MAGEST: ESTs and Gene Expression Pattern Database for *Halocynthia roretzi* Maternal cDNA

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

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

Hiroki Nishida [3]
hnishida@bio.titech.ac.jp

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

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

[1] Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan
[2] Department of Zoology, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
[3] Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501, Japan

1 Introduction

The Ascidian *Halocynthia roretzi* is a good model organism to investigate gene functions in developmental mechanisms. In the process of early embryogenesis, maternal factors stored in egg cytoplasms are known to play various significant roles [1]. Recent works have revealed that there exist cytoplasmic determinants that direct the formation of epidermis, muscle and endoderm as well as cytoplasmic factors are involved in the axis specification of the embryo and in gastrulation [1]. In addition, the embryonic cell lineage is almost completely revealed by intracellular marking [2]. For these reasons, we became interested in maternal mRNAs as candidates for the cytoplasmic determinants and initiated a cDNA project that collects mRNA tag-sequences and their expression data. From the data derived from our project, we are constructing a database, named MAGEST- Maboya the ascidian, *Halocynthia roretzi* Gene Expression patterns and Expression Sequence Tags- to analyze the data produced in our project. The MAGEST database can be accessed through the WWW (http://www.genome.ad.jp/magest/).

2 CONTENTS OF MAGEST

Currently MAGEST contains two types of data: the 3' and 5' ESTs by DNA sequencing and the gene expression data by whole-mount *in situ* hybridization (WISH). We registered more than 2,000 cDNA clones in the public databases. Each cDNA clone is given a unique genecode consisting of 6 alphanumerical characters. These sequences are used as query sequences for BLAST homology searches against GenBank at the nucleotide level and nr-aa, which is a non-redundant protein sequence database constructed from SWISS-PROT, PIR, PRF and GenPept. Up to ten entries above a given threshold are stored in MAGEST. They can be retrieved from the original databases by the DBGET/LinkDB system [3]. All 3' EST sequences are compared with each other to examine the numbers of redundant genes. Because we use an unnormalized cDNA library, redundant genes may be considered to reflect the population of maternal mRNAs. In addition, we annotate the EC number to a clone encoding an enzyme for linking to the KEGG pathway map. WISH was carried out for staged embryos to obtain information about localization and/or expression sites of the clones during embryogenesis. We adopt three developmental stages: the 8-cell stage, the 110-cell stage and early tailbud (eTb) stage. We classify the expression data according to each blastomere or tissue as shown in Table 1 . This

classification is based on the cell lineage analysis in the ascidian embryo [2]. We provide the original images of WISH in addition to the classification data.

Table 1: Classification scheme to evaluate localization and/or expression patterns.

8-cell	*110-cell*	*Early tail bud*
ISH done	ISH done	ISH done
overall	overall	overall
mitochondria like	mitochondria like	mitochondria like
a4.2	epidermis	epidermis
b4.2	brain	adhesive organ
A4.1	nerve cord	brain
B4.1	notochord	nerve cord
post plasmic RNA	muscle	notochord
	mesenchyme	muscle
	trunk lateral cell	mesenchyme
	trunk ventral cell	endoderm
	endoderm	endodermal strand

3 DATA RETRIEVAL SYSTEM

MAGEST, implemented in the Sybase relational database system, is accessible through the WWW. Its CGI programs are written in the Perl programming language with Sybperl, a Sybase extension module to Perl. We provide several facilities for data retrieval. One can retrieve the data by using keywords or by specifying an entry identifier. A homology search can be performed for the 3' and 5' ESTs in MAGEST. In addition, we provide a unique data retrieval system using our classification of gene expression data derived from WISH.

4 Acknowledgments

This work has been supported by the Grants-in-Aids from the Ministry of Education, Science, Sports and Culture, Japan and the "Research for the Future" Program from the Japan Society for the promotion of Science. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, kyoto University.

References

[1] Satoh, N., *Developmental Biology of Ascdians*, Cambridge University Press, 1994.

[2] Nishida, H., Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Developmental Biology*, 121:526–541, 1987.

[3] Fujibuchi, W., Goto, S., Migimatsu, H., Uchiyama, I., Ogiwara, A. Akiyama, Y., and Kanehisa, M, *Pacific Symposium on Biocomputing '98*, World Scientific, 175–186.