

Update of MAGEST: Maboya Gene Expression patterns and Sequence Tags

Takeshi Kawashima, Shuichi Kawashima¹, Yuji Kohara², Minoru Kanehisa¹ and Kazuhiro W. Makabe*

Department of Zoology, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan, ¹Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan and ²Japan Science and Technology Corporation, Genome Biology Laboratory, National Institute of Genetics, Mishima 411-8540, Japan

Received September 20, 2001; Revised and Accepted October 15, 2001

ABSTRACT

MAGEST is a database for maternal gene expression information for an ascidian, *Halocynthia roretzi*. The ascidian has become an animal model in developmental biological research because it shows a simple developmental process, and belongs to one of the chordate groups. Various data are deposited into the MAGEST database, e.g. the 3'- and 5'-tag sequences from the fertilized egg cDNA library, the results of similarity searches against GenBank and the expression data from whole mount *in situ* hybridization. Over the last 2 years, the data retrieval systems have been improved in several aspects, and the tag sequence entries have increased to over 20 000 clones. Additionally, we constructed a database, translated MAGEST, for the amino acid fragment sequences predicted from the EST data sets. Using this information comprehensively, we should obtain new information on gene functions. The MAGEST database is accessible via the Internet at <http://www.genome.ad.jp/magest/>.

INTRODUCTION

In the early development of many animals, maternal cytoplasmic factors are known to have various essential roles. Ascidian is a lower chordate, a group that includes humans, and is a good model system to investigate the characteristics of maternal factors because its developmental processes are shown to be largely dependent on the maternal factors (1,2), and its genome size is small (3). We are interested in the maternal mRNAs as candidates for cytoplasmic determinants, and thus initiated a cDNA project to collect mRNA expressed sequence tags (ESTs) and their localization or expression data. Thus, we are constructing a database, named MAGEST (Maboya Gene Expression patterns and Sequence Tags) to analyze the data gathered by our project. Here we present an update of the MAGEST database (4,5). We have improved the data retrieval systems, and have started work on the translated

MAGEST, which is a database for predicting amino acid fragment sequences.

CONTENTS OF MAGEST

Basic data update

To date, we have determined the tag sequences for more than 20 000 cDNA clones, and investigated the localization or expression patterns for more than 2000 genes of *Halocynthia roretzi* cDNAs. All of this information is deposited in public databases.

Translated MAGEST

Many computer programs for the prediction of gene function or protein structure require the amino acid sequence, not the nucleotide sequence. In some cases, the ESTs, which are nucleotide sequence fragments, may encode part of an amino acid sequence. Translated MAGEST is a database for the fragments of amino acid sequences that are translated from MAGEST EST entries. Each MAGEST EST entry undergoes a BLASTX similarity search against the SWISS-PROT protein database (6). We obtained the translated sequence to parse the result file of BLASTX. To date, about 3000 entries are registered in the Translated MAGEST.

Data retrieval system

MAGEST is implemented in the Sybase relational database management system. Access to data is provided primarily via World Wide Web-based query forms at <http://www.genome.ad.jp/magest/>. MAGEST has comprehensive links to other sequence resources using the DBGET/LinkDB database retrieval system (7). Using the MAGEST database search query form, one can rapidly identify the expressed tag sequences and the gene expression patterns from each cluster ID or clone ID, and select a distinct data set from keywords that include the predicted functions and motifs. In addition, one can execute sequence similarity searches for MAGEST entries using the BLASTN or the TBLASTX sequence similarity search programs (8). Using the MAGEST expression pattern search query form, one can locate the group of genes that share the same expression patterns. Using the MAGEST advanced

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

search query form, one can clarify the cluster size of each gene, and obtain a description of predicted functions.

FUTURE DIRECTIONS

The functions of ~50% of clones are unknown when only sequence similarity searches are carried out. To proceed with analyses of the functions of these genes, we attempt to translate the EST nucleotide sequences into the encoded amino acid fragment sequences using this database. For analyses of these gene structures, we used the SOSUI program (Classification and Secondary Structure Prediction System for Membrane Proteins) (9) against the Translated MAGEST database. SOSUI predicted that 1% of the cloned entries are membrane proteins. Recently, genome analyses of ascidians including EST analyses have been undertaken on a global scale. Comparison of MAGEST data with the data obtained from various projects should shed some light on the gene networks that control early development in chordates.

ACKNOWLEDGEMENTS

We thank Prof. Nori Satoh at Kyoto University for generous use of the facilities, and Dr Susumu Goto at Kyoto University for support with database management. This work has been supported by Grants-in-Aid from the Ministry of Education, Science, Sports and Culture, Japan (12680714, 13045018 and 13202024) to K.W.M. and the 'Research for the Future'

Program from the Japan Society for the promotion of Science (96L00404) to K.W.M. The computational resources are from the Super Computer Laboratory of the Institute for the Chemical Research, Kyoto University.

REFERENCES

1. Conklin, E.G. (1905) Mosaic development in ascidian eggs. *J. Exp. Zool.*, **2**, 145–223.
2. Nishida, H. (1997) Cell fate specification by localized cytoplasmic determinants and cell interactions in ascidian embryos. *Int. Rev. Cytol.*, **176**, 245–306.
3. Simmen, M.W., Leitgeb, S., Clark, V.H., Jones, S.J. and Bird, A. (1998) Gene number in an invertebrate chordate, *Ciona intestinalis*. *Proc. Natl Acad. Sci. USA*, **95**, 4437–4440.
4. Kawashima, T., Kawashima, S., Kanehisa, M., Nishida, H. and Makabe, K.W. (2000) MAGEST: MAboya Gene Expression patterns and Sequence Tags. *Nucleic Acids Res.*, **28**, 133–135.
5. Makabe, K.W., Kawashima, T., Kawashima, S., Minokawa, T., Adachi, A., Kawamura, H., Ishikawa, H., Yasuda, R., Yamamoto, H., Kondoh, K. *et al.* (2001) Large-scale cDNA analysis of the maternal genetic information in the egg of *Halocynthia roretzi* for a gene expression catalog of ascidian development. *Development*, **128**, 2555–2567.
6. Bairoch, A. and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.
7. Fujibuchi, W., Goto, S., Migimatsu, H., Uchiyama, I., Ogiwara, A., Akiyama, Y. and Kanehisa, M. (1998) DBGET/LinkDB: an integrated database retrieval system. *Pac. Symp. Biocomput.*, **98**, 683–694.
8. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
9. Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, **14**, 378–379.