

Fop Values, Start Codons, SD Sequence Conservation, and Their Correlation to Gene Expression Level

Hiroaki Sakai¹² **Chiaki Imamura**¹³ **Hiroshi Ohno**¹⁴
sakacho@sfc.keio.ac.jp s97113ci@sfc.keio.ac.jp t99540ho@sfc.keio.ac.jp
Takanori Washio¹² **Masaru Tomita**¹⁴
washy@sfc.keio.ac.jp mt@sfc.keio.ac.jp

¹ Laboratory for Bioinformatics
² Graduate School of Media and Governance
³ Department of Policy Management
⁴ Department of Environmental Information
Keio University
5322 Endo, Fujisawa-shi Kanagawa 252-8520, Japan

1 Introduction

In this study, we have analyzed the correlation between codon usage bias and various features of genes. Codons recognized by the most abundant tRNAs are used more frequently than other synonymous codons recognized by rare tRNAs. This bias is more significant in genes that are highly expressed and produce a large amount of protein, and is said to be related to the translation efficiency [1, 2].

We systematically calculated the *fop values* of all genes in the genome of *Escherichia coli* K-12, and analyzed their correlation to SD sequence conservation and start codon usage. The fop value in each gene is defined as the frequency of 'optimal codons', which are codons assumed to make translation most efficient [3].

We also calculated the fop values of all genes in the *Saccharomyces cerevisiae* genome, and analyzed the correlation between the fop value and temporal change in the transcription level, based on microarray data sets published by Patrick Brown's laboratory at Stanford University [4, 5].

2 Results

We found the following tendencies from the analysis on *E. coli*: (1) genes with a well-conserved SD sequence have higher fop values; (2) genes for abundant proteins have higher fop values; and (3) genes with a non-AUG start codon have lower fop values.

Fig. 1 shows results of the analysis on *S. cerevisiae*. It has been revealed that genes whose transcription level drops during the diauxic shift from fermentation to respiration tend to have high fop values. Genes with a large increase in their transcription level during sporulation, on the other hand, were found to have low fop values.

Acknowledgments

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas 'Genome Science' from The Ministry of Education, Science, Sports and Culture in Japan.

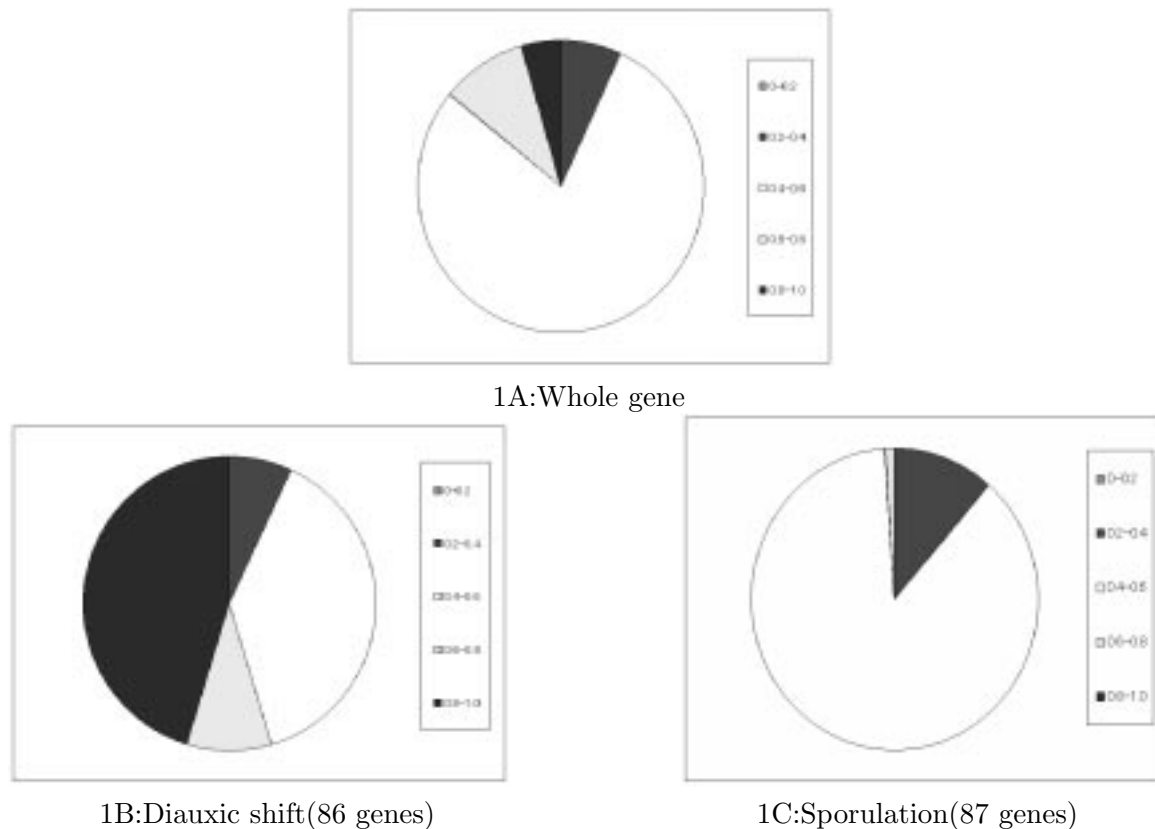


Figure 1: The number of genes classified by their fop values. (1A) all genes of *S. cerevisiae*; (1B) genes whose transcription level changes drastically during diauxic shift; and (1C) genes whose transcription level changes during sporulation. 55% of the genes in (1B) have their fop values greater than 0.6, which is much higher than (1A) (14%). On the other hand, only 1% of the genes in (1C) have fop values greater than 0.6.

References

- [1] Ikemura, T., Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs, *J. Mol. Biol.*, 158:573–597, 1982.
- [2] Ikemura, T., Codon usage and tRNA content in unicellular and multicellular organisms, *Mol. Biol. Evol.*, 2:13–34, 1985.
- [3] Ikemura, T., Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes, *J. Mol. Biol.*, 146:1–21, 1981.
- [4] DeRisi, J.L., Iyer, V.R., and Brown, P.O., Exploring the metabolic and genetic control of gene expression on a genomic scale, *Science*, 278:680–686, 1997.
- [5] Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O., and Herskowitz, I., The transcriptional program of sporulation in budding yeast, *Science*, 282:699–705, 1998.