

Xenbase: a *Xenopus* biology and genomics resource

Jeff B. Bowes¹, Kevin A. Snyder¹, Erik Segerdell², Ross Gibb¹, Chris Jarabek¹, Etienne Noumen¹, Nicolas Pollet³ and Peter D. Vize^{1,2,*}

¹Department of Computer Science, ²Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada and ³Laboratoire Développement et Evolution, CNRS UMR 8080, Université Paris-Sud, Orsay 91405, France

Received August 13, 2007; Revised September 18, 2007; Accepted September 20, 2007

ABSTRACT

Xenbase (www.xenbase.org) is a model organism database integrating a diverse array of biological and genomic data on the frogs, *Xenopus laevis* and *Xenopus (Silurana) tropicalis*. Data is collected from other databases, high-throughput screens and the scientific literature and integrated into a number of database modules covering subjects such as community, literature, gene and genomic analysis. Gene pages are automatically assembled from data piped from the Entrez Gene, Gurdon Institute, JGI, Metazome, MGI, OMIM, PubMed, Unigene, Zfin, commercial suppliers and others. These data are then supplemented with in-house annotation. Xenbase has implemented the Gbrowse genome browser and also provides a BLAST service that allows users to specifically search either *laevis* or *tropicalis* DNA or protein targets. A table of *Xenopus* gene synonyms has been implemented and allows the genome, genes, publications and high-throughput gene expression data to be seamlessly integrated with other *Xenopus* data and to external database resources, making the wealth of developmental and functional data from the frog available to the broader research community.

INTRODUCTION

Over the past decade the *Xenopus* experimental system has been supported by static community web sites such as the *Xenopus* Molecular Marker Resource and by specialized databases such as Axeldb (1) and XDB3 (N. Ueno, Personal communication). The latter databases contain massive amounts of *in situ* gene expression data but there is often little or no annotation and there is limited integration with other data types or external resources. Xenbase (<http://www.xenbase.org>) was designed to greatly broaden the range of data stored and to integrate this information both internally between different types of

information, e.g. the genome, gene function, gene expression and the literature, and externally to resources such as human congenital disease data from OMIM and mutant data from the mouse and the zebrafish.

Xenopus has a number of unique experimental advantages as a vertebrate model system (2). Paramount among these is the robustness of early embryos and their amenability to microinjection and microsurgery. This makes them a particularly attractive system for testing the ectopic activity of gene products and loss-of-function experiments using antagonizing reagents such as morpholinos (3), dominant-negatives and neomorphic proteins (4). Morpholinos are synthetic oligonucleotides (3) that can be used to inhibit hnRNA splicing or mRNA translation and are the common gene inhibition reagent in *Xenopus* as neither siRNA or miRNA have yet been shown to function in frog embryos. *Xenopus* embryos develop very quickly and form a full set of differentiated tissues within days of fertilization, allowing rapid analysis of the effects of manipulating embryonic gene expression. There have also been a number of high-throughput screens of gene expression patterns by wholemount *in situ* hybridization in *Xenopus* (5). Not only are these data critical to ongoing research efforts in *Xenopus*, they also serve researchers working in other systems by providing gene expression and function data not available in their own model.

Building a *Xenopus* database offered a variety of data handling challenges. In particular, seamlessly combining data from two species, *Xenopus laevis*—a tetraploid, and *Xenopus tropicalis*—a diploid (2), in an integrated environment is complex. There is a wealth of data in the literature on gene function, full-length cDNA sequences and *in situ* databases on *laevis*, but no genome sequence or genetics. *Xenopus tropicalis* on the other hand has a sequenced genome, genetics and strong EST support but little literature or gene expression data. Integrating data from both organisms in a common environment will allow these different strengths to complement each other. As *laevis* has larger embryos that are better suited to microsurgery and *tropicalis* has the advantages of a diploid in mutant and experimental gene

*To whom correspondence should be addressed. Tel: +403 220 8502; Fax: +403 289 9311; Email: pvize@ucalgary.ca

knockdown experiments, both systems will continue to be used in parallel and need ongoing support. A challenge that is common to all model organism databases is finding a balance between automation and the manual curation of data, yet another is to use tools and lessons from other model organism databases without inheriting flaws and limitations. The public release of Xenbase version 2.0 is a solid step towards achieving these various goals. We utilize elements of the generic model organism database (GMOD) schema CHADO (6), the GMOD genome browser Gbrowse (7), and elements of zfin (8) and MGI (9) database design in a customized manner to provide a comprehensive online resource of *Xenopus* developmental and genomics data.

GENE PAGES

At the heart of Xenbase are the gene pages—a catalog of over 11 000 genes derived from annotations of Joint Genome Institute (JGI) *Xenopus tropicalis* gene models (<http://genome.jgi-psf.org/Xentr4/>). At the moment, *Xenopus tropicalis* sequence assembly 4.1 has 27 934 associated models, generated by a variety of gene prediction algorithms. Annotation from both a community-driven jamboree and ongoing distributed annotation has resulted in approximately 2700 manually curated gene models. A total of 11 230 gene model annotations have also been generated by the metazome automated pipeline which uses synteny to assign orthology. Of the 2700 manually curated Xenbase gene pages, most also contain a large variety of other data including sequence data on *X. laevis* and human, mouse and zebrafish orthologs gathered from Entrez Gene (10) or from local BLAST.

Data from the two contributing *Xenopus* species are arranged in columns, with information in the left column representing *tropicalis* data and that in the right, *laevis* data (Figure 1). There are exceptions to this layout, for example data regarding orthologs from mammalian and non-mammalian species and GO terms common to the gene products of both species. At the moment only one of the two paralogs from the tetraploid species, *laevis*, is displayed and paired with data from the orthologous *tropicalis* gene. In the future, data on the *laevis* paralog will also be made available through a tab system. When *laevis* genomic data becomes available it will also be arranged on the right side of the page, although there has been no firm agreement on when this project will begin.

The *Xenopus* community voted to adopt HUGO gene nomenclature committee (HGNC) names and symbols (11) in surveys posted shortly before genome annotation was initiated. In the gene annotation process, HGNC names were applied to *tropicalis* gene models only if synteny or other forms of data support the assignment of a gene as an ortholog—genes are not named on BLAST hits only. Gene model annotation is performed by both community annotators and the metazome pipeline using the web-based JGI annotation toolset. The JGI database is then parsed for records matching a minimal set of cutoffs and a variety of *tropicalis* data from passing

records is transferred into Xenbase. Predicted protein and mRNA sequences from JGI gene models are then used to identify *laevis* orthologs locally via BLAST while gene names and symbols are used to identify orthologs from mammals and fish.


As a quality control step, a Xenbase staff annotator manually checks community annotations, ensuring that sufficient criteria were applied to select a specific gene name and that the approved HGNC name and symbol were applied. This is performed by Xenbase staff using the JGI gene annotation toolset on the JGI web site. As new records are added, or old records modified, they are flagged by a Xenbase weekly script, and then manually processed by a Xenbase annotator. The Xenbase annotations from the JGI database are promoted to gene pages, while machine annotated pages are used to build clearly identified temporary gene records.

As HGNC names constantly evolve and change (11), Xenbase runs scripts that ensure the *Xenopus* data remains synchronized with the values associated with the predicted human ortholog. When conflicts are identified these records are flagged and a curator reannotates the gene record at the JGI and this will then be processed as an updated record in the next download cycle at Xenbase keeping both *Xenopus* databases in synchrony with HGNC names.

In addition to gene names, gene symbols, mRNA and protein sequence, a variety of additional data is collected from the JGI web portal. This includes genomic scaffold mapping data, appropriate ontology terms from GO, KOG and EC, plus metazome and Unigene IDs. When information for a manually curated gene is not present in the JGI database, for example when a mouse gene name/symbol does not match the HGNC name/symbol, a custom curation tool allows an annotator to add this information manually. This tool also allows the two duplicated paralogs present in the zebrafish genome to be added to a single Xenbase gene page. A screenshot of a representative gene page is illustrated by Figure 1.

GENE NAMING AND SYNONYM MATCHING

Applying HGNC names to *Xenopus* genes has disconnected genome-generated gene models from the scientific literature, as the names used in these two environments rarely match. The previous gene naming custom in the *Xenopus* community was quite *ad hoc*—most genes were only given a symbol, the symbol usually began with an ‘X’, and the symbol could refer to orthology, homology, function, a fictional character, etc. Once new gene names and symbols were applied, it became important to associate these to the legacy names spread throughout the literature and used in previous DNA, mRNA and protein annotations. To alleviate this problem, Xenbase utilizes a synonym field populated with legacy names derived from the scientific literature. Synonyms are used to link gene records to legacy data in the literature, for example an article that mentions XPax-8 in a title or abstract will be matched to the pax8 record, and to help

 **Xenbase** [2.0] My Xenbase
[Login] [Register]

Blast Frogs Genes Literature Genomics Anatomy & Development Community Education & Research

Search Genes : For Search Give Feedback


Summary XB-FEAT-483692

Gene Symbol: pax8

Gene Name: paired box 8

Synonyms: XPax8, XPax-8, pax-8 [Add Xenopus synonyms](#)

Gene Function: paired box transcription factor



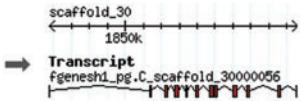
Molecules	tropicalis	laevis		
Gene	JGI XT Genome 4.1 Entrez Gene	Entrez Gene		
mRNA	Refseq ↗	Refseq ↗		
Protein	Refseq ↗	Refseq ↗		
Genomic	tropicalis			
Genome Browser	JGI Browser Xenbase G-Browse			
Syteny	Metazome			
Expression	tropicalis	laevis		
UniGene	Str.51436	Xl.2754		
Gurdon EST/FL	CABI11353	LNeuM.L045d19		
Orthology	MOD	Entrez Gene	Genetic Phenotypes	Morphants
human		PAX8	OMIM:167415 OMIM:218700	
mouse	MGI	Pax8	All(3) Targeted, knock Out(1) Targeted, other(2)	
zebrafish	ZFIN	pax8		
Publications	tropicalis	laevis		
Papers	Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. [Dev Dyn 2002]		Synergism between Pax-8 and lim-1 in embryonic kidney development. [Dev Biol 1999]	
Literature Search	Search Xenopus literature for pax8			
Functional Ontologies				
GO: Molecular:				
Biological:	development			
Cellular Component:	nucleus			
EC:				
KOG:	[K] Transcription factor PAX2/5/8, contains PAX domain			
Suppliers	tropicalis	laevis		
Clones	open biosystems ↗	open biosystems ↗		

Figure 1. Xenbase gene page screenshot. A screenshot of the display generated by a search for 'pax8'. *Xenopus tropicalis* is arranged on the left when species-specific data is displayed and *X. laevis*-specific data in the right column. Data common to the both species is not restricted to this format and is presented throughout the corresponding row.

users find genes for which they are familiar only with legacy names.

Given the number and diversity of legacy gene names, finding and curating them all is a difficult task. However, individual members of the *Xenopus* research community have this knowledge. To take advantage of this Xenbase

provides a user accessible interface allowing community members with a Xenbase account to add synonyms to gene records. While Xenbase also stores synonyms for human, mouse and zebrafish orthologous genes obtained from Entrez Gene (10), these cannot be used to link *Xenopus* legacy names to gene pages—there are simply too

Xenbase G-Browse

Showing 100 kbp from scaffold_334, positions 534,416 to 634,415

Instructions

[Hide banner] [Bookmark this] [Link to Image] [High-res Image] [Help] [Reset]

Search

Landmark or Region:

scaffold_334:534416..634415

Search

Reports & Analysis:

Download Decorated FASTA File

Configure...

Go

Data Source

Xenbase G-Browse

Scroll/Zoom: <<< << < > >> >>>

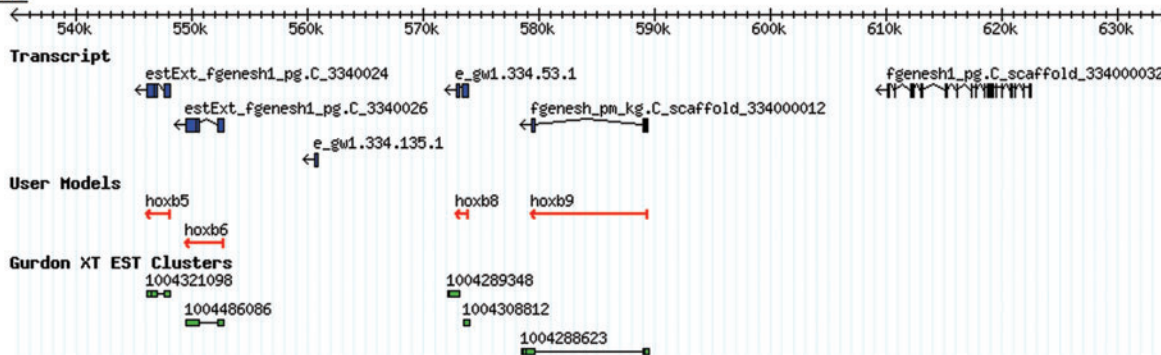
Show 100 kbp

Flip

Overview



Details



Clear highlighting

Update Image

Figure 2. Xenbase GBrowse screenshot. Three sets of models are displayed along genomic scaffolds—transcript, user and Gurdon ESTs—with the default settings. Clicking on a transcript model brings up the gff3 data used to build the model, clicking on the user models will bring up the corresponding Xenbase gene page, while clicking on Gurdon EST clusters will jump to the Gurdon web site with additional data on ESTs and clusters (12). The search field can be used to search for gene symbols or scaffold coordinate numbers such as those illustrated in this figure.

many synonyms and too many collisions to function in this manner. The *Xenopus* community and Xenbase curator synonyms can, however, be used in this manner as long as collisions are avoided, and scripts are run routinely to ensure that this is the case.

Presently the gene pages present a summary of the data types described above. The interface is a tabbed system and additional data pertinent to each gene will be added as tabs as new systems, for example *in situ* data, gene associated literature, morphant phenotypes, genetic mutants, available EST clones, etc. are completed.

GENOMICS

For genomics research, Xenbase provides a genome browser and *Xenopus*-specific blast tool. Xenbase implements the GBrowse (7) genome browser developed as part of GMOD. The Xenbase GBrowse tool can be used to visualize various features along the JGI *tropicalis* genomic scaffolds. GBrowse loads and displays gene models formatted in gff3 format. The JGI resource outputs gff1 formatted files of gene models. These were downloaded and converted to gff3 format using a custom script that inserts the additional data lines required for the

gff3 specifications. The gff3 files and file generation scripts are freely available upon request and will also be posted to the Xenbase ftp site.

At present, Xenbase GBrowse displays top-ranked JGI gene models, Gurdon EST clusters and Xenbase gene page matches (Figure 2). This clearly identifies models that have gene page support and allows users to switch back and forwards between the genome browser and gene pages corresponding to flanking models. Xenbase also maintains a link to the JGI Genome Browser from its gene pages to provide access to the variety of tracks available at this resource, including mouse and human ESTs.

A snapshot of the Xenbase genome browser is also integrated within the gene pages for each gene (Figure 1). The gene page displays a dynamically generated thumbnail view of the GBrowse gene model. Clicking on the thumbnail will open the genome browser, while clicking on the gene page model in the browser will jump back to the gene page.

A stand-alone BLAST interface is also provided by Xenbase to provide high speed *Xenopus*-specific queries of homology to a local database that contains only *Xenopus*-derived sequences. Currently, the *Xenopus* BLAST tool allows for searches against *tropicalis* and *laevis* protein, mRNA or EST sequences from NCBI. In addition,

searches may also be performed against the representative consensus sequences from *tropicalis* and *laevis* Gurdon EST clusters. An automated process regularly updates all of the offered databases without disrupting user queries and downloads of the sequence collections are provided through a Xenbase FTP site. The Xenbase BLAST tool then formats results and the display contains an overview list of hits, a detailed summary of each hit, links to matched Xenbase gene pages and NCBI records, and a set of BLAST statistics for the query. Gene pages display a small rocket icon next to each mRNA, EST and protein sequence. Clicking on this icon will load the appropriate FASTA sequence into the BLAST interface to allow for rapid searches for related genes.

COMMUNITY MODULE

Historically, Xenbase has acted as a hub for obtaining information on researchers working with the *Xenopus* system and continues this role with a revamped community section. The community section contains information on people, laboratories and organizations involved in *Xenopus* research. Personal profiles contain research interests, images, links and articles published by that person. Other features include an automated system for posting advertisements for positions available and announcements relevant to the *Xenopus* community (conferences, etc.). All information in this section is posted by community members themselves with the exception of publications. Publications present in the literature module are automatically associated with community member profiles by a name-matching algorithm that runs on new papers when they are uploaded from PubMed (described below). In addition to serving as a contact point for the community this section allows users to publicize their research interests, for users to find additional publications associated with an author and as a data attribution node.

LITERATURE MODULE

The Xenbase literature section provides a library of papers from NCBI's PubMed that pertain to *Xenopus* research. New papers are loaded into Xenbase on a weekly basis by an automated process using NCBI's E-Utilities and a relaxed search criterion for the term '*Xenopus*'. In addition, registered Xenbase users have the ability to upload additional papers from PubMed when the system fails to identify a relevant publication simply by submitting a PubMed ID number into a web form. Presently, over 35 000 literature records are available. Literature in Xenbase is searchable by author, abstract, title, journal, PubMed ID and the uniquely defined Xenbase accession number literature ID.

After each new import of PubMed articles into Xenbase, an automatic matching process is run to automatically associate publication authors to existing Xenbase users. This process creates links between individual paper pages in the literature module and user pages in the community module. Xenbase users have the

ability to confirm papers matched to them or remove links to any papers that were incorrectly associated with them as a result of the matching process. This tool is visible only to a user whose name has been matched to a publication by the automated system and only if logged in as a registered Xenbase user, so this feature can be ignored should the users prefer to do so.

The literature section also contains a curator accessible server-side Java tool for downloading images from the web sites of select journals with whom Xenbase has negotiated the right to reproduce images illustrating gene expression data. Due to the unique design of each journal's web site, a universal image scraper was not feasible. The tool is however extensible and simple plug-in modules have been developed for each target journal. Users can view downloaded images as thumbnails below the abstract of a paper, or click on the image to view a full resolution file. Figure legends are also collected and displayed. These images will serve as an important supplement to those derived from the large-throughput *in situ* screens that are presently being processed for inclusion in the gene expression component of the gene pages and the literature-derived expression images will also be displayed in this module.

The literature module provides an extraordinarily rich environment for exploring papers describing *Xenopus* research. As in PubMed, author names link out to additional publications by the same individual, though in the case of Xenbase they actually link to a community profile that will also contain a large amount of additional information on the individual and their research. Xenbase also links gene names and symbols in abstracts to Xenbase gene pages that in turn contain a wealth of associated data, which is not available from other literature databases such as PubMed. Images from papers are also uniquely available to Xenbase. Together these features make this section a powerful resource for exploring *Xenopus* data. Future development will allow users to browse weekly updates and receive email notifications of the addition of publications matching specific criteria.

ANATOMICAL ONTOLOGY

Xenbase is the center for development of the *Xenopus* Anatomical Ontology (XAO). This ontology describes *Xenopus* anatomy as a controlled vocabulary of terms organized in a graphical structure representing which tissues are 'part of' other tissues and during what stage of development each tissue exists. The XAO also represents morphogenesis of tissues by representing 'develops from' relationships between different tissues at different stages of development. The XAO is available as an Open Biological Ontologies (<http://obofoundry.org/>) formatted file on the Xenbase ftp site. It will play an essential role in annotating gene expression data from high-throughput *in situ* screens.

SEARCH FUNCTIONS

Two different locations offer search functions. Each module section selected from the horizontal

navigation banner present on every Xenbase page opens with a module-specific search page. The second option is a smaller search window directly below this horizontal banner, in the same position as the search window on NCBI pages, and is known as the minibar. While the module-specific search pages offer many more search selection criteria, for examples gene pages can be searched by all terms, name only, symbol only, synonyms, etc., the minibar only allows searching this module by the 'All' option. This later restriction was essential to keep the drop down of options in the minibar to a reasonable scale. Once Xenbase has annotated gene expression data from high-throughput screens and the literature, tools for performing more complex queries will be added such as site of expression, GO terms, etc.

ATTRIBUTION

Appropriate attribution of data is a critical requirement for model organism databases. With the advent of new methods for presenting scientific research results, such as web sites and direct submission to databases, published journal articles are no longer sufficient as a sole source of attribution. Therefore, Xenbase has a flexible database schema that allows any data item (i.e. any table record or even the value of a specific column in a record) to be attributed to a variety of possible sources such as papers, books, people, laboratories, organizations or other databases.

IMPLEMENTATION

The Xenbase schema is based on the CHADO database schema (6) produced as part of the GMOD project. Several extensions and modifications have been made to CHADO to improve performance and add functionality. Xenbase web interfaces are developed using Java servlets and JSPs, HTML, XML, CSS and AJAX. The majority of data pipelines are developed with Perl. Xenbase curation tools are transaction based and connected directly to the production database, therefore annotations made by curators or community members are instantly visible. The system runs on IBM WebSphere Application Server, IBM DB2 and Linux.

XENBASE CONTRIBUTIONS TO OTHER DATABASES AND DATABASE AVAILABILITY

Data from the Xenbase Gene Catalog is currently fed to the NCBI. The NCBI uses this information to update *Xenopus* gene names in Entrez Gene records and to provide links back to Xenbase. Xenbase also provides an ftp site where various downloads of data are available to the community and other databases. Existing downloads can be modified or new downloads can be created to meet the needs of other resources that want to create links to Xenbase, or for researchers who simply request customized reports.

The database is freely accessible. Users can utilize all data features and perform queries without an account or

logging in. The only time at which login is required is when database content is to be modified. Users have the ability to modify data in their own account in community, they can add publications in literature, or add synonyms to gene pages if logged in. Accounts are set up by users through an automated system that leads through a series of simple data fields, most of which are not required. Only fields required for some element of database functionality are mandatory, but the additional optional fields allow users to create much richer records for the Community module should they choose to utilize these features.

USER FEEDBACK

Community input is critical to the success of Xenbase. Each page provides a 'Give Feedback' button that encourages users to send comments and questions and to report bugs. Researchers can use the 'Suggest A New Gene' button on the gene search page to contact a curator and request additions to the gene catalog. General comments, requests and questions can be sent to xenbase@ucalgary.ca.

FUTURE DIRECTIONS

Several important new areas are currently under development. First, the next major addition to Xenbase will be support for *in situ* gene expression images. This section will incorporate *in situ* gene expression images from large-scale screens that are currently held in Axelddb (1,5) and XDB3 (N. Ueno, personal communication). Xenbase will enhance the value of these images by annotating gene expression images with tissues and developmental stages displaying expression that are represented in the *Xenopus* XAO. Gene expression data will also be drawn from the literature and associated with gene catalog entries and XAO terms using an expanded version of the literature image-loading tool. There will also be an interface allowing laboratories to submit gene expression data to Xenbase.

Second, as with gene expression data, adding annotated morpholino data on the effects of gene knockdowns (3) to gene pages is a high priority. One large-scale morpholino screen has been performed (13) and most *Xenopus* publications on gene function include a morpholino experiment. This data will be added by curator annotation of the literature.

Third, Xenbase has entered into a collaboration with the European *Xenopus* Resource Center (EXRC) to help develop the informatics support for the stock center. Xenbase will maintain information on wild type, transgenic and mutant lines of *Xenopus* held at the EXRC. These lines will be linked to gene pages, the XAO and other objects in Xenbase where appropriate.

Fourth, Xenbase is implementing a generic database schema for supporting the relationship between evidence and assertions in Xenbase. This will make it straightforward for users to quickly examine the evidence behind

various assertions presented in Xenbase—for example, why Xenbase indicates that a certain mouse gene is the ortholog of a *X. tropicalis* gene.

Fifth, there is too much diverse data being produced to model and represent fully in Xenbase or other databases. Therefore, Xenbase will be developing a WIKI that will be connected to the gene catalog. This will provide a medium for the community to supplement the content of gene pages or post research results that would be otherwise difficult to fit into Xenbase. This will be particularly useful in fulfilling the data distribution guidelines of many grant agencies.

ACKNOWLEDGEMENTS

Funding for Xenbase is provided through NIH R01HD045776-03 (P.D.V.) and by an AHFMR Senior Scholar award to P.D.V. We are indebted to the Xenbase advisory panel for providing invaluable feedback and the University of Calgary Department of Computer Science, and in particular Ken Barker, for space, guidance and infrastructure support. BLAST support is provided by the Center for Visual Genomics at the University of Calgary, Paul Gordon and Christoph Sensen. JGI provides data and database access that is essential to Xenbase functionality and in particular we would like to thank Paul Richardson, Astrid Terry and Igor Grigoriev and for their generous help. The Xenbase logo was designed by Victor Gerth. IBM kindly provided DB2 and Websphere software under their academic initiative program. Funding to pay the Open Access publication charges for this article was provided by NIH R01HD045776-03.

Conflict of interest statement. None declared.

REFERENCES

1. Pollet, N., Schmidt, H.A., Gawantka, V., Vingron, M. and Niehrs, C. (2000) Axeldb: a *Xenopus laevis* database focusing on gene expression. *Nucleic Acids Res.*, **28**, 139–140.
2. Amaya, E., Offield, M.F. and Grainger, R.M. (1998) Frog genetics: *Xenopus tropicalis* jumps into the future. *Trends Genet.*, **14**, 253–255.
3. Heasman, J. (2002) Morpholino oligos: making sense of antisense? *Dev. Biol.*, **15**, 209–214.
4. Kessler, D.S. (1997) Siamois is required for formation of Spemann's organizer. *Proc. Natl Acad. Sci. USA*, **94**, 13017–13022.
5. Pollet, N., Muncke, N., Verbeek, B., Li, Y., Fenger, U., Delius, H. and Niehrs, C. (2005) An atlas of differential gene expression during early *Xenopus* embryogenesis. *Mech. Dev.*, **122**, 365–439.
6. Mungall, C.J. and Emmert, D.B. (2007) FlyBase Consortium. A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics*, **23**, i337–i346.
7. Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
8. Sprague, J., Bayraktaroglu, L., Clements, D., Conlin, T., Fashena, D., Frazer, K., Haendel, M., Howe, D.G., Mani, P. *et al.* (2006) The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Res.*, **34**, D581–D585.
9. Blake, J.A., Eppig, J.T., Bult, C.J., Kadin, J.A. and Richardson, J.E. (2006) Mouse Genome Database Group. The Mouse Genome Database (MGD): updates and enhancements. *Nucleic Acids Res.*, **34**, D562–D567.
10. Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
11. Eyre, T.A., Ducluzeau, F., Sneddon, T.P., Povey, S., Bruford, E.A. and Lush, M.J. (2006) The HUGO Gene Nomenclature Database, 4006 updates. *Nucleic Acids Res.*, **34**, D319–D321.
12. Gilchrist, M.J., Zorn, A.M., Voigt, J., Smith, J.C., Papalopulu, N. and Amaya, E. (2004) . Defining a large set of full-length clones from a *Xenopus tropicalis* EST project. *Dev. Biol.*, **271**, 498–516.
13. Rana, A.A., Collart, C., Gilchrist, M.J. and Smith, J.C. (2006) Defining synphenotype groups in *Xenopus tropicalis* by use of antisense morpholino oligonucleotides. *PLoS Genet.*, **2**, 1751–1772.