Axeldb: a *Xenopus laevis* database focusing on gene expression

Nicolas Pollet*, Heiko A. Schmidt[1], Volker Gawantka, Martin Vingron[1] and Christof Niehrs

Department of Molecular Embryology and [1]Department of Theoretical Bioinformatics, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

ABSTRACT

Axeldb is a database storing and integrating gene expression patterns and DNA sequences identified in a large-scale *in situ* hybridization study in *Xenopus laevis* embryos. The data are organised in a format appropriate for comprehensive analysis, and enable comparison of images of expression pattern for any given set of genes. Information on literature, cDNA clones and their availability, nucleotide sequences, expression pattern and accompanying pictures are available. Current developments are aimed toward the interconnection with other databases and the integration of data from the literature. Axeldb is implemented using an ACEDB database system, and available through the web at http://www.dkfz-heidelberg.de/abt0135/axeldb.htm

INTRODUCTION

The description and analysis of gene expression patterns is crucial to elucidate the physiological functions of genes and to understand the network of genetic interactions that underly the process of normal development. Knowing the patterns of expression of large numbers of genes allows the identification of specific promoters, marker genes useful to monitor cells in a specific state and genes which are tightly co-regulated.

Currently, there are gene expression databases available for the worm *Caenorhabditis elegans* (1), the fly *Drosophila melanogaster* (2) and the mouse (3), but one is lacking for *Xenopus laevis* although it has been a major model system in embryology for some time. Some laboratories are engaged in large-scale *in situ* hybridization screening (4). Due to these and other non-systematic efforts we expect a dramatic increase in the number of characterised gene expression patterns in the future. The development of specialised databases will therefore be imperative to manage and interpret this biological information.

We carried out a systematic study of gene expression during the early development of the frog *X.laevis* where we described nucleotide sequences and expression patterns of 273 unique genes at three stages of development (5). In parallel, we developed a database to compile and integrate our results in a comprehensive manner. This database, called Axeldb (a *Xenopus laevis* database) should provide a major information resource to identify genes according to their expression, compare gene expression patterns and facilitate the elucidation of gene interactions during development. Since we have made the corresponding cDNA clones publicly available, Axeldb is also expected to provide the relevant important practical information to worldwide clone users. Other information will be made available as our project evolves to provide a convenient global resource for biologists using the frog as a model system. For those outside the field of *X.laevis*, the database would provide expression patterns of homologous genes from other species.

Axeldb DESIGN AND IMPLEMENTATION

Axeldb was implemented using ACEDB (A *Caenorhabditis elegans* database) software designed by Richard Durbin (Sanger Centre, UK) and Jean Thierry-Mieg (CNRS, France). The documentation, code and data of ACEDB are available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov.

ACEDB version 4.5e for UNIX was used along with Perl 5.004, CGI.pm 2.42, and tailored AcePerl 1.54 and AceBrowser 2.01 (6). The database is accessible on the World Wide Web at http://www.dkfz-heidelberg.de/abt0135/axeldb.htm

We chose the ACEDB software because of its flexibility, simplicity of set-up and usefulness in modelling biological objects as well as for its user interface which is both user-friendly and graphical. Moreover, a powerful Perl interface is available (6). While the original ACEDB database was developed to meet the needs of the worm genome project, it is now in use for many other genome projects and to compile all kinds of biological information. Axeldb, being an ACEDB database, is object-oriented, meaning that the data are divided into classes of objects such as gene, sequence, clone, expression pattern, picture and within each class there can be several objects each having a unique name.

The basic function of Axeldb is to link expression pattern information to pictures, cDNA clones, genes and sequences. Further information concerning genes, such as their usefulness as differentiation markers, their functional category based on sequence information and their appurtenance to a synexpression group (see below) is integrated. This allows search and display of all information available for any given gene.

All information for Axeldb is initially in the form of ASCII flat files, and it is converted into the ace format files required by the database using shell and Perl scripts. In case of problems experienced by users concerning the database or the quality of

*To whom correspondence should be addressed. Tel: +49 6221 424694; Fax: +49 6221 424692; Email: n.pollet@dkfz-heidelberg.de

the information, a feedback form is available to send reports or data corrections directly to the curator.

BIOLOGICAL PROBLEMS AND DATABASE COMPROMISES

When having to compile gene expression information in a comprehensive and curated way one has to decide on the kind of information to track. One important issue is the use of controlled nomenclature, since it enables efficient searches and global analysis. With this in mind, we decided to describe expression patterns using a comprehensive list of 17 anatomical terms for the tailbud stage embryos, and a short textual description for all stages studied. A score between 0 (no expression detected) and 4 (very strong expression detected) was attributed for each structure. We described a class entitled synexpression groups to store the information relevant to sets of genes tightly co-expressed. Another class called expression_domain was described to represent the territory of expression of genes considered as differentiation markers. We incorporated modifications in the original description of object classes in ACEDB to meet our special needs, and the resultant models.wrm file is available at http://www.dkfz-heidelberg.de/abt0135/axeldb/misc.html

QUERY INTERFACES

The combination of AcePerl and AceBrowser interfaces allows complex queries to be passed on the Axeldb server transparently for the user, and to display information in a comprehensive way.

Axeldb can be browsed and searched in various ways:

Browsing

All objects of Axeldb can be accessed by browsing through the classes of objects. Alternatively, browsing can start based on the domains of expression. In one option the criteria of expression are defined using a self-explained form, in the second option, a list of expression domains is proposed. Finally, browsing can be picture-oriented. In this case the selection of pictures is made according to either genes, expression domains, gene class or life stages. All corresponding thumbnail pictures and links to gene information are returned.

Searching

Various search forms enable the selection of genes, clones, pictures, sequences or synexpression group according to their name in Axeldb. The database can be searched using text keywords or ACEDB query language for more complex queries. A sequence similarity search allows the comparison of DNA or protein sequence with those present in Axeldb using BLAST or FASTA.

Documentation of search forms and examples of displays are available at http://www.dkfz-heidelberg.de/tbi/axeldb_images/docs/help.html

DATA

Currently, most of the information available in Axeldb comes from studies performed in our laboratory. Some external data were extracted from other databases, such as MEDLINE for literature references and the *Xenopus* Molecular Marker Resource for *Xenopus* life stages pictures (7) (http://vize222.zo.utexas.edu/) and EMBL for nucleotide sequences. Since our large scale gene expression study combined mRNA *in situ* hybridization and cDNA sequencing, we had to cope with sequence data, gene expression patterns and picture documentation. At present, comprehensive and curated information on gene names, sequences, expression patterns and cDNA clones for 273 genes is available. Moreover, Axeldb contains all nucleotide sequences from *Xenopus* from the most recent EMBL release.

FUTURE DIRECTIONS

While the vast majority of the expression data in Axeldb comes from the screen performed in our laboratory, we anticipate much more detailed information on the expression of specific genes. We would like to see these also included in Axeldb to present to the *Xenopus* community a comprehensive gene expression database. Expression pattern data or comments can be submitted to Axeldb through feedback or submission forms. The submission form incorporates a list of controlled vocabulary terms for expression pattern description. A brief text description and the associated pictures is deemed sufficiently detailed to encourage submissions and brief enough not to discourage potential contributors. We believe that computer publication of gene expression data has many advantages since information can be updated or discussed by the experts in the field.

We would like to link the data of expression patterns concerning known orthologous genes. As a short term goal, the next releases of Axeldb will contain more gene expression information as our large-scale study continues. As a mid to long term goal Axeldb can serve as a resource for the *Xenopus* field, acquiring, managing and releasing data about the biology of the frog.

REFERENCES

1. Martinelli,S.D., Brown,C.G. and Durbin,R.S. (1997) *Semin. Cell Dev. Biol.*, **8**, 459–467.
2. Janning,W. (1997) *Semin. Cell Dev. Biol.*, **8**, 469–475.
3. Ringwald,M., Mangan,M.E., Eppig,J.T., Kadin,J.A., Richardson,J.E. and the Gene Expression Database Group (1999) *Nucleic Acids Res.*, **27**, 106–112. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 115–119.
4. Niehrs,C. (1997) *Genes Funct.*, **1**, 229–231.
5. Gawantka,V., Pollet,N., Delius,H., Vingron,M., Pfister,R., Nitsch,R., Blumenstock,C. and Niehrs,C. (1998) *Mech. Dev.*, **77**, 95–141.
6. Stein,L. and Thierry-Mieg,J. (1999) *Genome Res.*, **8**, 1308–1315.
7. Nieuwkoop,P.D. and Faber,J. (1994) *Normal Table of Xenopus laevis* (Daudin). Garland, New York