

Research article

Open Access

Genomic resources for a commercial flatfish, the Senegalese sole (*Solea senegalensis*): EST sequencing, oligo microarray design, and development of the Soleamold bioinformatic platform

Joan Cerdà*¹, Jaume Mercadé², Juan José Lozano³, Manuel Manchado⁴, Angèle Tingaud-Sequeira¹, Antonio Astola⁵, Carlos Infante⁴, Silke Halm⁵, Jordi Viñas⁶, Barbara Castellana⁷, Esther Asensio⁴, Pedro Cañavate⁴, Gonzalo Martínez-Rodríguez⁵, Francesc Piferrer⁶, Josep V Planas⁷, Francesc Prat⁵, Manuel Yúfera⁵, Olga Durany², Francesc Subirada², Elisabet Rosell² and Tamara Maes²

Address: ¹Laboratory of the Institut de Recerca i Tecnologia Agroalimentàries (IRTA)-Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas (CSIC), 08003 Barcelona, Spain, ²Oryzon Genomics, Scientific Parc University of Barcelona, 08028 Barcelona, Spain, ³CIBERehd, Hospital Clínic, 08036 Barcelona, Spain, ⁴IFAPA Centro "El Toruño", Junta de Andalucía, 11500 El Puerto de Santa María, Cádiz, Spain, ⁵Instituto de Ciencias Marinas de Andalucía, CSIC, 11510 Puerto Real, Cadiz, Spain, ⁶Institut de Ciències del Mar, CSIC, 08003 Barcelona, Spain and ⁷Department of Physiology, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain

Email: Joan Cerdà* - joan.cerda@irta.es; Jaume Mercadé - jmercade@oryzon.com; Juan José Lozano - juanjo.lozano@ciberehd.org; Manuel Manchado - manuel.manchado.ext@juntadeandalucia.es; Angèle Tingaud-Sequeira - tingaud@icm.csic.es; Antonio Astola - antonio.astola@uca.es; Carlos Infante - carlos.infante.ext@juntadeandalucia.es; Silke Halm - silke.halm@icman.csic.es; Jordi Viñas - jvinas@icm.csic.es; Barbara Castellana - bcastellana77@gmail.com; Esther Asensio - esther.asensio.ext@juntadeandalucia.es; Pedro Cañavate - josep.canavate@juntadeandalucia.es; Gonzalo Martínez-Rodríguez - gonzalo.martinez@icman.csic.es; Francesc Piferrer - piferrer@icm.csic.es; Josep V Planas - jplanas@ub.edu; Francesc Prat - francisco.prat@icman.csic.es; Manuel Yúfera - manuel.yufera@icman.csic.es; Olga Durany - odurany@oryzon.com; Francesc Subirada - fsubirada@oryzon.com; Elisabet Rosell - erosell@oryzon.com; Tamara Maes - tmaes@oryzon.com

* Corresponding author

Published: 30 October 2008

Received: 25 July 2008

BMC Genomics 2008, 9:508 doi:10.1186/1471-2164-9-508

Accepted: 30 October 2008

This article is available from: <http://www.biomedcentral.com/1471-2164/9/508>

© 2008 Cerdà et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The Senegalese sole, *Solea senegalensis*, is a highly prized flatfish of growing commercial interest for aquaculture in Southern Europe. However, despite the industrial production of Senegalese sole being hampered primarily by lack of information on the physiological mechanisms involved in reproduction, growth and immunity, very limited genomic information is available on this species.

Results: Sequencing of a *S. senegalensis* multi-tissue normalized cDNA library, from adult tissues (brain, stomach, intestine, liver, ovary, and testis), larval stages (pre-metamorphosis, metamorphosis), juvenile stages (post-metamorphosis, abnormal fish), and undifferentiated gonads, generated 10,185 expressed sequence tags (ESTs). Clones were sequenced from the 3'-end to identify isoform specific sequences. Assembly of the entire EST collection into contigs gave 5,208 unique sequences of which 1,769 (34%) had matches in GenBank, thus showing a low level of redundancy. The sequence of the 5,208 unigenes was used to design and validate an oligonucleotide microarray representing 5,087 unique Senegalese sole transcripts. Finally, a novel interactive

bioinformatic platform, Soleamold, was developed for the Senegalese sole EST collection as well as microarray and ISH data.

Conclusion: New genomic resources have been developed for *S. senegalensis*, an economically important fish in aquaculture, which include a collection of expressed genes, an oligonucleotide microarray, and a publicly available bioinformatic platform that can be used to study gene expression in this species. These resources will help elucidate transcriptional regulation in wild and captive Senegalese sole for optimization of its production under intensive culture conditions.

Background

In recent years, genomic research employing laboratory fish models, such as zebrafish (*Danio rerio*), medaka (*Oryzias latipes*) and pufferfish (*Tetraodon nigroviridis* and *Takifugu rubripes*), has contributed to understanding the mechanisms of vertebrate development and has shed light on poorly understood evolutionary phenomena such as genome duplication and duplicate gene silencing [1]. The use of genomic technologies on fish is also essential to improve aquaculture research for the mass production of target species, which may have a profound impact on human welfare since fish is a major food source for humans. These studies may be very useful in providing a physiological and genetic basis for marker-assisted selection of strains with enhanced growth or disease resistance, or for identifying key genes and genetic networks involved in the production of viable gametes.

Salmonids are important commercial fish and also unique models for evolutionary studies as they are recent tetraploids [2]. A large expressed sequence tag (EST) database and several cross-species and species-specific cDNA microarrays have recently been developed [3-8]. However, although efforts have been made to partially sequence the transcriptome of specific tissues in some species [9-15], there are still few genomic tools available for commercially relevant marine fish, particularly for functional genomics studies.

Flatfish, members of the order Pleuronectiformes, are a relatively large group of teleosts with about 570 extant species [16]. They are benthic and carnivorous, and most are marine species, although about four probably occur only in freshwater. These fish undergo a unique development process, known as metamorphosis, during which one eye migrates across the top of the skull to lie adjacent to the other eye on the opposite side, while the body flattens and lies on the eyeless side [17]. This unique developmental event involves drastic morphological and physiological changes, although its molecular regulation is still poorly understood [18,19]. Flatfish have long been a choice food, with many members of the group (e.g., halibut, flounder, sole, turbot, plaice) being important for commercial fisheries. With the general worldwide decline in wild fishery [20], with an essentially stable, wild catch

of flatfish, research into producing them in aquaculture have been underway for the last fifteen to twenty years. Aquaculture of Japanese flounder (*Paralichthys olivaceus*), turbot (*Psetta maxima*), Atlantic halibut (*Hippoglossus hippoglossus*) and others have been successfully achieved, although improvements in the efficiency of the industrial production of most of these species are still needed.

The Senegalese sole, *Solea senegalensis* (Kaup, 1858), is a highly prized flatfish of which intensive culture has increased over the last decade, particularly in Southern Europe [21,22]. However, partly due to the lack of knowledge of the physiological and genetic mechanisms involved, the development of aquaculture is impaired by the lack of methods to control reproduction in captivity and those to improve larval growth and disease resistance. The Pleurogene™ project, an initiative funded by Genome Spain and Genome Canada, was set up to improve understanding of physiological and evolutionary processes influencing the reproduction and larval development and survival of both Senegalese sole and Atlantic halibut in natural environments and aquaculture [23]. This project has developed genomic and proteomic research tools to help achieve these goals [12,24]. Here we report the establishment of an EST database for Senegalese sole containing 5,208 cDNA sequences, and the construction and validation of an oligo-based microarray for the detection of 5,087 putative transcripts from this species. In addition, an interactive bioinformatic platform termed Soleamold, developed to accommodate the EST database and results from microarray and in situ hybridization (ISH) experiments, is presented. Currently, this platform only contains data on Senegalese sole gonad development but it could be extended in the future with data from other tissues and organs, as well as from other flatfish, to become a powerful tool for flatfish genomic research.

Results and discussion

EST survey

Ten non-normalized, directionally-cloned (5' *EcoRI*, 3' *XhoI*) cDNA libraries were constructed from different adult tissues and larval stages (Table 1). For all libraries, the percentage of clones with inserts (recombinant efficiency) was high (97-98%) with average insert sizes between 0.9-1.7 kb. To minimise redundancy before

Table 1: Statistics of *Solea senegalensis* cDNA libraries

Tissue (library)	Recombination efficiency ^a	Titre (pfu/ml) ^b	Average insert (kb)	% Representation in master library
Larvae/Juvenile				
Premetamorphosis	97.0	4.2 × 10 ⁹	1.0	1.27
Metamorphosis and postmetamorphosis	97.0	4.0 × 10 ⁹	0.9	1.93
Abnormal stages ^c	97.0	2.5 × 10 ⁹	1.0	2.38
Undifferentiated gonads ^d	98.4	1.8 × 10 ⁸	1.4	46.10
Adult				
Brain (including pituitary gland)	97.3	3.0 × 10 ⁹	1.2	5.35
Stomach	97.0	6.7 × 10 ⁹	1.2	0.97
Intestine	97.0	4.1 × 10 ⁹	1.0	1.88
Liver	97.0	6.6 × 10 ⁹	1.1	1.17
Ovary ^e	98.2	2.0 × 10 ⁸	1.0	32.68
Testis ^f	97.0	0.5 × 10 ⁹	1.7	6.26

^aThe recombination efficiency represents the percentage of clones containing inserts; ^bTitre after one round of amplification; ^cLarvae showing developmental alterations: albinism, malpigmentation, osteological malformations, and incomplete eye migration; ^dThree different primary libraries were pooled before amplification; ^eMixed tissue from ovaries at pre-vitellogenesis, vitellogenesis and maturation stages; ^fMixed tissue from testes at mid and late spermatogenesis, and at functional maturation stage (i.e., spermiating).

sequencing, aliquots of each library, depending on the titre after one round of amplification (Table 1), were pooled into a master library, which was normalized through three rounds of hybridization. About 10,400 clones from this master library were sequenced with T7 primer for sequences corresponding to the 3' UTR in correctly oriented inserts. Because the general low conservation in 3' UTRs, and genome duplication events that have occurred in teleosts [2], we focused on 3' sequencing to design specific oligos for a microarray to distinguish between potential paralogues arising from gene duplications. After trimming and vector and contaminant removal, 10,185 high-quality sequences were obtained with an average read length of approximately 600–700 bp, with the majority of reads being > 700 bp (Figure 1A). These sequences have been submitted to GenBank (accession numbers: [FF281814-FF291996](#)).

The EST sequences were assembled using PHRAP under high stringency to identify clusters (contiguous sequences or contigs) representing redundant transcripts. Clustering analysis yielded 5,208 unique sequences, indicating a redundancy factor of 1.9 (Figure 1B). This compares positively with that obtained in other normalized cDNA libraries available in the UniGene database [25] in which at least 5,000 clones were sequenced from libraries prepared from a mixture of different tissues (e.g., [5,26]). Lower redundancy could possibly be obtained by applying the duplex-specific nuclease method to the cDNA [27], prior to cloning and library generation, which requires mixing all initial RNA samples. However, this method uses a PCR amplification step which may introduce bias against long transcripts, thus selecting for partial

(shorter) transcripts. For the libraries to contain as many expressed full-length clones as possible, we consider our normalized library a good compromise between large clone size and efficient normalization.

The most abundant ESTs (represented 30 times or more) in the sole normalized library corresponded to cytoskeletal proteins, vitellogenin Aa and Ab, apolipoprotein A1, zona pellucida (ZP) proteins, elongation factors, creatine kinase 1, lectin, ribosomal proteins and heat shock protein 70 (Table 2). The high abundance of transcripts encoding cytoskeletal proteins, creatine kinase 1, ribosomal proteins and elongation factors may be a consequence of the enrichment of cDNAs of larval origin in the mixed library (30%) since there is a complex reorganization of somatic tissues in sole larvae at metamorphosis and post-metamorphosis [28-32]. Different isoforms of vitellogenins and ZP proteins are known to be highly expressed in the liver of fish females at ovarian recrudescence [33,34], and some ZPs are also highly expressed in the ovary [5,34-36]. Therefore, high levels of vitellogenin and ZP ESTs is consistent with the fact that sexually maturing females were used for the construction of the primary libraries. The over-representation of apolipoprotein A1 and lectin, also observed in cDNA libraries from liver, ovarian follicles and larvae in other fish [12,15,36], may also be related to the physiological stage of the females. Finally, the low but significant abundance of ESTs encoding the stress-related heat shock cognate protein 70 probably reflects the widespread expression of this gene in both larval and adult tissues [37], although it may also suggest a degree of stress caused by the capture and confinement of fish prior to sampling (e.g. [38,39]).

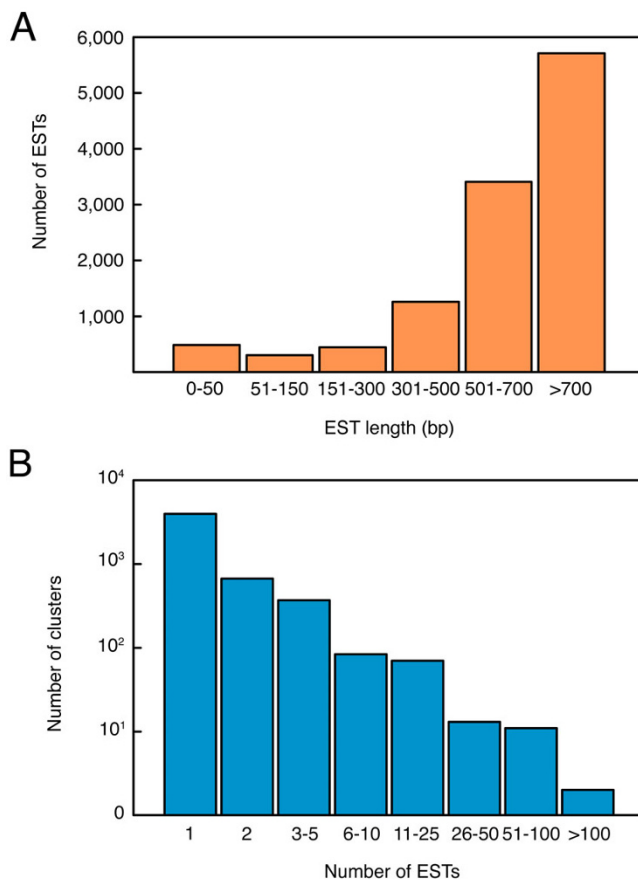


Figure 1
Distribution of Senegalese sole expressed sequence tags (ESTs) in the normalized multi-tissue cDNA library. (A) Sequencing read length of ESTs. Read lengths were binned in 100–200 bp increments. Most of the ESTs were more than 700 bp in length. (B) Number of ESTs in each cluster.

EST annotation and gene ontology

The BLAST2GO program uses BLAST to find homologous sequences for input sequences and extracts gene ontology (GO) terms to each hit using existing annotations. These GO terms are assigned to the query sequence to give an assessment of the biological process, the molecular function and the cellular compartments represented. In our case, 3,086 sequences out of 5,208 ESTs (59.2%) did not show significant similarity to any known protein after BLASTX, while providing product or gene names for 1,769 sequences (33.9%) and unassigned protein matches (hypothetical proteins) for 353 sequences (6.7%). The number of identified ESTs in sole was lower when compared with EST projects carried out on other fish species. This could be related to sequencing from the 3' end, with most of the nucleotide sequences corresponding to the 3' UTR, and artifacts produced during cDNA synthesis and cloning.

The majority of the functionally annotated sequences from sole had GO assignments for biological process (1,526 ESTs), molecular function (1,613 ESTs) and cellular component (1,395 ESTs) categories (Figure 2). Sequences with GO terms corresponding to biological process fell into 10 categories, with most of the ESTs being dedicated to cellular and metabolic processes (57%), and lower, similar amounts dedicated to biological regulation (9%), multicellular organism processes (8%), localization (8%) and developmental processes (7%) (Figure 2A). Regarding molecular function in the GO resource, defined as 'what a gene product does at the biochemical level', we found that 71% of the ESTs were dedicated to binding and catalytic functions, while the remaining ESTs were mostly dedicated to structural molecule activity (8%) and transporter activity (6%) (Figure 2B). Most of the binding functions were at the intracellular level rather than external and included nucleotide and nucleic acid binding, protein binding and ion binding. Catalytic activities included transferase, hydrolase and oxidoreductase activities. In relation to cellular components, Senegalese sole gene products were generally found associated with the intracellular space (50%) or in organelles (27%), with more associated with membrane bound organelles such as mitochondria (Figure 2C). Almost 9.5% were found in protein complexes, such as the ribonucleoprotein complex.

BLASTX and GO analyses, however, did not give full information on the tissue-specific transcription profiles as the ESTs were derived from a multi-tissue cDNA library. Future sequencing of ESTs in a tissue-specific fashion is necessary for this purpose. Nevertheless, the analyses performed did give information on the nature of the ESTs in the current Senegalese sole database, and identified a number of mRNAs of potential interest during gametogenesis and larval development (Table 3). For some of the ESTs, sequencing information from the 3' end allowed to identify isoform-specific ESTs or paralogues (Table 3). The expression pattern of a number of these ESTs, including the complete set of 40S and 60S ribosomal proteins, has been further characterized during spermatogenesis and larval development [29-32,40,41].

Design and validation of a Senegalese sole oligo microarray

We developed an oligonucleotide microarray for studies on *S. senegalensis* gene expression given the greater reproducibility of the data collected with oligonucleotide as opposed to spotted cDNA microarrays [42]. *Tethys*, the proprietary oligo design software from Oryzon Genomics, was used to design an oligo microarray based on the ESTs sequenced. As mRNA labelling procedures such as the Eberwein protocol [43] are biased towards the 3' polyA tail of the transcripts, oligo design was equally biased

Table 2: Largest EST clusters (with ≥ 30 ESTs) in the normalized multi-tissue cDNA library of Senegalese sole

GenBank hit acc. no ^a	Gene identification (species) of top BLAST hit ^a	BLAST E-value ^a	Length (% identity) ^b	No. ESTs
ABQ58114	Vitellogenin Ab (<i>Hippoglossus hippoglossus</i>)	1E-109	185 (78%),	150
CAC45059	Type I keratin S8 (<i>Oncorhynchus mykiss</i>)	1E-137	299 (83%)	136
ABQ58113	Vitellogenin Aa (<i>Hippoglossus hippoglossus</i>)	0	1293 (73%)	112
CAH59609	Apolipoprotein AI precursor (<i>Platichthys flesus</i>)	2E-70	260 (52%)	98
P68140	Actin, alpha skeletal muscle A (<i>Takifugu rubripes</i>)	0	377 (100%)	94
BAB12571	Myosin heavy chain (<i>Pennahia argentata</i>)	0	779 (95%)	82
AAAY21007	Zona pellucida protein Bb (<i>Sparus aurata</i>)	3E-139	353 (77%)	78
BAA85157	Elongation factor 1 alpha (<i>Seriola quinqueradiata</i>)	0	448 (95%)	64
AAAY21009	Zona pellucida protein Ba (<i>Sparus aurata</i>)	5E-164	333 (68%)	54
NP_956752	Eukaryotic translation elongation factor 2, like (<i>Danio rerio</i>)	0	738 (86%)	52
ABN80443	Creatine kinase I (<i>Poecilia reticulata</i>)	0	359 (94%)	47
P61155	40S ribosomal protein S19 (<i>Pagrus major</i>)	8E-64	121 (95%)	46
NP_571231	Keratin 5 (<i>Danio rerio</i>)	2E-138	270 (75%)	44
P84335	Tropomyosin (<i>Pennahia argentata</i>)	6E-139	279 (98%)	44
AAO43607	Serum lectin isoform 3 precursor (<i>Salmo salar</i>)	7E-17	45 (34%)	39
CAA63709	Chorion protein (<i>Sparus aurata</i>)	2E-149	283 (63%)	37
ABBI7040	Heat shock cognate 70 (<i>Fundulus heteroclitus</i>)	3E-100	198 (96%)	32
BAA95131	Myosin light chain 2 (<i>Pennahia argentata</i>)	2E-87	162 (98%)	30

^aMost significant BLASTX hit is reported; ^bExtent of BLASTX hit aligned region (in amino acids), and percent identity over the aligned region.

towards the 3' end of the sequences. Specific 50 to 60-mer oligonucleotide probes were successfully designed for 5,087 out of the 5,208 unigenes assembled from the ESTs (Additional file 1). We also designed probes for control mRNAs, which were spiked into each mixture prior to hybridization, for quality control monitoring and data processing. The final design of each slide, comprising 2 arrays with gene specific probes for all 5,087 *S. senegalensis* unigenes as well as negative and positive control oligos, was submitted for synthesis to Agilent using the eArray platform.

Two independent experiments were conducted to estimate the rate of false-positive expression in the microar-

ray. In the first experiment, self-to-self hybridization, total RNA from two different aliquots of the same sample of sole ovary were used to produce either Cy3 or Cy5 differentially labelled amplified RNA (aRNA). The aRNAs were mixed in equal amounts, with the fluorescent cyanine dyes Cy3 and Cy5, and hybridized in triplicate to the microarray (Figure 3A and 3B). In the second experiment, mRNA from sole larvae was used, from 16-day post-hatching to produce Cy3 labelled aRNA and from 22-day post-hatching to produce Cy5 labelled aRNA, mixed in equal amounts and hybridized in triplicate to the sole microarray (Figure 3C and 3D). We assessed whether the variation of all three fold change (FC) values were within the limits of expected variation, and any considered as an outlier

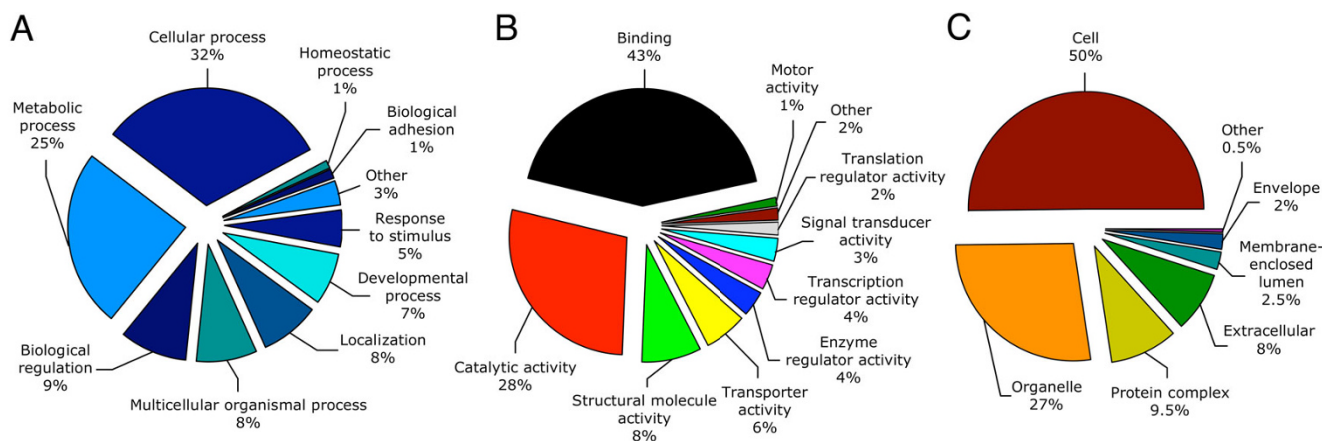


Figure 2
BLAST2GO categories of Senegalese sole ESTs. ESTs were analyzed using the BLAST2GO software. Level 2 categories are shown for biological processes (A), molecular function (B), and cellular component (C).

Table 3: Examples of some ESTs encoding mRNAs of potential interest during gametogenesis and development in Senegalese sole.

GenBank acc. no	Gene identification (species) of top BLAST hit ^a	GenBank hit acc. No	BLAST E-value	Length (% identity) ^b
FF291770	Steroidogenic acute regulatory protein (<i>Equus caballus</i>)	O46689	2E-25	55 (50%)
FF285271	17-beta hydroxysteroid dehydrogenase type 12B (<i>Danio rerio</i>)	Q6QA33	6E-42	87 (72%)
FF288093	Inhibin alpha subunit (<i>Fundulus heteroclitus</i>)	AAW02847	9E-43	76 (73%)
FF282090	Cyclin B1 (<i>Orizyas latipes</i>)	Q9IBG1	1E-29	66 (81%)
FF287053	Cyclin B2 (<i>Oncorhynchus mykiss</i>)	Q09IZ1	3E-43	93 (72%)
FF287848	Cyclin D1 (<i>Danio rerio</i>)	Q90459	3E-32	89 (78%)
FF284385	Gonadotropin I beta subunit (<i>Paralichthys olivaceus</i>)	AAK58601	4E-50	89 (74%)
FF285638	Gonadotropin II beta subunit (<i>Morone saxatilis</i>)	AAC38019	9E-35	64 (81%)
FF291056	Gonadotropin alpha subunit precursor (<i>Epinephelus coioides</i>)	AANI18039	1E-48	86 (86%)
FF283815	Forkhead box F2 (<i>Xenopus tropicalis</i>)	A4IIG1	2E-75	156 (64%)
FF282163	Transcription factor jun-B (<i>Cyprinus carpio</i>)	AAB39939	6E-28	65 (92%)
FF283387	Testis-specific gene A2-like protein (<i>Cyprinus carpio</i>)	Q6VTH5	3E-10	24 (46%)
FF290593	Creatine kinase testis-isozyme (<i>Oncorhynchus mykiss</i>)	P24722	6E-84	162 (90%)
FF290243	Vitellogenin C (<i>Morone americana</i>)	AAZ17417	4E-40	90 (66%)
FF283702	Apolipoprotein A-I (<i>Morone saxatilis</i>)	ACH90229	2E-25	48 (53%)
FF285567	Apolipoprotein A-IV1 (<i>Takifugu rubripes</i>)	Q5KSU4	8E-70	139 (61%)
FF283106	Apolipoprotein A-IV4 (<i>Takifugu rubripes</i>)	Q5KSU1	5E-35	73 (64%)
FF282216	Elongation factor I alpha (<i>Oryzias latipes</i>)	Q9YIC0	3E-35	76 (93%)
FF287281	Elongation factor I beta (<i>Oryzias latipes</i>)	CAB40840	3E-32	67 (90%)
FF287587	Elongation factor I gamma (<i>Danio rerio</i>)	Q8JIU6	9E-90	148 (84%)
FF285240	Elongation factor-I, delta isoform 1 (<i>Danio rerio</i>)	B055L3	6E-61	122 (78%)
FF282440	Myosin, heavy polypeptide 1, skeletal muscle (<i>Danio rerio</i>)	Q7T1B7	1E-124	249 (90%)
FF291822	Myosin 10 (<i>Xenopus laevis</i>)	Q694W8	6E-32	65 (83%)
FF287487	Myosin, heavy polypeptide 11, smooth muscle (<i>Danio rerio</i>)	Q4U0S2	1E-94	217 (92%)
FF289234	Heat shock protein 90-alpha (<i>Danio rerio</i>)	Q90474	3E-121	225 (95%)
FF291891	Heat shock protein 90 beta (<i>Paralichthys olivaceus</i>)	AAQ95586	3E-57	163 (91%)
FF288187	Heat shock protein 70 (<i>Poecilia reticulata</i>)	ABN80448	2E-20	51 (68%)
FF281871	Keratin type I, CK18 (<i>Danio rerio</i>)	Q7ZTS4	2E-104	195 (84%)
FF283334	Keratin type I, CK19 (<i>Pongo abelli</i>)	Q5R8S9	3E-53	104 (70%)

^aMost significant BLASTX hit is reported; ^bExtent of BLASTX hit aligned region (in amino acids), and percent identity over the aligned region.

(e.g., due to imperfections in the array), were eliminated. From this analysis we calculated the mean FC value for each oligo on the array and the distribution (Figure 3B and 3D).

As expected, the narrow distribution obtained with the identical RNA sources in self-to-self hybridization indicated very few differences in gene expression (Figure 3B). In contrast, a much broader distribution was observed for the gene specific probes in the differential gene expression experiment (Figure 3C and 3D), with distribution of the control oligos similar to that in the self-to-self experiment. Individual error rates are a function of the FC and the reproducibility of the values for a single oligo in the experiment, but comparison of the distribution of the gene specific data between self-to-self and differential gene expression experiments gives an estimation of the overall false discovery rate (FDR). The FDR of a test is defined as the expected proportion of false positives in the significant results, and can be estimated by dividing the number of data points or surface under the data curve of the self-to-self (or control) hybridization in the histogram beyond a chosen FC cut-off (e.g., Figure 3B), by the datapoints or

surface beyond the same FC cut-off under the data curve from the differential gene expression (e.g., Figure 3D). Overall, the FDR is expected to increase with lower cut-off FC values, and decrease when the dispersion of the self-to-self data is low, i.e. in function of the overall reproducibility of the microarray platform. In the differential gene expression experiment shown in Figure 3C, we identified 21 up-regulated genes and 71 down-regulated genes, with a $p\text{-value} \leq 0.01$ for genes considered significantly expressed and an FC cut-off $\pm 3\sigma$, in this case equal to ± 1.62 . In the same interval, we identified 1 up-regulated and 2 down-regulated genes for the self-to-self experiment, giving an estimated overall FDR of 3.26%. These results indicate excellent technical reliability of the platform.

The microarray developed may be of great interest in research on Senegalese sole gene expression, given the range of questions that it can be used to address. This array, has already been employed within the Pleurogene program to investigate changes in the transcriptome profiling during gonad differentiation, growth and maturation, and during larval metamorphosis and development.

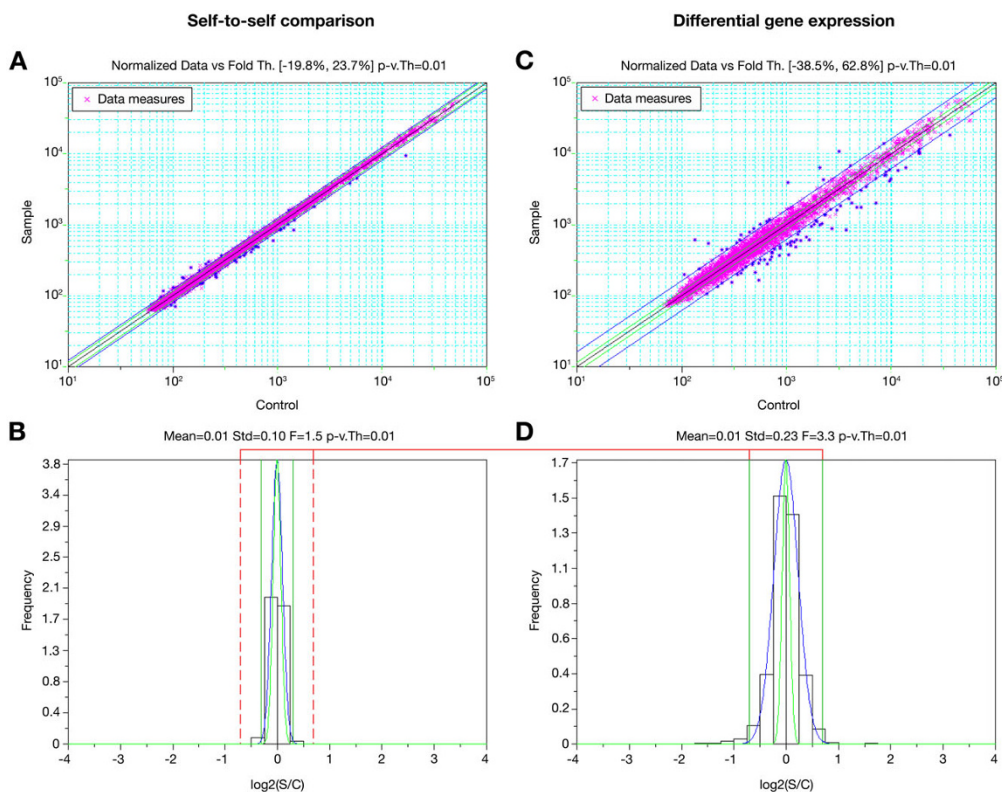


Figure 3

Senegalese sole microarray platform. Scatter plot of the signal intensities in the Cy3 and Cy5 channel of replica analysis of self-to-self (A) and differential gene expression (C) experiments. In both panels, the green and blue lines parallel to the black diagonal line represent the $\pm 3\sigma$ limits of the data from control and *S. senegalensis* specific oligos, respectively. The histograms of the distribution of fold changes (FC) as $\log_2(S/C)$ for control and *S. senegalensis* specific oligos in the self-to-self (B) and differential gene expression (D) experiments are also shown. In these panels, the green and blue curves represent the $\pm 3\sigma$ limits on the data from control and *S. senegalensis* specific oligos, respectively, and the red dotted lines in the self-to-self histogram represent the cut-off chosen in the differential gene expression experiment to estimate overall FDR.

Additional microarrays can be obtained from Agilent using the EST-specific oligos listed in the Additional file 1.

Development of the Soleamold bioinformatic platform

The use of functional genomics approaches can be highly implemented towards the discovery of new genes and genetic pathways by determination of the cellular localization of gene expression using ISH. However, most current databases that integrate both approaches on an anatomical basis in vertebrates are based on model organisms, such as mouse [44] or zebrafish [45] embryos. The Soleamold platform was developed as a bioinformatic resource able to manage and store all experimental data

on gene expression levels and gene localization in *Solea senegalensis* [46]. This platform is based on the Orymold software [47] which integrates microarray experimental data through an ontological description of organs, tissues and cell types of a given organism hierarchically. The current ontology of Soleamold describes *Solea senegalensis* male and female reproductive organs using a total of 94 terms, from 'ovary' and 'testes' to 'late spermatogenesis' and 'follicle maturation stages', but can be extended to cover many different organs and experimental treatments (Figure 4). The ontology is complemented with histological images, forming a coherent, user-navigable atlas that backs up the meaning of the terms included in the hierar-

chical description of Senegalese sole. The user can input new data to both atlas and ontology.

The Soleamold platform integrates experimental data from DNA microarrays and ISH based on the ontology (Figure 5). Once integrated, data can be easily browsed and retrieved. The genome of *Solea senegalensis* has not been sequenced yet, and therefore it was not possible to build a genomic knowledge database to serve as backend support for experimental data. Instead, a BLAST-able [48] database derived from the clustered and annotated sole EST collection was incorporated into the tool. This database dynamically maps and annotates each of the probes contained in a DNA microarray or employed for ISH. Once probes have been introduced into the system, experimental data is inserted by selecting a term in the ontology and uploading the expression values that are automatically linked to the backend database through probe names. Users are also able to determine the quality of the introduced data and comparability of the results. When dealing with microarray data, the same reference sample must be used for all hybridization experiments, ensuring future cross-comparability of data.

The information in Soleamold is retrieved while maintaining constant visual contact with the ontology and the 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?', based on ontology terms that allow for the isolation of genes whose expression values meet specific criteria spatially and temporally. 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 current experimental database of Soleamold includes nineteen normalized and scaled microarray hybridizations to compare the global pattern of gene expression during ovarian and testicular development, including data from males treated with different hormones, and few ISH experiments. Detailed analysis of these experiments, as well as validation of the FC variations in expression by real-time quantitative PCR, will be published elsewhere. The Soleamold can be extended to additional tissues and organs of Senegalese sole, as well as of other flatfish, after normalization of microarray gene expression data.

Conclusion

The lack of genomic resources for non-model fish species of commercial interest has been in stark contrast to the increasing importance of aquaculture to counteract the world-wide decline in wild fishery. By sequencing and

analyzing 10,185 *Solea senegalensis* ESTs, we have generated new genomic resources for this commercially important flatfish. In addition to contributing to the characterization of fish genes of unknown function, the EST survey will be useful to identify molecular markers for future genetic studies in Senegalese sole and to provide probes for ISH cellular localization studies. The DNA microarray can be used for further investigation of physiologically and ecologically relevant transcriptional regulation in *Solea senegalensis*. Finally, the Soleamold interactive morphoanatomical atlas of gene expression may aid in the identification of crucial genes and regulatory pathways involved in different physiological process in flatfish, as such, it may be a powerful research tool for both academia and industry.

Methods

Animals and sampling

Adult, sexually mature Senegalese sole males and females ($n = 25$; 749 ± 36 g in weight) were captured in the salt marshes of the bay of Cadiz (Spain). Fish were maintained in 28 m² round fibre glass tanks in the laboratory for approximately 30 days under natural conditions of photoperiod and temperature as described [49]. After this period, fish were killed by sedation in 100 ppm 2-phenox-yethanol and decapitation, and samples of the gonads, liver, brain (including pituitary), stomach and gut were rapidly dissected. Larvae were raised from fertilized eggs spawned by wild-caught broodstock adapted to captivity and maintained under ambient conditions as described [40]. Samples of larvae were collected before metamorphosis (up to day 12 postfertilization; 6–7 mm) and during and after metamorphosis (from day 14 up to day 23 postfertilization; 8–9 mm). Samples of larvae with albinism, malpigmentation, osteological malformations, and incomplete eye migration, were also taken. Samples from undifferentiated gonads were taken from approximately 120 fish between 138 and 205 days post-fertilization (43–112 mm total length). In all cases, tissues were flash-frozen in liquid nitrogen and stored at -80°C prior to RNA extraction. The protocols for the sacrifice of fish were approved by the Ethics and Animal Experimentation Committee from IRTA (Spain) in accordance with the Guiding Principles for the Care and Use of Laboratory Animals.

cDNA libraries and normalization

Total RNA was extracted from frozen adult tissues (from 3–5 different fish) and larvae using the RNeasy kit (Qiagen), and poly(A)⁺ RNA purified using the Oligotex Direct mRNA Mini Kit (Qiagen). For larvae, samples at the same developmental stage were pooled to obtain 1 g total tissue for RNA extraction. Ten different cDNA libraries (Table 1) were unidirectionally constructed (5' *EcoRI*, 3' *XhoI*) using the ZAP Express[®] cDNA Synthesis Kit following the manu-

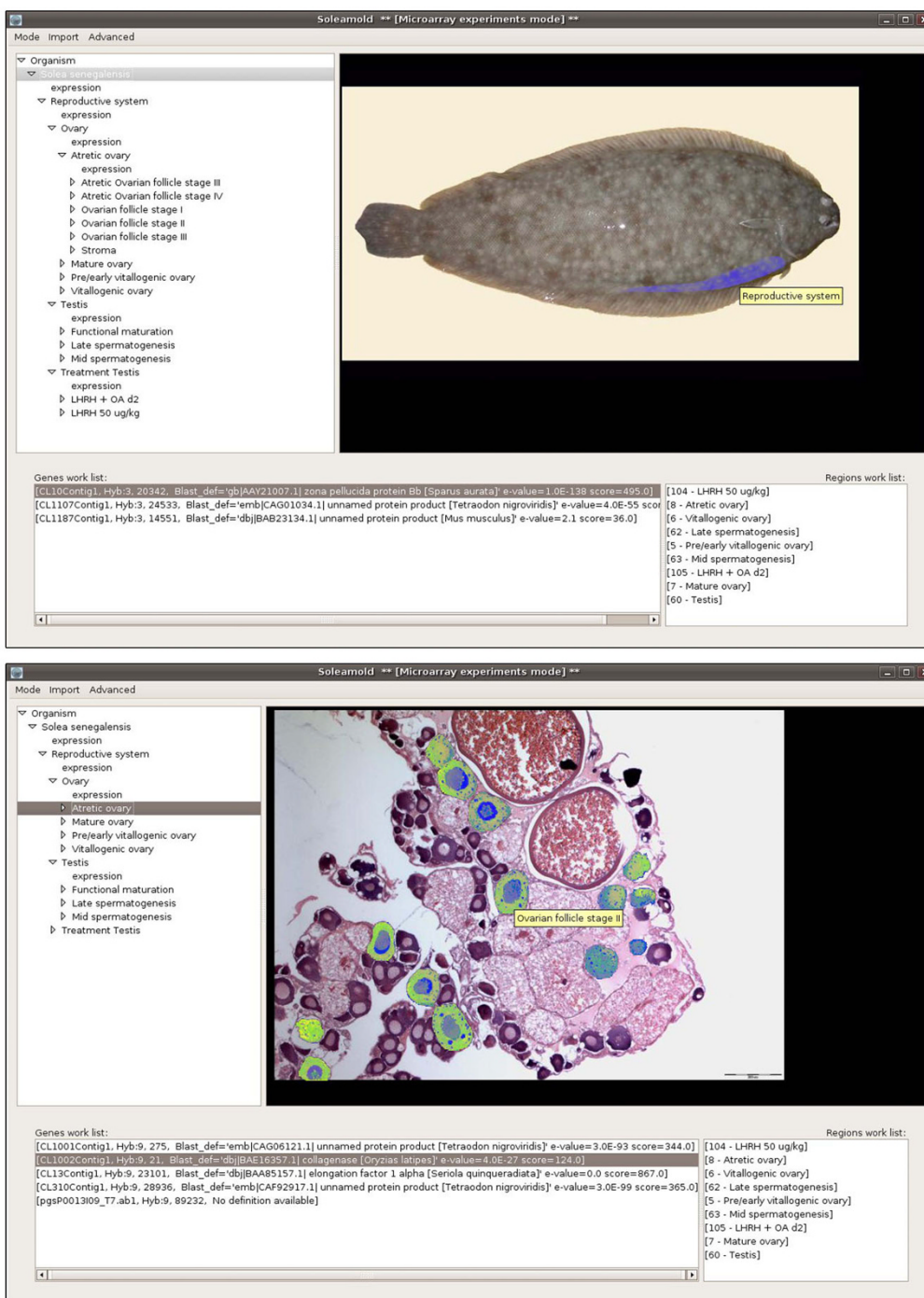


Figure 4
Soleamold main interface and navigation of the atlas. The upper panel shows the Soleamold main interface which is divided into three interrelated sections: a hierarchical tree with the ontology describing the organism (upper left), an atlas coherent with the terms of the ontology (upper right), and two work lists for the regions (i.e., samples) and genes of interest. All parts of the interface are interactive and/or navigable through contextual menus and have been designed to be visually intuitive, allowing for constant contextualization of the data. The lower panel shows that the construction of an atlas coherent with the ontology allows for the construction of maps that, when overlaid with the original picture, highlight the different regions that can be identified in the picture. Thus, users can slide the cursor over any image and rapidly identify and/or select each region in it. In the figure, follicles at stage II from an atretic ovary are highlighted.

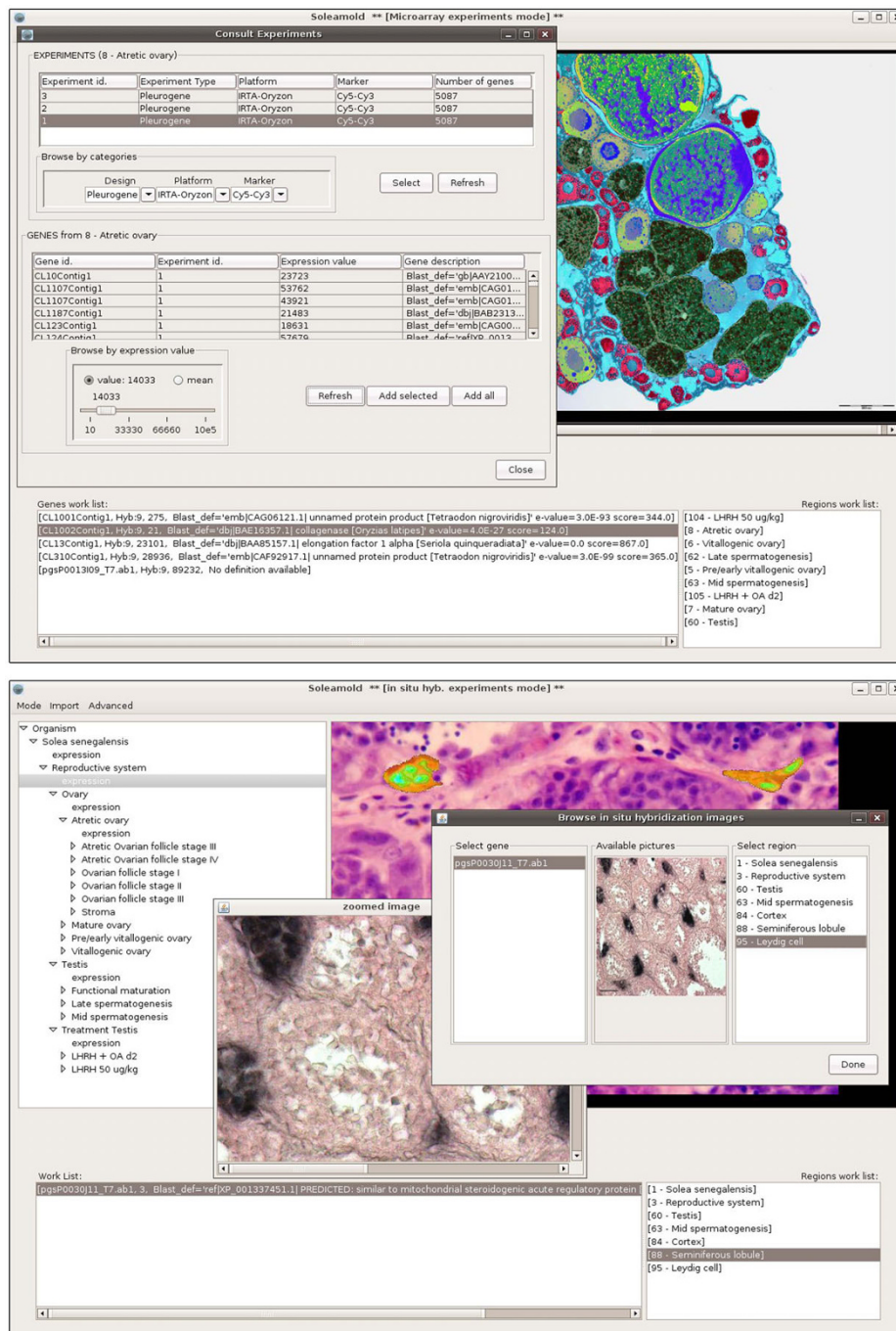


Figure 5
DNA Microarray and ISH modules. By consulting a term of the ontology a pop-up frame allows the user to browse DNA microarray experiments present in the database, as shown in the upper panel. Experiments can be browsed using several parameters: chip design, hybridization platform and labelling. Once an experiment is selected, those genes showing expression appear in the lower list. Expression values can then be selected according to a lower threshold value and sent to the work list of the main interface for further consultation. The lower panel shows that Soleamold also allows for the storage, consultation and interpretation of ISH experiments. The figure shows the distinct browsing interfaces for this type of data. In the foreground, the original experimental pictures browsing interface shows steroidogenic acute regulatory (StAR) mRNA expression in Leydig cells, and its zoomed image. In the background, there is the automatically built, atlas image, of the same experiment showing the expression in Leydig cells as interpreted by the user at the insertion time.

facturer's instructions (Stratagene). Analysis of size-fractionated *Xho*I-digested cDNAs, performed immediately prior to ligation into the vector, was done using 5% non-denaturing acrylamide gel electrophoresis or the 2100 Bioanalyzer (Agilent) and DNA 7500 LabChip kit (Agilent). For normalization, an aliquot of each library, depending on its titre after one round of amplification (Table 1), was pooled into a master library which was mass-excised according to the manufacturer's instructions, and subsequently normalized through three successive rounds of plating and hybridization. In the first round, colonies were plated on large LB+ kanamycin plates and 384 clones selected randomly and transferred to 384-well plates. A ³²P labelled probe was prepared from the inserts of the pool of these selected clones using PCR and primers corresponding to either end of the multicloning site of the vector. For the second hybridization round, we plated the library and obtained colony blots on Hybond N+ filters (Amersham), following the manufacturer's instructions. These were hybridized with the first PCR probe. From the second round of screening, 3,456 non-hybridizing clones were selected and transferred to 384-well plates. The pool of these selected clones was then used to prepare a second radioactive labelled PCR probe which was hybridized to additional colony blots. From this third round of normalization, 6,528 non-hybridizing colonies and all selected clones were sequenced.

Sequencing, sequence analysis and contig assembly

The normalized master library (10,368 cDNA clones) was arrayed in Genetix X7001 flat 384-well plates. Glycerol stocks of overnight cultures were prepared in 384-well format. Plasmid DNAs were extracted from cultures grown overnight in deep 384-well plates containing 200 µl of 2YT Media. Plasmids were isolated and purified automatically by a modified alkaline lysis procedure. Quality control of plasmids was done by randomly testing eight samples of each plate on agarose gels. Sequences were determined using the T7 universal primer, ABI Big Dye Terminator 3 chemistry and ABI 3730XL capillary sequencer systems (ABI, Weiterstadt, Germany). Base-calling from chromatogram traces was performed using PHRED [50,51]. Vector, *Escherichia coli* contamination, and low-quality regions were trimmed from EST sequences.

Before clustering, EST sequences were subjected to an extensive quality control procedure. Firstly, sequences were clipped at both ends as long as PHRED quality values stayed below 20 in a window of 15 bp in size. Secondly, putative vector and repeat sequences were masked for all sequences. PHRAP [52], under stringent clustering parameters (minimum score, 100; repeat stringency, 0.99), was used to assemble ESTs into contigs. Consed (Version 14.00) [53] was used for final editing of the sequence.

Contig consensus sequences and singleton sequences were aligned with nonredundant GenBank nucleotide and amino acid sequence databases using BLASTN and BLASTX, respectively [48,54].

Functional characterization of ESTs

Senegalese sole contig consensus sequences and singleton sequences were submitted for GO annotation to the online version of the BLAST2GO v1 program [55,56]. Annotated accession numbers and GO numbers were derived with NCBI's QBLAST, with an expectation E-value $\leq 10^{-3}$ and an HSP length cut-off of 33. Contig sequences were then annotated according to the following parameters: a pre-E-value-Hit-Filter of 10^{-6} , a pro-Similarity-Hit-Filter of 15, an annotation cut-off of 55, and a GO weight of 5. Directed acyclic graphs (DAG's) were generated using a sequence filter of 5, an alpha score of 0.6 and a 0 node score filter. A score was computed by Blast2GO to highlight areas of high annotation. This score was computed at each node according to the following equation:

$$\text{Score} = \sum_{GOs} seq \times \alpha^{dist}$$

where *seq* is the number of different sequences annotated at a child GO term, *dist* the distance to node of the child GO term, and α the parameter entered by the user. The data presented represent the level 2 analysis, illustrating general functional categories.

Oligo design, microarray fabrication and quality control

The EST unigenes were used to design microarray oligo-probes using *Tethys*, Oryzon's proprietary software which is based on the thermodynamic simulation of hybridization *in silico*. The final array design contained 5,087 gene specific *Solea senegalensis* probes, as well as positive and negative control probes (maize expansin, ZmMYB42, and xyloglucan endo-transglycosylase), a total of 700 per array. These were used to assess detection limits and range, to verify spatial homogeneity, and to determine experimental within-array variation. Microarray slides, each containing two microarrays, were synthesized by Agilent using ink-jet printing and *in situ* oligonucleotide synthesis.

RNA extraction, microarray hybridization and analysis

Total RNA was extracted from Senegalese sole ovaries and testes using the RNeasy extraction kit (Qiagen) followed by DNase treatment. The quality and concentration of the RNA was analyzed using the Agilent 2100 bioanalyzer and NanoDrop™ ND-1000 (Thermo Scientific). Samples with RNA integrity number (RIN; [57]) < 6.0 were discarded. Total RNA (0.5 µg) amplification and labelling with Cy3 or Cy5 was carried out using the Eberwein mRNA amplification procedure [43] employing the MessageAmp™

aRNA amplification kit from Ambion (Applied Biosystems) following the manufacturer's instructions with minor modifications.

The Cy3- and Cy5-labelled cRNAs were combined and hybridized to the microarray for 17 h at 60°C using Agilent's gaskets G2534-60002, G2534A hybridization chambers and DNA Hybridization Oven G2545A, according to the manufacturer's instructions. Raw data were obtained using Agilent's DNA Microarray Scanner G2505B and Feature Extraction software (v10.1). The raw fluorescence intensity data were processed using the *Polyphemus*[™] software, developed at Oryzon Genomics, which includes spatial data compensation, non-significant expressed data filtering, and data normalization. Data normalization was carried out by an improved version of the nonlinear *Q-splines* normalization method [58], enhanced with robust regression techniques. Normalized and log-transformed data were used to calculate FC values. Differential expression was assessed with *Polyphemus*[™] analysing biological replicates based on repeated experiments using robust statistics on average technical replicates removing the outlier points (caused by dust or array imperfection). The *p*-values were calculated based on the absolute value of the *regularized t-statistic* [59], which uses a Bayesian framework to derive the algorithm, using internal replicated controls to assess the minimum technical variability of the process. The *regularized t* approach is more sensitive than the significance analysis of microarrays (SAM) method [60,61]. Cut-offs for significant changes were chosen in various ways, but they were always greater than the inherent experimental variation as assessed by the FC of internal controls and/or self-to-self hybridizations.

Histology and ISH

Samples of gonads were transferred into Bouin fixative for approximately 48 h at room temperature, and subsequently dehydrated and embedded in Paraplast (Sigma). In situ hybridization of eosin-haematoxylin stained sections (7 µm) was carried out with digoxigenin-alkaline phosphatase (DIG-AP) incorporated cRNA probes as described [41]. DIG-AP riboprobes were synthesized with T3 and T7 RNA polymerases using the DIG RNA Labeling Kit (Roche). Sections were examined and photographed with a Leica DMLB light microscope.

Construction of Soleamold

Soleamold was designed following a two-tier architecture: a desktop standalone client in charge of data presentation and processing, and a relational database management system (DBMS) for handling data organization, storage and retrieval. This architecture allows for concurrent remote access by users in a fairly scalable manner. Standalone applications are more difficult to implement but

they allow a higher degree of interactivity, and in general, for a much richer user experience than those that are web based. Java J2SDK [62] was chosen for client implementation to ensure portability through operative systems while MySQL [63] was chosen as DBMS. MySQL proprietary Java Database Connectivity (JDBC) driver was used for client connection. Soleamold is available through the Spanish National Institute of Bioinformatics website [44].

Abbreviations

EST: Expression sequence tag; ISH: In situ hybridization; ZP: Zona pellucida; GO: Gene ontology; FDR: False discovery rate; SAM: Significance analysis of microarrays; DMS: Database management system.

Authors' contributions

JC conceived and designed the study. MM, CI, AA y GMR constructed the larvae cDNA libraries, JV and FP the undifferentiated gonads library, SH, FP and GMR the brain library, CI, AA, MY and GMR the stomach, intestine and liver libraries, ATS and JC the ovarian library, and BC and JVP the testis library. JC and JVL performed the bioinformatic analyses of ESTs. OD, FS and ER, assisted by TM, normalized the sole multi-tissue cDNA library, designed microarray oligonucleotides and carried out microarray hybridizations and bioinformatic analyses. JM and TM, assisted by JC and ATS, designed and constructed the Soleamold. JC and TM wrote and edited the manuscript. All authors read and approved the manuscript.

Additional material

Additional file 1

Sequence of oligos for each *S. senegalensis* unigene employed in the microarray.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-508-S1.xls>]

Acknowledgements

This research was supported by Genome Spain and Genome Canada within the framework of the international consortium Pleurogene[™]. We thank J. L. Gelpí and Dmitry Repchevsky for technical assistance to install the Soleamold software in the website of the Spanish National Institute of Bioinformatics.

References

1. Crollius HR, Weissenbach J: **Fish genomics and biology.** *Genome Res* 2005, **15**:1675-1682.
2. Meyer A, Peer Y van de: **From 2R to 3R: evidence for a fish-specific genome duplication (FSGD).** *Bioessays* 2005, **27**:937-945.
3. Rise ML, von Schalburg KR, Brown GD, Mawer MA, Devlin RH, Kuipers N, Busby M, Beetz-Sargent M, Alberto R, Gibbs AR, Hunt P, Shukin R, Zeznik JA, Nelson C, Jones SR, Smailus DE, Jones SJ, Schein JE, Marra MA, Butterfield YS, Stott JM, Ng SH, Davidson WS, Koop BF: **Development and application of a salmonid EST database**

- and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Res* 2004, **14**:478-490.
4. Von Schalburg K, Rise ML, Cooper GA, Brown GD, Gibbs AR, Nelson CC, Davidson WS, Koop BF: **Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray.** *BMC Genomics* 2005, **6**:126.
 5. Govoroun M, Le Gac F, Guiguen Y: **Generation of a large scale repertoire of Expressed Sequence Tags (ESTs) from normalized rainbow trout cDNA libraries.** *BMC Genomics* 2006, **7**:196.
 6. Bobe J, Montfort J, Nguyen T, Fostier A: **Identification of new participants in the rainbow trout (*Oncorhynchus mykiss*) oocyte maturation and ovulation processes using cDNA microarrays.** *Reprod Biol Endocrinol* 2006, **4**:39.
 7. Adzhubei AA, Vlasova AV, Hagen-Larsen H, Ruden TA, Laerdahl JK, Hoyheim B: **Annotated expressed sequence tags (ESTs) from pre-smolt Atlantic salmon (*Salmo salar*) in a searchable data resource.** *BMC Genomics* 2007, **8**:209.
 8. Von Schalburg KR, Leong J, Cooper GA, Robb A, Beetz-Sargent MR, Lieph R, Holt RA, Moore R, Ewart KV, Driedzic WR, Ten Hallers BF, Zhu B, de Jong PJ, Davidson WS, Koop BF: **Rainbow Smelt (*Osmerus mordax*) genomic library and EST resources.** *Mar Biotechnol (NY)* 2008, **10**(5):487-491.
 9. Gong Z, Hu Z, Gong ZQ, Kitching R, Hew CL: **Bulk isolation and identification of fish genes by cDNA clone tagging.** *Mol Mar Biol Biotechnol* 1994, **3**:243-251.
 10. Inoue S, Nam B-H, Hirono I, Aoki T: **A survey of expressed genes in Japanese flounder (*Paralichthys olivaceus*) liver and spleen.** *Mol Mar Biol Biotechnol* 1997, **6**:376-380.
 11. Douglas SE, Gallant JW, Bullerwell CE, Wolff C, Munholland J, Reith ME: **Winter flounder expressed sequence tags: establishment of an EST database and identification of novel fish genes.** *Mar Biotechnol* 1999, **1**:458-464.
 12. Douglas SE, Knickle LC, Kimball J, Reith ME: **Comprehensive EST analysis of Atlantic halibut (*Hippoglossus hippoglossus*), a commercially relevant aquaculture species.** *BMC Genomics* 2007, **8**:144.
 13. Sarropoulou E, Power ME, Magoulas A, Geisler R, Kotoulas G: **Comparative analysis and characterization of expressed sequence tags in gilthead sea bream (*Sparus aurata*) liver and embryos.** *Aquaculture* 2005, **243**:69-81.
 14. Zhou L, Yao B, Xia W, Li C-J, Wang Y, Shi Y-H, Gui J-F: **EST-based identification of genes expressed in the hypothalamus of male orange-spotted grouper (*Epinephelus coioides*).** *Aquaculture* 2006, **256**:129-139.
 15. Chini V, Rimoldi S, Terova G, Saroglia M, Rossi F, Bernardini G, Gornati R: **EST-based identification of genes expressed in the liver of adult seabass (*Dicentrarchus labrax* L.).** *Gene* 2006, **376**:102-106.
 16. Nelson JS: *Fishes of the World* 4th edition. New York: John Wiley & Sons; 2006.
 17. Okada N, Takagi Y, Seikai T, Tanaka M, Tagawa M: **Asymmetrical development of bones and soft tissues during eye migration of metamorphosing Japanese flounder, *Paralichthys olivaceus*.** *Cell Tissue Res* 2001, **304**:59-66.
 18. Power DM, Llewellyn L, Faustino M, Nowell MA, Bjornsson BT, Einarsson IE, Canario AV, Sweeney GE: **Thyroid hormones in growth and development of fish.** *Comp Biochem Physiol Part C Toxicol Pharmacol* 2001, **130**:447-459.
 19. Schreiber AM: **Asymmetric craniofacial remodeling and lateralized behavior in larval flatfish.** *J Exp Biol* 2006, **209**:610-621.
 20. Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, Jackson JB, Lotze HK, Micheli F, Palumbi SR, Sala E, Selkoe KA, Stachowicz JJ, Watson R: **Impacts of biodiversity loss on ocean ecosystem services.** *Science* 2006, **314**:787-790.
 21. Dinis MT, Ribeiro L, Soares F, Saraquete C: **A review on the cultivation potential of *Solea senegalensis* in Spain and in Portugal.** *Aquaculture* 1999, **176**:27-38.
 22. Imsland AK, Foss A, Conceicao LEC, Dinis MT, Delbare D, Schram E, Mastroratti A, Rema P, White P: **A review of the culture potential of *Solea solea* and *S. senegalensis*.** *Rev Fish Biol Fish* 2003, **13**:379-407.
 23. **Pleurogene** [<http://www.pleurogene.ca>]
 24. Cerdà J, Douglas SE, Buesa C, Cañavate P, López-Barea J, Manchado M, Martínez G, Navas JJ, Piferrer F, Planas J, Prat F, Reith M, Ruiz-Rejón M, Yúfera M: **PLEUROGENE: flatfish genomics and proteomics for aquaculture.** *Comp Biochem Physiol* 2005, **141A**:s87.
 25. **UniGene database, National Centre for Biotechnology Information (NCBI)** [<http://www.ncbi.nlm.nih.gov/UniGene>]
 26. Smith TP, Grosse WM, Freking BA, Roberts AJ, Stone RT, Casas E, Wray JE, White J, Cho J, Fahrenkrug SC, Bennett GL, Heaton MP, Laegreid WW, Rohrer GA, Chitko-McKown CG, Perrea G, Holt I, Karamycheva S, Liang F, Quackenbush J, Keele JW: **Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle.** *Genome Res* 2001, **11**:626-30.
 27. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, Shagin DA: **Simple cDNA normalization using kamchatka crab duplex-specific nuclease.** *Nucleic Acids Res* 2004, **32**:e37.
 28. Fernández-Díaz C, Yúfera M, Cañavate JP, Moyano FJ, Alarcón FJ, Díaz M: **Growth and physiological changes during metamorphosis of Senegal sole reared in the laboratory.** *J Fish Biol* 2001, **58**:1086-1097.
 29. Manchado M, Infante C, Asensio E, Cañavate JP, Douglas SE: **Comparative sequence analysis of the complete set of 40S ribosomal proteins in the Senegalese sole (*Solea senegalensis* Kaup) and Atlantic halibut (*Hippoglossus hippoglossus* L.) (Teleostei: Pleuronectiformes): phylogeny and tissue- and development-specific expression.** *BMC Evol Biol* 2007, **7**:107.
 30. Manchado M, Infante C, Asensio E, Planas J, Cañavate JP: **Thyroid hormones down-regulate thyrotropin beta subunit and thyroglobulin during metamorphosis in the flatfish Senegalese sole (*Solea senegalensis* Kaup).** *Gen Comp Endocrinol* 2008, **155**:447-455.
 31. Infante C, Asensio E, Cañavate JP, Manchado M: **Molecular characterization and expression analysis of five different elongation factor 1 alpha genes in the flatfish Senegalese sole (*Solea senegalensis* Kaup): differential gene expression and thyroid hormones dependence during metamorphosis.** *BMC Mol Biol* 2008, **9**:19.
 32. Matsuoka MP, Infante C, Reith M, Cañavate JP, Douglas SE, Manchado M: **Translational Machinery of Senegalese Sole (*Solea senegalensis* Kaup) and Atlantic Halibut (*Hippoglossus hippoglossus* L.): Comparative Sequence Analysis of the Complete Set of 60S Ribosomal Proteins and their Expression.** *Mar Biotechnol (NY)* 2008. doi: 10.1007/s10126-008-9104-y.
 33. Babin PJ, Carnevali O, Lubzens E, Schneider WJ: **Molecular aspects of oocyte vitellogenesis in fish.** In *The Fish Oocyte: From Basic Studies to Biotechnological Applications* Edited by: Babin P, Cerdà J, Lubzens E. The Netherlands: Springer; 2007:39-76.
 34. Modig C, Westerlund L, Olsson P-E: **Oocyte zona pellucida proteins.** In *The Fish Oocyte: From Basic Studies to Biotechnological Applications* Edited by: Babin P, Cerdà J, Lubzens E. The Netherlands: Springer; 2007:113-139.
 35. Zheng S, Gong Z: **Expressed sequence tag analysis of expression profiles of zebrafish testis and ovary.** *Gene* 2002, **294**:45-53.
 36. Knoll-Gellida A, Andre M, Gattegno T, Forgue J, Admon A, Babin PJ: **Molecular phenotype of zebrafish ovarian follicle by serial analysis of gene expression and proteomic profiling, and comparison with the transcriptomes of other animals.** *BMC Genomics* 2006, **7**:46.
 37. Poltronieri C, Maccatrozzo L, Simontacchi C, Bertotto D, Funkenstein B, Patrino M, Radaelli G: **Quantitative RT-PCR analysis and immunohistochemical localization of HSP70 in sea bass *Dicentrarchus labrax* exposed to transport stress.** *Eur J Histochem* 2007, **51**:125-135.
 38. Geist J, Wernner I, Eder KJ, Leutenegger CM: **Comparisons of tissue-specific transcription of stress response genes with whole animal endpoints of adverse effect in striped bass (*Morone saxatilis*) following treatment with copper and esfenvalerate.** *Aquat Toxicol* 2007, **85**:28-39.
 39. Bermejo-Nogales A, Benedito-Palos L, Saera-Vila A, Calduch-Giner JA, Sirtjà-Bobadilla A, Pérez-Sánchez J: **Confinement exposure induces glucose regulated protein 75 (GRP75/mortalin/mtHsp70/PBP74/HSPA9B) in the hepatic tissue of gilthead sea bream (*Sparus aurata* L.).** *Comp Biochem Physiol Part B Biochem Mol Biol* 2008, **149**:428-438.
 40. Manchado M, Infante C, Asensio E, Cañavate JP: **Differential gene expression and dependence on thyroid hormones of two glyceraldehyde-3-phosphate dehydrogenases in the flatfish Senegalese sole (*Solea senegalensis* Kaup).** *Gene* 2007, **400**:1-8.

41. Cerdà J, Chauvigne F, Agulleiro MJ, Marin E, Halm S, Martínez-Rodríguez G, Prat F: **Molecular cloning of Senegalese sole (*Solea senegalensis*) follicle-stimulating hormone and luteinizing hormone subunits and expression pattern during spermatogenesis.** *Gen Comp Endocrinol* 2008, **156**:470-481.
42. Brennan C, Zhang Y, Leo C, Feng B, Cauwels C, Aguirre AJ, Kim M, Protopopov A, Chin L: **High-resolution global profiling of genomic alterations with long oligonucleotide microarray.** *Cancer Res* 2004, **64**:4744-4748.
43. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH: **Amplified RNA synthesized from limited quantities of heterogeneous cDNA.** *Proc Natl Acad Sci USA* 1990, **87**:1663-1667.
44. **Edinburgh Mouse Atlas Project (emap)** [<http://genex.hgu.mrc.ac.uk/>]
45. **The Zebrafish Model Organism Database (ZFIN)** [<http://zfin.org>]
46. **The Soleamold bioinformatic platform** [<http://inb.bsc.es/pleurogene/>]
47. Mercadé J, Espinosa A, Adsua JE, Adrados R, Segura J, Puigdomenech P, Maes T: **Orymold: ontology based gene expression data integration and analysis tool applied to rice.** *BMC Bioinformatics* in press.
48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
49. Anguis V, Cañavate P: **Spawning of captive Senegal sole (*Solea senegalensis*) under a naturally fluctuating temperature regime.** *Aquaculture* 2005, **243**:133-145.
50. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using Phred. I. Accuracy assessment.** *Genome Res* 1998, **8**:175-185.
51. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8**:186-194.
52. **Laboratory of Phil Green, Genome Sciences Department, University of Washington** [<http://www.phrap.org>]
53. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8**:195-202.
54. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: A new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
55. **Bioinformatics Centro de Investigación Príncipe Felipe CIPF, Centro de Genómica Instituto Valenciano de Investigaciones Agrarias, Universidad Politécnica de Valencia** [<http://www.Blast2GO.de>]
56. Conesa A, Gotz S, García-Gómez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**:3674-3676.
57. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T: **The RIN: an RNA integrity number for assigning integrity values to RNA measurements.** *BMC Mol Biol* 2006, **7**:3.
58. Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielsen HB, Saxild H-H, Nielsen C, Brunak S, Knudsen S: **A new non-linear normalization method for reducing variability in DNA microarray experiments.** *Genome Biol* 2002, **3**:research0048.
59. Baldi P, Long AD: **A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes.** *Bioinformatics* 2001, **17**:509-519.
60. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
61. Cui X, Hwang JT, Qiu J, Blades NJ, Churchill GA: **Improved statistical tests for differential gene expression by shrinking variance components estimates.** *Biostatistics* 2005, **6**:59-75.
62. **Java Programming Language** [<http://www.java.com>]
63. **MySQL** [<http://www.mysql.com>]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

