**Evolution of Gene Regulation between Two Yeast Species**

Alexis Weinnig

**Introduction**

Phenotypic divergence is ultimately controlled by the central dogma of biology, meaning a phenotype is determined by what genes are being transcribed and then translated to protein and at what *rate* this processes is happening. Transcription can be quantified by analyzing the amount of mRNA present in a cell. The more mRNA present for a certain gene, the more that gene was transcribed from the DNA.1 Translation can then be quantified through ribosomal profiling and determining ribosomal occupancy. Ribosomal occupancy estimates the relative number of ribosomes translating mRNA from each gene, so the higher the ribosomal occupancy, the higher the protein synthesis from that mRNA2. Analyzing mRNA expression and translation efficiency (ribosomal occupancy/mRNA) can help to tease apart the evolutionary forces at each stage of gene regulation (transcription and translation) and shed light on the mechanisms behind speciation.

The subphylum Saccharomycotina (Hemiascomycete yeasts), with its short and functional gene-rich genomes, has provided a platform for comparative genomic studies since the early 2000s3,4. Two species within this subphylum, *Saccharomyces cerevisiae* and *Saccharyomyces paradoxus*, diverged ~5 million years ago, yet they now occur in sympatry.5 Scannell *et al.* (2011) furthered the field of comparative genomics by generating three new genomes (*S. bayanus var. uvarum, S. kudriavzevii*, and *S. mikatae*) and comparing them to both *S. cerevisiae* and *S. paradoxus*. This advanced the annotation of the existing genomes and provided an invaluable tool for comparative genomic studies through the genus *S. sensu strico*6. Numerous studies have utilized this data to perform analyses on metabolism, gene function, and evolution; including ribosome profiling work by McManus *et al*. (2014)7. We utilized part of the data set (mRNA-Seq and ribosomal profiling data) from McManus *et al*. (2014) to investigate how evolution occurred at the transcriptional and translational level between *S. cerevisiae* and *S. paradoxus.* Taking it a step further, we investigate which “mode” of regulatory evolution (coordinated or compensatory) has acted upon each gene, aiming to improve the understanding of the roles played by the evolution of gene regulatory mechanisms, which contributed to the divergence of these two species.

**Methods**

*Sequence data acquisition and alignment*

mRNAseq and ribosomal profiling data from two yeast strains, *S. cerevisiae* and *S. paradoxus* was provided from McManus et al (2014). Transcriptomes for each yeast strain were generated from the complete genomes using bed files containing gene coordinates. The was accomplished using a python script that read the genome in as a dictionary, using the chromosome number as the key. Then each gene was extracted from its respective chromosome using the start and stop positions identified in the bed file. All genes names and sequences were concatenated in a fasta file. The transcriptome files were then indexed and the indexed transcriptomes were used to quantify expression of both the mRNA and ribosomal profiling data of both yeast strains using kallisto8.

*Gene expression analyses*

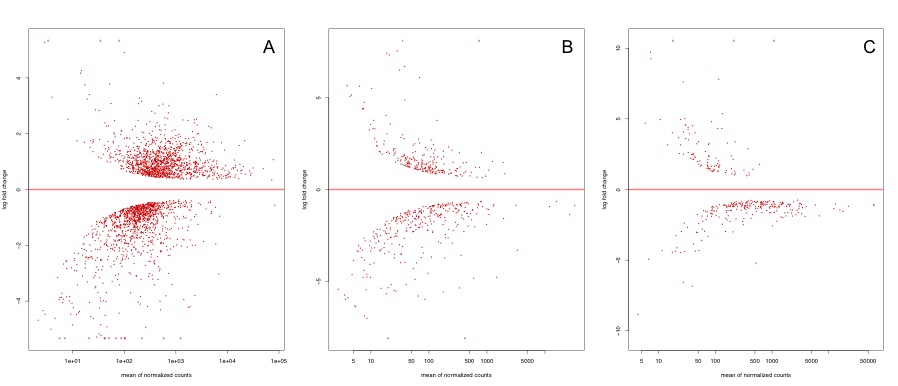
The mRNA differential gene expression and differential ribosomal occupancy (RPF) between species were determined using DESeq2 using R (version 3.4.0)9,10. DESeq2 was also used to detect differences in translation efficiency (RPF/mRNA) between species. However, a likelihood ratio test (LRT) was used to account for this. For mRNA differential expression, RPF, and translation efficiency (TE) significant differences were identified using and adjusted false discovery rate (FDR) of 10%.

*Determining compensatory or coordinated evolution*

The number of genes undergoing compensatory and coordinated evolution were determined using R. If a gene met the condition of having opposite mRNA expression and TE (ex. mRNA up & TE down) directions of effect it was categorized as compensatory evolution. If a gene met the condition of having the same mRNA expression and TE (ex. mRNA up & TE up) direction of effect it was categorized a coordinated evolution.

**Results**

Overall, out of 5438 genes total, 2472 genes were significantly differentially expressed between S. *cerevisiae* and *S. paradoxus* (FDR: 10%, Sup. Material: Table 1). Of those genes 1284 (52%) were up regulated (LFC > 0) and 1188 (58%) were down regulated (LFC < 0). 556 genes exhibited differential ribosomal occupancy (FDR: 10%, Sup. Material: Table 2). Of those genes 232 (42%) had an increased number or ribosomes per gene (LFC >0) and 324 (58%) had a decreased number of ribosomes per gene (LFC < 0). Lastly, 332 genes showed differential translation efficiency (FDR: 10%, Sup. Material: Table 3). 98 genes (30%) had increased translation efficiency (LFC > 0) and 234 genes (70%) had decreased translation efficiency (LFC < 0).



**Figure 1.** MA plots showing the log2 fold change (y-axis) vs the mean of normalized counts (x-axis) for A) differentially expressed genes, B) genes with differential ribosomal occupancy, and C) genes with differential translation efficiency between S. *cerevisiae* and *S. paradoxus*. Each red point represents a single gene that passed the significance threshold (FDR: 10%).

Generally, significant log fold change (LFC) in mRNA abundance, ribosomal occupancy and translational efficiency decreases as the mean of normalized counts increases (Fig 1). Each red dot in Figure 1 represents a gene that was significantly different (FDR: 10%) between the two yeast species. Differential mRNA expression had the highest number of genes with significant differences (Fig 1A, 2472 genes), followed by ribosomal occupancy (Fig 1B, 556 genes), and translational efficiency had the least number of significant differences between species (Fig 1C, 332 genes).



Additionally, the log fold change of mRNA abundance and log fold change of translation efficiency are negatively correlated (Fig 2). However, the majority of genes are clustered in the center of the plot, exhibiting relatively low log fold changes. This negative trend suggests that the majority of genes have undergone compensatory evolution instead of coordinated evolution. Occurrences of compensatory evolution (1,310 genes, Table 1) are more frequent than occurrences of coordinated evolution (1,162 genes, Table 1), which aligns with Figure 2.

**Figure 2.** Scatter plot of log fold change of translation efficiency vs log fold change of mRNA abundance between species. Each black circle represents a gene that passed the significance threshold (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 | 931 | 379 | 353 | 809 |

**Table 1.** Genes undergoing compensatory evolution (mRNA and translation efficiency (TE) changes in opposite directions) and coordinated evolution (mRNA and TE changes in the same direction) for all genes and genes that passed the significance threshold (FDR: 10%).

**Discussion**

Several previous studies have noted that, while evolutionary changes in mRNA abundance are definitely important, they may not always account for interspecies differences in gene expression.7,11,12 Overall, these studies suggest that differences in translation regulation play a considerable role in the evolution of gene expression. However, in this study, only roughly 6% of the total genes (332 out of 5438) exhibited differential translational efficiency (TE). When observing just the significantly differently expressed genes the ratio of genes experiencing differential TE increases to ~13% (332 out of 2472), but this is still less than the 25% observed in previous studies.7

In the pool of significantly differentially expressed genes, ~53% of genes were experiencing compensatory evolution (1,310 genes) and ~47% of genes were experiencing coordinated evolution (1,162 genes). This is different than expected and it is speculated that the low number of genes experiencing differential TE (~13% of genes) is affecting this distribution. However, it appears that when TE is decreased significantly, many more genes pass the significance threshold regardless of mRNA expression. When mRNA is up regulated and TE is down regulated ~37% of genes are experiencing compensatory evolution, compared to when mRNA is down regulated and TE is up regulated (~15% of genes). Also, when mRNA and TE are downregulated ~33% of genes are experiencing coordinated evolution, compared to when mRNA and TE are upregulated (~14% of genes). This could suggest that negative changes in TE could affect divergence more than mRNA abundance.

**Supplemental Material**

1. ‘mRNA\_diff\_exp.pdf’ – Table of differentially expressed genes between S. *cerevisiae* and *S. paradoxus* that pass the significance threshold (FDR: 10%)
2. ‘rRNA\_diff\_exp.pdf’ – Table of genes with differential ribosomal occupancy between S. *cerevisiae* and *S. paradoxus* that pass the significance threshold (FDR: 10%)
3. ‘trans\_eff\_diff\_exp.pdf’ - Table of genes with differential translational efficiency between S. *cerevisiae* and *S. paradoxus* that pass the significance threshold (FDR: 10%)

Supplemental material and all scripts can be found at:

<https://github.com/aweinnig/HTS-Seminar/tree/master/Final>

**References**

1. Li, J. J. & Biggin, M. D. Statistics requantitates the central dogma. *Science (80-. ).* **347,** 1066–1067 (2015).

2. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–23 (2009).

3. Dujon, B. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet.* **22,** 375–387 (2006).

4. Piskur, J. & Langkjaer, R. B. Yeast genome sequencing: the power of comparative genomics. *Mol. Microbiol.* **53,** 381–9 (2004).

5. Sniegowski, P. D., Dombrowski, P. G. & Fingerman, E. 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 (2002).

6. Scannell, D. R. *et al.* The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3 (Bethesda).* **1,** 11–25 (2011).

7. McManus, C. J., May, G. E., Spealman, P. & Shteyman, A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res.* **24,** 422–430 (2014).

8. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34,** 525–527 (2016).

9. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

10. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4,** 1521 (2015).

11. Foss, E. J. *et al.* Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLoS Biol.* **9,** (2011).

12. Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499,** 79–82 (2013).