Comparative Analysis of Location of First Intron

Atsushi Sakurai[1,2] s97446as@sfc.keio.ac.jp
Shigeo Fujimori[1,2] s97831sf@sfc.keio.ac.jp
Hiromi Kochiwa[1,3] t97357hk@sfc.keio.ac.jp
Takanori Washio[1,4] washy@sfc.keio.ac.jp
Masaru Tomita[1,3] mt@sfc.keio.ac.jp

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

Keywords: eukaryote, intron, computer analysis

1 Introduction

It is known that introns tend to be situated near the 5' end of genes in *S. cerevisiae.* Although it has been suggested that this locational bias may be related to the intron-loss process [2], the cause of the locational bias of introns is not yet clearly understood. One purpose of this work is to investigate whether a similar tendency of intron positions is observed in eucaryotes other than *S. cerevisiae.* We conducted comprehensive analyses on the locational distribution of all introns in seven species.

2 Methods and Results

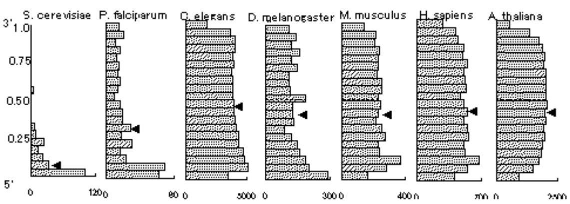

Figure 1: Intron distribution (all genes)

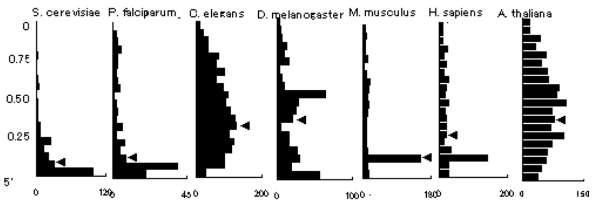

Figure 2: genes with one intron

Figure 3: genes with two introns

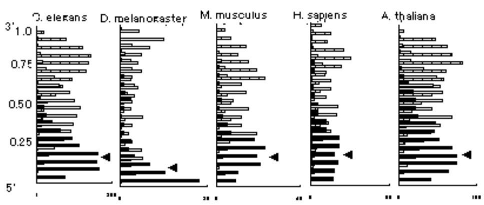

Figure 4: genes with three introns

The genome sequences of the following organisms were downloaded from the National Center for Biotechnology Information (NCBI) (ftp://ncbi.nlm.nih.gov/): *S. cerevisiae, P. falciparum, C. elegans, D. melanogaster, M. musculus, H. sapiens, A. thaliana.* We used Cleanup1.8.3 for the GenBank files in order to exclude redundant entries. Only introns located in coding regions were considered, and introns located in 5'or 3' untranslated regions are excluded from our analysis. Figure 1 shows the number of introns in each relative location for all of the seven species. The Y axis is the relative position within coding region, and the X axis is the number of introns located in the position. Triangles represent the median of all intron positions, showing the bias that more introns are located toward the 5' end of genes than expected. Figure 2 shows the result of the same analysis with genes containing only one intron. Comparatively, these single-intron genes show a stronger bias of intron positions. The results of analyses on genes with two and three introns are presented in Figures 3 and 4, respectively.

3 Discussion

From our results, it appears that first introns have a stronger tendency to be located towards the 5' end than do second or third introns. We also found, as shown in Figure 5, that first introns (the bottom) are on average longer than 2nd, 3rd, 4th, and 5th introns. It has been reported that certain genes have enhancer and silencer sequences within first introns, which may explain their longer length. If first introns contain signal sequences important for transcription initiation, it may be preferable for them to be located closer to the initiation sites. Genes coding ribosomal proteins often possess a single intron at the very beginning of their coding region. We compared these genes with homologous genes of other species. We found that many of the homologous genes also contain an intron at the beginning of their coding region. In summary, we found a strong indication that first introns have a special role in cellular metabolism and that this role may explain their longer length and the locational bias specific to them.

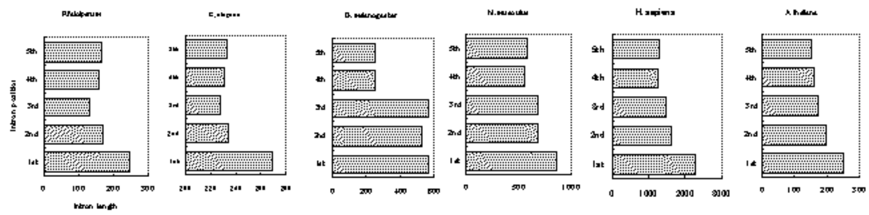

Figure 5: Average length of introns

4 Acknowledgments

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

References

[1] Derr, L. and Strathern, J., A role for reverse transcripts in gene conversion, *NATURE*, 361:170–173, 1993.

[2] Fink, G., Pseudogenes in yeast?, *Cell*, 49:5–6, 1987.

[3] Maxine, L., Creation of processed pseudogene by retroviral infection, *Cell*, 49:93–102, 1987.

[4] Tourmente, S., Chapel, S., Dreau, D., Drake, M., Bruhat, A., Couderc, J., and Dastugue, B., Enhancer and silencer elements within the first intron mediate the transcriptional regulation of the beta 3 tubulin gene by 20-hydroxyecdysone in Drosophila Kc cells, *Insect Buichem. Molec. Biol.*, 23(1):137–143, 1993.