

Marseille, May 3rd, 2007

Welcome to the 3rd ANISEED newsletter!!

We are pleased to announce the release of ANISEED v3.0 (<http://crfb.univ-mrs.fr/aniseed/aniseedV3/index.php>), platform for Ascidian In Situ Expression and Embryological Data. As before the system is composed of three main modules:

The ANISEED database and web interfaces (<http://crfb.univ-mrs.fr/aniseed/aniseedV3/index.php>), the ANISEED genome browser (<http://crfb.univ-mrs.fr/ciona-bin/gbrowse/intestinalis/>), and the 3D virtual embryo (software downloadable at http://crfb.univ-mrs.fr/aniseed/aniseedV3/virtual_embryo.php).

The main features of the new release are:

Gene models from JVI v1.0, Kyotograil2005 and ENSEMBL are included, functionally annotated and linked to the corresponding EST and cDNA clones.

Thanks to a collaboration with Yutaka Satou, Nori Satoh, and the Japanese Ciona consortium, all Ghost data are now included in ANISEED including the zn finger screen of the Satou lab.

These data have been partially manually curated to take profit of the more detailed ANISEED anatomical ontology.

Literature articles start to be entered (30 articles currently entered).

All data can now be downloaded.

The system has a new skin....

To keep the system growing, the Aniseed team has also expanded. It is now constituted of Fabrice Daian (software engineer; daian@ibdm.univ-mrs.fr), Delphine Dauga (database curator and redactor of this letter; dauga@ibdm.univ-mrs.fr), Renaud Schiappa (contribution to the annotation of the in situ data and development of the embryo scheme; schiappa@ibdm.univ-mrs.fr), Pierre Khoeiry (development of the ANISEED Genome Browser; khoeiry@ibdm.univ-mrs.fr) and Daniel Sobral (development of Cis-regulatory region interfaces; sobral@ibdm.univ-mrs.fr). Moreover, five part-time undergraduate students annotate, interpret and enter ISH and Cis-regulatory sequence data into ANISEED.

You will find below a summary of the features and data introduced since the last letter to the community sent last year.

Our statistical analysis shows an increasing number of ANISEED users during the last months (more than 6000 connections in 2006) and we are very happy by this form of feedback. In addition, we would much appreciate direct comments or suggestions (send e-mail at aniseed@ibdm.univ-mrs.fr).

- **New DATA**
 - JGI V1.0 Genome assembly supported only.

In agreement with Yutaka Satou, we decided to stick to the JGI v1.0 genome assembly as v2.0 may not provide much improvement over v1.0.

➤ **New gene models supported**

The JGI v1.0 gene models are not bad and everyone in the community is used to them. Yet, they are not perfect as, for instance, they do not take all EST evidence in consideration. Over the past couple of years, three additional sets of gene models were released: JGI v2.0, Kyotograil2005, and ENSEMBL. We found that all these gene model sets are often complementary and therefore decided to support JGI v1.0, Kyotograil2005 and ENSEMBL gene models in ANISEED and the associated ANISEED browser. Each of these gene models was functionally annotated via the ANISEED pipeline with additional ENSEMBL functional annotation for this set. Each set of gene models was then clustered with available ESTs so that it can be linked to expression data.

Each gene model is linked to the ANISEED genome browser (see below) and to the external browser it comes from (Ghost, ENSEMBL, JGI).

➤ **Conversion of Ghost high throughput in situ data into ANISEED**

Thanks to a collaboration with Nori Satoh, Yutaka Satou and the members of the Japanese ISH consortium, all Ghost data have now been included into ANISEED. Currently, the high throughput ISH available on ANISEED total:

- Satou/Satoh Groups (Ciona tb-stage embryos; TF/SM screen; zn finger screen): 12 071
- Nishikata group (Ciona maternally expressed genes): 1904
- Kusakabe (Ciona larval stage embryos): 971
- Ogasawara (Ciona juvenile stage) : 894
- Fujiwara (Ciona cleavage stage embryos) : 4495
- MAGEST (Halocynthia embryos) : 5825
- Lemaire lab unpublished screen : 427

All these data are labelled high throughput on their gene expression web page to highlight that data quality may be inferior to small scale data.

➤ **Manual reannotation/curation of high throughput data**

The first step of the inclusion of Ghost data into ANISEED automatically inserts both annotations and pictures from the flat file provided in Ghost (<http://hoya.zool.kyoto-u.ac.jp/download.html>). As a second step, we partially reannotated these data to take advantage of the richer anatomical ontology proposed by ANISEED. Genes expressed in subregions of the epidermis or the nervous system can thus be readily identified. This partial reannotation first focused on the TF/SM expression data (Imai et al., Development, 2004) and is carried out by a team of Marseille undergraduate students (currently, most of the Imai data has been reannotated). In this reannotation process, we also decided to make no distinction between maternal and zygotic expression. Thus, maternally expressed genes are treated differently in Ghost and ANISEED and depending on what you are looking for you may want to scan one or the other database. Finally, most of TF/SM of the Ghost pictures has been

reprocessed so as to show for each stage a representative, oriented embryo picture. This reannotation/Curation effort will continue with the other classes of high throughput data.

A mention of the curation status is displayed for each entry:

- “Not curated”: Automatic conversion of Ghost into ANISEED format
- “Curated”: Manual reannotation/curation completed
- “Being manually curated”: Manual reannotation in progress.

➤ **Integration of curated small scale ISH data**

Thanks to the hiring of biological annotators, ISH annotation and pictures from the literature have started to be entered into the system, with the agreement of their authors who also provided high resolution image files (Thanks!!!). If you wish to enter your own articles, please send a message at aniseed@ibdm.univ-mrs.fr and we will open an annotator account at your name and grant you access to the ANISEED curator tool. In some cases, the ANISEED team can also enter your published data. All the articles present in ANISEED can be found at <http://crfb.univ-mrs.fr/aniseed/show-article.php>.

For each of these articles, a hub webpage (“article card”) displays the abstract, the pictures, a link to Pubmed and to each experimental data (see below).

➤ **Integration of curated small scale cis-regulatory region data**

Ascidians are a powerful model organism for Cis-regulatory analysis. We started entering cis-regulatory data from 14 articles. The display of this kind of information has been reorganised (see below). A list of all the regulatory region entered can be found at <http://crfb.univ-mrs.fr/aniseed/aniseedV3/reg-region-search-results.php>.

➤ **Morpholino sequences**

In each gene card a new button lists the sequence of published Morpholinos corresponding to this gene when available. This morpholino sequenced is linked to experiments in which the morpholino was used, and to the relevant literature references.

➤ **Embryo schemes**

To allow a more intuitive approach to the anatomy of embryos, we have developed interactive embryo schemes, on which you can click on individual cells to have access to a shortened anatomical tree, and to the fate, lineage, and biometry information on the selected cell.

Up to the 110-cell stage, the whole embryo is represented by a vegetal and an animal scheme. Between the mid-gastrula and early tailbud, the scheme represents only the neural plate and its derivatives.

You have access to these schemes by two different ways through the "Explore Anatomy" menu:

- Find an anatomical territory in the ontology
(<http://crfb.univ-mrs.fr/aniseed/aniseedV3/anatomy-search.php>)
- Find an anatomical territory on an embryo scheme
(<http://crfb.univ-mrs.fr/aniseed/aniseedV3/anatomy-search-embryo.php>)

TOOLS

- **ANISEED Genome browser**

<http://crfb.univ-mrs.fr/ciona-bin/gbrowse/intestinalis>

A Gbrowse-based Genome Browser was made public last year to display various information on the genome. The ANISEED genome browser is cross linked to the database. Information displayed includes JGI, KyotoGrail2005 and ENSEMBL gene models, TIGR EST clusters and predicted non-coding miRNA. It also displays the conservation profile between the two sequenced ciona genomes (*Ciona intestinalis* and *Ciona savignyi*), as well as Cis-regulatory sequences supported by ANISEED.

- **New or improved ANISEED web pages**

- **New layout**

As you probably discovered, ANISEED has changed its skin and organization to make its use easier. The menu is now on the top of each page and we tried to provide more informative titles than in the previous versions. This menu also allows you to access the ANISEED 3D virtual embryo software, to explore 3D models of embryos and the ANISEED genome browser to display gene models (JGI v1.0, KyotoGrail2005 and ENSEMBL), ESTs, conservation between *Ciona intestinalis* and *savignyi* etc...

- **Tutorials**

A list of tutorials will gradually be put online, with little movies to show you how you can access information in ANISEED.

- **Download section**

<http://crfb.univ-mrs.fr/aniseed/aniseedV3/Download.php>

All ANISEED data can now be downloaded from the download section. This includes the anatomical ontologies; fate, lineage and morphometry information; gene models with their functional annotations; EST/cDNA sets; molecular tools including morpholinos; cis-regulatory sequences; and Cis-reg/ISH whole mount annotations and pictures.

To exchange ISH data, with other systems we defined a new standard MISFISHIE-compliant XML format, the InSitu Markup Language. We have also defined a generic and complete import/export tool for ISH, supported by this new data format.

- **Anisearch**

Anisearch is a Google-like search engine, which scans all tables of the database for the desired information (this includes comment fields!). You can find Anisearch on the banner page and in the top right corner of each web page.

Just type in the keywords you are interested in and select the “and” and “or” or “complete sentence” operators to link your keywords. The results comes out classified according to various fields (GO term, molecules, ISH comments, articles ...) and cross linked to ANISEED and to the genome browser.

➤ **Gene model card page**

Eg: <http://crfb.univ-mrs.fr/aniseed/aniseedV3/molecule-gene.php?name=ci0100133709>

The gene model card page recapitulates all major properties of each gene model. This page gives access to the functional annotation (name, scaffold, orthologues in mouse, human and drosophila, best blast hit in Swissprot, interpro domains and Gene Ontology classification). At the top of this page, you can find different tabs corresponding to the sequence (nucleotide and aa), associated ESTs, EST counts in the various sequenced libraries, in situ data in WT and manipulated conditions, molecular tools (morpholinos, full length Gateway clones), microarray probes (to be implemented). The genomic annotations can be exported as a flat file via the export tab.

To access this page, you can click on the gene model name found on any page you come to. You can also use the “find a gene model by its name” search interface in the genes/cDNA clones section of the website. Alternatively, you can type the gene name in Anisearch.

➤ **Cis-regulatory region page**

Eg: <http://crfb.univ-mrs.fr/aniseed/reg-region-edit.php?id=2751807>

This page gives access to all data describing a cis-regulatory region and its activity. In ANISEED, a Regulatory Region is a genomic locus for which the regulatory potential has been tested. Regulatory regions for a given gene are arranged hierarchically to provide an overview of the various regions tested. This page also describes the identified regulatory motifs. While a regulatory region is an abstract piece of the genome, the system also presents the actual constructs that were experimentally tested and links these to the corresponding spatial and temporal reporter activities (stored in ANISEED as in-situ records). A construct thus contains a DNA sequence (one of the region's sequences), and a reporter gene (e.g. *LacZ*). It may also contain a standard basal promoter (chosen among the regions defined as such). A link will soon allow you to visualise cis-regulatory regions on the ANISEED Genome Browser.

To access the cis-regulatory region page for a given gene, you can use the “find a cis-regulatory region” in the genes/cDNA clones section of the home page. You can search by species, by region's name (e.g. searching "bra" will retrieve all regions that include "bra" in the name), by gene regulated, by publication (using PUBMEDID), by author or annotator name, or by date of entry.

You can also retrieve all cis-reg pages active in a given territory through the “Find cis-regulatory sequences active in a given territory” interface of the expression data section on the mainpage.

➤ **Article card page**

Eg: <http://crfb.univ-mrs.fr/aniseed/aniseedV3/biblio.php?pmid=9510546>

This page provides gives access all ANISEED data linked to a given article (abstract, authors, experimental data, molecules, morpholinos, etc...).

To find an article by author, gene name, etc..., use the “keyword search into article corpus” page in the “explore literature” section of the main page, which is essentially a pubmed-like search interface restricted to articles entered into ANISEED.

➤ **BLAST search**

<http://crfb.univ-mrs.fr/aniseed/aniseedV3/blast-search.php>

This page allows you to blast a nucleotide or protein sequence against all cDNA clones, Ciona or Halocynthia EST/cDNA clones, the Ciona gene models or the Ciona intestinalis JGI v1.0 whole genome assembly.

➤ **3D virtual embryo**

http://crfb.univ-mrs.fr/aniseed/aniseedV3/virtual_embryo.php

This software, which was already present in ANISEED V2.0 uses interactive three dimensional digital embryo reconstructions to display information from the database as well as to enter into the database quantitative descriptions of the geometry of individual embryonic territories and their interactions. It has been improved to avoid crashing when too many embryos were successively visualised.

Available embryos cover the following stages: 2-, 4-, 8-, 16-, 24-, 32-, and 44-cell for *Ciona intestinalis* and 32-cell for *Halocynthia roretzi*. Ongoing work is carried out to continue the collection of reconstructed embryos through gastrulation.

➤ **Archived V2.0 database**

In some circumstances, you may want to get back to the results you obtained while searching the ANISEED V2.0 database. This database has not disappeared: it will remain accessible at the following address. Note however that this is a freeze: the data do not evolve anymore and all new V3.0 data will never be included in the archived freeze.

FUTURE PROSPECTS.

Most of the hard core development of the ANISEED system is now complete and the future versions will mainly focus on data collection. For this your input will be crucial, for instance

by entering your own published data into the system. In many labs, lots of data that are not really crucial and hence never published are generated while looking for precise features (i.e. validation by ISH of biocomputationally predicted expression profiles, or of microarray data). These data which may not be interesting for you may be of crucial interest for someone else in the community, and vice versa. By requesting an annotator account at aniseed@ibdm.univ-mrs.fr, you will be able to actively contribute to the project.

Delphine will also be happy to visit your lab (Yeah, travels, people!!!!). Please contact me at dauga@ibdm.univ-mrs.fr if you would be happy to host me to have a tour of the system and its submission/curation tools. The Lemaire lab will cover travel expenses, but I would appreciate if you could contribute to my local living cost.

Some more minor development will also continue. For instance, we are working on the “gene concept” that integrates all gene models for the same gene. In cases of transcription factors, we are integrating the expected DNA-binding specificity either determined experimentally or inferred from Jaspar. We will also establish better links to established databases in the field (eg. the Ascidian Body Atlas by Kohji Hotta, or TBTGR by Take Kusabake). Some version 2.0 tools (eg. STEPBLAST, search for co-expressed genes) are currently being redesigned and will come online later this year.

And of course your suggestions for further developments are most welcome!!! Do not hesitate to ask us (aniseed@ibdm.univ-mrs.fr) if you cannot place a request you would like on ANISEED. Several people in the community asked us such requests and in most cases, we could explain how to use the existing search tools to place the desired often complex request.