinated (i.e., mRNA abundance going up and translation efficiency going up, or mRNA abundance going down and translation efficiency going down) evolution was also computed using R.

All python and R scripts are available on GitHub: https://github.com/SMG17/Final-Project.git.

## Results

First, out of 5474 genes, 2472 genes were differentially expressed between *S. cerevisiae* and *S. paradoxus* at a FDR of 10% (**Figure 1A**; **Table S1**). Among these significantly differentially expressed genes, 1188, or 51.9%, were down-regulated (log2 fold change < 0), and 1284, or 48.1%, were up-regulated (log2 fold change > 0). Second, out of the same total number of genes, 556 genes had differential ribosome occupancy levels between the two species at a FDR of 10% (**Figure 1B**; **Table S2**). 324, or 58.3%, of those genes exhibited decreased ribosome occupancy (log2 fold change < 0), and 232, or 41.7%, showed increased ribosome occupancy (log2 fold change > 0). Finally, 332 genes were characterized by differential translation efficiency between the two species at a FDR of 10% (**Figure 1C**; **Table S3**). 98, or 29.5%, of those genes had decreased translation efficiency (log2 fold change < 0), and 234, or 70.5%, had increased translation efficiency (log2 fold change > 0).

**Figure 1.** MA plots showing the log2 fold change as a function of the mean of normalized counts for A) differentially expressed genes, B) genes with differential ribosome occupancy, and C) genes with differential translation efficiency between S. cerevisiae and S. paradoxus. Points in red represent genes with an adjusted p-value below 0.1 (FDR: 10%). Open triangles represent genes which fall out of the window.

We also examined the relationship between the log2 fold change in translation efficiency and the log2 fold change in mRNA abundance (**Figure 2**). Although most genes showed low log2 fold changes and were found close to the origin (0,0), the two variables appeared to be negatively correlated, which would be consistent with compensatory evolution. In line with this result, we found a greater number of genes under compensatory evolution (**Table 1**; mRNA up/TE down, or mRNA down/TE up) than under coordinated evolution (**Table 1**; mRNA up/TE up, or mRNA down/TE down), when comparing changes in mRNA abundance and translation efficiency levels. More precisely, 3522 genes (or 216 at a FDR of 10%) were undergoing compensatory evolution, versus 1916 genes (or 42 at a FDR of 10%) undergoing coordinated evolution.

**Figure 2.** Scatterplot showing the log2 fold change in mRNA abundance between *S. cerevisiae* and *S. paradoxus* as a function of the log2 fold change in translation efficiency.

**Table 1.** Number of genes undergoing compensatory evolution (offset mRNA abundance and translational efficiency (TE) levels) and coordinated evolution (both mRNA abundance and translational efficiency (TE) levels act in concert) for all genes, and for genes with an adjusted p-value below 0.1 (FDR: 10%).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Compensatory evolution** | | **Coordinated evolution** | |
|  | mRNA up │ TE down | mRNA down │ TE up | mRNA up │ TE up | mRNA down │ TE down |
| All genes | 1815 | 1707 | 892 | 1024 |
| 10% FDR | 169 | 47 | 15 | 27 |

## Discussion

Variation in mRNA abundance has been shown to be a poor proxy for protein synthesis (e.g., Foss et al. 2007, Ghazalpour et al. 2011, Wu et al. 2013), and as such, may not accurately represent interspecific differences in gene expression. Variation in mRNA translation, however, has been suggested to significantly affect the evolution of gene expression (McMAnus et al. 2014). Nevertheless, only about 6% (332 out of 5474) of all expressed genes in this study exhibited differential translation efficiency, while more than 45% (2472 out of 5474) of the genes were differentially expressed between *S. cerevisiae* and *S. paradoxus*. Therefore, contrary to previous studies, our results seem to indicate that divergence in mRNA abundance affects considerably more genes than divergence in translation regulation.

When assessing the mode of regulatory evolution acting upon genes, we found evidence for a greater number of genes under compensatory evolution than coordinated evolution, and most importantly in cases where mRNA abundance was up-regulated, and translation efficiency was down-regulated. These results are consistent with the hypothesis that post-transcriptional processes, such as translation regulatory divergence, act as buffers to interspecific gene expression divergence by mitigating the effects of variation in mRNA levels. Indeed, while only about 2.5% (74 out of 3002) of genes with conserved mRNA abundance exhibited divergent translation efficiency, more than 10% (258 out of 2472) of differentially expressed genes between *S. cerevisiae* and *S. paradoxus* showed differential translation efficiency. This trend is also consistent with the negative correlation observed between genome-wide divergence in mRNA abundance and translation efficiency.

## Supplemental material

* **Table S1:** "mRNA\_diff\_exp\_sig.csv" - A table of differentially expressed genes between *S. cerevisiae* and *S. paradoxus* at an FDR of 10%.
* **Table S2:** "Ribo\_diff\_occup\_sig.csv" - A table showing genes with differential ribosome occupancy between *S. cerevisiae* and *S. paradoxus* at an FDR of 10%.
* **Table S3:** "TE\_diff\_sig\_csv" - A table showing genes with differential translation efficiency between *S. cerevisiae* and *S. paradoxus* at an FDR of 10%.
* **Scripts**: https://github.com/SMG17/Final-Project.git

## Literature cited

Carroll SB. 2005. Evolution at two levels: On genes and form. PLoS Biol 3: e245.

Foss EJ, Radulovic D, Shaffer SA, Ruderfer DM, Bedalov A, Goodlett DR, Kruglyak L. 2007. Genetic basis of proteome variation in yeast. Nat Genet 39: 1369–1375.

Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, et al. 2011. Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet 7: e1001393.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218-223.

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

Laurent JM, Vogel C, Kwon T, Craig SA, Boutz DR, Huse HK, Nozue K, Walia H, Whiteley M, Ronald PC, et al. 2010. Protein abundances are more conserved than mRNA abundances across diverse taxa. Proteomics 10: 4209–4212.

McManus CJ, May GE, Spealman P, Shteyman, A. 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Research: 24:422-430.

Rifkin SA, Kim J, White KP. 2003. Evolution of gene expression in the *Drosophila melanogaster* subgroup. Nat Genet 33: 138–144.

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

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, et al. 2003. Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297–302.

Schrimpf SP, Weiss M, Reiter L, Ahrens CH, Jovanovic M, Malmström J, Brunner E, Mohanty S, Lercher MJ, Hunziker PE, et al. 2009. Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. PLoS Biol 7: e48.

Sniegowski PD, Dombrowski PG, Fingerman E. 2002. *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEMS Yeast Res 1:299–306.

Tirosh I, Wong KH, Barkai N, Struhl K. 2011. Extensive divergence of yeast stress responses through transitions between induced and constitutive activation. Proc Natl Acad Sci 108: 16693–16698.

Wu L, Candille SI, Choi Y, Xie D, Jiang L, Li-Pook-Than J, Tang H, Snyder M. 2013. Variation and genetic control of protein abundance in humans. Nature 499: 79–82.