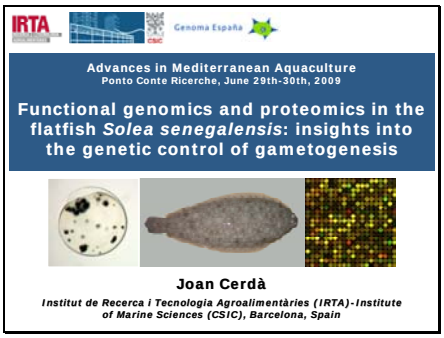
Advances in Mediterranean Aquaculture
Ponto Conte Ricerche, June 29th-30th, 2009

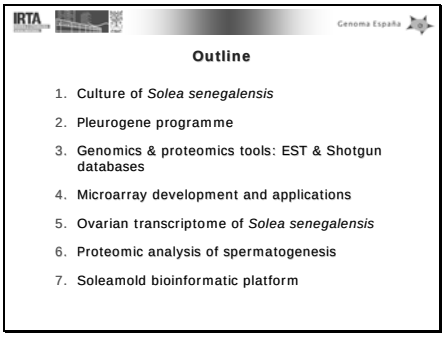
Functional genomics and proteomics in the flatfish *Solea senegalensis*: insights into the genetic control of gametogenesis

Joan Cerdà

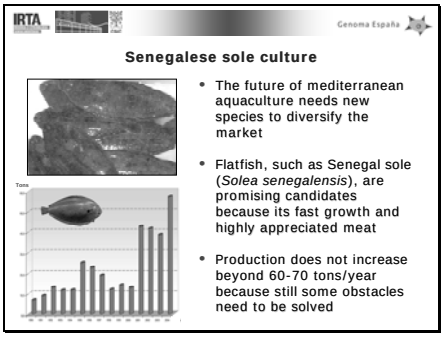
Institut de Recerca i Tecnologia Agroalimentàries (IRTA)-Institute of Marine Sciences (CSIC), Barcelona, Spain



My talk today will be on recent work that we have carried out using functional genomics and proteomics on *Solea senegalensis*, a flatfish of increasing commercial interest in Southern Europe. This work was part of the PLEUROGENE programme financed by Genome Spain, and has contributed with new genomic resources for this species that may help to optimize its production in aquaculture.

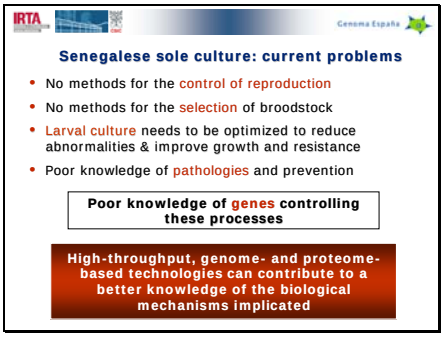


The talk will be divided in seven sections. First, I will summarize the current status of the culture of Senegalese sole as well as the development of different genomics and proteomics databases and resources for this species, such as EST and shotgun databses, and a microarray. Then, I will show some experiments in which these technologies have been employed to gain insights into the genetic control of gametogenesis of sole. Finally, I will present a novel bioinformatic platform that integrates genomic information.



The current production of fish in European aquacuture is mostly restricted to sea bass, sea bream and turbot, and therefore production needs to be diversified into new species in order to increase the market. The Senegalese sole is one of the candidate species because of its relatively fast growth and highly apreciated meat.

The culture of Senegalese sole started in South Spain during 1980's, and in the 90's the first spwnings in captivity were reported. Since then, the production has slowly increased until 60-70 tons per year. However, this production is not increasing further because there are still some biological obstacles that have not been solved yet.



These problems are summarized here.

The main obstacle is the absence of **methods to control reproduction**, and therefore, spawnings and egg production are spontaneous and can not be controlled.

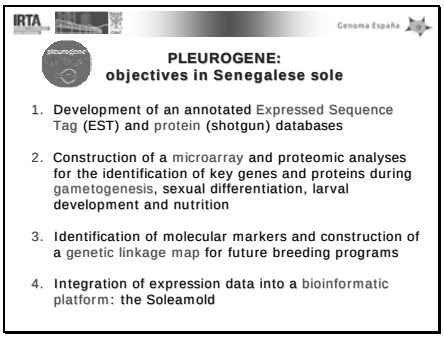
Consequently, this has prevented the development of controlled breeding methods to select for strains with fast growth or increased resistance to diseases.

Larval culture is not specially problematic but it still requires improvements to reduce malformations and to increase the resistance of alevins.

And finally, there is still a poor knowledge of the pathologies of this species and its prevention.

Many of these obstacles still remain because the biological processes implicated, specially their genetic control, are largely unknown.

The project PLEROGENE aimed at contributing to improve this knowledge by using high-throughput, genome- and proteome-based technologies.



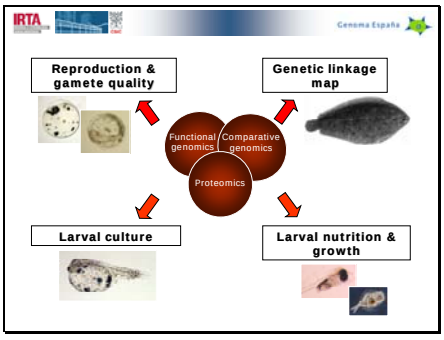
The objetives of PLEUROGENE were the following:

The obtention of an annotated **EST and protein databases** for *Solea senegalensis*.

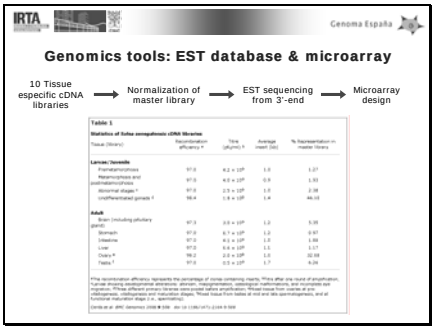
The construction of a **microarray** for the study of the global pattern of gene expression during reproduction, larval development and nutrition.

The identification of **molecular markers** for the construction of a preliminary genetic linkage map

The development of a **bioinformatic platform** for the integration of expression data.



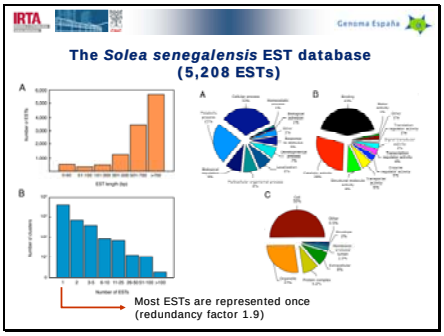
Therefore, the project combined functional and comparative genomics and proteomics to improve reproduction, gamete quality, and larval culture and growth in Senegalese, as well as to construct a genetic linkage map.



As I mentioned, the genomics tools developed in the project were the construction of a EST database and a microarray.

To obtain the ESTs, 10 different cDNA libraries were constructed, mixed into a master library, and normalized. About 11,000 ESTs were sequenced from the 3' end, analyzed, and used to design the microarray.

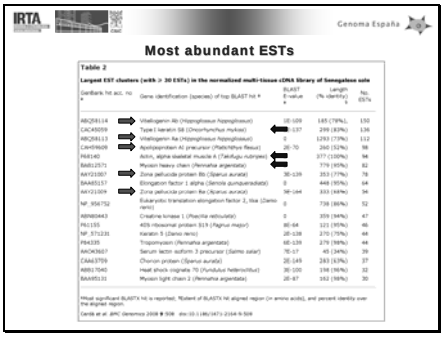
In this table, you can see the larval stages and adult tissues employed. In the table you can also note that the master library contained different amounts of the individual libraries according to their concentration after one round of amplification.



The quality and composition of the EST database is shown in these graphs.

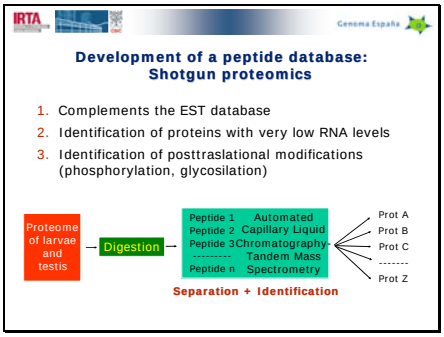
5,208 unique ESTs were obtained, and most of the ESTs were longer than 500 bp. The clustering indicated that most of the ESTs were represented once, which resulted in a redundancy factor of 1.9.

The analysis of GO terms of the database indicated that many ESTs were related to metabolic and cellular processes, and binding and catalytic activities.



Some ESTs were over-represented, as vitellogenins and apolipoprotein transcripts, which may reflect the elevated expression of these genes in the livers of the females sacrificed which were at the reproductive period.

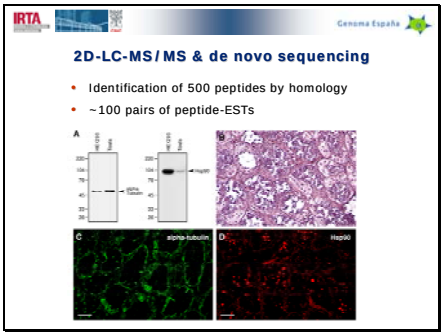
Other cytoskeleton-related transcripts were also over-represented which may be the result of the larval stages used for the libraries, since it is known that larvae at metamorphosis express high levels of cytoskeletal transcripts.



The EST database was complemented with a peptide database obtained by shotgun proteomics.

This methodology has the potential of identifying proteins with very low RNA levels that may be not present in the EST database, and also may identify potential posttraslational modifications such as phosphorylation or glycosilation.

Our approach was based on the digestion with trypsin of a protein extract from testis and larvae, and further identification of peptides by capillary liquid chromatography combined with tandem mass spectrometry.

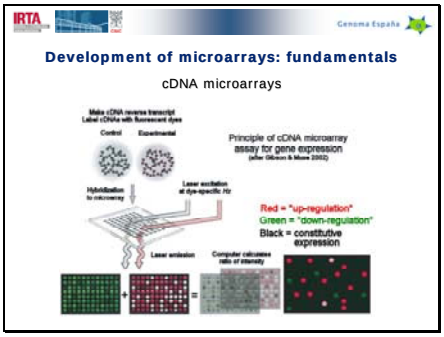


Shotgun identified about 500 different peptides.

The combined analysis with the EST database revealed the presence of about 100 pairs peptide-EST.

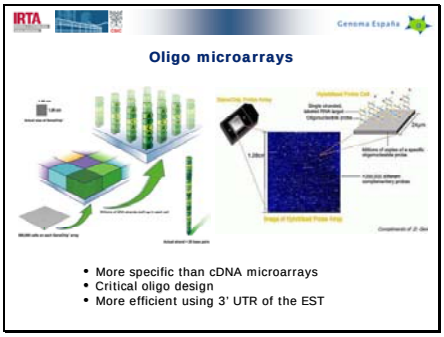
Among these, alpha-tubulin and Hsp90 were tested by Western blot and immunocytochemistry on testis, using antibodies raised against mammalian proteins, confirming the expression of these peptides in the testis.

This method is therefore an interesting approach to find proteins in an specific tissue.

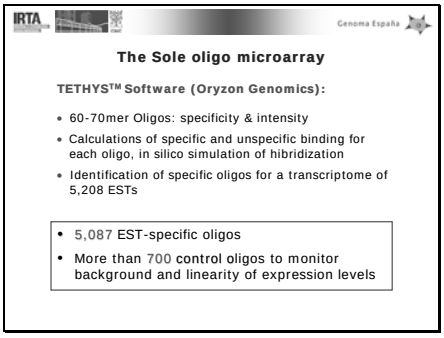


For the analysis of global gene expression in sole we developed a microarray. This technology allows to determine which genes are activated and which genes are repressed when two populations of cells or tissues are compared.

A microarray or DNA chip is a solid matrix that contains thousands of spots with DNA, corresponding to cDNA synthesised from each EST. The methology is based on the hybridization of each RNA sample that wants to be compared labelled with different fluorescent dyes (Cy5 or Cy3). The different fluorescent intensity is proportional to the amount of RNA present in the sample. Therefore, changes in gene expression in cells at two different physiological conditions can be identified.



Recently, microarrays containing long oligos specific for each gene, rather than cDNAs, are preferred since these microarrays are more specific. However, oligo design is critical to asure specificity, and therefore usually the nucleotide sequence of the 3'end of the ESTs is used to design the oligos.

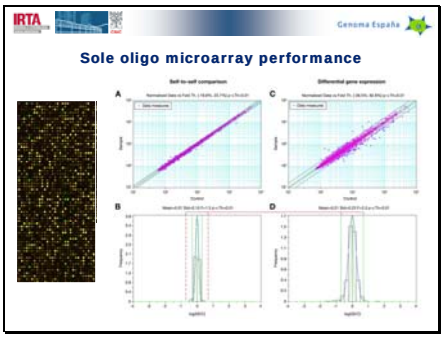


For *Solea senegalensis* an oligo microarray was designed.

The array contained 60-70 mer oligos with a good compromise between specifity and intensity.

The oligos were designed using a specific software developed by Oryzon Genomics, a Spanish biotech company participating in the project, which calculates the specific and unspecific binding for each oligo using the whole transcriptome (5,208 ESTs).

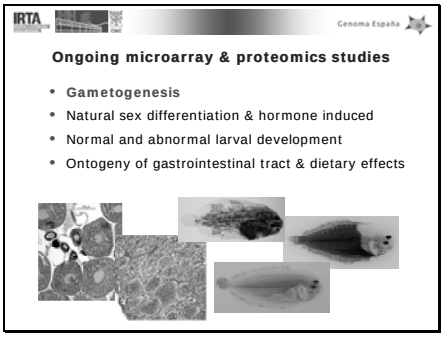
The Thethys software was able to design 5,087 specific oligos with were printed in the microarray by Agilent. The array also contained 700 oligos to control the background and the linearity of the expression levels.



In this slide, it is shown the performance of the microarray developed for sole during two different experiments: a self-to-self comparison and a differential gene expression experiment.

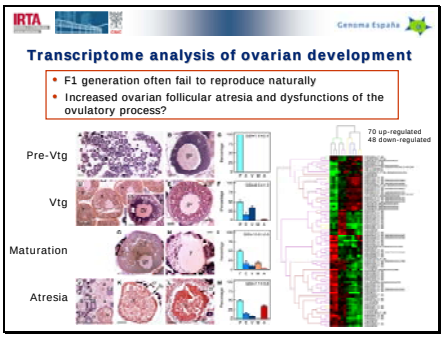
In a self-to-self experiment, the same RNA sample is labelled with Cy5 or Cy3, and therefore no changes in gene expression are expected. Accordingly, a narrow distribution of data is observed. In contrast, in the differential gene expression experiment, in which larvae at different developmental stage were compared, a much broader distribution is observed.

The comparison of the distribution of gene expression data in both experiments gave a false discovery rate of 3.2%, indicating a good technical reliability of the microarray.



The sole microarray has been used to investigate changes in gene expression during reproduction and larval development.

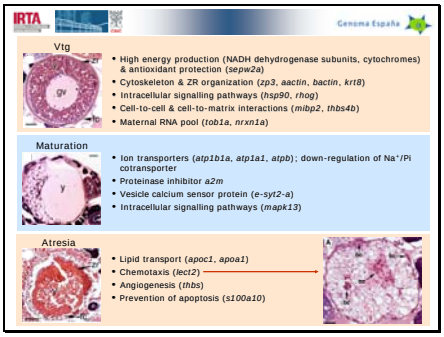
As an example, I will now present to you an experiment carried out to identify changes in the ovarian transcriptome of the Senegalese sole during the anual reproductive cycle.



As I mentioned before, the reproduction of Senegalese sole in captivity can not be controlled. In the F1 generation, the situation is even worse since these animals often fail to reproduce naturally. An increased atresia and dysfunctions of the ovulatory process are factors that may be involved.

To obtain information on the genes involved in ovarian growth an maturation in sole, we used the microarray. Four different ovarian stages were compared: previtellogenesis, vitellogenesis (growth phase), maturation and atresia (degeneration of vitellogenic and mature ovarian follicles). Since Senegalese sole has a group-synchronous ovary, the different proportion of ovarian follicles at each stage defined the four types of ovaries.

The microrray analysis detected the significant variation of 118 genes which were clustered as indicated in this figure.



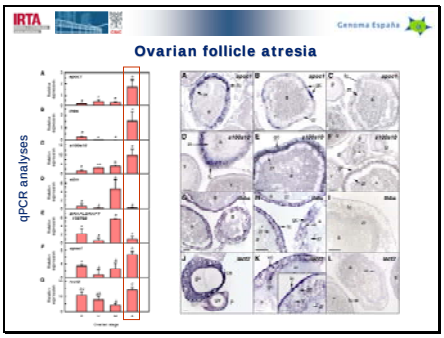
Microarray analysis revealed that during vitellogenesis, many up-regulated ovarian transcripts had putative mitochondrial function/location suggesting high energy production (NADH dehydrogenase subunits, cytochromes) and increased antioxidant protection (selenoprotein W2a), whereas other transcripts were related to cytoskeleton and zona radiata organization (zona glycoprotein 3, alpha and beta actin, keratin 8), intracellular signalling pathways (heat shock protein 90, Ras homolog member G), cell-to-cell and cell-to-matrix interactions (beta 1 integrin, thrombospondin 4b), and the maternal RNA pool (transducer of ERBB2 1a, neurexin 1a).

At maturation, regulated transcripts included ion transporters (Na^+-K^+-ATPase subunits), probably required for oocyte hydration, as well as a proteinase inhibitor (alpha-2-macroglobulin) and a vesicle calcium sensor protein (extended synaptotagmin-2-A).

Finally, during follicular atresia, we found transcripts with inferred roles in lipid transport (apolipoprotein C-I), chemotaxis (leukocyte cell-derived chemotaxin 2), angiogenesis (thrombospondin), and prevention of

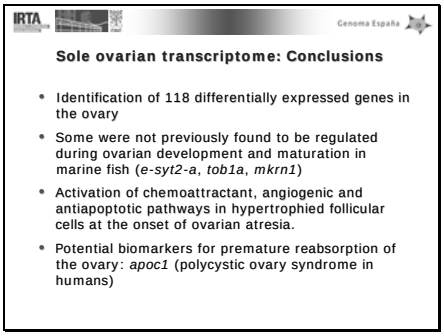
apoptosis (S100a10 calcium binding protein).

Interestingly, the finding of lect2 is consistent with the invasion of atretic ovarian follicles by blood cells.



Real-time PCR of some of the genes regulated during ovarian atresia confirmed the microarray data, and showed a marked accumulation of *apoc1*, *thbs*, and *s100a10* mRNAs in the ovary.

Interestingly, in situ hybridization revealed that most of genes up-regulated during atresia were expressed by follicle cells associated with the oocyte. These observations therefore suggest that follicle cells play an important role during this process, and provide evidence, for the first time in fish, of the secretion of chemoattractant signals for leukocytes by follicle cells.

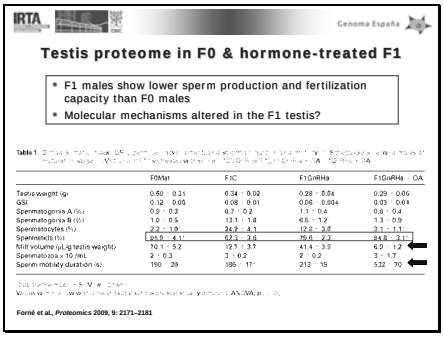


In summary, the microarray experiment carried out has revealed the differential expression of 118 genes during ovarian development in sole. This number is probably higher since the microarray available contained only a limited number of genes.

However, this analysis was able to identify some genes that were not previously found to be regulated during ovarian development and maturation in marine fish which function is yet unknown (*e-syt2-a*, *tob1a*, *mkrn1*).

Evidence was also found for the activation of chemoattractant, angiogenic and antiapoptotic pathways in hypertrophied follicular cells at the onset of ovarian atresia.

Finally, the data obtained also provided potential biomarkers that can be used to detect premature reabsorption of the ovary ('atretic females'), such as *apoc1*. Interestingly, in humans APOC1 has been suggested a a marker fo polycystic ovary syndrome.

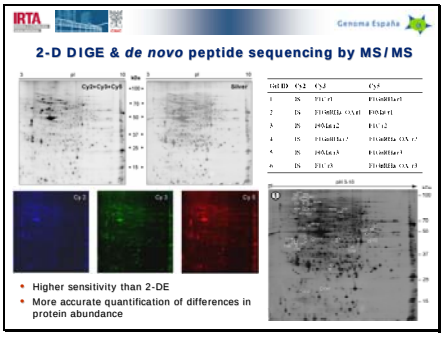


Protemic approaches have been also employed in PLEUROGENE for the study of gametogenesis, particularly during spermatogenesis, as in the example I will show you now.

One observed phenomena in cultured Senegalese sole is the low sperm production and fertilization capacity of F1 males. The causes of this are not known but it is possible that molecular mechanisms during spermatogenesis are altered in the testis of F1 males.

To investigate this possibility, we carried out a study in which F1 males were treated with GnRHa (a neuropetide that releases gonadotropins from the pituitary) in the presence or absence of 11-ketoandrostenedione (OA), a precursor or 11-KT, the main androgen in the fish testis.

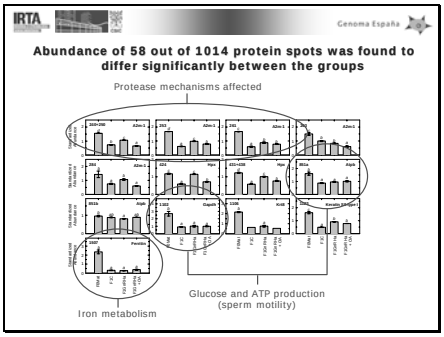
As you can see in the table, GnRHa+OA stimulates spermatogenesis in F1 males, but the production of spermatozoa is reduced with respect to that of F0 males. However, an increased motility of spermatozoa is observed in the sperm from the males treated with GnRHa+OA.



The approach that we used to investigate potential changes in the testicular proteome of these fish was 2 dimensional differential gel electrophoresis.

In this method, protein extracts from the two samples to be compared are labelled with Cy3 or Cy5, mixed together, and ran in a 2-D gel together with an internal standard. Differences in fluorescence intensity of specific spots will indicate different protein abundance. This method is more sensitive than conventional 2-DE methods, and allows a more accurate quantification of differences in protein abundance.

In this table, you can see the comparisons made for this particular experiment. Below, you can also see a typical 2-D DIGE gel after silver staining, prior to the collection of the spots for de novo peptide sequencing.



2-D DIGE revealed differences in the abundance of 58 protein spots between the testis of F0 males and that of F1 males treated or not with hormones.

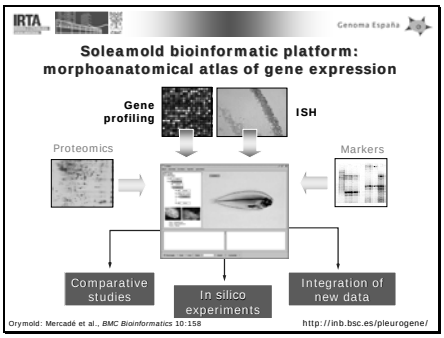
Here, a brief summary with some relevant proteins is shown.

As you can see, some proteins were represented in more than just one spot which indicates the presence of posttraslational modifications. This is the case por instance for alpha-2-macroglobulin. However, in all the F1 groups, this protein in the testis was less abundant which indicates that protease mechanisms in the testis of F1 fish may be affected.

Similarly, the levels of ferritin in the testis of F1 males were lower than in F0 males possibly indicating a deficiency in iron metabolism.

Interestingly, we also found low levels of ATP synthase and glyceraldehyde 3-phosphate-dehydrogenase (involved in glucose production), which are proteins involved in sperm motility.

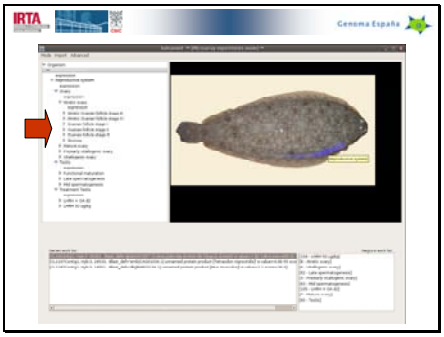
These data thus suggest that alterations in protease inhibition, iron and glucose metabolism, and an increased oxidative stress (peroxiredoxin and hemopexin are more elevated in GnRHa+OA treated fish), may be mechanisms underlying the low production and poor fertilization capacity of the sperm produced by sole F1 males.



One of the tasks also undertaken in the PLEUROGENE programme was the development of a bioinformatic platform to integrate the genomic data (microarray and ISH) in a morphoanatomical fashion. This platform is based on the Orymold software recently developed for rice, and is called Soleamold.

The Solemold can be used to perform comparative studies and in silico experiments, as well as to integrate new data. These data can be genomic, but also proteomic and genetic (molecular markers).

This platform may also be expanded in the future to integrate data from other flatfish.



Currently, the Soleamold integrates microarray data from experiments on gametogenesis, but in the future it can be implemented to additional organs and tissues, including larval development.

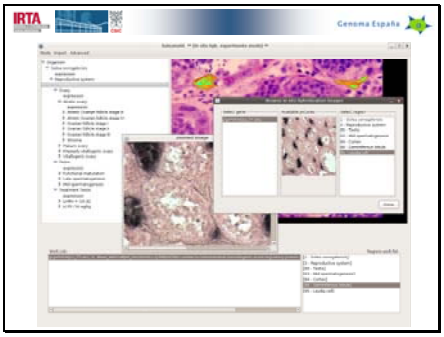
The platform works through an ontological description of organs, tissues and cell types that is organized hierarchically.

The ontology is complemented with histological images forming a user-navigable atlas.

The information is retrieved while maintaining constant visual contact with the ontology and atlas, demonstrating the biological significance of the gene expression data.

Users can make semantic queries, such as 'Where or when do gene X and gene Y show differential expression, so that gene X is overrepresented while gene Y has a low level of expression?'.

Semantic queries allow for the construction of lists of elements of interest based on the model ontology, therefore checking and validating both the experimental data and the model.



The platform also integrates spatial distribution of gene expression through in situ hybridization images.

The Soleamold platform might be therefore a powerful tool for research in both academia and industry and is freely available for non-commercial users.

To finish my talk I would like to mention that the PLEUROGENE programme has been carried out by a consortium formed by different labs in Spain and Canada belonging to different research institutions. A biotech company, Oryzon Genomics, was also part of the consortium and provided the facilities and expertise in functional genomics.

Thank you very much for your attention.