 Nucleic Acids Research, 2000, Vol. 28, No. 1 **133–135**

MAGEST: MAboya Gene Expression patterns and Sequence Tags

Takeshi Kawashima, Shuichi Kawashima[1], Minoru Kanehisa[1], Hiroki Nishida[2] and Kazuhiro W. Makabe*

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

Received August 31, 1999; Revised and Accepted October 22, 1999

ABSTRACT

MAGEST is a database for newly identified maternal cDNAs of the ascidian, *Halocynthia roretzi*, which aims to examine the population of the mRNAs. We have collected 3′ and 5′ tag sequences of mRNAs and their expression data from whole-mount *in situ* hybridization in early embryos. To date, we have determined more than 2000 tag-sequences of *H.roretzi* cDNAs and input them into public databases. The tag sequences and the expression data as well as additional information can be obtained through MAGEST via the WWW at http://www.genome.ad.jp/magest/

INTRODUCTION

In the phylum Chordata, genomic duplication events occurred twice in the process of vertebrate evolution. Ascidian, which is lower chordate, however, has a non-duplicated genome that can be regarded as a basic set of chordate-type genome (1). This suggests that ascidian is a good model system to investigate the functions of chordate genomes. 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 process of early embryogenesis, maternal factors stored in egg cytoplasms are known to play various significant roles (2). Since the last century, the ascidian egg has been well-known as a mosaic egg in which many blastomeres in the early embryo differentiate autonomously. Recent works have revealed 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 (3,4). In addition, the embryonic cell lineage is almost completely revealed by intracellular marking (5). For these reasons, we became interested in maternal mRNAs as candidates for cytoplasmic determinants and initiated a cDNA project that collects mRNA tag-sequences and their expression data. The ESTs determined in this work were obtained from eight different plates of a library made from fertilized ascidian eggs. cDNA libraries were constructed in the STRATAGENE pBluescript vector following the manufacturer's instructions. The libraries were arrayed in 384-well plates in a Genetix Q-Bot robot. The information regarding gene expression and localization is important for understanding the outline of gene functions in developmental mechanisms. Indeed, there are precedents of similar research and related databases that have been applied to several organisms (6,7). Thus, 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.

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). Each cDNA clone is given a unique gene code consisting of six alphanumerical characters. This gene code reflects our way of handling plasmid DNAs using a 384 (16 lines × 24 columns) well plate. For example, clone 001B03 represents line B and column 03 in plate No.001. 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 (8) at the nucleotide sequence level and also against nr-aa, a non-redundant protein sequence database constructed from SWISS-PROT (9), PIR (10), PRF (11) and GenPept (12) (translated GenBank), at the amino acid sequence level. Up to 10 entries above a given threshold are stored in MAGEST. They can be retrieved from the original databases by the DBGET/LinkDB system (13). 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 each clone. In addition, we annotate the EC number to a clone encoding an enzyme. Using the EC numbers, these clones are linked to the KEGG pathway map (14). The list of contents in MAGEST is shown in Table 1.

WISH is 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. At the 8-cell embryo stage, it is easy to identify the orientation of the

*To whom correspondence should be addressed. Tel: +81 75 753 4102; Fax: +81 75 705 1113; Email: kwmakabe@ascidian.zool.kyoto-u.ac.jp

Table 1. List of contents in MAGEST

Data Field	Content
Information from ESTs	
Gene code	ex. 001A03
Gene name	ex. HrWnt-5
Description	Brief description of the sequence
Overlap gene	Redundant gene identifiers
EC number	
3′ EST	3′ tag sequence
5′ EST	5′ tag sequence
GenBank	Results of a similar search against GenBank
dbEST	Results of a similar search against dbEST
NRAA	Results of a similar search against nr-aa
Information from WISH	
8-cell	Evaluation according to our classification of expression patterns in the 8-cell stage embryo
110-cell	Evaluation according to our classification of expression patterns in the 110-cell stage embryo
eTb	Evaluation according to our classification of localization patterns in the eTb stage embryo
image	Images from WISH

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 the lineage-specificity of gene expression. In the eTb embryo, the basic body plan is established. The embryo at this stage is used to see the tissue specificity of gene expression. Many genes that are isolated from the cDNA library are also 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 blastomere or tissue as shown in Table 2. This classification is based on the cell lineage analysis in the ascidian embryo (5). 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.

Table 2. Classification scheme to evaluate the localization and/or expression patterns, which is based on the cell lineage analysis in ascidian embryo (5)

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
	post plasmic RNA	post plasmic RNA

post plasmic RNA, expressed in the posterior cytoplasms in a pair of posterior-vegetal cells; mitochondria_like, similar to the expression pattern of mitchondrial 16S ribosomal RNA; overall, expressed throughout the embryo; ISH_done, WISH was carried out.

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 provided 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 for the 3′ and 5′ ESTs in MAGEST. In addition, we provided a unique data retrieval system using our classification of gene expression data derived from WISH. One can search the clone using an arbitrary combination of expression patterns, for example, ‘Which clones are expressed in the brain and nerve cord in the eTb stage?’ or ‘Are there clones expressed in a4.2 in the 8-cell embryo but unexpressed in eTb embryo?’.

FUTURE DIRECTIONS

In this project, we aim at an all-inclusive and systematic description of maternal transcripts of fertilized Japanese ascidian

eggs, *H.roretzi*: cDNA sequences from ~10 000 different genes and their expression patterns during embryogenesis. Currently, we have registered more than 2000 cDNA clones in the public databases. In maternal cDNA clones, which are registered in MAGEST, we expect many genes that encode protein to be involved in the intrinsic genetic programs triggered by cytoplasmic determinants or gene regulatory cascades, which play important roles in early development. We are going to approach such molecules from the following angles. First, we will collect homologs of signal transducing components that may function in cleavage stages from MAGEST and investigate the molecular signaling cascades (15). Second, from the WISH results, we will direct our attention to the clones that have a positive signal in specific embryonic regions, which may control the developmental fates of the regions. We have already found a number of clones that are sequestered in the posterior-vegetal region of cleaving embryos. Now, we are studying the localization mechanisms and the function of these clones (16,17). Finally, we will analyze some clones deduced to encode DNA binding motifs. These clones may function as transcription factors that regulate the very early zygotic transcription of the embryos.

MAGEST will also enable us to understand molecular mechanisms to establish embryonic body plans of chordates and evolution from invertebrate to vertebrate.

ACKNOWLEDGEMENTS

This work has been supported by Grant-in-Aids from the Ministry of Education, Science, Sports and Culture, Japan (10168213 and 11149212) to K.W.M. and the 'Research for the Future' Program from the Japan Society for the promotion of Science (96L00404) to H.N. and K.W.M. The computational resources are given by the Super Computer Laboratory of the Institute for Chemical Research, Kyoto University.

REFERENCES

1. Holland,P.W.H., Garcia-Fernandez,J., Williams,N.A. and Sidow,A. (1994) *Development* Suppl., 125–133.
2. Satoh,N. (1994) *Developmental Biology of Ascdians.* Cambridge University Press, Cambridge.
3. Conklin,E.G. (1905) *Biol. Bull.*, **8**, 205–230.
4. Nishida,H. (1997) *Int. Rev. Cytol.*, **176**, 245–306.
5. Nishida,H. (1987) *Dev. Biol.*, **121**, 526–541.
6. Tadasu,S. and Kohara,Y. (1998) *Genome Informatics*, **9**, 224–225.
7. Gawantka,V., Pollet,N., Delius,H., Vingron,M., Pfister,R., Nitsch,R., Blumenstock,C. and Niehrs,C. (1998) *Mech. Dev.*, **71**, 151–163.
8. Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J., Ouellette,B.F.F., Rapp,B.A. and Wheeler,D.L. (1999) *Nucleic Acids Res.*, **27**, 12–17. Updated article in this issue: *Nucleic Acids Res*. (2000), **28**, 15–18.
9. Bairoch,A. and Apweiler,R. (1999) *Nucleic Acids Res.*, **27**, 49–54. Updated article in this issue: *Nucleic Acids Res*. (2000), **28**, 45–48.
10. Barker,W.C., Garavelli,J.S., McGarvey,P.B., Marzec,C.R., Orcutt,B.C., Srinivasarao,G.Y., Yeh,L.L., Ledley,R.S., Mewes,H., Pferffer,F., Tsugita,A. and Wu,C. (1999) *Nucleic Acid Res.*, **27**, 39–43.
11. Protein Research Foundation, Osaka, Japan. http://prf.or.jp/
12. Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J. and Ouellette,B.F.F. (1998) *Nucleic Acids Res.*, **26**, 1–7. Updated article in this issue: *Nucleic Acids Res*. (2000), **28**, 15–18.
13. Fujibuchi,W., Goto,S., Migimatsu,H., Uchiyama,I., Ogiwara,A. Akiyama,Y. and Kanehisa,M. (1998) In Altman,R.B., Dunker,A.K., Hunter,L. and Klein,T.E. (eds), *Pacific Symposium on Biocomputing '98.* World Scientific, Singapore, pp. 175–186.
14. Ogata,H., Goto,S., Sato,K., Fujibuchi,W., Bono,H. and Kanehisa,M. (1999) *Nucleic Acids Res.*, **27**, 29–34. Updated article in this issue: *Nucleic Acids Res*. (2000), **28**, 27–30.
15. Kobayashi,A., Sasakura,Y., Ogasawara,M. and Makabe,K.W. (1999) *Dev. Growth Differ*. **41**, 419–427.
16. Sasakura,Y., Ogasawara,M. and Makabe,K.W. (1998) *Int. J. Dev. Biol.*, **42**, 573–579.
17. Sasakura,Y., Ogasawara,M. and Makabe,K.W. (1998) *Mech. Dev.*, **76**, 161–163.