Published in final edited form as:

Methods Mol Biol. 2012 ; 917: 541–562. doi:10.1007/978-1-61779-992-1_31.

On-line Resources for *Xenopus*

Jeff Bowes

Department of Biological Sciences, University of Calgary

Jeff Bowes: bowes@ucalgary.ca

Summary

Since the advent of computational methods in biology the quantity of biological data has grown exponentially. These data support genomic, genetic, developmental, and other forms of biological experimentation. The number of on-line resources has kept pace with the growth in data. *Xenopus* has perhaps lagged some of the other model organisms in developing resources but is now quickly catching up. There are now a number of well-established and developing resources for *Xenopus*. This chapter looks beyond the widely-known public databases, Genbank and the EBI, and describes how the researcher can use a number of central sites such as Xenbase, UniProtKB, and major genome browsers to navigate to a variety of different resources.

Keywords

Database; Website; Genome Browser; gene model; Xenbase; ECR

1. Introduction

Xenbase (1), the model organism database for *Xenopus laevis* and *tropicalis* is the natural starting point for the researcher to explore on-line resources for *Xenopus*. It is the only *Xenopus* specific resource to integrate different experimental data. Xenbase includes genomic, gene expression, tissues and development, literature and community information. It incorporates a gene catalog, gene expression search, a GBrowse genome browser (2), Textpresso (3) literature search and mining tool, an anatomy browser, community directory and a wiki. Importantly, Xenbase ties into many related resources.

Major institutional resources such as the NCBI (4) and EBI (5) are also very powerful integrated resources that focus upon multiple organisms, including *Xenopus*. The reader is expected to already have some experience with these resources. Therefore, they will not be covered in this chapter except for some specific tools such as the Ensembl genome browser.

The resources section of this book will introduce a number of on-line resources, some widely used and some interesting resources that are a little off the beaten path. Because of its

¹⁴To best present the view of a gene. The *X. tropicalis* variant of a gene and the *X. laevis* variant (or both a and b paralogs) of the gene are all presented on the same gene page. However, each gene variant has its own accession number shown at the top of the gene column. Gene accessions take the form XB-GENE-##### where ##### is a number. There is also an accession for the whole gene page at the top (XB-GENEPAGE-#####). If searching for the gene page by Xenbase accession you can get to the gene page by linking to any of these accessions.

importance to *Xenopus* research, gene expression is excluded here and covered in detail in Chapter 19.

The methods section describes how to research a gene of interest using Xenbase as a hub. From Xenbase UniProtKB (6) will be used to link out to protein and other resources, and the genome browsers will be used to link out to other genomics resources. As well, several effective literature mining tools will be demonstrated.

2. Resources

2.1. Xenbase (www.xenbase.org)

Xenbase accesses a wealth of integrated data for *Xenopus* research both internal to the site and linked to external resources. Xenbase uses a combination of computational and manual curation, with a staff of two to three curators.

The body of the front page of Xenbase provides an excellent summary of the resources available in Xenbase. A sidebar provides announcements and links to other relevant sites. The Xenbase header (Fig. 1) is found on every page and provides the principal means of navigating through Xenbase. The menu items in the navigation bar mirror the menu sections of Xenbase described on the home page. These are as follows:

1. The **Blast Tool** provides access to a number of different *Xenopus*-specific databases. Results provide links to Xenbase gene and clone pages as well as to external resources.
2. The **gene catalog** is the heart of Xenbase. The gene pages provide detailed information on genes, gene expression evidence, literature associated with a gene (*see* Note 1), sequences and potential interactants (*see* Note 2) and the gene wiki. Each gene page provides information for *X. tropicalis* and *X. laevis* orthologs and *X. laevis* A and B paralogs, if known. Many external resources are also linked in. At the time of writing the gene catalog included 15,865 genes of which 4,711 have manual curation (*see* Note 3). Users may contribute to Xenbase gene pages by contributing synonyms and entering information about the gene in the Wiki tab.

¹Papers are matched to genes by the Xenbase link matching algorithm. The link matching algorithm searches paper titles, abstracts and captions for gene mentions by finding gene symbols or synonym names using database searches and regular expressions. (Using gene mentions from paper bodies resulted in too many false positives due to mentions in the references section.) Symbols or synonyms likely to produce false positives (e.g., “t” for brachyury) are excluded from the search. Terms that match several genes have disambiguation pages allowing the user to choose the correct gene. *Xenopus* Anatomical Ontology (XAO) terms and development stages are matched in the same way in literature.

²Potential gene interactants are identified based on co-occurrences of gene mentions in the same paper. The interactants tab indicates how many documents two genes appear in together. Jenssen *et al.* (26) demonstrated that the co-occurrences of genes references in an abstract are indicative of meaningful biological interactions and the more often genes are mentioned together the more likely this is true.

³Genes in Xenbase are created from JGI or Ensembl gene models, or *Xenopus* Entrez Gene (27) records. At the time of writing, ~4,000 genes were manually curated and named based on synteny with humans. Manually curated genes are marked as provisional if there is no synteny evidence but only alignment-based putative orthology. Automatically curated genes are marked as “[provisional-machine annotated only]”. If a gene has a putative human ortholog, the gene symbol and name will be annotated as “[Predicted: Source is Human Entrez Gene]”. Otherwise the naming will come from *Xenopus* Entrez gene records and it will be annotated as “[Predicted: Source is *Xenopus* Entrez Gene]” or “[Predicted: Source is *Xenopus* Entrez Gene]”. Finally, there are a large number of genes whose symbols cannot be determined either via human orthology or *Xenopus* Entrez gene records. They all have the symbol “unnamed” and have a long name drawn from JGI gene model define records that are in turn based on protein domain annotation and homology derived with a Metazome pipeline.

3. The **expression search** section provides an interface that allows users to search for evidence of gene expression filtered by any combination of gene symbol /clone name, sequence or accession, Nieuwkoop and Faber (NF) development stages (7), tissues (*see* Note 4), and other variables. Xenbase's gene expression search is described in more detail in Chapter 19.
4. **Anatomy and development** allows users to explore data from the perspective of tissues or development stages. This section includes images of embryos, development stage temperature charts and cell fate maps. Users can also search for information on each NF development stage or *Xenopus* Anatomical Ontology (XAO) tissue. The page for each tissue includes information on their relationship with other tissues in the XAO (*see* Note 4), listings of genes and clones expressed in that tissue and papers containing that evidence. Users can also add information by updating the wiki tab on any anatomy term or development stage page.
5. Xenbase contains a comprehensive **literature** section containing over 40,000 *Xenopus* related articles that have been imported from Pubmed. The main component of the literature sections allows users to search articles by title, figure legend, abstract, journal, grant funding agency or Pubmed id. Each article entry contains an abstract entry that has references to genes and XAO terms hyperlinked (*see* Note 1). Additionally, Xenbase has permission to display gene expression images from a number of journals (*see* Note 5). In many cases, users can directly view images (often curated for gene expression) from the paper entry in Xenbase. Additionally, Xenbase has implemented the powerful Textpresso text search and mining tool (*see* **Section 2.8**).
6. The Xenbase **gene browser** is compared with other genome browsers in **Section 2.2**.
7. **Community** provides listings of people, labs, organizations (e.g., companies) and jobs in the *Xenopus* community. Each person's page also includes papers they have published. A person's papers are automatically detected but may be manually modified (8). Anyone can register and create an entry for themselves or their lab in Xenbase.
8. **Reagents and Protocols** include a search of *Xenopus* clones, clone libraries and vector information. Like gene pages, clone pages have an expression tab indicating where they have been expressed and links to suppliers. The expression tab is particularly useful for clones that have not been mapped to a gene. This section also

⁴Selecting tissues in the gene selection search is powered by the *Xenopus* Anatomical Ontology (XAO) (28). The XAO stores tissue types, the development stages the tissues generate and terminate at, and the three different types of relationships among them. The anatomy ontology includes **is-a** relationships such as a forelimb is a limb; **part-of** relationships such as the endocardium is part of the heart; and **develops-from** relationships such as prechordal plate develops from the mesoderm. These relationships enable powerful anatomy-based queries. For example, if a user searches for a gene expressed in the heart the search would return not only genes annotated as being expressed in the heart but also genes annotated as being expressed in any of the heart's subparts, such as the endocardium. However, a search for genes expressed in endocardium would not return genes annotated as being expressed in the heart because they could be expressed in other parts of the heart excluding the endocardium.

⁵Xenbase has permission to display gene expression images from the journals Cell, Development, Developmental Biology, Developmental Dynamics, Gene Expression Patterns, the Journal of Cell Science, Mechanisms of Development, and PNAS.

includes pages in the Xenbase wiki for antibodies and protocols. All Xenbase users can add antibodies or protocols.

9. The **Stock Centre** link contains wild type, mutant and transgenic frog lines; clones; and transgenes carried at the European *Xenopus* Resource Centre. In the future it will also contain resources from the U.S. *Xenopus* Stock Centre that is currently under development.

2.2. Genome Browsers

Genome browsers are a critical tool for finding and analysing genomic features and comparing them between organisms. Different genome browsers have different focuses and strengths. Table 1 compares tools and current genome version (4.1 or 4.2) tracks supported by different genome browsers. At the time of writing, version 7.1 of the *Xenopus* genome had not yet officially been released. However, Xenbase GBrowse already supports the 7.1 assembly.

1. The **JGI browser** is operated by the Joint Genome institute (JGI), the organization that sequenced *X. tropicalis* and performed the 4.1, 4.2 and earlier assemblies (9). It has gene models produced by numerous different gene model prediction algorithms; a track called “FilteredModels” containing the best of those models; and a track called “UserModels” that are derived from filtered or other models but have added human annotation. Also, models have GO, KEGG and KOG annotation. Importantly, as the JGI was not involved in the 7.1 assembly they will not be supporting it on their genome browser.

<http://genome.jgi-psf.org/cgi-bin/browserLoad/?db=Xentr4>

2. **Xenbase GBrowse’s** strengths are its integration with Xenbase and focus on development. Xenbase has 4.1 gene model, JGI filtered model, Ensembl and many preliminary JGI gene model prediction tracks. The Xenbase 4.1 gene models track directly corresponds to Xenbase gene pages which may be derived from JGI or Ensembl models, or Entrez Gene records (*see* Note 3). Users can click on Xenbase 4.1 gene models track, or either *X. laevis* or *X. tropicalis in situ* clone tracks and pull up a balloon with thumbnails of gene expression images, allowing them to jump immediately from the track to gene expression images. Xenbase is also the only major genome browser to have tracks for the Veenstra laboratory’s methylation data (10). While not a browser primarily designed for interspecies comparison, it contains evolutionarily conserved regions from ECRBase (*see* **Section 2.3**). At the time of writing, Xenbase GBrowse was the leading genome browser to support the prerelease 7.1 assembly. Xenbase already had preliminary gene models, Xenbase gene pages, Gurdon clusters, *X. laevis* and *X. tropicalis* clusters and Affymetrix probes mapped against the 7.1 assembly, and was in the process of mapping additional tracks to the 7.1 assembly.

Assembly 4.1: http://www.xenbase.org/fgb2/gbrowse/Xenopus_tropicalis_4_1/

Assembly 7.1: http://www.xenbase.org/fgb2/gbrowse/Xenopus_tropicalis_7_1/?source=Xenopus_tropicalis_7_1

3. **Ensembl** (11) is a dedicated genome browser, supporting many organisms, with some of the more powerful genome browsing features. Ensembl has a wide choice of tracks including the most genome features mapped to the *X. tropicalis* assembly of any of the genome browsers. Like the UCSC browser it has strong support for comparative genomics. Table 2 in note 6 compares comparative genomics tracks between Ensembl and the UCSC browser. However, Ensembl really stands out in comparative genomics with its tool for comparing two separate organisms side by side. Finally, Ensembl incorporates Biomart (12), which allows users to download the Ensembl data tracks.

www.ensembl.org/Xenopus_tropicalis/

4. The **UCSC genome browser** (13) is a dedicated genome browser, supporting many organisms. It has a wide variety of genomics tracks and, like Ensembl, has strong support for comparative genomics (*see* Note 6). The UCSC genome browser also has some unique tools. In particular, the table browser tool allows the user to perform powerful queries against tracks on the genome. For example, the user can query intersecting regions of two tracks.

[http://genome.ucsc.edu/cgi-bin/hgGateway?
hgsid=206930983&clade=vertebrate&org=X.+tropicalis&db=0](http://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=206930983&clade=vertebrate&org=X.+tropicalis&db=0)

2.3. Other Nucleotide Resources

1. **ECRbase** (14) is a database of evolutionarily conserved regions (ECRs) constructed from whole genome alignments. The database features the ECR Browser, which allows users to browse or search a base genome and look for evolutionarily conserved regions. At the time of writing *X. tropicalis* genome could be compared to cow, *Fugu*, chicken, human (hg18 and hg19), mouse, opossum, chimpanzee, Rhesus macaque, rat, *Tetraodon* and zebrafish. The ECR Browser also features refseq, Genescan mRNA predictions and mammalian genome tracks. ECRbase offers downloadable files for promoter, synteny and ECRs. Xenbase features ECRbase ECR and promoter prediction tracks.

Browser: <http://ecrbrowser.dcode.org>

Home page: <http://ecrbase.dcode.org>

2. **Gurdon Clusters** (15) contains clusters of overlapping ESTs. The page for each cluster contains the consensus sequence and a graph showing how it was assembled from component ESTs. This is a useful tool for *X. laevis*, which currently lacks a sequenced genome. A virtual northern blot graph shows the relative abundance of clones from a cluster in a library or libraries drawn from the same tissue or stage.

[http://genomics.nimr.mrc.ac.uk/cgi-bin/public.exe?
dbe=http&dbs=INFOPUBLIC&uid=guest&species=Xt¤t=ESTs&full=Guest](http://genomics.nimr.mrc.ac.uk/cgi-bin/public.exe?dbe=http&dbs=INFOPUBLIC&uid=guest&species=Xt¤t=ESTs&full=Guest)

⁶The Ensembl and USCS genome browsers have the most comprehensive cross species comparison. Table 2 contains a comparison of features between these two browsers for comparing genomes from different species.

+User&src=public&tgt=main&menu=main_intro&project_key=1000000190&version=1)

3. **Rfam** (16) is a database of non-coding cis-regulatory RNA elements. Rfam entries for each family contain descriptions from Wikipedia including clan associations and descriptions of its function and clinical significance. Other sections show species containing the gene family, sequences for each species, visualization of secondary structure, ontology annotations, and literature referencing the ncRNA gene family. Currently, there are approximately 100–200 RNA families directly annotated to *Xenopus*. ncRNA genes can also be viewed as a track on Ensembl.

<http://rfam.sanger.ac.uk/>

4. **miRNA** (17) is the central on-line repository for microRNA. It is searchable by name and browsable by organism. At the time of writing there were 186 *X. tropicalis* entries and 22 *X. laevis* entries. Each entry describes the stem loop sequence and structure and mature sequence. This database also acts as a registry for miRNA naming. *Xenopus* miRNA records are available as a track in the Ensembl gene browser.

<http://www.mirbase.org>

2.4. Synteny tools

1. In addition to the genome browsers and ECRbase, Metazome (18) and Genomicus (19) are valuable tools for viewing synteny. They are described in detail in Chapter 4.

2.5. Protein and Molecular Resources

There are many different protein resources. The following reflects a sample of some of the more popular ones.

1. **UniProtKB** is a central, comprehensive resource for proteins. UniProtKB is composed of the manually curated UniProtKB/Swiss-Prot and computationally annotated UniProtKB/Trembl (*see* Note 7) entries. Each record indicates whether it comes from the Swiss-Prot or Trembl immediately after the accession. Annotation includes descriptions of protein function, subunit structure, subcellular localisation, expression in terms of tissues and development stages and Gene Ontology (GO)(). Additionally, there are sequence annotations for molecule processing, repeat regions, cleavage sites, amino acid modifications and experimental issues. Each protein entry also includes references to curated literature and many other on-line protein resources. Thus, UniProtKB is an excellent hub for jumping off to other protein resources.

⁷UniProtKB was divided into two sections because the number of protein sequences being generated, especially from translation of coding sequences, exceeded UniProtKB's capacity to provide quality human annotation for everything. Thus, Swiss-Prot contains high quality manually curated sequences. Furthermore, all proteins for the same gene are merged into a single Swiss-Prot entry with one canonical sequence. Trembl sequences are curated using a computational pipeline and are not merged into a single entry for each gene.

<http://www.uniprot.org/>

2. **InterPro** (20) integrates PROSITE, PRINTS, ProDom, Pfam, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, Gene3D and PANTHER. It contains protein alignments, relationships with other proteins in InterPro, GO annotation, domain annotation, links to literature and other resources.

<http://www.ebi.ac.uk/interpro/>

3. **PFAM** (21) is a database of protein families and superfamilies determined using Hidden Markov Models (HMM). PFAM provides its own summary as well as summaries from Wikipedia and InterPro. It also provides protein domain annotation, HMM model visualizations, interactions, and structure visualizations.

<http://pfam.sanger.ac.uk/>

4. **PDB** (22) is a 3D structure database describing experimentally determined protein and other biological macromolecules including nucleic acids and protein-nucleic acid complexes. It contains the atomic coordinates describing the molecule and various tools for visualizing protein structure. At the time of writing, PDB contained approximately 270 *Xenopus* proteins. It also contains a variety of other annotations, literature resources and links to other on-line protein resources.

<http://www.pdb.org>

2.6. Genetic Resources

1. The **Genetic Map of *Xenopus tropicalis*** (23) is a genetic map of unique simple sequence repeats (SSRs) onto the *X. tropicalis* chromosomes.

<http://tropmap.biology.uh.edu/map.html>

2.7. Phenotype Resources

1. **Smith Laboratory Morpholino Screen** provides a site where users can search a screen of morpholinos for 202 genes by gene name, phenotypic class or specific defects. Each page presents images of the embryo with the morpholino and controls.

<http://www.gurdon.cam.ac.uk/~smithlab/screens/Xenopus-morpholino-pilot/>

2.8. Literature

1. **Textpresso** is a powerful literature search and text mining tool that is implemented by, and integrated into, Xenbase. Users can search by keywords and/or ontology terms such as all *Xenopus* genes, XAO terms, gene regulation terms, select GO terms and many others. Unlike almost all other literature search resources, Textpresso searches the actual paper content and not just the title and abstracts. Textpresso returns sentences from documents that match the user's query. This allows users to do searches like "find at least two genes mentioned in a sentence with a regulatory term" or "find sentences with pax8 and a morpholino-related term" (see **section 3.10**).

<http://www.xenbase.org/cgi-bin/textpresso/xenopus/home>

2. **iHOP** (24) or **information hyperlinked over proteins** is a database of Pubmed abstracts that have been annotated with gene names. It is useful for manually researching literature for gene interactions. Gene entries in iHOP are organized by pages for different organism's version(s) of each gene. These pages have sections for interactions, definitions and recent information; with each section containing sentences from article abstracts referencing the subject gene. Sentences are included based on gene references found by matching symbols and synonyms. For each sentence, the user can pull up the full abstract. They can also click on the name of another gene to navigate to the page for that gene. This provides a similar experience to using the literature section in Xenbase, minus the figures and figure legends that are only found on Xenbase. However, in iHOP the user is able to navigate to articles about the same gene in other organisms. Also, for any sentence that contains two gene references the user is able to add a link to a graph that keeps track of the relationships they are finding between genes.

<http://www.iHOP-net.org>

2.9. Other Resources

1. **Comparative Toxicology Database** (25) is a databases of gene, chemical and disease interactions in vertebrates, including *Xenopus*, that are curated from literature.

<http://ctd.mdibl.org/>

3. Methods

While far too many resources have been described above to provide detailed instructions for using them, this section provides a basic guide on navigating to and making basic use of the above resources. Searching Xenbase for a gene and linking out through the gene page, to UniProtKB or one of the genome browsers provides access to the majority of these resources. There are also instructions on accessing unconnected resources: ECRbase, Gene Map of *Xenopus tropicalis*, and iHOP.

3.1. Review a Gene

1. Go to the Xenbase homepage (www.xenbase.org). In the search minibar (Fig. 1) select Genes and enter an accession number (*see* Note 8). If searching by symbol, using the human gene symbol is most effective (*see* Note 9). As you type, the search will pull up matching symbols as suggestions. You can choose one of the suggestions or keep typing. Instead of using the minibar, you can also click on the

⁸You can search for a gene using a number of criteria such as symbol, name, synonym, orthologs' synonyms and symbol, NCBI Accession, Entrez Gene Id, Affymetrix ID, Unigene Id, OMIM ID, GO ID or Xenbase ID.

⁹Xenbase gene naming follows the human genome naming committee (HGNC) conventions. This is consistent with the direction followed by other major model organism databases. The downside of this is that it makes it more difficult to find genes or match genes in literature based on historical names. Therefore, it is especially useful if users help keep synonym lists updated (*see* Note 12).

genes option in the navigation bar and perform the regular or advanced (*see* Note 10) search on the genes landing page.

2. If your search matched only one gene you will be taken straight to that gene page. Go to step 3. Otherwise, you will have a list of genes to choose from. A camera icon beside a gene indicates that there are images associated with this gene (*see* Note 11).
3. You will land on the summary tab for the gene page. At the top of the gene page you will find tabs for different types of information about the gene: Summary, Expression, Gene Literature, Sequence, Interactants and the Wiki. The top of the page contains information on the gene symbol, name, synonyms (*see* Note 12), function and potential interactants and some archetypal expression images for the gene. Below that, the summary tab is divided into sections. The page contains data for *X. tropicalis* and the *X. laevis* gene (*a* and *b* variants if both are known.) The data for each gene are organized in columns. There is an accession for the gene page at the top of the page and accessions for each individual gene heading the columns for each gene.
4. The molecule section has links to Entrez Gene, Ensembl, UniProtKB, and links to Blast or view sequences (*see* Note 11). Click on the link marked UniProtKB to explore your gene's protein product (*see* **Section 3.2**).
5. In the genomic section you can see a snapshot of your gene in the genome browser. You can click on JGI, Ensembl, UCSC, or Xenbase GBrowse links to view your gene in any of those browsers. *See* **Section 3.3** and **3.4** for a brief overview of using Xenbase GBrowse and Ensembl respectively. Chapter 4 contains detailed information on using the UCSC and the JGI genome browser.
6. In the orthology section you can link to human, mouse and zebrafish orthologs in OMIM, MGI and ZFIN respectively. There are direct links to phenotype and morphant data from these organisms when such data exists.
7. If the gene has phenotype data there will be a phenotype section on the summary page with a link to the Smith laboratory's morpholino screen. (nog2 is an example of such a gene.)
8. To view expression data, click on the expression tab. This page will provide a summary of anatomy terms and the range of development stages for which there is evidence of expression. There are images summarizing expression (the same as the summary images), community submitted images (*see* Note 13) or images drawn from literature. Click on an image to view more information on it. Summary and

¹⁰The Xenbase advanced search allows you to add or subtract criteria. You can add or remove items from search by clicking a plus or minus icon. You can also limit your search to pages that are manually curated or have gene expression data.

¹¹A number of icons are used throughout Xenbase. A camera icon means an object has images associated with it; a rocket takes the user to a blast search populated with the sequence for that object. Clicking a magnifying glass will bring up a popup window with an object's sequence. The user can click on a thumbs-up or thumbs-down icon to indicate whether the quality of an image is high or low.

¹²If you do not see a name that you know is commonly used click on the "add synonym" link to add it. This will aid future searches. You need to be a registered Xenbase user and log in to use this function.

¹³Community images include 2,500 and 18,000 images from AxelDB (29) and the Ueno laboratory's NIBB high throughput *in situ* gene expression screens respectively. These are described further in Chapter 19.

literature images are the most likely to be curated. Clicking on the images will open the image, revealing gene expression curation. Users are encouraged to click the thumbs-up or thumbs-down to rate an image's quality. Xenbase uses image quality ratings to better order images as they are presented to the user.

9. The sequences tab has a variety of sequences associated with each gene. The literature tab features articles. Users are encouraged to use the Wiki tab to enter additional useful information about their gene (*see* Note 15).

3.2. Finding Protein resources using Xenbase and UniProtKB

1. Link to the UniProtKB from a Xenbase gene page. (Use the *X. laevis* variant of trove2 for this example.)
2. Check by the accession to see if the UniProtKB resource is from Swiss-Prot (manually curated) or Trembl. On the pages you will notice a wealth of information about the resource.
3. Under the **Cross-references** heading you will find links to the InterPro, Pfam, PDB and the Comparative Toxicogenomics Databases described in the Resources section of this chapter. You will also find links to a wealth of other protein resources.
4. As an example, try going to PDB. Under **Cross-references**, under **3D structure databases** select **RCSB PDB** in the radiobox and click a link under the column: entry.
5. On the PDB page, look in the right hand column, in the **Biological Assembly** box to see the 3D structure. Click on the button that says "View in Jmol". You may be asked if you want to run an application. If so, select yes. You will have a 3D image appear in a box (Fig. 2). You can rotate the image around by dragging it with your mouse.

3.3. Using Xenbase GBrowse

1. From a Xenbase gene page (use pax3 for this example) click on the link to GBrowse or click on genome browser snapshot. This will take you to the portion of the genome where the gene in question is located. Alternatively, from the Xenbase home page, select **Genome Browsers -> Xenbase Xentr 4.1** and enter either the gene name or scaffold region in the **Landmark or Region** field.
2. The page is organized from top to bottom with navigation controls, followed by a scaffold and region overviews, detailed view tracks and track display controls (Fig. 3).
3. To turn tracks on or off go to the track control section. For this example turn on the following tracks: Xenbase 4.1 Gene Models, Gurdon XT 7.1 Clusters, Gurdon XL 3 EST Clusters, Fosmids, any ECR tracks and Promoter predictions. After you

¹⁵The wiki tab in Xenbase exists on gene, anatomy and development stage pages. On a wiki page simply click on the edit tab to add or modify data. You do not need to be logged in. The process used to edit a Xenbase wiki article is the same as is used to edit a Wikipedia article (*see* <http://en.wikipedia.org/wiki/Wikipedia:Cheatsheet>).

open tracks and move them around, GBrowse will open the same tracks the next time you use the browser.

4. You can rearrange tracks by holding the mouse down over a track header and then drag it up or down to where you want it. Icons on the side of the track header allow you to control it.
5. The Xenbase 4.1 Gene Models track corresponds to Xenbase gene pages and may be derived from JGI or Ensembl gene models, or Entrez Gene entries (*see* Note 3). Click this track to open a balloon with links to the gene page, source model and genome details. If there are gene expression images associated with the gene then there will be a thumbnail of one of the images in the balloon. Click on the thumbnail to view it or use the arrows to page through additional gene expression images.
6. Click on one of the Gurdon cluster links to jump to the Gurdon EST cluster site.
7. Click on one of the *in situ* clone tracks to view expression images for the clone.
8. Look at the methylation tracks which present the level of methylation in a wiggle format.
9. Click on a fosmid that was used in sequencing the *X. tropicalis* genome. One of the options in the balloon that pops up will be a link to order the fosmid from the European *Xenopus* Resource Centre (EXRC).
10. Examine the positioning of the predicted promoter and evolutionarily conserved regions relative to the gene model.

3.4. Using Ensembl

1. Click on the Ensembl link on a Xenbase gene page (for this example use pax3 again).
2. To choose what tracks you want, open click the **Configure this page** link in the left hand menu. To access the RFAM and miRNA database described in the resources section of this paper select **Genes** from the **configure view** and click the ncRNA checkbox to select how you want the track displayed. Then exit the configure browser image by clicking the checkmark in the upper right-hand corner of the window.
3. To compare the *X. tropicalis* genome side by side with another genome select **Comparative Genomics -> Multi-species view** from the **Location-based displays** menu on the left side of the screen.
4. Add species to the view by pressing the **Select species** button on the left. The window that pops up will show selected species on the left and unselected species on the right. Press the plus button beside a species to add it or the minus button to remove a species.

3.5. Using ECR Browser

1. From the ECR Browser home page (<http://ecrbrowser.dcode.org/>) select “Frog” in the base **genome field** and enter a scaffold position or gene name in the **feature or position** field. Press “submit”.
2. This will take you to a view of that genomic region (Fig. 4). Adjust the focus controls until you can view the full area around the gene or region examined.
3. Add or remove species using the controls on the right side of the screen. Add additional reference tracks (e.g., refseq mRNA) using controls in the bottom right corner.
4. If you are looking at the gene, look for ECRs within the gene boundary as indicators of exons and in the region before the gene as indicators of potential promoter or enhancer regions.
5. Click the **instructions** link on the navigation bar for further information on using ECRbase.

3.6. Using The Genetic Map of *Xenopus tropicalis*

1. From the Genetic Map of *Xenopus tropicalis* home page (<http://tropmap.biology.uh.edu/map>), select a linkage group or cluster from the drop boxes and press the “Retrieve_Map” button.
2. This will take you to a page with the map of the chromosome showing the locations of SSRs. Click on an SSR’s accession number to view details for it.

3.7. Browsing Anatomy in Xenbase

1. From the Xenbase home page select **Anatomy & Development -> Xenopus Anatomy Ontology** from the navigation bar.
2. Select the **Search Anatomy** tab.
3. On the search anatomy page there are two ways you can search for the anatomy item you want. First, you can use the search field by tissue name, XAO id, or Xenbase Id. If you search by name a suggestion box will provide options that match what you are typing. You can select an option from the suggestion box or you can search for the string you have typed in. If there is more than one match they will be displayed in a list. Alternatively you can use a tree. For example, to find mesoderm you would select **embryo -> primary germ layer -> mesoderm**.
4. The page for the anatomy item will have several tabs at the top: Summary, Anatomy Item Literature, Expression, Attributions and Wiki. You initially start on the summary page which has the anatomy terms name, synonyms, definition and relationships with other anatomy terms (*see* Note 4). You can explore the tabs or link to developmentally related or component tissues.

3.8. Viewing Antibodies in Reagents and Protocols

1. From the Xenbase home page select the chevron next to the **Reagents and Protocols** and select **Antibodies** from the menu.
2. This will take you to a Xenbase wiki page that has an alphabetical list of curated antibodies. Choose one of the antibodies and you will be taken to a wiki page with additional information.
3. Users are encouraged to add or edit antibodies. You can edit an existing antibody page by selecting wiki edit tab for an antibody. (*see* Note 15 for more information on editing wikis.)
4. If you wish to add an antibody you would edit the antibody list page and add a link with the name of your antibody. There is a link at the top of the page called **Antibody Reagent Template**. This provides the template that an antibody entry should follow.

3.9. Using the Xenbase to search for publications

1. From the Xenbase home page select **Papers** from the Literature menu on the navigation bar. This will take you to the paper search page.
2. On the literature search page you can search by many criteria (for this example, search Journal for “Dev Dyn”).
3. Select a paper whose link has a camera beside it (indicating it has imported images).
4. On the paper page you will see the abstract with other genes references, anatomy terms and development stages hyperlinked. You can click on any of these links to explore these items. If authors are registered Xenbase users their names will also be hyperlinked and clicking on the name will take you to the author’s Xenbase person page.
5. Click the Pubmed link to open the Pubmed entry.
6. Go to the **Article images** section and click “show captions”. This will allow you to peruse image captions as well as abstracts.
7. As in other sections, clicking on the thumbnail image will open a larger version of the image with annotations of expression.

3.10. Using Textpresso

1. From the Xenbase home page click the chevron to the right of the **literature** menu. Then select **Textpresso**. This will take you to the Textpresso search page.
2. As an example, we will perform a query to find articles with evidence of a regulatory interaction between pax8 and any other gene. In the **Keywords** field type “pax8” (Fig. 5). Ensure the **Exact Match** and **Search Synonyms** are checked. This adds the criterion of “pax8” or synonyms to the search. Next, place your cursor in the first field and press the **List>** button. This will pop-up a menu. Hover

over **relationships** and then select the **regulation** option. This adds the criteria that a regulation term must also appear in the sentence. Finally, place your cursor in the next category box. Click the **List >** button again. Select **Biological Concepts -> gene-> gene (*Xenopus*)**. This requires that the sentence contain an additional gene term. Press the **Search** button.

3. Browse through the results.

3.11. Using iHOP

1. Go to the iHOP home page (www.iHOP-net.org) and enter the name of a gene in the **Search Gene** field. (Use *pax8* for this example.)
2. This will take you to a page of genes that match your search. These will include *X. laevis*, *X. tropicalis* and other species' genes. Select the gene that you want.
3. This will take you to the interactions page for your gene. The page will contain sentences that mention your gene and another gene.
4. Press the view abstract icon to view the whole abstract marked up with other gene mentions and MeSH terms.
5. Click the Add to gene model icon beside the sentence and it will add the relationship described in the sentence to a graphical visualization of a gene network.
6. Click a link to another gene to investigate further interactions.

Acknowledgments

This work is funded by the NICHD grant 1P41HD064556-01. Thanks to Kevin Snyder for assistance in this work.

References

1. Bowes JB, et al. Xenbase: Gene Expression and Improved Integration. *Nucleic Acids Res.* 2010; 38:D607–12. [PubMed: 19884130]
2. Donlin MJ. Using the Generic Genome Browser (GBrowse). *Curr Protoc Bioinformatics.* 2009 Chapter 9, Unit 9.9.
3. Muller HM, Kenny EE, Sternberg PW. Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature. *PLoS Biol.* 2004; 2:e309. [PubMed: 15383839]
4. Sayers EW, et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2011; 39:D38–51. [PubMed: 21097890]
5. Brooksbank C, Cameron G, Thornton J. The European Bioinformatics Institute's Data Resources. *Nucleic Acids Res.* 2010; 38:D17–25. [PubMed: 19934258]
6. UniProt Consortium. Ongoing and Future Developments at the Universal Protein Resource. *Nucleic Acids Res.* 2011; 39:D214–9. [PubMed: 21051339]
7. Nieuwkoop, PD.; Faber, J. Normal Table of *Xenopus laevis* (Daudin). Garland; New York: 1994.
8. Bowes JB, et al. Xenbase: A *Xenopus* Biology and Genomics Resource. *Nucleic Acids Res.* 2008; 36:D761–7. [PubMed: 17984085]
9. Hellsten U, et al. The Genome of the Western Clawed Frog *Xenopus tropicalis*. *Science.* 2010; 328:633–636. [PubMed: 20431018]
10. Akkers RC, et al. A Hierarchy of H3K4me3 and H3K27me3 Acquisition in Spatial Gene Regulation in *Xenopus* Embryos. *Developmental cell.* 2009; 17:425–434. [PubMed: 19758566]

11. Flicek P, et al. Ensembl 2011. *Nucleic Acids Res.* 2011; 39:D800. [PubMed: 21045057]
12. Smedley D, et al. BioMart—Biological Queries made Easy. *BMC Genomics.* 2009; 10:22. [PubMed: 19144180]
13. Fujita PA, et al. The UCSC Genome Browser Database: Update 2011. *Nucleic Acids Res.* 2011; 39:D876. [PubMed: 20959295]
14. Loots G, Ovcharenko I. ECRbase: Database of Evolutionary Conserved Regions, Promoters, and Transcription Factor Binding Sites in Vertebrate Genomes. *Bioinformatics.* 2007; 23:122–124. [PubMed: 17090579]
15. Gilchrist MJ, et al. Defining a Large Set of Full-Length Clones from a *Xenopus Tropicalis* EST Project. *Dev Biol.* 2004; 271:498–516. [PubMed: 15223350]
16. Gardner PP, et al. Rfam: Updates to the RNA Families Database. *Nucleic Acids Res.* 2009; 37:D136–40. [PubMed: 18953034]
17. Griffiths-Jones S, et al. MiRBase: Tools for microRNA Genomics. *Nucleic Acids Res.* 2008; 36:D154–8. [PubMed: 17991681]
18. Dirks B, et al. Zomes: Tools for Phylogenomic Exploration and Analysis of Plants, Fungi, and Animals. *Phytopathology.* 2006; 96
19. Muffato M, et al. Genomicus: A Database and a Browser to Study Gene Synteny in Modern and Ancestral Genomes. *Bioinformatics.* 2010; 26:1119. [PubMed: 20185404]
20. Mulder NJ, et al. The InterPro Database, 2003 Brings Increased Coverage and New Features. *Nucleic Acids Res.* 2003; 31:315. [PubMed: 12520011]
21. Finn RD, et al. The Pfam Protein Families Database. *Nucleic Acids Res.* 2010; 38:D211–22. [PubMed: 19920124]
22. Rose PW, et al. The RCSB Protein Data Bank: Redesigned Web Site and Web Services. *Nucleic Acids Res.* 2011; 39:D392. [PubMed: 21036868]
23. Wells DE, et al. A Genetic Map of *Xenopus Tropicalis*. *Dev Biol.* 2011
24. Hoffmann R, Valencia A. Implementing the iHOP Concept for Navigation of Biomedical Literature. *Bioinformatics.* 2005; 21(Suppl 2):ii252–8. [PubMed: 16204114]
25. Davis AP, et al. The Comparative Toxicogenomics Database: Update 2011. *Nucleic Acids Res.* 2011; 39:D1067–72. [PubMed: 20864448]
26. Jenssen TK, et al. A Literature Network of Human Genes for High-Throughput Analysis of Gene Expression. *Nat Genet.* 2001; 28:21–28. [PubMed: 11326270]
27. Maglott D, et al. Entrez Gene: Gene-Centered Information at NCBI. *Nucleic Acids Res.* 2011; 39:D52. [PubMed: 21115458]
28. Segerdell E, et al. An Ontology for *Xenopus* Anatomy and Development. *BMC Dev Biol.* 2008; 8:92. [PubMed: 18817563]
29. Pollet N, et al. Axeldb: A *Xenopus Laevis* Database Focusing on Gene Expression. *Nucleic Acids Res.* 2000; 28:139–140. [PubMed: 10592204]
30. Kent WJ, et al. Evolution's Cauldron: Duplication, Deletion, and Rearrangement in the Mouse and Human Genomes. *Proc Natl Acad Sci U S A.* 2003; 100:11484. [PubMed: 14500911]

Fig. 1.

The header for Xenbase is on every page in Xenbase. It consists of the (a) navigation bar, (b) search minibar, (c) title and version, (d) wiki and (e) My Xenbase console. The navigation bar contains links to different topic areas. Clicking the (f) chevron to the right of a menu item will open submenu items. The search minibar allows the user to search Xenbase for a variety of different accessions or terms. Clicking the title always takes one to the *Xenopus* homepage and clicking the version will take the user to a description of everything in the current release. The My Xenbase bar allows users to register or login. The wiki icon links to an integrated wiki. There are prominent buttons for (g) submitting data and giving feedback.

Fig. 2.

The JMOL viewer in PDB allows you to manipulate 3D protein assemblies.

Fig. 3.

Xenbase GBrowse genome browser tracks (example pax3 gene): **a**: Lines demarcate the area shown in the detailed view on the scaffold overview and the region tracks **b**: a track header that can be dragged to move the track. It also contains controls to open/close, remove from the display, broadcast with DAS or access help **c**: A balloon with detailed information is opened when clicking on a track (in this case the Xenbase 4.1 gene track) **d**: You can navigate through *in situ* images for the gene using arrow controls or click on the thumbnail to view a larger image with annotations.

Fig. 4.

The ECR Browser **a**: the *X. tropicalis* model being compared **b**: the list of organisms that the *X. tropicalis* model is being compared to **c**: the icon for adding additional species to the comparison **d**: the icon to remove a species from the comparison **e**: icons for adding and removing reference tracks **f**: navigation and zoom controls **g**: A set of evolutionarily conserved regions that line-up with exons in the *X. tropicalis* gene **h**: Upstream ECRs that are potentially promoters or enhancers.

Textpresso

About Textpresso	Categories/Synonyms	Copyright	Document Finder	Downloads	Feedback	Home	Query Language	Search	User Guide
---------------------	---------------------	-----------	--------------------	-----------	----------	------	-------------------	--------	---------------

Search for keywords or categories or both

Keywords

pax8

☒ Exact match ☐ Case sensitive ☒

Categories ?

[List >](#)

regulation

Select category 2 from list above

Select category 3 from list above

Select category 4 from list above

Select category 5 from list above

Advanced Search Options : [on](#) | [off](#)

[Search!](#) [Undo current changes](#)

- [Biological Concepts](#) >
- [Descriptions](#) >
- [Relationships](#) >
- [allele](#)
- [anatomy](#)
- [biological process \(GO\)](#)
- [cellular component \(GO\)](#)
- [cellular component curation \(Caltech\)](#)
- [clone](#)
- [developmental stage](#)
- [disease](#)
- [drugs](#)
- [entity feature](#)
- [gene](#)
- [life stage](#)
- [molecular function \(GO\)](#)
- [molecular function curation \(Caltech\)](#)
- [morpholino](#)
- [mutants](#)
- [nucleic acid](#)
- [organism](#)
- [phenotype](#)
- [reporter gene](#)
- [restriction enzyme](#)
- [second messenger](#)
- [sequence \(SO\)](#)
- [sex](#)
- [strain](#)
- [transgene](#)
- [transposon](#)
- [vector](#)

gene (all)

gene

gene (Xenopus)

..., filtering (author, journal,...)]

Fig. 5.
Textpresso search for pax8 appearing in a sentence with another *Xenopus* gene and a regulation term.

Table 1

Genome Browser Comparison

	JGI	Xenbase	Ensembl	UCSCD
Assembly Versions	3, 4.1	4.1, 7.1	4.1, 4.2	3, 4.1
4.1 Tracks				
Gene Model Tracks				
JGI filtered	Yes	Yes		
Ensembl	Yes	Yes	Yes	
Xenbase gene models		Yes		
Refseq genes				Yes
Other	Many	Fgenesh, genewise, GeneWisePlus from JGI	Genewise, EST base predictions	Genewise
mRNA Tracks				
ESTs			Yes	Yes
Genbank mRNA	Yes			Yes
Xtropicalis cDNA	Yes	ISH probes	Refseq, EMBI	<i>Xenopus</i> Gene Collection
Xlaevis cDNA	Yes	ISH probes		
Predicted transcripts	Yes			
Unigene		Yes	Yes	
Gurdon clusters		Yes		
Proteins				
UniProtKB			Yes	
NCBI Refseq	Yes		Yes	
Other Tracks				
Start stop codons			Yes	
Transcription start sites			Yes	
tRNA			Yes	
ncRNA			Yes	
CpG islands			Yes	Yes
Restriction enzymes		Yes		Yes
Predicted promoters		Yes		
Affymetrix probes			Yes	
Repeats	Yes		Yes	Yes
BAC End Pairs	Yes			
Methylation tracks		Yes		
Other features				
Add track	Yes	Yes (DAS)	Yes	Yes
Alignment tools	BlastN	Blast (Xenbase)	Blast/Blat	Blat
Query tools			Biomart	Table Browser

Table 2

Genome Browser Cross-Species Comparison Features

Features	Ensembl	UCSC
Genomic		
Human	Yes	Yes
Chicken	Yes	Yes
Opossum		Yes
Mouse	Yes	Yes
Rat		Yes
Tetraodon	Yes	
Zebrafish	Yes	
Lizard		Yes
Genes		Transmap (30) Refseq Genes
mRNA	European Nucleotide Archive Vertebrates	Transmap mRNA Mappings
Protein	UniProtKB	Human