Analysis of the Maternal mRNAs Encoding DNA-Binding Motifs in the Ascidian Egg Using SSDB

Shuichi Kawashima[1]
shuichi@kuicr.kyoto-u.ac.jp

Takeshi Kawashima[2]
takeshi@develop.zool.kyoto-u.ac.jp

Akihiro Nakaya[1]
nakaya@kuicr.kyoto-u.ac.jp

Kazuhiro W. Makabe[2]
makabewk@ascidian.zool.kyoto-u.ac.jp

Susumu Goto[1]
goto@kuicr.kyoto-u.ac.jp

Minoru Kanehisa[1]
kanehisa@kuicr.kyoto-u.ac.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, 611-0011, Japan
[2] Molecular and Evolutionary Developmental Biology, Depertment of Zoology, Graduate School of Science, Kyoto University, Sakyoku, Kyoto 606-8502, Japan

Keywords: maternal factor, transcription factor, DNA binding motif, Pfam, KEGG, MAGEST

1 Introduction

In the early development, the same animal species develop into the same adult body through correct developmental processes. If correct developmental steps can be described by mathematical functions, the maternal cytoplasmic factors should be the initial condition of the functions. The maternal factors mainly consist of mRNAs and proteins stored in the egg. These factors regulate the earliest developmental processes, such as egg cleavage and basic metabolisms, and also activate the zygotic transcription events that are the first steps of the transcription of genes coded on a set of chromosomes in the embryo. Therefore, we believe that cataloging of these factors will be important to system-level understanding of developmental processes.

In this study, we analyzed the population of maternal genes carrying DNA binding motifs in the egg of the ascidian *Harocynthia roretzi*, in comparison to the populations of genes encoded in other genomes. In order to identify possible transcription factors, we use DNA binding motifs defined in Pfam. We then perform an exhaustive search for DNA binding factors in the four complete genomes and also in the MAGEST database, which currently contains about 2600 tag sequences of maternal cDNA of *H.roretzi* and their expression data obtained from whole-mount in situ hybridization [1].

2 Method and Results

Sequence motifs in Pfam are precomputed against all protein genes in the compelete genomes and stored in SSDB. Given Pfam entries representing DNA binding motifs, SSDB reports all protein genes carrying such motifs in each genome of the KEGG/GENES database. We examined the following four eukaryotic genomes: *Saccharomyces cerevisiae* (abbreviation: S), *Drosophila melanogaster* (D), *Arabidopsis thaliana* (A), and *Caenorhabditis elegans* (C). On the whole, the frequency of observing each DNA binding motif tends to be similar through the four genomes; for examplem, genes carrying C2H2 and C3HC4 zinc fingers apear with high percentages. In contrast, HLH motifs exist in four organisms but the number of such genes is small. The total numbers of genes obtained are 195 in *S.cerevisiae*, 1538 in *A.thaliana*, 777 in *D.melanogaster* and 845 in *C.elegans*. In decreasing order of

observed frequency, motifs are ranked as: C2H2 zinc finger (22%), C3HC4 zinc finger (14%), and bZIP zinc finger (10%) in *S.cerevisiae*; C3HC4 zinc finger (23%), myb CCHC zinc finger (13%), and CCHC zinc finger (8%) in *A.thaliana*; C2H2 zinc finger (38%), homeobox (12%), and C3HC4 zinc finger (11%) in *D.melanogaster*; and C4 zinc finger (23%), C2H2 zinc finger (12%), and C2HC4 zinc finger (11%) in *C.elegans*. Thus, the frequency of each type of DNA binding motif varies a little depending on the organism.

Next, we analyzed orthologous relations, if any, among four organisms for each type of motif. An ortholog is defined by the best-best relation in SSDB [2]. In the case of C2H2 zinc finger, no orthologous relations were identified among four organisms, possibly because there were so many paralogous genes carrying C2H2 motifs in each genome. In contrast, for NFX1, ARID, TFIIS, and transcript_fac2 ortholog groups could be identified among four organisms because the number of genes carrying these motifs is small. When orthologous relations are examined between only two organisms, there are numerous orthologous pairs especially between *D.melanogaster* and *C.elegans*.

Table 1: DNA binding motifs found in MAGEST and comparison with four genomes.

Pfam ID	SADC	SAD-	SA-C	S-DC	-ADC	SA--	S-D-	S--C	-AD-	-A-C	--DC	S---	-A--	--D-	---C
GATA	0	0	0	0	0	1	0	1(1)	0	0	2	0	26	4	7
zf-C2H2	0	0	0	2	1	3	5	1	4	3	24(3)	32	98(1)	258(8)	74
zf-C3HC4	3	4	3	1	8	5	1	2	9	5	9(2)	9	318	47	65
zf-CCHC	0	(2)	2	0	0	1	0	0	3	1	3	0	110	11(1)	20
ZZ	0	1	0	0	1	0	0	0	0	0	3	1	15(1)	7	7
HLH	0	0	0	0	2(1)	0	0	1	1	2	12(1)	5	101	33	14(1)
CSD	0	0	0	0	0	0	0	0	0	0	2(1)	0	4	2	4
T-box	0	0	0	0	0	0	0	0	0	0	2	0	0	6(1)	13

In order to compare these results with *H.roretzi*, we selected maternal genes carrying DNA binding motifs from MAGEST. The result is summarized in Table 1. The number in parentheses represents the number of the *H.roretzi* genes which have the corresponding DNA biniding motif and belong to the orthologue group. Zinc finger motifs, such as C2H2 and C3HC4, also appear with high frequency in Ascidian maternal genes and the tendency is similar to the genomes of the four organisms. Many of the genes carrying C2H2 zinc finger motifs have homologs to the orthologous group obtained from *D.melanogaster* and *C.elegans*. On the other hand, C4, bZIP, myb and homeobox don't exist maternally. Consequently, it is suggested that the population of maternal factors is considerably different from the repertoire of transcription factors encoded in the genomes.

3 Acknowledgements

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

References

[1] Kawashima, T., Kawashima, S., Kanehisa, M., Nishida, H., and Makabe, W.K., MAGEST: Maboya Gene expression patterns and Expression Sequence Tags, *Nucleic Acids Res.*, 28(1):133–135, 2000.

[2] Sato, Y., Nakaya, A., Shiraishi, K., Kawashima, S., Goto, S., and Kanehisa, M., SSDB: Sequence Similarity Database in KEGG, *Genome Informatics*, 12, 2001.