ANISEED CURATOR MANUAL

How to enter data into ANISEED?

Delphine Dauga, Fabrice Daian and Patrick Lemaire

IBDML, UMR6216 CNRS/Université de la méditerranée,
Campus de Luminy, case 907
F-13288 Marseille cedex 9
France
Contact: aniseed@ibdm.univ-mrs.fr

March 2010 version

INTRODUCTION

In parallel to the further development of the ANISEED system, a major issue is to optimise data collection. For this, your input is crucial.

For example, a large fraction of experimental data collected in your lab will probably remain unpublished. These data, which may not be very interesting to you, could however be of crucial interest for someone else in the community, and vice versa. Also, entering data into ANISEED is a great way to back up high-resolution data including images, to organize data, to compare data with those obtained by other groups… And your data can stay private if you wish, so that no one else can see them. Not even our curator team.

Entering published literature is also critical, for the system, and for the visibility of the papers. You have probably already been in the situation in which you remember that someone carried out a crucial experiment for your project. But where was it? Who did it? What was precise result obtained? Scanning Pubmed will not give you the answer if it is not in the article's abstract but a simple interrogation of ANISEED will return what you are looking for…. That is, provided the paper has been entered in the database. Conversely, any paper entered into ANISEED has an increased impact because it pops up when looking for regulators of a given gene, or genes involved in a process, or lineage tracing experiments, etc…Even if the result is mentioned in the main text, but not the abstract. Even if it is in a supplemental figure!

This manual will help you enter your own data, unpublished or published, into the system using the submission/curation tools.

TABLE OF CONTENT:

a. Example: describing the pattern of activity of a transgene containing the -1135/-840 regulatory region of FoxF, at tailbud stage

V) ENTERING A PUBLISHED ARTICLE

a. Example: entering the paper from Yasuo et al. 2006, FGF8/17/18 functions together with FGF9/16/20 during formation of the notochord in Ciona embryos.

VI) FINDING EXPRESSION DATA

1. Finding my expression data
    a. Example: you want all the in situ data you have entered concerning Ciona intestinalis 110-cell stage embryos
    b. Example: you want all the in situ data with genes expresses in muscle but not in mesenchyme.
    c. Example: you want all the in situ data you have entered stained with Brachyury in manipulated conditions
    d. Example: you want all the in situ data you have entered concerning the paper from Yasuo et al. 2007
    e. Understanding the results

2. Finding public expression data

VII) FINDING REGULATORY REGION

VIII) FINDING ARTICLE CARDS

IX) DISPLAYING ARTICLE CARDS

X) DISPLAYING MOLECULAR TOOLS

GENERAL RULES

1. Authentification:

- For privacy, each user is requested to get an individual password before entering data. Please contact Delphine Dauga, the Aniseed curator at dauga@ibdm.univ-mrs.fr or aniseed@ibdm.univ-mrs.fr
- By default, you will appear as author and annotator for all data you enter, with the date of entry at the top of each entry page. If you wish to enter your buddies data, you can also do so. You then need to indicate the name of your buddy in the Author field.
- For security reasons, you can edit, but you can't delete entries. If you make a mistake (you will….), contact the curator who will delete the erroneous entry.

2. What can I submit?

- Expression patterns during normal development as well as in response to experimental manipulation (embryological or molecular), morphological phenotypes, cis-regulatory elements, molecular tools (inhibitors, constructs, morpholinos).
- Unpublished data can be entered independently of one another (cf sections 4, 5, 6). For published data, you need to adopt a more rigorous format to make sure you are not forgetting pieces of information. This is achieved via the "*article card*" concept (section 7), which ensures that your article has the best visibility.
- The system only accepts clones (ESTs, cDNAs) that have been submitted to Genbank. If this is not your case, contact the curator to find a solution.
- The developmental stages are as in the standard ascidian developmental table (Hotta et al 2007 Developmental Dynamics), except that ANISEED compresses tailbud stages into 4 stages only (initial, early, mid, late).

3. Can I keep some data "private"?

- All new data are "*Private*" and "*Not curated*" by default, meaning that they cannot be viewed from the public website or by the Aniseed curation team (but your lab could see them). When you agree that your data appear on the public website, you will click on the "*Submit to curation pipeline*" button in the curator tool. Your data will not yet be available on the public website. They will be put in the curation pipeline ("*Destined to be published*"). After verification by the Aniseed curator, your data will be "*Available on the public website*" and "*Curated*".

4. What should I prepare before entering data?

- Before entering expression data, prepare the images, the clones IDs used for the experiments and the JGI Version 1.0 sequence and the true one for regulatory region. The system accepts high-resolution jpeg and png files, **not** gif and tiff. For consistency, the orientation of images should be standardised: anterior is up till the late gastrula stage; then to the left from the early neurula stage onwards. Pictures are not compulsory but they provide useful information.

II) LOGIN TO ANISEED

- Access the Aniseed web interface : http://crfb.univ-mrs.fr/aniseed/CURATOR/index.php and click on the "*login*" link.

- Login using your login name and password (obtained from the curator, see above) and click on the "*submit*" button.

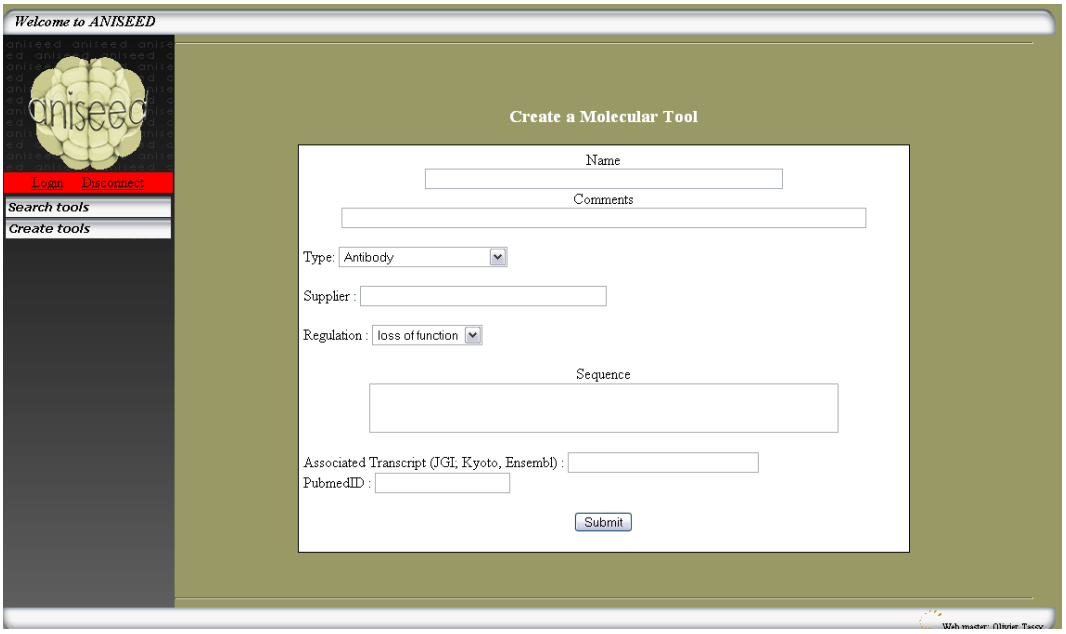
III) DEFINING THE TOOLS AND CONSTRUCTS USED IN AN EXPERIMENT

1. Creating a molecular tool

Aniseed supports the description of expression patterns and phenotypes in response to perturbations of gene function. To define these perturbations, it uses the concept of "Molecular tool" which can be morpholinos, overexpression contructs (Electroporation or mRNA injection), or pharmacological reagent. Each molecular tool is linked to the gene(s) that it regulates and to all articles and experiments where it was used/described, if published.

c. Example: Morpholino against FGF9/16/20, from the lemaire Lab

- To enter a molecular tool, choose "*Create a molecular tool*" from the menu of the left hand aniseed Curator "*Create tools*" menu section. An empty "*create a molecular tool*" form appears:



- Enter the "*name*" of the molecular tool. For morpholino, specify the person who designed it or the laboratory it comes from. *E.g. FGF9/16/20-MO (Lemaire Lab)*

- Enter a description of the tool. *E.g. morpholino against the ATG region of FGF9/16/20 designed in the Lemaire lab.*

- Choose the "*type*" of the molecular tool from the menu by clicking on the arrows and scrolling down. The molecular tool could be antibody, morpholino, construct, chemical inhibitor, mRNA, RNAi…

- Enter the "*supplier*" which designs the molecular tool if know it. *E.g. Gene Tools for a morpholino or the lab for the construct*

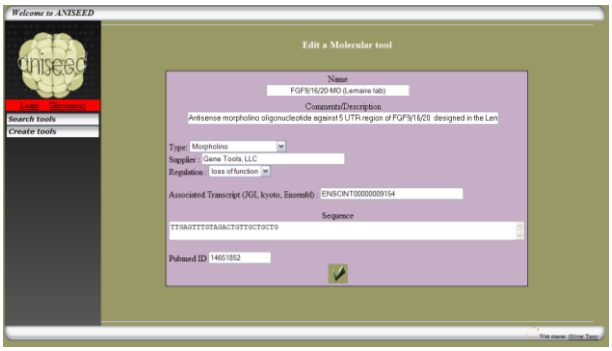
- Choose the "*type of effect*" of the molecular tool from the menu by clicking on the arrows and scrolling down. It could be a loss of a gain of function. *Here loss of function*

- Enter the "*sequence*". This is usually irrelevant for a whole overexpression construct, but is critically important for a morpholino. Also, in case of the overexpression of a mutated ORF, this information may be crucial if the mutation has not been previously published.

- Enter the "*Transcript model*" (JGI, Kyoto or Ensembl) which is deregulated by the molecular tool. *Here ENSCINT00000009154 (Ensembl transcript model)*

- Enter the "*pubmed ID*" of the article where the tool was first described. *Here the Pubmed ID associated to Vincent et al. 2007 is 14651852*

- Valid your entry by clicking on the *green tick*.



You are now done. You could in a similar manner enter an overexpression construct, in which case the comment should be more precise: which driver? Which ORF? Mutated or not, if mutated how, etc….

d. Example: electroporation construct mesp::FoxF-VP16

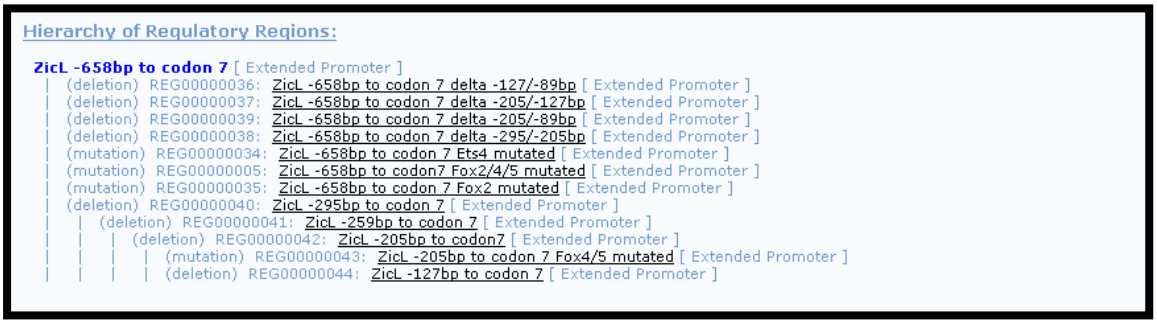
Same steps to enter constructs.
Note that the colour change according to the type of molecular tool ;-)

2. Defining cis-regulatory regions

a. Definitions

In ANISEED, a **Regulatory Region** is a **segment** of DNA that has been tested for its capacity to drive expression in vivo. It can have a *natural* (genomic) or an *artificial (mutations, synthetic sequence)* **origin**. The 5' and 3' boundaries of the sequence are defined by a set of **coordinates**. In ANISEED, we normalize a region's coordinates with respect to the expected mature mRNA start site (noted +1). For some regions, the ANISEED coordinates may thus differ from the coordinates defined in the **publications** describing them. The comments field allows to indicate this. By definition, a region is represented by its sequence on the JGI version 1 browser, even if the individual sequences tested in electroporation constructs differ from the "canonical" sequence because of polymorphisms. This info is not lost but will be included in the description of the "construct" used to test the region.

Regulatory regions are organised hierarchically in ANISEED. A regulatory region may thus derive from a **parent** region by *deletion*, *mutation*, or/and *other artificial modification* (e.g. addition of a binding site).



A region has a specific **regulatory activity**. ANISEED classification of these activities is based on the ongoing Sequence Ontology, and currently includes the following concepts:

- *basal promoter* (only includes sites for the basal transcription machinery)
- *extended promoter* (basal promoter plus other regulatory elements, this usually corresponds to a 5' flanking region)
- *enhancer* (a unique enhancer without basal promoter)
- *silencer* (a unique silencer without basal promoter)
- *insulator* (a unique insulator without basal promoter)
- *complex regulatory region* (several enhancers or silencers, etc… but no basal promoter)
- *No Activity*
- *Unknown activity* (Other cases we did not think about…..)

A region may contain **regulatory motifs**. These are (usually short) sub-sequences of a regulatory region that play a role in the region's capacity to drive expression (e.g. transcription factor binding sites). Their location in the region's sequence is defined

by **motif coordinates**. A regulatory motif is an instance of a **motif class,** which represents the high-level abstract notion of a motif (e.g. GATA motif). Each individual motif can be associated to the **upstream transcription factor(s)** that binds it, when known.

Finally, the regulatory activity of a region is tested through **constructs**. These are **molecular entities** used to store spatial and temporal data of expression assays (stored in ANISEED as **in-situ** records). They contain a **sequence** (if possible the precise one that was tested, if known, otherwise its JGI v1.0 representative) and a **reporter element** (e.g. *LacZ*). They may also contain a **basal promoter** (chosen among the regions that are defined as such), if the region tested does not contain an endogenous promoter.

In this case, all regions are fully described in the paper. It may happen that a region being described in a paper derives from a region described in another paper. One should follow the dependencies of regions until an "original" region from which all others derive. Each region should be associated to the publication(s) where they are described. Entering data from one article may therefore sometimes require entering a previous article if they are not yet in ANISEED (we hope this should soon become rare…).

b. Example: two Ci-RAR cis-regulatory regions (Fujiwara et al, 2005)

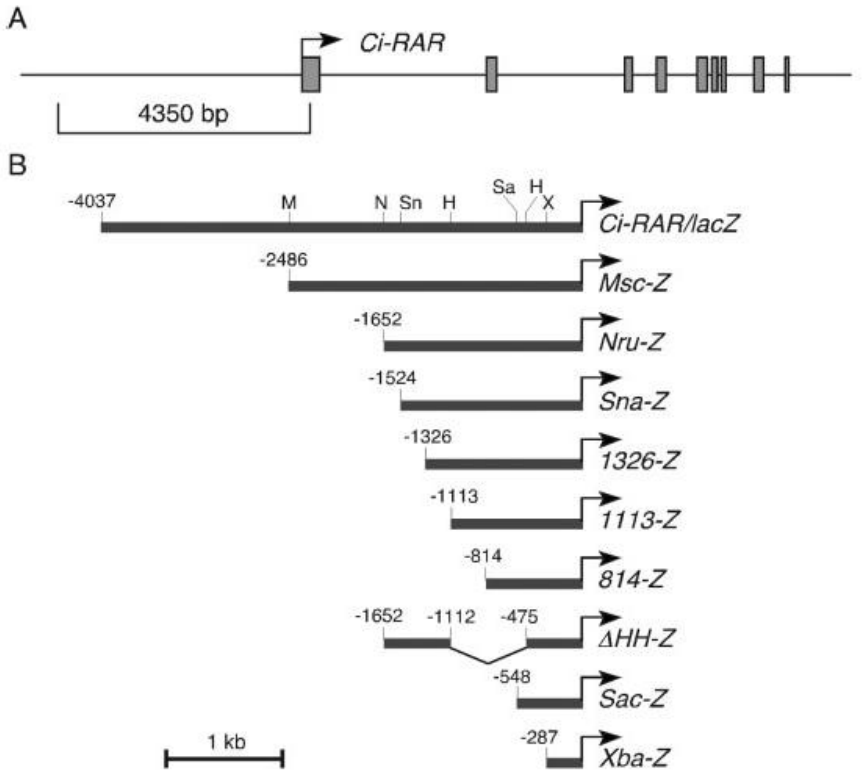
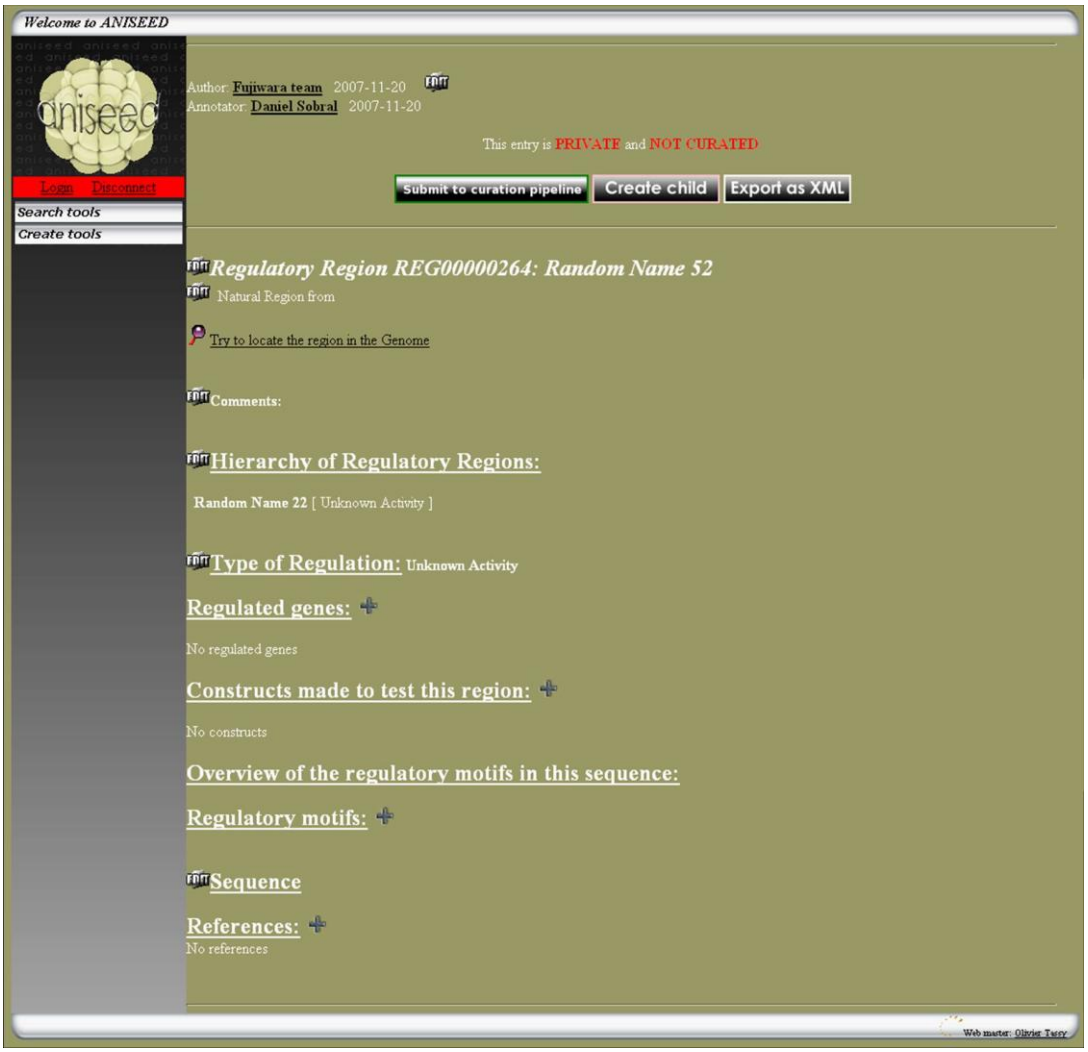


Fig. 1. The 5′ flanking region of the *Ciona intestinalis* retinoic acid receptor gene (*Ci-RAR*). **A:** The genomic structure of the *Ci-RAR* gene. Exons are indicated by boxes. The transcription start site is indicated by an arrow. **B:** Diagrams show different 5′ flanking regions used in this study. Names of transgenes are on the right side. Restriction sites and nucleotide positions are indicated. H, *Hin*dIII; M, *Msc*I; N, *Nru*I; Sa, *Sac*II; Sn, *Sna*BI; X, *Xba*I.

We will take the example of the largest cis-regulatory region defined in Fujiwara et al. 2005 (-4350bp), and then of the child region defined by the construct Msc-Z. As you will see, the process is simple as long as the precise coordinates of the region are precisely defined. Depending on the information available this may need some thought….

i. *Step 1: creating a parental region*

To create the region, you just have to go to the "*Create a regulatory region*" from the menu of the Aniseed "*Create tools*" section. An interactive page appears displaying the new region (entry *Private* and *Not curated* by default, cf general rules section). In this page you can edit all fields to further describe the region.
As indicated in the general rules, a new entry is by default "*Private*": it is not posted on the public website and cannot be viewed by the Aniseed Curator and other annotators outside of your own research group. To release it to the curation pipeline, tick on the "*submit to curation pipeline*" button. The data now appear as "*destined to be published*", which allows visualisation and validation of your data by the curator. After curation, your entry will be posted on the public website.



Here is a description of the different fields:

- **Author**: by default, the author and the annotator of the entry will be you. But you can modify it by clicking on the "*edit*" button. So, the author can be you or the first author of a paper if you are entering published literature. Avoid indicating simply the lab.

- **Name**: by default, a random name will be given to the region. You have to modify it by clicking on the "*edit*" button. The name of the region… must be **unique**! Remember the name of a region is not the name of the construct that was used to test it….. We ask you to respect the following format for region names: ci-"gene name" "start of region "bp" end of region "bp *e.g. ci-ZicL -4037/313.* No reporter indicated, as this is a region, not a construct!!! Do not add "+" in front of 313 as this will interfere with the display of the full name of the region on the Aniseed browser.

- **Origin**: the region is either Natural (if it can be associated to the genome of a given Species) or Artificial (in which case the field Species is ignored).

- **Comments:** enter a general comment describing the region, the different modules and motifs that are important to regulate the gene.

ii. *Step 2: adding type of regulation, regulated gene, and constructs*

- **Type of Regulation**: type of regulatory activity displayed by this region.

- **Gene regulated (when known):** use a transcript model to describe the gene (*e.g. ci0100147513*).

- **Constructs:** constructs are used in reporter assays to test a given region. We distinguish the concept of regulatory and construct so that several constructs can be associated to one same region (e.g. by changing the reporter gene or by changing the basal promoter). Also, this distinction will

permit you to associate other types of evidence to a regulatory region (e.g. IP, gel-shift, etc...).

- To create a construct, you just need to provide a name (give a name what is different from that of the region). If possible, use a name derived from the construct name in the paper or add this name to the comment field. You should add a "p" (for plasmid) in front of the name, the basal promoter name and the reporter *e.g. pfog -214/-74 pbra::NLS LacZ.*
  After entering the name, provide basal promoter end reporter gene information. You can also provide any comments relatively to this construct. Thus, to add new basal promoters you just need to create a region with a "*Basal Promoter*" type of regulatory activity.
  The sequence tested in electroporation constructs could be enter in the "*sequence associated to this construction*" box even if this one differ from the "canonical" sequence because of polymorphisms. In the case of Ciona, which has a highly polymorphic genome sequence, the sequence actually tested differs in most cases from the JGI V1.0. This information can be very important to identify functional transcription factor binding sites.

The activity of a region is described through the activity of a construct following electroporation. This is done in form of an "in situ record" (cf section IV/2). Below there is an example of a reporter assay:

iii. *Step 3: adding sequence, motifs and references*

- **Sequence**: DNA sequence of the regulatory region… This should always be the JGI v1.0 version of the sequence, NOT the sequence that you cloned or that was described in the article you are entering. These sequences should be added the constructs that were used to test the region. This strategy may sometimes create some difficulties: i.e. the JGI sequence has a different length than the reported sequence, or lacks a TF binding site, but it is the best way we found to simplify future updates when the genome sequence will improve. The **End** coordinate of the region is calculated on the basis of the length of the sequence (note that the position 0 does not exist, coordinate go from -1 to +1!). After entering the sequence, you can visualize it in the genome browser by clicking on the "*Try to locate the region in the Genome*" button. Thus, you can check if the sequence you have entered is the good one.

- **Coordinates**: an integer that describes the beginning of the sequence (5' end). Usually is the number of base pairs upstream the Transcription Start Site (+1). This number may differ between papers (e.g. by a different definition of +1) and such incoherence should be noted in the comments.

- **Comments**: any particular remarks regarding the region, namely how it was obtained, if there are conflicts between different papers, etc...

- **Motifs:** a motif is a small sequence of DNA that has an experimentally role in the regulatory ability of the region (*i.e. through mutational analysis*). It is defined by its abstract **class** (e.g. GATA binding motifs), by its relative

position in the region (thus defining its specific sequence) and by the genes that bind to the motif when known. Its position is always defined 5'-3' in the region. **Comments** related to this particular motif can also be added. New classes of motifs can be added in the link "*Create a new Regulatory Motif*" in the CURATOR page of ANISEED.

After adding motifs, their overall localization in the sequence can be visualized. A graphic displays the overall disposition of the motifs in the region. They are also highlighted in colour in the sequence.

- **References:** references are simply added by giving the Pubmed ID of the articles.

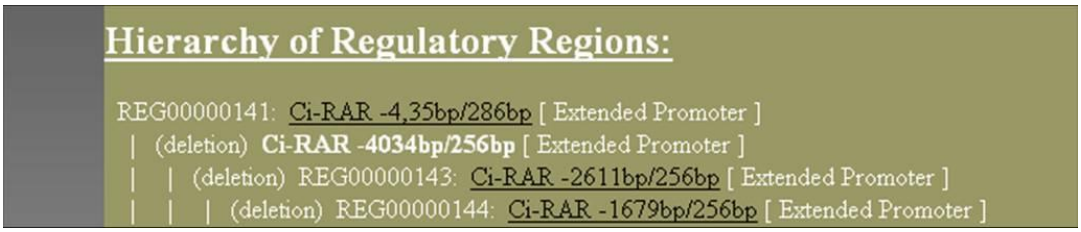
iv. Step 4: creating a child region

Now that we have created the base region from which all others derive, we can start creating the derivatives. For this, we can create children of this region, by pressing the "*Create Child*" button. This will make a new child region, clone of the current region. All data is cloned, except the type of modification, which is cleared. Constructs are also not cloned. Since the name must be unique, the parent region's name is used with a random extension, which should be modified.

After creating a child, you need to edit it to reflect the changes from the parent region, as the name, the type of modification, the sequence and the motifs (if relevant). When you enter that the relationship between parent and child regions is a mutation or *other artificial modification,* the region is automatically declared *artificial*. This does not happen in the case of *deletion*, since it remains a part of the genome. A mutation usually disrupts motifs, in which case the proper motif should be deleted. In the case of internal deletions, the region's coordinates may be affected. A possible option to keep the same coordinates is to replace the deleted base pairs by some symbol (*e.g.* "-"), although no particular.

The full hierarchy of dependencies between regions is displayed in the editing area. The current region is displayed in white, while the others are in black (with links to

them). You could change the dependencies between regions by clicking on the "*edit hierarchy*" button and add the parental region you would like to associate to the region edited. For that, you have to enter the identifier of the parental region in the corresponding box (*e.g. REG00000141*).



IV) ENTERING INDIVIDUAL EXPRESSION DATA

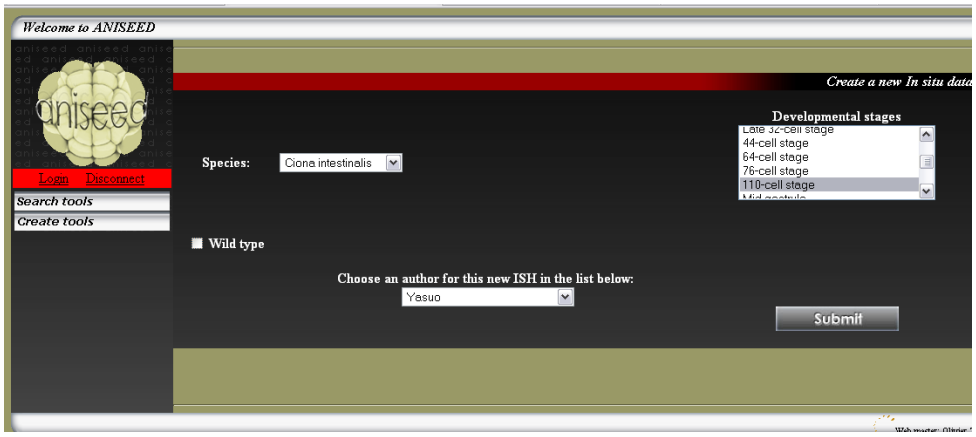
1. Expression pattern of a gene in response to the injection of a morpholino against another gene

a. Example: expression of Ci-bra by In Situ hybridization at 110-cell stage, derived from an egg injected with an FGF9/16/20 morpholino

Let's imagine you want to enter an ISH experiment showing a 110-cell stage embryo, derived from an egg injected with an FGF9/16/20 morpholino, and probed for Ci-Bra expression. You will have to go through several steps to enter this experiment. The specific choices you will have to make to enter the pattern above are in pink in the text.

ii. Step 1: creating an in situ data card

- You first need to create a new in situ expression data card. For this choose "*Create an in situ data*" from the menu of the Aniseed "*Create tools*" and the "*In situ: step 1*" window will open:



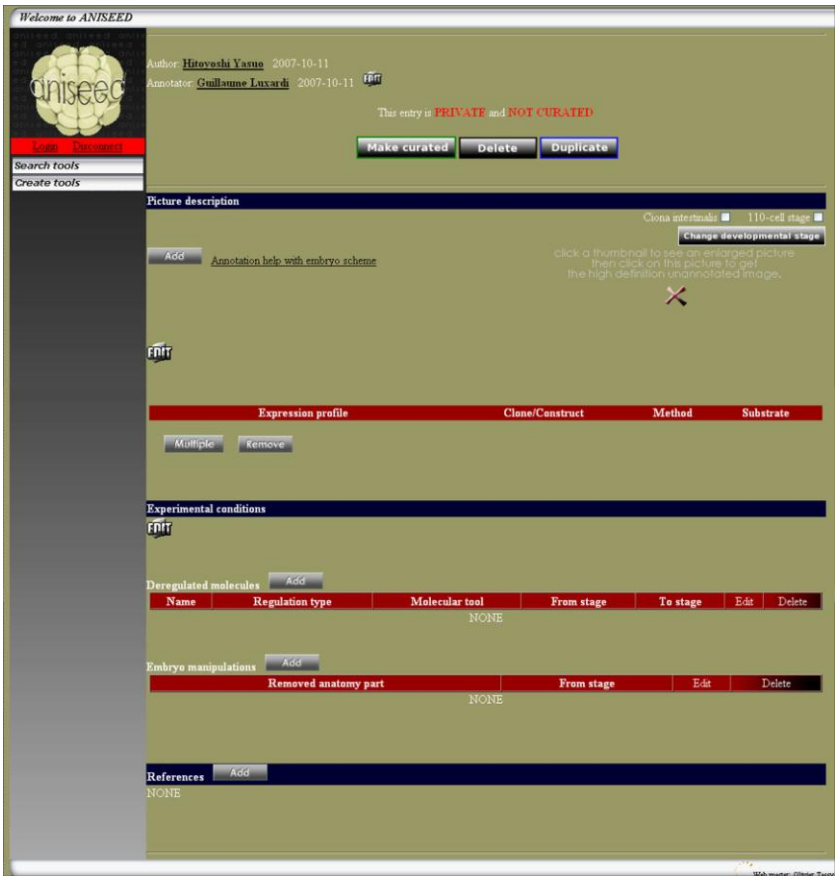
- Choose an Ascidian "*species*" from the menu by clicking on the arrows and scrolling down. *Ciona intestinalis* and *Halocynthia roretzi* are currently supported. *Here select Ciona intestinalis*

- Choose a "*Developmental stage*" according to Hotta et al. 2007. Click on the arrows and scroll down. (Developmental stages will not be listed unless you have chosen an Ascidian species). *Select 110-cell stage.*

- Specify if the experiment is in wild type conditions or not by ticking on the "*wild type*" box. *For our example, do not tick" wild type".*

- Choose an "*Author*" for this new ISH by clicking on the arrows and scrolling down. If you enter data from an article, choose the first author. If you enter your data or data from an author who is not present in the list, contact the ANISEED curator who will add the name in the database. Then, click on the "*submit*" button.

- A new page appears, the "*in situ data card*", on which you will complete the annotation: you will add and describe embryo picture(s) and specify the experimental conditions (see below). At the bottom of the page, you will be able to link your entry to one or several references. This is not relevant for unpublished data, but becomes crucial when dealing with published material.



- As indicated in the general rules, a new entry is by default "*Private*": it is not posted on the public website and cannot be viewed by the Aniseed Curator and other annotators outside of your own research group. To release it to the curation pipeline, tick on the "*submit to curation pipeline*" button. The data now appear as "*destined to be published*", which allows visualisation and validation of

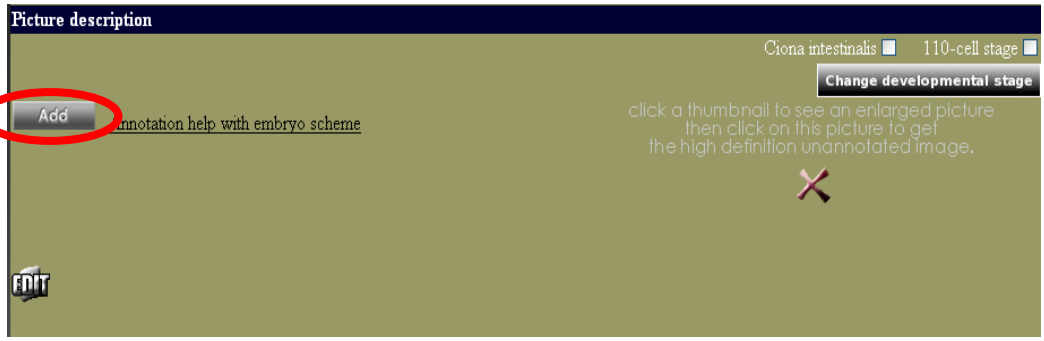
your data by the curator. After curation, your entry will be posted on the public website. *In our example click submit to curation pipeline.*

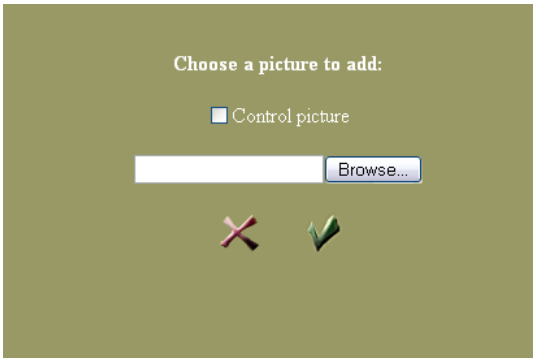
- You will also find at the top of the page, the "*duplicate*" button. It allows you to create a new entry with the same information as the original (species, stages, clones, reference, molecular tools, etc… After creation of the "copy", you can modify the features as necessary. This is a great help when entering a series of data with the same clone but at different developmental stages, in different experimental conditions, etc…..

iii. *Step 2: importing picture(s) and adding comments*

If you have a picture for the experiment reported, you should import these image(s) and if you want, you can add a comment to describe it. Although adding a picture as supporting evidence is highly recommended, it is not compulsory and not always necessary (e.g. no expression, ubiquitous expression, etc…). Remember pictures should be in JPEG, png, but not Tiff.

- To select and load your image into the database, click the "*add*" button, select the required file and validate your choice by clicking on the *green tick* or cancel by clicking on the *red cross* in the pop up window. For wild type expression patterns, ignore the "*control*" box. In contrast, when entering pictures describing an expression pattern in an experimentally manipulated embryo, you HAVE TO include control image(s) (WT expression patterns) in your data entry and tick the "*control*" box to indicate this is the associated control experiment. Linking control and experimental data is key to the reconstruction of trustworthy gene regulatory networks. The "*Edit button*" allow you to add a comment: you should specify the angle of view of your image(s), the treatment used for the experiment and, when applicable, the difference between wild type and manipulated expression. The comment will appear on the thumbnail summary of the ANISEED expression card and is scanned by ANISEARCH.



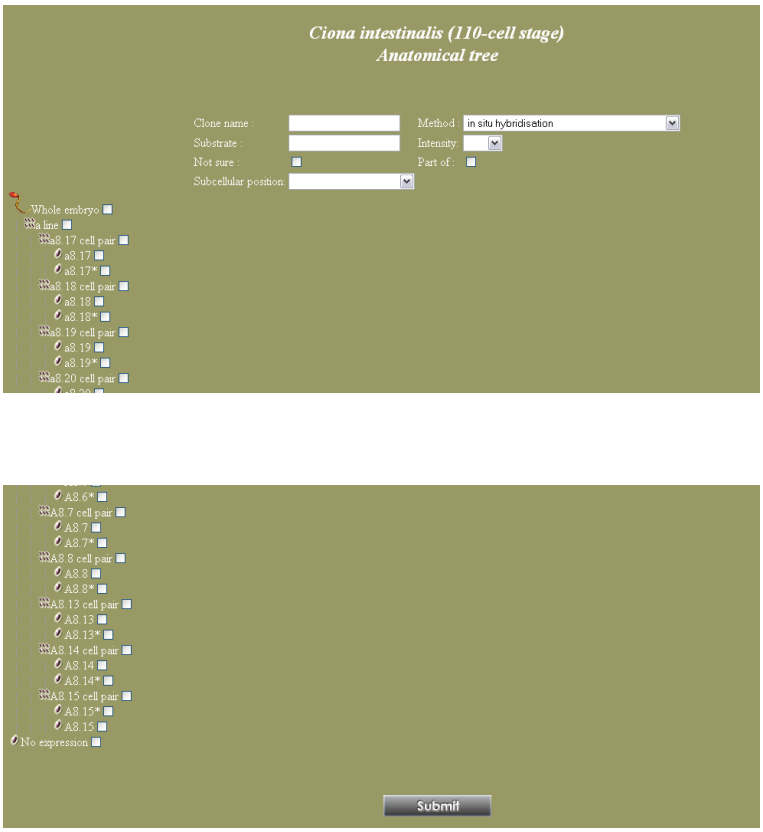


For our example, first add the control picture by clicking on the "add" button and by ticking the "control" box. Then, add the experimental picture by clicking on the "add" button. Refresh your page.

iv. *Step 3: adding description of the expression profile*

Next, you will describe the territories of expression of your gene of interest using the ANISEED controlled vocabulary.

- Click on the "*add territories of expression*" button to access to the annotation page, on which you will specify the clones used in the ISH experiment, the type of experiment, and the territories labelled.



- Clones are identified using either Genbank accession numbers or clone numbers (Satoh and Lemaire library clones). The system only accepts clones that have been submitted to Genbank. If you know the clone ID, enter it in the "*Clone name*" field (eg: *cicl50d03 in our case for Brachyury*). If seeing all possible clones helps you remember which clone was used, go to the *Aniseed menu/Search tools* and click on the "*gene/clones*" tab in which you can search for a clone by the biological name of the corresponding gene and retrieve the clone ID. When the correct clone ID is found enter it in the "*Clone name*" field.

- Choose the "*Method*" used from the menu by clicking on the arrows and scrolling down ("*in situ hybridisation*", "*immunohistochemistry*" or "*reporter gene*"). *In our case use "in situ hybridisation".*

- Select "*Intensity*" using the "*+*". The following stain intensities are given based on room temperature revelation, +++ (less than 6 hours at 37°C for an intense signal), ++ (7 to 36 hours), + (longer than 36 hours). This is quite rough. *In our case leave default as the gene is not expressed following FGF9/16/20 MO injection.*

- Enter the "*Substrate*" used in the experiment (*e.g. NBT-BCIP*).

- Tick the "*Not sure*" box if you have a doubt concerning the localisation of the staining. *Do not tick in our case.*

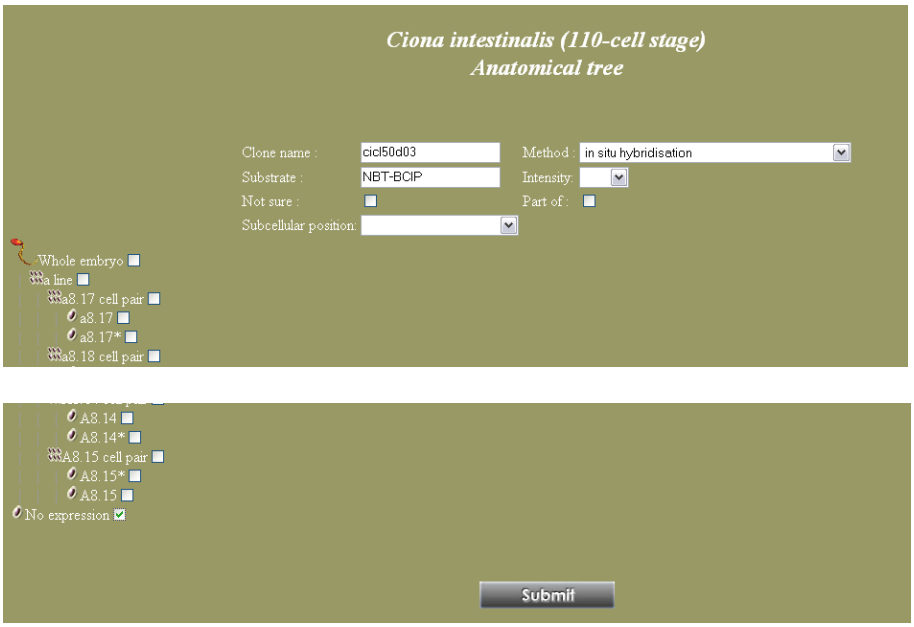
- Tick the "*Part of*" box if the staining is restricted to part of a territory. *Do not tick in our case.*

- Select "*subcellular localisation*" of staining if it is restricted to part of cell (ex: *perinuclear, endoplasmic reticulum, membrane, CAB*). *Leave empty in our case.*

- The anatomical dictionary corresponding to the developmental stage of your experiment allows you to determine precisely the localisation of the staining. Tick the anatomical part(s) stained. To account for no expression a "territory" called "no expression" has been added. *Tick this one in our example.*

- Click on the "*Submit*" button.

  In summary, in our example, the clone used is cicl50d03; the method, in situ hybridization; the substrate, NBT-BCIP and no expression is observed in the embryo.



v. *Step 4: adding experimental conditions*

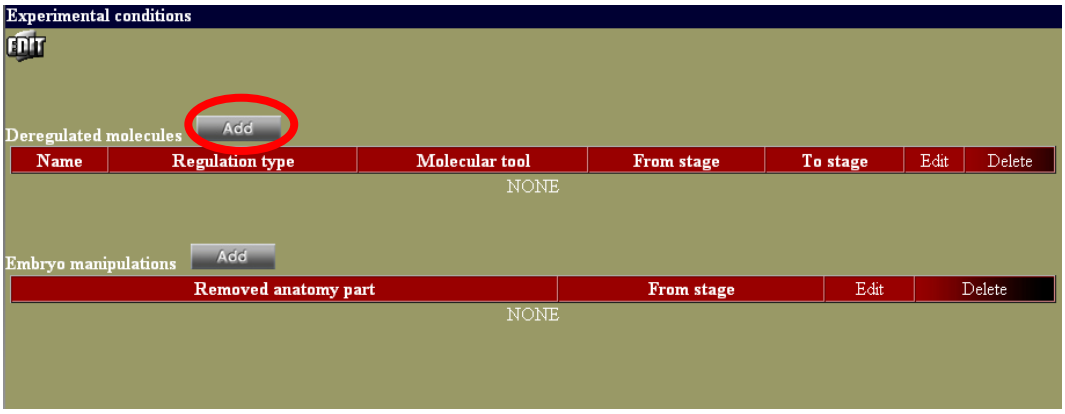
The system allows submission and annotation of wild type expression data and experimental expression data (in response to embryological manipulation or molecule deregulation).

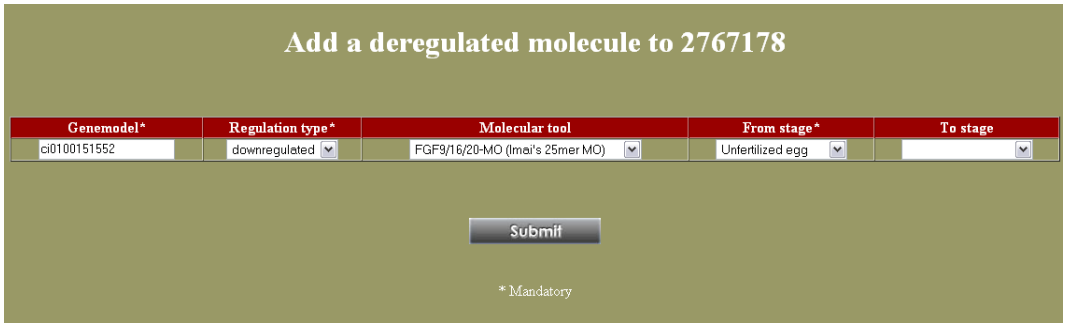
- *Wild type expression patterns:* optionally, you can add a comment by clicking on the "*edit*" button. This can relate to the origin of the embryos,….

- *Experimentally modified expression patterns:* you can describe both molecular and embryological perturbations. You could also add a general comment on experimental conditions by clicking on the "*edit*" button (*e.g. concentration of the morpholino used*).

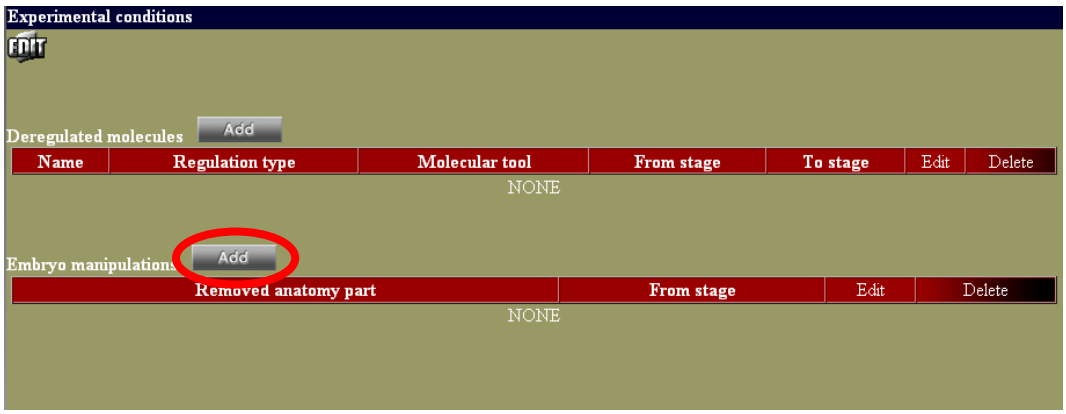
    - Deregulated molecule: click on the "*add*" button to specify the features of the deregulation. Enter the transcript model, *ci0100151552 in our case* (use JGI gene model if possible. Do not use an ANISEED gene model v3.0 IDs here), the type of regulation (*downregulated*), the molecular tool used to deregulate the gene (*FGF9/16/20 Imai's 25 mer MO*), and the stages when the perturbation started and ended (*unfertilised egg to 110-cell stage*). If the molecular tool doesn't appear in the list, you have to create it using the "*Create molecular tool*" from the menu of the Aniseed "*Create tools*". Then, click on the "*submit*" button.

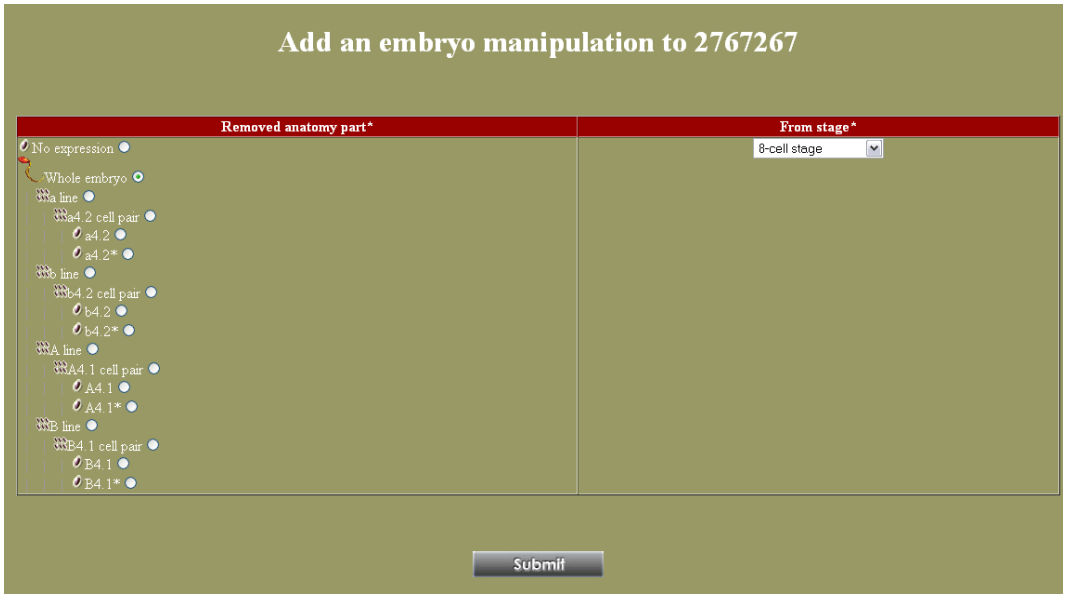
      For our example, FGF9/16/20-MO was used to deregulate FGF9/16/20 function and was injected into unfertilized eggs.



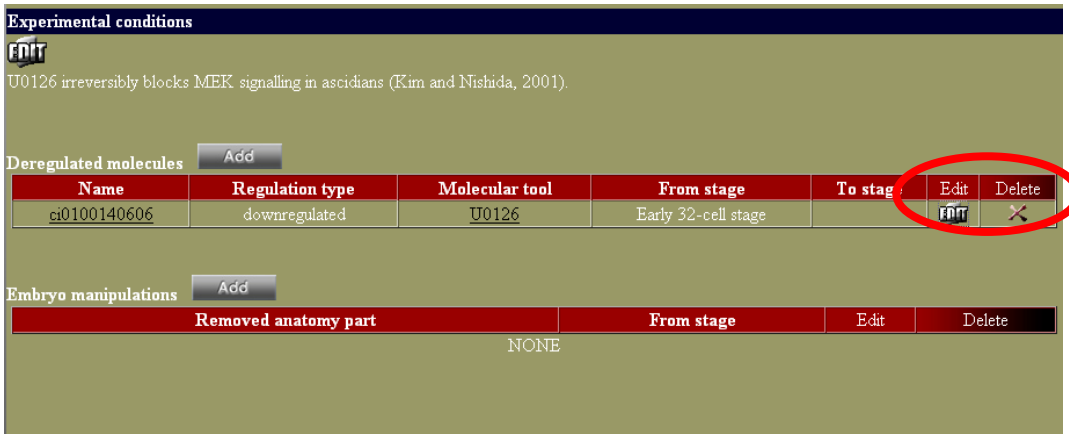


- Embryo manipulation: click on the "*add*" button, select the developmental stage when the embryo was manipulated by clicking on the arrows and scroll down. The corresponding anatomical tree will appear on the left of the page. Choose the removed anatomy part and click on the "*submit*" button.



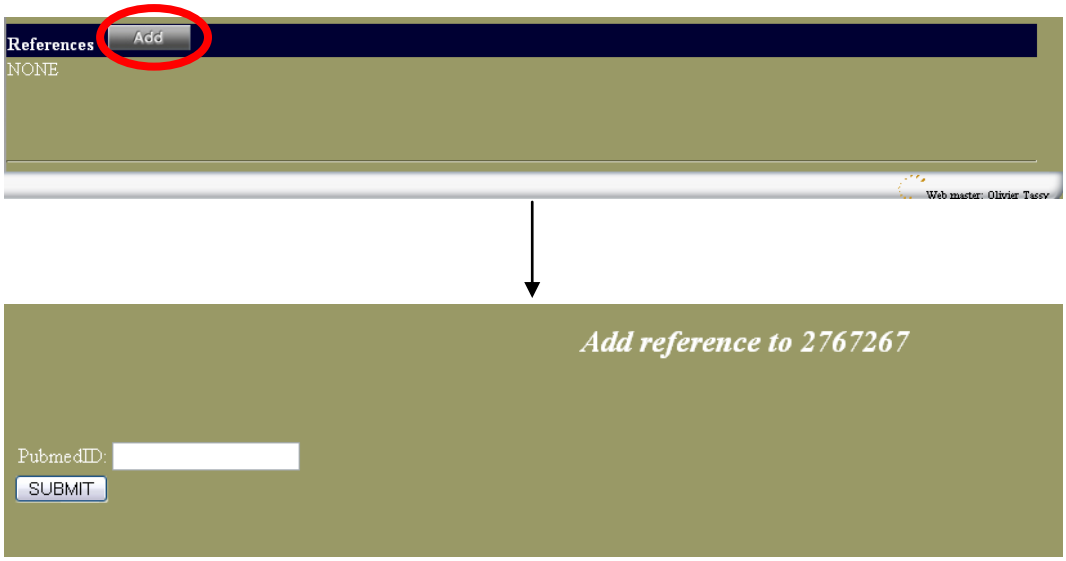


You can edit or delete features you have entered by clicking on the "*edit*" button or on the *Red Cross*.

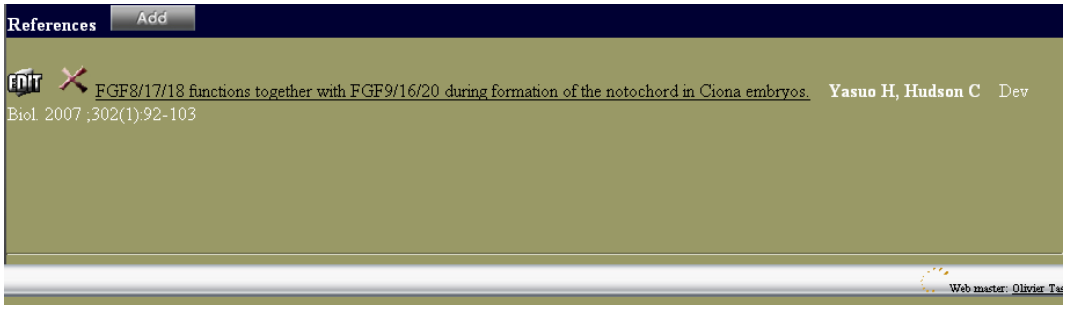


vi. *Step 5: adding references*

To enter reference(s) associated to the experiment if applicable. Click on the "*add*" button.

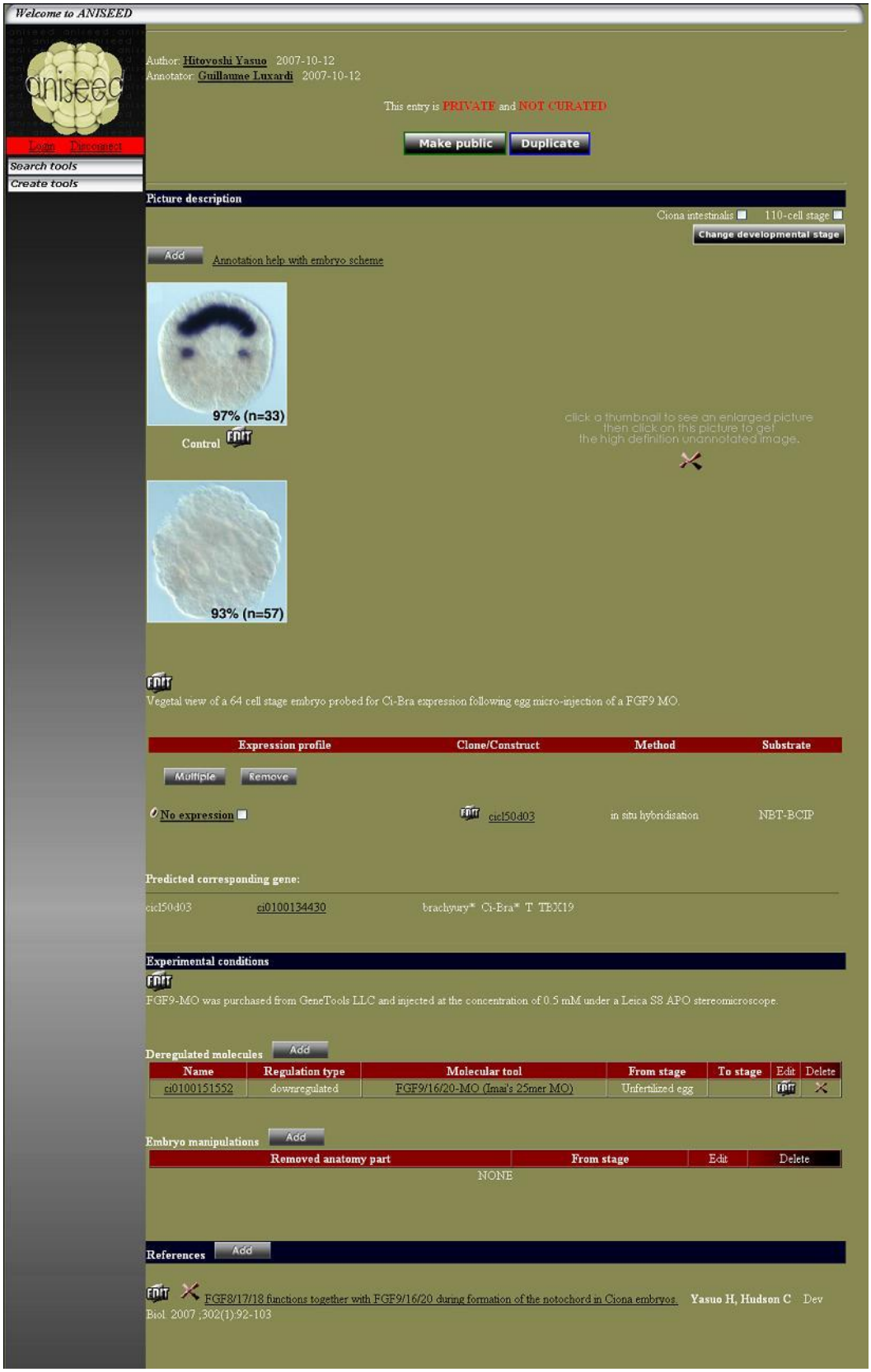


Enter the Pubmed ID of the article *(in our case, 17022960)*, click on the "*submit*" button and a "*main info*" window with article features will open. Check, complete it and click on the "*submit*" button.



You can edit or delete features you have entered by clicking on the "*edit*" button or on the *Red Cross*.

vii. Overview of the filled in "expression card"



2. Description of the pattern of transcriptional activity of a cis-regulatory region

a. Example: describing the pattern of activity of a transgene containing the -1135/-840 regulatory region of FoxF, at tailbud stage

You can describe a pattern associated to a transcriptional activity of a cis-regulatory region just after creating the construct, thanks to the "*Create a new In-Situ Record*" link in the page where you edit the different feature of a construct:

You will be redirected to the "*in situ data card*". Then, step 1, 2, 4 and 5 are the same as the example above. Step 3 is not exactly the same: instead of using a clone, use the identifier of the construct as target that you will find in the construct features (*i.e. CONS00000193*). Furthermore, the method should be related to the assay (e.g. *reporter gene X-gal* or *LacZ insitu,* etc...).

V) ENTERING A PUBLISHED ARTICLE

The article card will give you an access point to all the data present in the paper.

a. Example: entering the paper from Yasuo et al. 2006, FGF8/17/18 functions together with FGF9/16/20 during formation of the notochord in Ciona embryos.

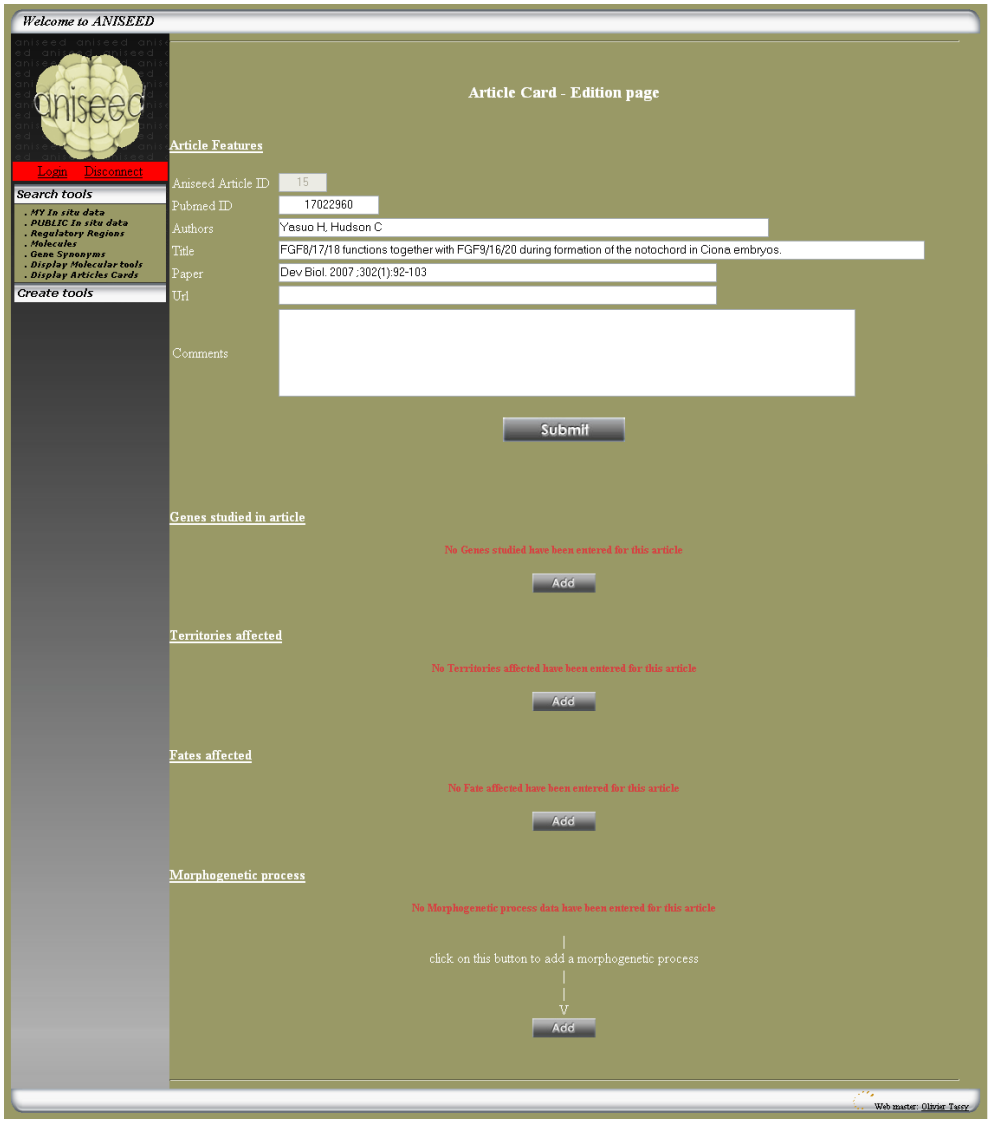
i. *Creating the article card*

To create an article card, choose the "*Create an Article Card*" from menu of Aniseed "*Create tools*" and a "main info" window will open:

- Enter the article's Pubmed ID and click on the "*submit*" button.

- Check, complete the information concerning the article and validate by clicking on the "*submit*" button.

- **The article card**: different features to complete in the article card:
  - article features
  - genes studied in the article
  - territories and fates affected by genes found in the different experiment
  - morphogenetic process studied.



ii. *Step 2: checking the article features*

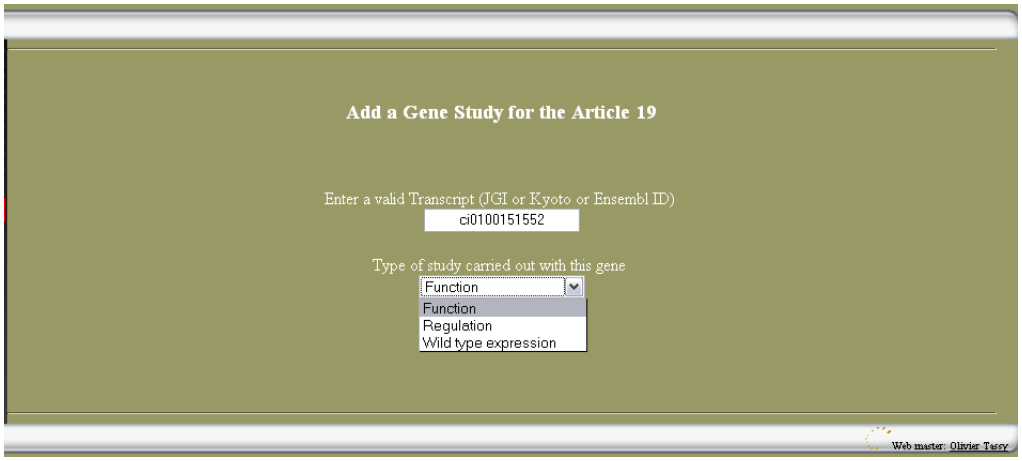
Check/modify/complete the article features. You could enter in the "*comments*" box the goals of the authors. In comments, you could describe what was the goal of the authors. This could also allow you to indicate features of the article not entered in Aniseed.

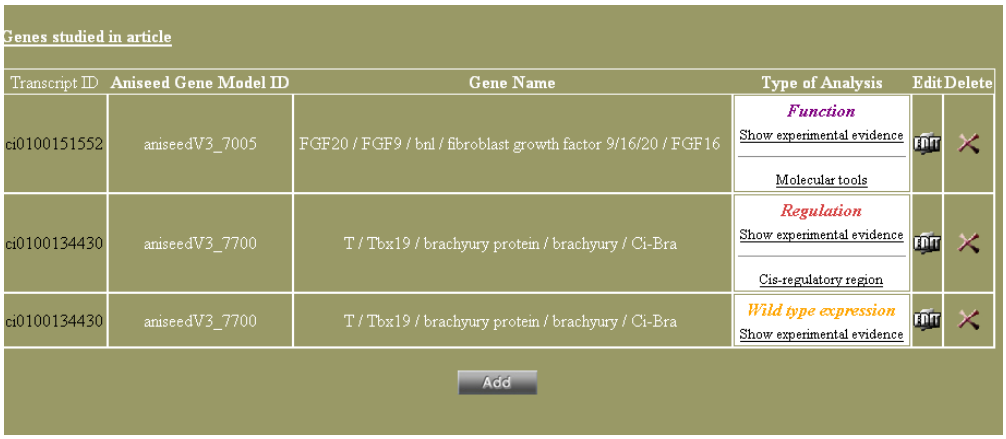
iii. *Step 3: entering the genes studied in the article and associated evidences*

Enter the different genes studied in the article by clicking on the "*add*" button:

Enter a valid transcript model (JGI, Kyoto or Ensembl), associated to the type of study carried out with the gene (function, regulation or wild type expression) and submit your request.

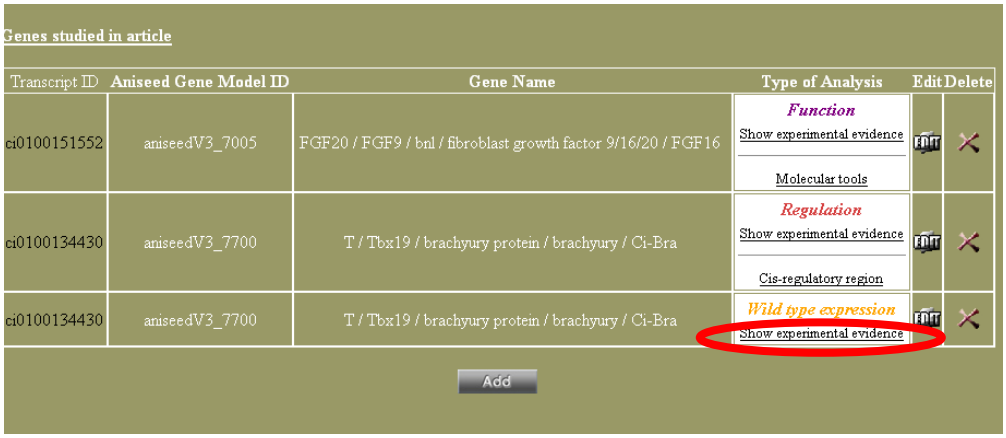
e.g.the authors have studied the wild type expression and the regulation of Brachyury, and the function of FGF9/16/20.

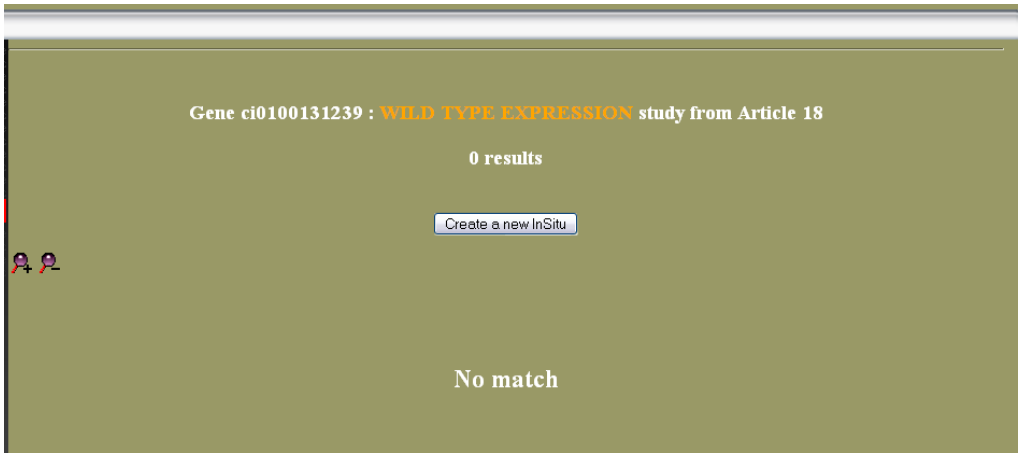




Enter experiment evidences associated to the gene studied:

- Wild type expression
  - Click on the "*show experimental evidence*" link to check what already exists in the database concerning wild type experiment associated to the article.
  - If the information you whish to enter is not already in the database, click on create the "*new In Situ*" button and the first window to create a new in situ data will appear.

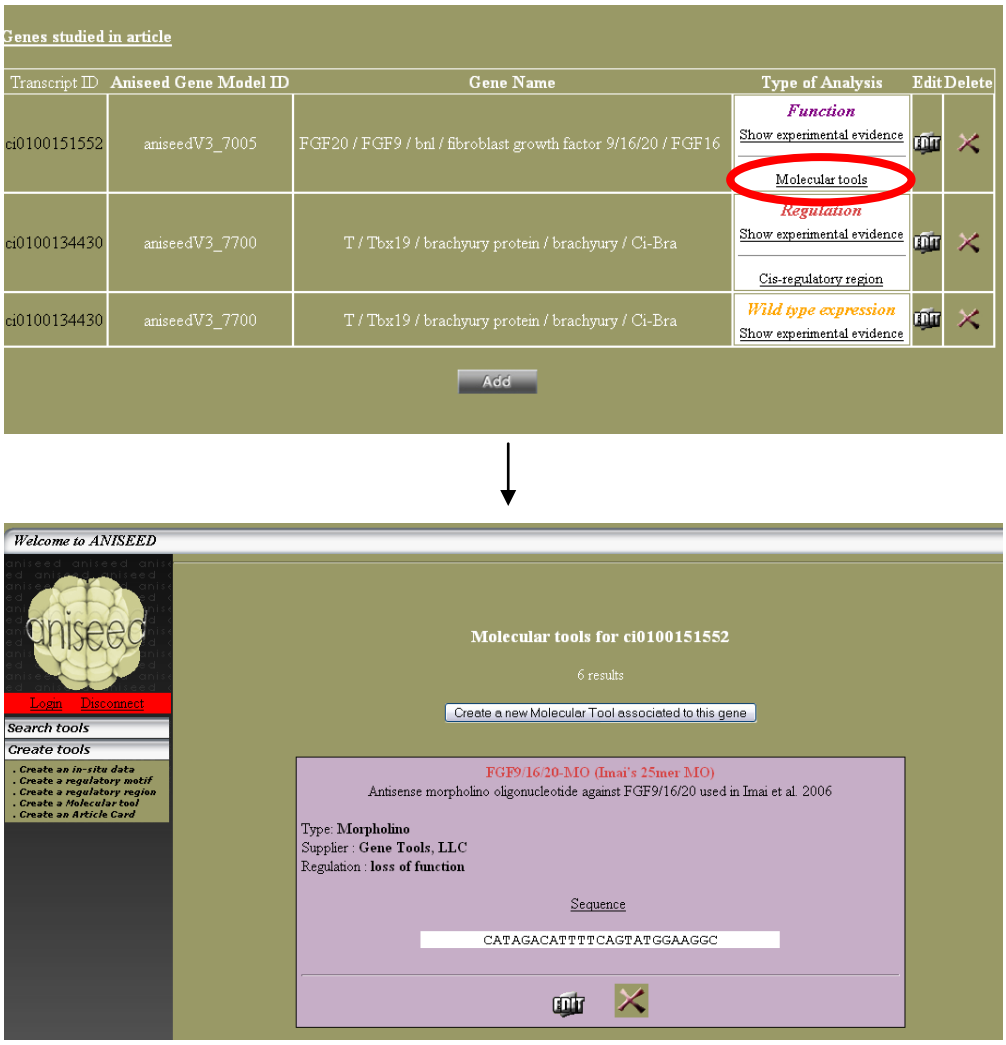




- Enter the experiment evidence as it is shown below (cf section IV)

- Function: the function of a gene could be studied by loss or gain of function of it.
  - Check if the molecular tool used to alter the gene studied already exists or if you have to create it: click on the "*molecular tool*" link. If your tool is not listed, you must create it. click on the "*create a new molecular tool*

associated to this gene" link and you will access to main page of the creation of tool (cf section III/1)



- Once you have the molecular tool used in the experiment, click on the "*show experimental evidence*" link and click on "*create a new In Situ*" button so as to enter the corresponding experiment (cf section IV).

- Regulation: the regulation of a gene can be studied either through the analysis of its cis-regulatory region or by use effect of gain/loss of function of other genes on its expression.

Ex1: the authors used a construct containing Ci-Bra cis regulatory region from -483bp to the first 17 codons with a lacZ reporter to electroporate embryos and to study the activity of this transgene.

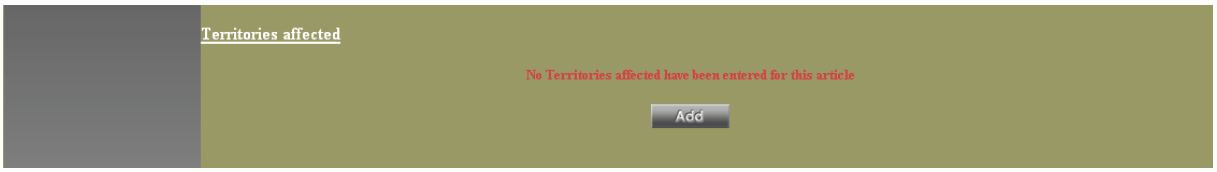
Ex2: the authors studied the expression of Ci-bra following injection of FGF9/16/20 morpholino.

- If you enter experiment concerning cis-regulatory region study, check if data concerning the regulatory element are entered by clicking on the "*Cis regulatory region*" link. If not, you have to enter the information concerning this region by clicking on "*create a new regulatory region*" button (cf section III/2). Then, you can create the experiment card (cf section IV).

- If you enter experiment with a down/up regulation of another gene, check if you have well enter it before with function as type of analysis. Then, you can create the experiment card (cf section IV).

iv. Step 4: entering the territories affected

The idea of this feature is to resume what are the genes implicated in a modification of expression in a given territory. For that, you have to "*add*" the different territories affected and the genes responsible of this modification.

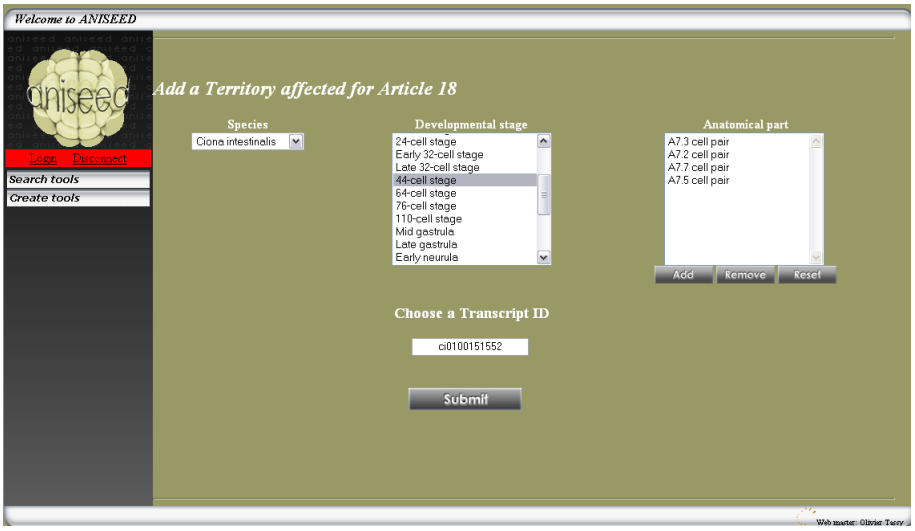


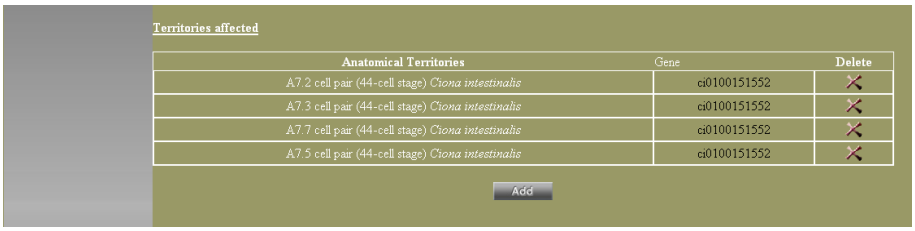
- Choose an Ascidian "*species*" from the menu by clicking on the arrows and scrolling down. *Ciona intestinalis* and *Halocynthia roretzi* are currently supported. *Here select Ciona intestinalis.*

- Choose a "*Developmental stage*" according to Hotta et al. 2007. Click on the arrows and scroll down. (Developmental stages will not be listed unless you have chosen an Ascidian species). *Select 44-cell stage.*

- Choose the "*Anatomical part*" affected by clicking on the "*add*" button. *Here select A7.2, A7.3, A7.5, A7.7, which are the cells that don't express Ci0100134330 when Ci0100151552 is deregulated at the 44-cell stage.* You can remove selected anatomical part by clicking on the "*remove*" button or delete them all by clicking on the "*reset*" button if you want to change or correct a mistake.

- Enter the transcript model responsible of the perturbation. *Here enter Ci0100151552* and click on the "*submit*" button.



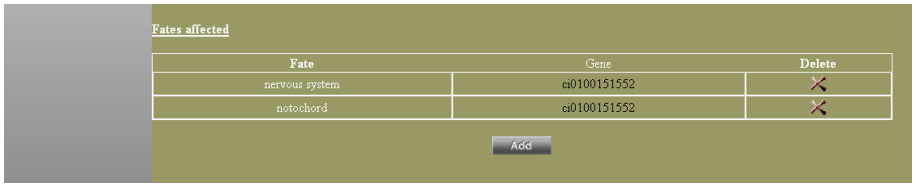


v. *Step 5: entering the fates affected*

Here you have to "*add*" the different fates affected and the genes responsible of this modification.

- Choose the "*Fates*" affected. *Here select Nervous system and notochord which are the fates that don't express Ci0100134330 when Ci0100151552 is deregulated.* You can remove selected anatomical part by clicking on the "*remove*" button or delete them all by clicking on the "*reset*" button if you want to change or correct a mistake.

- Enter the transcript model responsible of the perturbation. *Here enter Ci0100151552* and click on the "*submit*" button.



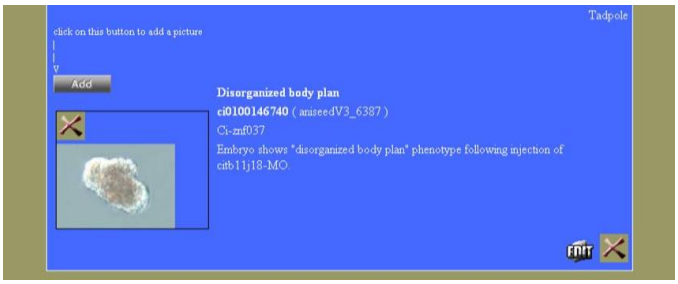
vi. *Step 6: entering the morphogenetic affected*

The idea here is to be able to describe phenotypes that cannot be described in terms of alteration of genetic programs.

By clicking on the "*add*" button, you will access to the "morphogenetic process card":

Add a name to describe the morphogenetic process observed (*e.g. short tail, disorganised body plan…*), the developmental stage in which the observation was done, the gene responsible of the affectation and a comments to describe in detail

the experiment. After filling the form, submit your entry and this last will appear on your article card. To illustrate your entry by a picture, you can click on the "*add*" button.



VI) FINDING EXPRESSION DATA

1. Finding the expression data I have entered

Thanks to the "My In situ data" link from Aniseed Search tools menu, you will access to all the in situ data you have entered (individual expression data or expression data from an article card). You could make precise query in your batch of data: you could retrieve expression profiles labelling defined territories, and excluded from others, find the in situ expression patterns associated to a gene or EST/cDNA clone... By default only the wild type expression patterns are shown but you can access expression patterns in manipulated conditions (gene deregulation, embryo manipulation) by ticking/filling the relevant boxes/fields. You can restrict your query to uncurated data, private data, and reporter gene essays.

a. Example: you want all the in situ data you have entered concerning Ciona intestinalis 110-cell stage embryos

- Choose an "*Ascidian species*" from the menu by clicking on the arrows and scrolling down. *Here Ciona intestinalis.*

- Choose a "Developmental stage" from the menu by clicking on the arrows and scrolling down. *Here 110-cell stage.*

b. Example: you want all the in situ data with genes expresses in muscle but not in mesenchyme.

- Search by territories
  - Define you territories of interest either via the detailed anatomical ontology for each stage (use the "*Add*" button) or a simpler list of fates.

c. Example: you want all the in situ data you have entered stained with Brachyury in manipulated conditions.

- Search by molecule
  - Enter the clone, the gene model or the construct of interest in the "*stained molecule*" box. If you don't know the ID of the gene

model, you can search your molecule ID using the "*Gene synonyms*" interface. You can enter several molecules separated by ":". *If you click on the "and" button, the result will give you all the ISH stained with the several molecules you have entered. If you click on the "or" button, you will have all the ISH stained with one or the others molecules*

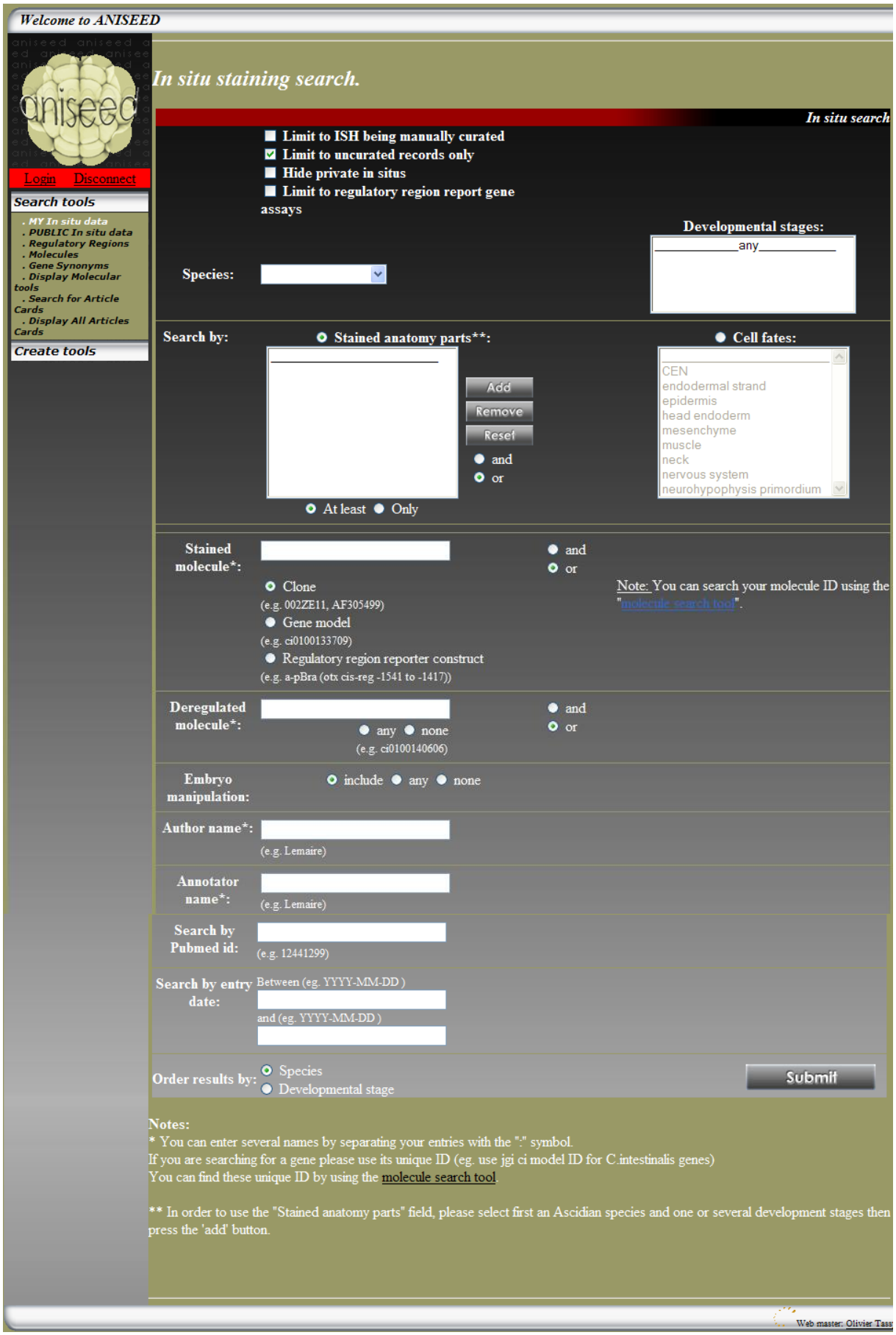
- If you just want to access expression pattern in manipulated conditions like gene deregulation, click on the corresponding "*any*" button. You can specify by entering a gene model ID in the corresponding box, a gene which could be deregulated in an experiment. Moreover, like the stained molecule, you can enter several gene models ID separated by ":" and use the "*and*" and "*or*" button.
- If you just want to access expression pattern in manipulated conditions like embryo manipulation, click on the corresponding "*any*" button. You can include (with the "*include*" button), these manipulated conditions in others complex queries.

d. Example: you want all the in situ data you have entered concerning the paper from Yasuo et al. 2007

- Search by authors or PubmedID
  - You can also restrict your search to the in situ experiments from a given author or found in a paper identified by its Pubmed ID by filling the relevant fields.

e. Understanding the results

You will have a list of all the experiment data corresponding to your query ordered by developmental stages (default option). The number of the experiment is specified on the top right of the window. You can click on the top left of the window to zoom in or to zoom out the pictures. If you want to know more about an experiment, you can click on the corresponding "more" button.
You can order the results according to the developmental stages or according to the species.

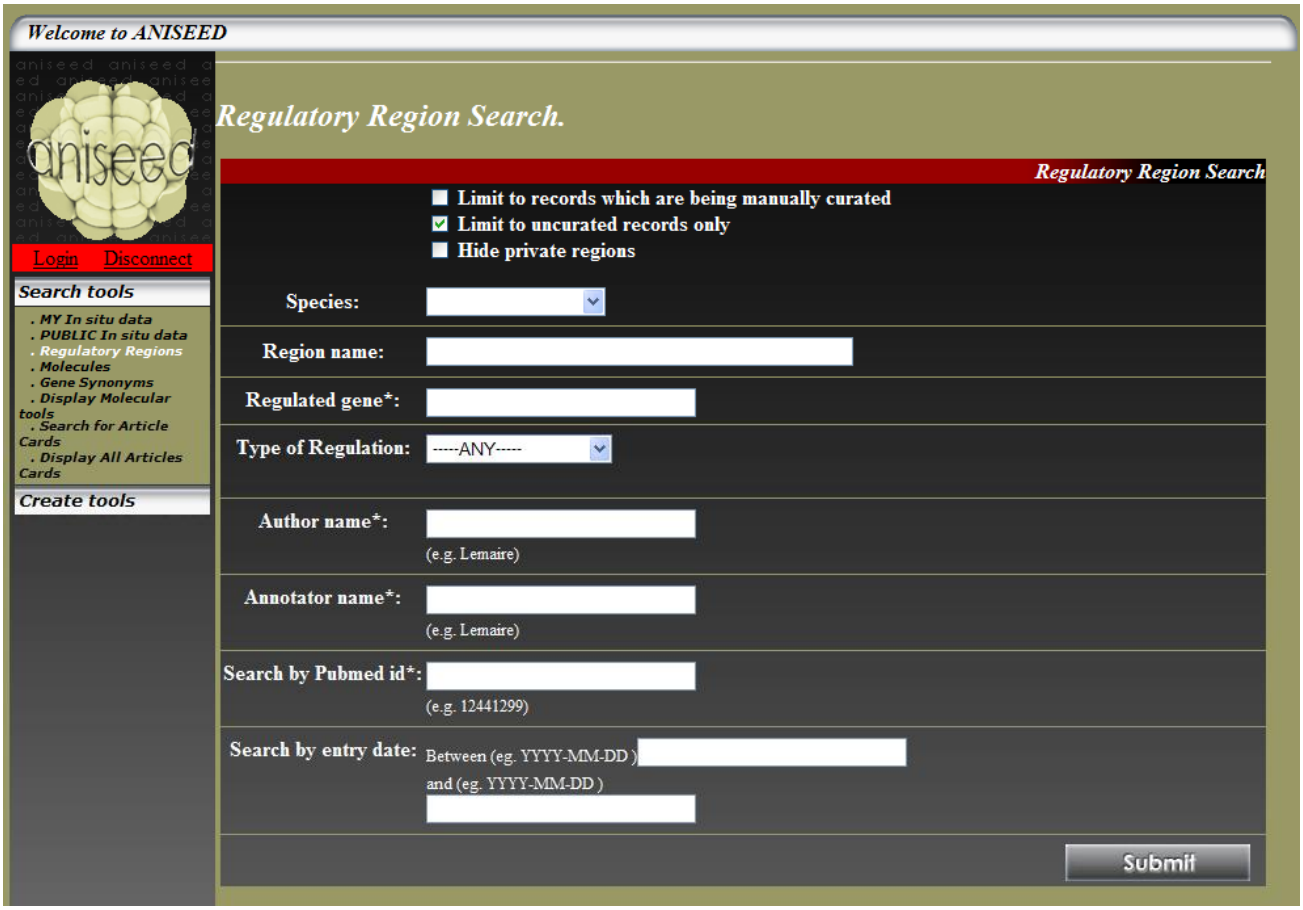

Welcome to ANISEED
aniseed
Login Disconnect
Search tools
. MY In situ data
. PUBLIC In situ data
. Regulatory Regions
. Molecules
. Gene Synonyms
. Display Molecular tools
. Search for Article Cards
. Display All Articles Cards
Create tools
In situ staining search.
In situ search
Limit to ISH being manually curated
Limit to uncurated records only
Hide private in situs
Limit to regulatory region report gene assays
Species:
Developmental stages:
any
Search by:
Stained anatomy parts**:
Add
Remove
Reset
and
or
At least
Only
Cell fates:
CEN
endodermal strand
epidermis
head endoderm
mesenchyme
muscle
neck
nervous system
neurohypophysis primordium
Stained molecule*:
and
or
Clone
(e.g. 002ZE11, AF305499)
Gene model
(e.g. ci0100133709)
Regulatory region reporter construct
(e.g. a-pBra (otx cis-reg -1541 to -1417))
Note: You can search your molecule ID using the "molecule search tool".
Deregulated molecule*:
any
none
(e.g. ci0100140606)
Embryo manipulation:
include
any
none
Author name*:
(e.g. Lemaire)
Annotator name*:
(e.g. Lemaire)
Search by Pubmed id:
(e.g. 12441299)
Search by entry date:
Between (eg. YYYY-MM-DD)
and (eg. YYYY-MM-DD)
Order results by:
Species
Developmental stage
Submit
Notes:
* You can enter several names by separating your entries with the ":" symbol.
If you are searching for a gene please use its unique ID (eg. use jgi ci model ID for C.intestinalis genes)
You can find these unique ID by using the molecule search tool.
** In order to use the "Stained anatomy parts" field, please select first an Ascidian species and one or several development stages then press the 'add' button.
Web master: Olivier Tass

2. Finding public expression data

Thanks to the "PUBLIC In situ data" link from Aniseed Search tools menu, you will access to the same interface as below except that the request will be done on the public data. Thus, you can compare your data with the public one.

VII) FINDING REGULATORY REGION

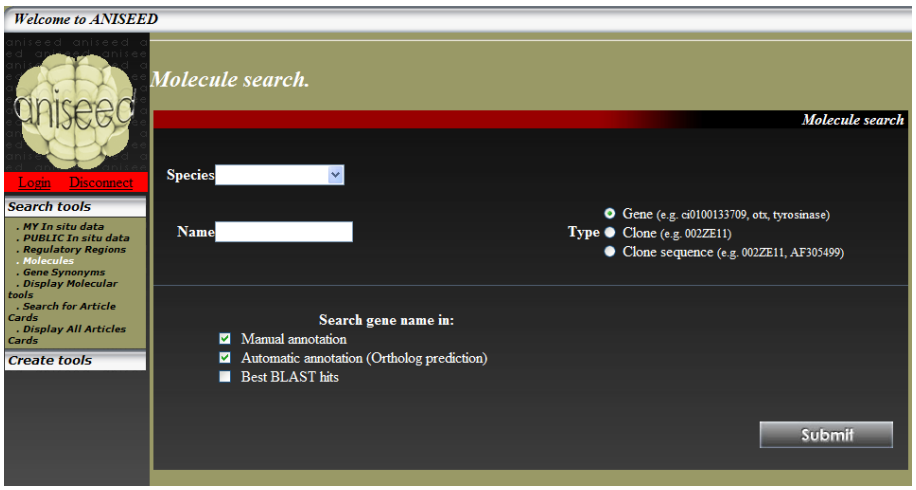
Thanks to the "Regulatory region" link from the Aniseed Search tools menu, you could access regulatory region you have entered. You can specify you query by selecting specie, the name of the region, the regulated gene (transcript model), the type of regulation (enhancer, extended promoter…), the name of the author, a Pubmed ID or the date of the entries.



VII) FINDING MOLECULES

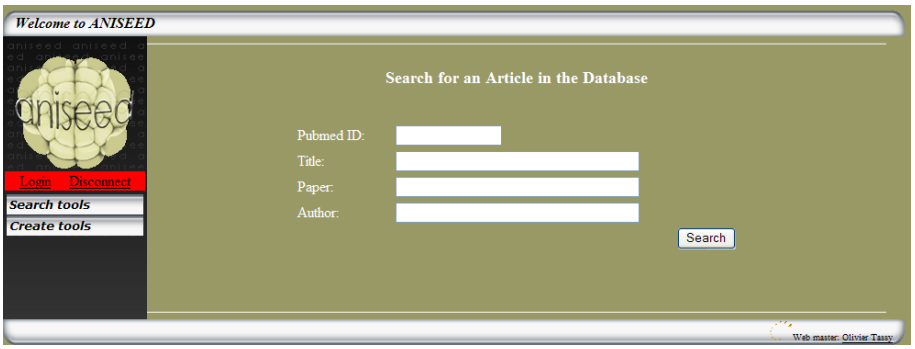
You can search for clones and genes present in ANISEED by clone name, gene name or transcript model.

- Choose an "Ascidian species" from the menu by clicking on the arrows and scrolling down.
- Enter the identifier of the clone or the identifier or name of the gene you are looking for.
- Don't forget to specify what you are looking for (gene, clone, clone sequence).
- Click on the "Submit" button.



VIII) FINDING ARTICLE CARD

Thanks to this interface, you can perform an article search by *Pubmed ID, title, Paper and Authors.*



IX) DISPLAYING ARTICLE CARDS

By choosing this option, you will access to all the article cards present in the database. You can modify it by clicking on the "*edit*" button and delete… NON

X) DISPLAYING MOLECULAR TOOLS

By choosing this option, you will have all the molecular tools present in the database. You can modify it by clicking on the "*edit*" button.