Constructing Expression Pattern Database of Ascidian mRNAs

Shuichi Kawashima

MAGEST is a database for ascidian, *Halocynthia roretzi*, maternal mRNAs that have been newly identified by our cDNA project. We have collected 3' and 5' tag sequences of mRNAs and their expression data from whole mount *in situ* hybridization in early embryo. To date we determined more than 2,000 tag-sequences of *Halocynthia roretzi* cDNAs and deposited in the public databases. The tag sequences and the expression data with additional information can be obtained through MAGEST via the World Wide Web at URL http://www.genome.ad.jp/magest/.

keywords: maternal mRNA / Ascidian / Database / Bioinformatics

Fertilized eggs cleave many times to give rise to multicellular organisms. Within embryos, embryonic blastomeres develop into various types of tissues such as epidermis, muscles and nervous systems. In the processes of early embryogenesis, maternal factors stored in the egg cytoplasms are known to play various significant roles (1). Since the last century, ascidian egg has been well-known as a mosaic egg in which many blastomeres in the early embryo differentiate autonomously. Recent works have revealed that there exist cytoplasmic determinants that direct formation of epidermis, muscle and endoderm as well as cytoplasmic factors involved in axis specification of the embryo and in gastrulation(2,3). In addition, embryonic cell lineage is almost completely revealed by intracellular marking (4). For these reasons, we became interested in maternal mRNAs as candidates for the cytoplasmic determinants and initiated a cDNA project which collects mRNA tag-sequences and their expression data. The information regarding gene expression and localization is important for understanding the outline of gene functions in developmental mechanisms. Thus, we are constructing a database, named MAGEST— Maboya ,the Japanese ascidian, (*Halocynthia roretzi*) Gene Expression patterns and Sequence Tags— for analysis of the data produced in our project.

CONTENTS OF MAGEST

Currently MAGEST contains two types of data: the 3' and 5' ESTs by DNA sequencing and the gene expression

RESEARCH FACILITY OF NUCLEIC ACIDS

Scope of research

The following is the current major activities of this facility.

With emphasis on regulatory mechanisms of gene expression in higher organisms, the research activity has been focused on analysis of signal structures at the regulatory regions of transcriptional initiation and of molecular mechanisms involved in post-transcriptional modification by the use of eukaryotic systems appropriate for analysis. As of December 1994, studies are concentrated on the molecular mechanisms of RNA editing in mitochondria of kinetoplastids

Assoc Prof
SUGISAKI, Hiroyuki
(D Sc)

Instr
KAWASHIMA, Shuichi

Techn
YASUDA, Keiko

data by whole-mount *in situ* hybridization (WISH). Each cDNA clone is given a unique gene code consisting of 6 alphanumerical characters. This gene code reflects our way of handling plasmid DNAs by 384 (16 lines times 24 columns) well plate. For each clone, we remove cloning vector sequences and ambiguous regions that contain stretches of N (ambiguous nucleotide) from raw sequence data and register the processed sequences in the MAGEST database. These sequences are used as query sequences for BLAST homology searches against GenBank at the nuceotide sequence level and also against nr-aa, which is a non-redundant protein sequence database constructed from SWISS-PROT, PIR, PRF and GenPept (trunslated GenBank), at the amino acid sequence level. Up to ten entries above a given threshold are stored in MAGEST and they can be retrieved from the original databases by the DBGET/LinkDB system (5). 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. Based on these search results, we annotate the EC number to a clone coding for an enzyme. Using the EC numbers, these clones are linked to the KEGG pathway map (6).

WISH was carried out for staged embryos to obtain informations 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. At the 8-cell embryo, it is easy to identify the orientation of the embryo along the animal-vegetal and anterior-posterior axes which are utilized to see the distribution of maternal RNAs. In the 110-cell embryo, the developmental cell fate of every blastomere is almost completely restricted. The 110-cell embryo is used to see lineage-specificity of gene expression. In the eTb embryo, the basic body plan is established. Embryo at this stage is used to see tissue-specificity of gene expression. Many genes that are isolated from the cDNA library are zygotically expressed, and the data of both the maternal and/or the zygotic expression pattern of these genes are registered in MAGEST. We classify the expression data according to each blastmere or tissue. We classify the expression data according to each blastomere or tissue. This classification is based on the cell lineage analysis in ascidian embryo (4). A clone not showing a specific expression pattern is classified as "overall" when it is expressed over the entire embryo, or only "ISH_done" when no signal is detected. We provide the original images of WISH in addition to the classification data.

DATA RETRIEVAL SYSTEM

MAGEST is 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 similar search can be performed against 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.

FUTURE DIRECTIONS

In this project, we aim at all-inclusive and systematic description of maternal transcripts of Japanese ascidian fertilized eggs, *H. roretzi*: cDNA sequences from ca. 10,000 different genes and their expression patterns during embryogenesis. Currently we registered more than 2,000 cDNA clones in the public databases. From a survey of these data, we found several genes that may play important roles during early development. These clones include key elements for development such as some transcription factors, signal transducing molecules, or RNA binding proteins (7, 8, 9). Finally MAGEST will also enable us to understand molecular mechanisms for establishment of embryonic body plans of chordates and more generally the evolution from invertebrate to vertebrate.

REFERENCES

1 Satoh, N., (1994) Developmental Biology of Ascdians, CAMBRIDGE UNIVERSITY PRESS]

2 Conklin, E.G. (1905) *Biol.Bull*., 8, 205-230.

3 Nishida, H. (1997) *Int.Rev. Cytol.*, 176, 245-306.

4 Nishida, H. (1987) *Dev.Biol*.,121,526-541.

5 Fujibuchi, W., Goto, S., Migimatsu, H., Uchiyama, I., Ogiwara, A. Akiyama, Y. and Kanehisa, M. (1998) *Pacific Symposium on Biocomputing '98* (Altman, R. B., Dunker, A. K., Hunter, L. and Klein, T. E., eds.), World Scientific, 175-186.

6 Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono,H., and Kanehisa, M. (1999) *Nucleic Acid Res*., 27, 29-34

7 Sasakura, Y., Ogasawara, M., and Makabe, K.W., (1998) *Int.J.Dev.Biol*. 42, 573-579

8 Sasakura,Y.,Ogasawara,M.,and Makabe,K.W.,(1998) *Mech.Dev*.,76,161-163

9 Kobayashi, A., Sasakura, Y., Ogasawara, M., and Makabe,K.W.,(1999) *Develop. Growth. Differ*. 41, 419-427