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 [1]. The genome consists of six chromosomes whose total size is about 100 Mbp and total number of genes is estimated to be about 15,000. All the genome will be sequenced by the consortium of the Sanger Centre and Washington University by the end of 1998 citews95.

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, name 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 Hardware

For data acquisition, processing and displaying, the following machines are used. ABI377 DNA sequencers connected with Power Mac computers. Zeiss AxioPlan photomicroscopes equipped with CCD cameras that are connected with Power Mac computers. Sun Ultra30 workstation, in which NEXTDB is created and operated, equipped currently altogether GB hard drives. All the computers are connected by the Ethernet.

2.2 Processing of ESTs

Raw data of one-pass sequencing were produced and stored in Macintosh computer connected with ABI377 DNA sequencers. The ABI chromatogram files and sequence files were transferred to NEXTDB in the SUN workstation. NEXTDB automatically removes ambiguous regions which contain larger rate of "N" (ambiguous base) than the threshold and cloning vector sequences. The processed 3'-tag sequences were classified into unique cDNA groups applying a cumulative method using FASTA, since the 3'-tags should be unique among the cDNAs derived from the same gene. Both 5'- and 3'-tag sequences were mapped to cosmid sequences by use of BLASTN. Then the results were compared to the cDNA groups to confirm the positions of the groups on the genome.

2.3 Processing of expression patterns

Images of whole mount in situ hybridization [2] are taken by CCD cameras equipped on the microscopes. The image data are processed (selection of good images, trimming of the images to reduce file size, etc.) on Macintosh computers and transferred to NEXTDB. Operators annotate individual images with respect to developmental stages and expression patterns using a Web browser. Currently, the images for embryogenesis are classified in 10 stages. The images for post-embryogenesis are classified in 4 stages. Once annotated, the images are arranged properly along development on the screen.

2.4 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 describes 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 sequence data. Sequences and homologies of the probe cDNA and all the in-situ images were arranged by making links to corresponding cDNA clones. All the data are integrated based on WWW, and the information of maps and their relations are depicted visually by use of JAVA applets.

3 Result and discussion

As of early October 1998, tag sequences of about 47,000 cDNA clones have been incorporated into NEXTDB. The database classifies the clones into about 10,000 unique cDNA groups, which correspond to two thirds of the total gene number. About 5,500 groups hit to the predicted CDSs in the genomic sequences, and about 3,500 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 2,500 cDNA groups mainly from chromosome 3 and chromosome X. The latest version is available through the URL "http://watson.genes.nig.ac.jp:8080/db/index.html." 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.

Acknowledgements

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] 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.

[2] 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.

[3] Waterston, R. and Sulston, J., The genome of *Caenorhabditis elegans*, *Proc. Natl. Acad. Sci. USA*, 92:10836–10840, 1995.