NEXTDB: The Expression Pattern Map Database for *C. elegans*

Tadasu Shin-i **Yuji Kohara**
tshini@genes.nig.ac.jp ykohara@lab.nig.ac.jp
CREST, JST and Genome Biology Laboratory, National Institute of Genetics
1,111 Yata Mishima 411-8540, Japan

1 Introduction

The nematode *Caenorhabditis elegans* (*C. elegans*) is a good model system to study functional genomics with respect to animal development, nervous system and behavior at the level of single cells. Although *C. elegans* has the basic structure of animals, it has only about 1,000 somatic cells. This simplicity has led to the description of entire cell lineage from embryo to adult, which has allowed us to study gene functions in individual cells [5]. The genome consists of six chromosomes whose total size is about 100 Mbp and total number of genes is estimated to be about 19,000. All the genome has been sequenced by the consortium of the Sanger Centre and Washington University by the end of 1998 [1].

In this laboratory, the systematic analysis of cDNA clones of *C. elegans* with respect to tag-sequences, map positions, pattern of expression during development and gene functions, has been carried out. To integrate all the information, we have developed a WWW-based database, named NEXTDB. In this report we describe an overview of the current version of the database NEXTDB and future plans to develop new functions of the database.

2 Materials and methods

2.1 Processing of ESTs

Raw data of one-pass sequencing were transferred to SUN Ultra30 workstation and processed automatically to classify the sequences into unique cDNA groups by comparing 3'-tag sequences [4]. Clones whose 3' tags were not qualified by the sequence clean-up process [4] were classified by mapping positions of their 5' tags to cosmid sequences by use of BLASTN. The confirmation of positions of the cDNA groups on the genome was also done by BLASTN.

2.2 Processing of expression patterns

Images of whole mount in situ hybridization [6] were loaded to NEXTDB and then annotated them with respect to developmental stages [4]. Images of immunostaining taken on Zeiss LSM510 confocal microscope [3] and the data of RNAi (RNA mediated interference) phenotypes [2] were also stored and arranged properly in the database. Operators added the information of expression patterns to the images by use of WWW-based interfaces.

2.3 Links to the genome map

In order to link NEXTDB with the genome map, we applied a hierarchical model to arrange all the clones and clusters; 1) chromosome, 2) cosmid clone, 3) CDS, 4) cDNA group, and 5) cDNA clone. The cosmid map data which connects 1) and 2) were obtained from the *C. elegans* genome database

AceDB which describe the relationships among cosmid clones, predicted genes or CDS and genetically defined genes. The information about cosmids and their CDS was retrieved from the annotations of the Sanger Centre cosmid sequence data and WormPep database. All the information, such as sequences, homologies of the probe cDNA, in situ images, immunostaining images, and RNAi phenotypes, are arranged based on the corresponding cDNA clones. They are integrated with the genome map based on WWW, and the information of maps and their relations are depicted visually by use of JAVA applets.

3 Result and discussion

Now tag sequences of about 56,000 cDNA clones have been incorporated into NEXTDB. The database classifies the clones into about 10,000 unique cDNA groups, which correspond to a half of the total gene number. About 7,500 groups hit to the predicted CDSs in the genomic sequences, and about 2,500 hit to the non CDS region. Comparing some ESTs and the corresponding genomic sequences has revealed the presence of alternative splicings and differential termination, and sometimes bridging cosmids at gaps of cosmid contigs. Therefore, close comparison of ESTs and the genomic sequences is very important to identify genes precisely. NEXTDB incorporates in situ images of about 3,700 cDNA groups. The latest version is available through the following URL.

http://watson.genes.nig.ac.jp:8080/db/index.html

It can also be accessed through the home page of DDBJ (http://www.ddbj.nig.ac.jp). Users can retrieve information by chromosomes, cosmid, gene names, clones, expression patterns, or sequences as similarity search queries. Visual description of relationships among cosmids, predicted genes, and clusters of cDNA clones can be seen, and expression pattern images which are arranged along the genome map are retrievable.

Acknowledgments

This work has been supported by CREST of JST (Japan Science and Technology Corporation) and by a Grant-in-aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan.

References

[1] The *C. elegans* Sequencing Consortium, Genome sequence of the nematode *C. elegans*: a platform for investigating biology, *Science*, 282:2012–2017, 1998

[2] Hirono, K., Onami, S., and Kohara, Y., Systematic RNAi experiments with maternal genes, *12th International C.elegans Meeting*, 393, 1999.

[3] Onami, S., Nagaoka, T., and Kohara, Y., Protein expression pattern analysis of maternal mRNAs, *12th International C.elegans Meeting*, 650, 1999.

[4] Shin-i, T. and Kohara, Y., NEXTDB: the expression pattern map database for *C. elegans*, *Genome Informatics 1998*, 224–225, 1998.

[5] Sulston, J.E., Schierenberg, E., White, E.J., and Thomson, J.N., The embryonic cell lineage of the nematode *C. elegans*, *Dev. Biol.*, 100:64–119, 1983.

[6] Tabara, H., Motohashi, T., and Kohara, Y., A multi-well version of *in situ* hybridization on whole mount embryos of *Caenorhabditis elegans*, *Nucl. Acids Res.*, 24:2119–2124, 1996.