Project n° 513692

Combined genetic and functional genomic approaches for stress and disease resistance marker assisted selection in fish and shellfish

AQUAFIRST

Instrument	STREP
Thematic Priority	1.3 Modernisation and sustainability of fisheries, including aquaculture-based production systems

Final Report (Part. 1)

Period covered:	2004-01-11 to 2008-31-04
Date of preparation:	2008-15-09
Contract Start Date :	2004-01-11
Duration:	42 months
Project coordinator name:	**Patrick Prunet**
Project coordinator organisation name:	**INRA Rennes**

DG MARE	**Project funded by the European Community**

Table of contents

Publishable executive summary

Type of contract: STREP		**Total project cost** (in euro): **5 821 320**
Contract number **513692**	**Duration** (in months) 42 Months	**EU contribution** (in euro) 3 799 954
Start date 01 November 2004		**Period covered by the project** 01 November 2004- 1 June 2008

PROJECT COORDINATOR

Name :
Dr Patrick PRUNET

Address :
INRA/SCRIBE
campus de Beaulieu
35042 RENNES, cedex
France

Telephone : **(33)223-48-50-14**	**Telefax :** **(33)223-48- 50-20**	**E-mail address :** Patrick.Prunet@rennes.inra.fr

World wide web address: http://aquafirst.vitamib.com/

List of participants

CO01 (Coordinator) : Institut National de la Recherche Agronomique (INRA)
Responsible scientist : Patrick PRUNET
INRA/SCRIBE, campus de Beaulieu, 35042 RENNES Cedex, France.
Tel : (33)223-48-50-14
Mail : prunet@rennes.inra.fr

CO01.1: Institut National de la Recherche Agronomique (INRA)
Responsible scientist : Edwige QUILLET
INRA/Laboratoire de génétique des poissons, 78350 Jouy-en-Josas, France.
Tel : (33)1 34 65 23 93
Mail : edwige.quillet@jouy.inra.fr

CR02 : University of Aberdeen (UNIABN)
Responsible scientist : Chris SECOMBES
Department of Zoology, University of Aberdeen, 2 Tillydrone Avenue, AB24 2TZ Aberdeen, UK
Tel: (44)1224-272872
Mail: c.secombes@abdn.ac.uk

CR03 : The Natural Environment Research Council (NERC)
Responsible scientist: Tom POTTINGER
Centre for Ecology and Hydrobiology, Windermere Laboratory, Far Sawrey, LA22 0LP Ambleside, Cumbria, UK
Tel: (44)1524 595854
Mail: tgp@ceh.ac.uk

CR04 : Uppsala University
Responsible scientist: Svante WINBERG
Evolutionary Biology Centre (EBC), Department of Comparative Physiology, Uppsala University, Norbyvaegen 18 A, SE-752 36 Uppsala, Sweden.
Tel: (46)18-471-26-49
Mail: svante.winberg@ebc.uu.se

CR05: Ifremer La Tremblade
Responsible scientist : Sylvie LAPEGUE
Laboratoire de génétique et pathologie, Ronce les Bains, 17390 La Tremblade, FRANCE
Tel: (33)546762631
Mail: sylvie.lapegue@ifremer.fr

CR05.1: Ifremer Brest
Responsible scientist : Jeanne MOAL
Laboratoire de Departement de Physiologie fonctionnelle, Laboratoire de Physiologie des invertébrés, BP 70, 29280 Plouzané, France
Tel : (33)2 98 22 43 00
Mail: jmoal@ifremer.fr

CR06: CSIC
Responsible scientist : Jaume PEREZ SANCHEZ
Fish Growth Endocrinology, Instituto de Acuicultura de Torre de la Sal (CSIC), 12595, Ribera de Cabanes, Castellón, SPAIN
Tel : (34) 964319500
Mail: jperez@iats.csic.es

CR07: ICCM
Responsible scientist : Daniel MONTERO
Instituto Canario de Ciencias Marinas, Gobierno de Canarias, P.O. Box 56, 35200 Telde, Las Palmas, Canary Islands, SPAIN
Tel : (34) 928 132 900
Mail: dmontero@iccm.rcanaria.es

CR08: Wageningen
Responsible scientist : Hans KOMEN
Wageningen University, leerstoelgroep Fokkerij en Genetica, Bode 27, Postbus338, 6700 AH Wageningen, THE NETHERLANDS
Tel : (31) 317 4 82465
Mail: jmoal@ifremer.fr

CR09 : NUIG
Responsible scientist: Michael CAIRNS
Gene regulation group, National Diagnostic Centre, National University of Ireland, University road, Galway, IRELAND.
Tel: (353)91-512405
Mail: michael.cairns@nuigalway.ie

CR10: PML
Responsible scientist : Anthony J S HAWKINS
Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, PL1 3DH, UK
Tel : (44)1752633443
Mail: ajsh@pml.ac.uk

CR11: CCMAR
Responsible scientist :Adelino CANARIO
Comparative and Moleculare Endocrinology, Universidade do Algarve, Centro de Ciências do Mar, 8000-117, Faro, PORTUGAL
Tel : (351)289800925
Mail: acanario@ualg.pt

CR12: KULeuven
Responsible scientist : Filip VOLCKAERT
Katholieke Universiteit Leuven, Laboratory of Aquatic Ecology, Research Group on Fish Genetics, Ch. De Bériotstraat 32, B-3000 Leuven, BELGIUM
Tel : (32)16 32 39 72
Mail: filip.volckaert@bio.kuleuven.ac.be

CR13: HCMR
Responsible scientist : Giorgios KOTOULAS
Hellenic Centre for Marine Research Thalassocosmos, Former US base at Gournes, Gournes Pediados, 71500 Heraklion, Crete, GREECE
Tel : (30) 2810 337852
Mail: kotoulas@imbc.gr

CR14: UNIPD
Responsible scientist : Luca BARGELLONI
Department of public health, Comparative Pathology and Veterinary Hygiene, University of Padova, Viale dell'Università 16, Agripolis, 35020 Legnaro (PD), ITALY
Tel : (39)-049-8272506
Mail: luca.bargelloni@unipd.it

CR15: ROSLIN
Responsible scientist : Chris HALEY
Roslin Institute , Genetics and Genomics, Roslin, Midlothian, EH25 9PS, UK
Tel : (44)1315274432
Mail: chris.haley@bbsrc.ac.uk

CR16: CNRS
Responsible scientist : Francis GALIBERT
Faculté de médecine, Université de Rennes1, UMR-CNRS Génétique et Dévelopement, 35000 Rennes, FRANCE
Tel : (33)2 23 23 44 86
Mail: francis.galibert@univ-rennes1.fr

CR17: Nireus
Responsible scientist : Costas BATARGIAS
Molecular Population & Quantitative Genetics, Koumanioti 17, 26222 Patras, GREECE
Tel : (30) 2610 334626
Mail: batc@imbc.gr

Summary

The overall aims of the project is to identify, in sea bream, sea bass, oyster, and rainbow trout, genes of which expression is associated with disease and stress resistance and to develop genetic approaches that allow characterisation of genetic markers for marker-assisted selective breeding of disease and/or stress resistant individuals.

For such project, the following main objectives have to br reached:

- Characterisation of stress- and disease-responsive genes as potential candidate gene markers for desirable traits.
- Identification of new genetic markers (SNP); mapping new markers (microsatellites and SNP) in linkage and gene maps.
- Seeking associations between (i) variations in response to stress and resistance to pathogen and (ii) selected candidates genes and microsatellites markers by segregation analysis in appropriate families (QTL analysis);

In order to characterize disease and stress-responsive genes in seabream, sea bass, trout and oyster, a functional genomic approach using microarray technology have been developed. These studies allowed us to analyze gene expression profiles in relevant tissues exposed to stress situation (confinement stress and pathogen for fish species, summer mortality, hypoxia and pathogen for oyster). For each experiment, genes of which expression is significantly modified have been identified and clustered in order to identify potential biological functions modified by stress exposure.

This project also bring new information related to genetic markers and their mapping: This included identification of Single Nucleotide Polymorphism (SNP) in trout, oyster and sea bass but also characterization of new microsatellite markers in sea bass and sea bream. A RH panel has been constructed which bring in that species an very interesting new tool for mapping genetic markers. Based on the different genetic map available in sea bass and sea bream, it was possible to develop a high resolution mapping of genetic markers, a set of information which become very important for QTL identification in these species.

Finally, QTL studies have been carried out in the 4 species: Associated traites were stress responsiveness (confinement) and disease resistance in sea bream, sea bass and trout, summer mortality resistance in oyster. Relevant protocols for QTL identification were developped and analysis of phenotyping and genotyping data oallowed identification of several new QTLs in trout (confinement stress, salinity stress, pathogens), oyster (mrotality) and sea bream (aptogen exposure). Finally, the project presents an outine of the advantages/disadvantages of operational genetic protocols incorporating identified QTL and traditional breeding approaches.

General project objectives

Under intensive aquaculture conditions, fish are exposed to various stressors which are unavoidable components of the fish aquaculture environment. These stressful conditions (confinement, pathogen exposure, temperature...) lead to an overall reduction in performance, including poor acclimatizing and growth performance, impaired reproduction and increased susceptibility to disease. Moreover, to address these problems, growers often resort to an increased use of antibiotics and drugs. In shellfishery, significant mortality have been reported to occur during the summer months among both juveniles and adult Pacific oyster in several countries (Cheney et al., 2000) and are a major concern of oyster farmers (Goulletquer et al., 1998). Although stressful conditions have also been linked to mortality among oysters, the causes underlying this phenomenon are complex and physiological, environmental and pathological causes have been suggested.

Modification of the stress and disease responses by selective breeding has long been considered as feasible in vertebrates of economic significance. For oyster, which presents a large genetic diversity, selective breeding has led to the production of lines that are divergent for high or low survival during the summer period. With respect to aquacultured fish, several studies published in recent years clearly indicated the feasibility of increasing the tolerance of fish to stress by selective breeding. This may lead to a reduction in the adverse effects of stress, i.e. improved efficiency of the food conversion or reduced incidence of disease. However, while a recent EU-funded project has clearly demonstrated that it is possible to manipulate stress responsiveness in rainbow trout it also highlighted the practical limits to developing such a familial selection protocol in the long term. Similarly, in oysters, although the multidisciplinary program “Morest” (2001-2005), established in France to understand the causes of the summer mortality in *Crassostrea gigas* juveniles, allowed the observation of a very high heritability of survival, but still, long term selection protocols appear difficult to establish for such destructive character.

The overall aim of the project is to identify, in sea bream, sea bass, oyster, and rainbow trout, genes of which expression is associated with disease and stress resistance and, from this information, to develop genetic approaches that allow characterisation of genetic markers for marker-assisted selective breeding of disease and/or stress resistant individuals.

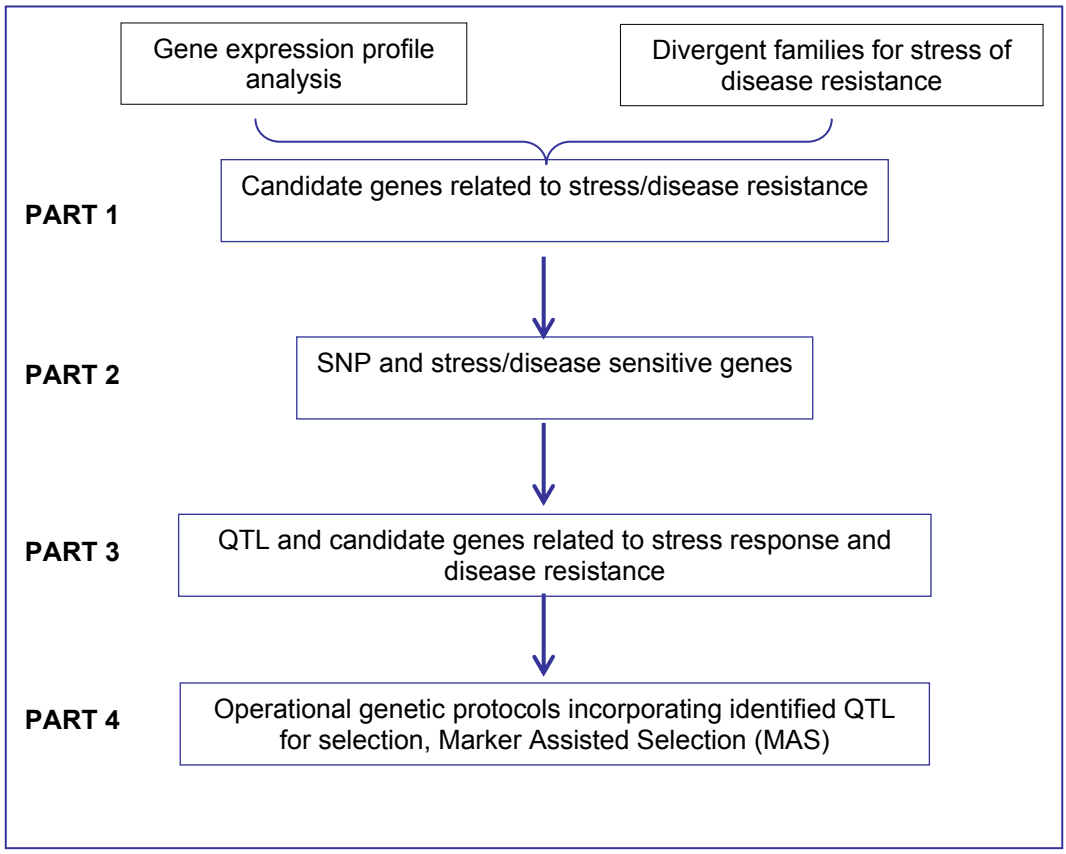
The originality of this approach relies on the successful association of
(i) functional genomic approach, i.e. development of microarray technology, to characterise disease and stress-responsive genes in the four species (trout, oyster, seabream, sea bass),
(ii) improvement of the genetic map for seabream and seabass associating information from microsatellite and SNP linkage map and radiation hybrid map.
(iii) evaluation of quantitative trait loci which have an effect on disease and/or stress resistance in the four species.
This requires mobilisation of a multidisciplinary research team gathering geneticists, physiologists and immunologists and a relevant coordination between the different research groups. The scientific strategy which have been employed in this project will involve four complementary parts which have developed over three and half years:

Part 1: Identification and characterisation of genes involved in functional responses to stress or pathogen exposure.

Part 2: Characterisation of Single Nucleotide Polymorphism (SNP) in the candidate genes previously identified.

Part 3: Using a QTL analysis, identification of association between stress or disease resistance traits and these candidate genes and also linked microsatellites markers.

Part 4: Outline operational genetic protocols incorporating identified QTL and traditional breeding approaches in oyster, sea bream, sea bass and trout.



This strategy, which aims to associate functional genomic data and QTL analysis in order to characterise genetic markers, is developed in trout and oyster for which a QTL mating design have been based on divergent families selected for stress or pathogen resistance. In sea bream and sea bass, a similar strategy is carried out and QTL analysis will carried out on fish taken from broodstocks. For genotyping analysis in these 2 species, the project takes advantage of the tools developed in flagship projects funded by EC “BRIDGEMAP” and “BASSMAP” addressing sea bass and sea bream. Based on low to medium density linkage and gene maps already obtained or under construction in other project (Marine Genomic Europe NoE), the project aimed to increase the density of sea bream and sea bass genetic maps. EST characterized by functional genomic analysis as being involved in stress or disease responses also go onto the linkage and gene maps. With this approach and radiation hybrid technology in these species, we have the most dense gene maps of any aquaculture species and exceptional tools for QTL approaches and marker assisted selection (MAS).

Sucessful accomplishment of these approaches requires establishment of genomic tools which were more or less present at the start of the project. Whereas a large EST collection and microarrays was already available for trout owing to the french AGENAE genomic program and to the european STRESSGENES project, development of such tools in oyster, sea bass and sea bream were more limited. A significant part of the efforts of the partners are then related to the development of these functional genomic tools and also to the training of partners to the use of these tools and relevant analysis of the transcriptome data.

Finally, these studies are carried out on stress/pathogen situations related to aquaculture problems. Thus, in fish, we propose to work on acute confinement stress or exposure to pathogens specific of each fish species. In oyster, a major problem encountered in production plans is summer mortality which is a complex phenomenon related to possible environmental, physiological (gonad maturation) or pathological causes. For these stress situations, several biological responses (traits) will be followed (for some of them, segregation between different families has already been observed and heritability measured). Finally, development of these physiological and genetic studies on similar stress situations in four marine species will favour the possibility of a comparative approach.

In order to facilitate overall understanding of the project, the presentation of the work carried out will be organized in 4 parts, including:

1. Development of functional genomic resources for sea bass, sea bream, oyster and trout.
2. Transcriptome analysis in trout, sea bass, sea bream and oyster exposed to stress conditions.
3. Single Nucleotide Polymorphism (SNP) characterization in trout, oyster and sea bass.
4. High resolution mapping of the sea bass and sea bream genome.
5. Charcaterization of QTL related to stress and disease responses in seabream, sea bass, oyster and trout.

Part 1: Development of functional genomic resources for sea bass, sea bream, oyster and trout.

In order to characterise disease and stress-responsive genes in sea bream, sea bass, trout and oyster, a common strategy using microarray technology has been developed. For each of these 4 species, the same strategy has been engaged: (i) construction of relevant EST collections using SSH cDNA libraries (ii) these EST are further spotted on glass microarrays (iii) and finally, using these microarrays, gene expression profiles can be carried out in various tissues of animals exposed or not to stress or disease.

Construction of relevant stress specific EST collections

Collection of the biological material in the four species (for more details, see annex1).

Detailed description of the work carried out in the four species exposed to stressful conditions is presented bellow in the annex 1. This can be summarized as follow:

In oyster, Families with high (R) and low (S) survival to mortality which show a high heritability for this character have been exposed in the natural environment to summer conditions and tissues were collected at several dates near the mortality outbreak. Mortality event occurred in June as planned and was higher for S than for R oysters. RNA were extracted from various tissues and distributed to partners in charge of constructing SSH libraries.

In trout, (i) In vitro, head-kidney leukocytes will be exposed to activators of specific leucocyte cell types (cytokines) suck as IL-15 (a known T-cells activator) and IL-1β/TNF for stimulation of isolated leukocytes (Partner 2) and (ii) in vivo virus exposure to VHS sensitive and resistant rainbow trout taken from selected lines have been performed. The subtraction procedure will be carried out on cells and various tissues (head-kidney, gills, liver) between exposed/non-exposed cells and between resistant/sensitive fish.

In seabream, in vivo exposure to confinement stress or pathogen have been carried out and physiological parameters characterizing exposure to such situations have been analyzed. For pathogen exposure, fish were exposed to the myxosporean parasite *Enteromyxum leei* by contact with infective effluent.

In sea bass, in vivo experiments have been also performed: Fish were exposed or to confinement stress or to pathogen (Fish were exposed to sub-lethal dosages of *Pasterella piscicida* by intraperitoneal injection). Biochemical analysis of relevant physiological parameters confirms that the experiment was carried out properly.

Construction of SSH cDNA libraries in the four species (for details see annex 1).

EST collections were constructed following two strategies:

- For seabream, sea bass and oyster, we have been developing stress and disease-selective cDNA libraries using a suppressive subtraction hybridization protocol (SSH libraries). This led to obtaining of enriched collection of differentially expressed cDNA. For each species, cDNA SSH libraries were constructed using two tissues and by subtracting control and stress samples. In some case, this subtraction was carried out at two time points, early and late.
- For trout, we have been using the EST collections developed during the Stressgenes project and which was also obtained from several SSH cDNA libraries following the same strategy. In order to enrich this trout EST collection, further SSH libraries have been constructed within Aquafirst project: This include SSH cDNA libraries constructed with leucocytes exposed or not to cytokines in vitro and SSH libraries constructed from tissues collected on VHS sensitive or resistant trout exposed or not to the virus.

This strategy leads us to obtain a rich EST collections in seabream, sea bass, trout and oyster. In trout, this EST collection was added to the one prepared during the Stressgene project. Thus, during this project, cDNA libraries from the following tissues have been prepared:

- Trout: leucocytes (cells exposed to cytokines), spleen (fish exposed to VHSV).

- Oyster: digestive glands, mantle, gonads, hemocyte, muscle, gills (oyster exposed to summer mortality).
- Sea bass: head-kidney, brain (fish exposed to confinement stress), gill, brain (fish exposed to pathogens)
- Seabream: liver, head-kidney, brain, gill (fish exposed to confinement stress), head-kidney, intestin (fish exposed to pathogens).

Sequencing of the EST collections and selection of EST for microarrays.

All the sequencing work has been carried out by the Max-planck Institute in Berlin. Sequences were further sent to the Aquafirst Bioinformatic Centre (SIGENAE team, INRA Toulouse, france) for quality control and selection of non-redundant clones and identification of contigs. Thus, this work allowed us to obtain the following results:

- In trout: 3298 new EST sequences coming from the SSH cDNA libraries constructed from VHSV exposure trout and from haed kidney trout leucocyte stimulated by cytokine..
- In oyster: 5533 new EST sequences.
- In sea bass: 6403 new EST sequences.
- In seabream: 8756 new sequences.

These sequences have been added to EST sequences previoulsy obtained in other EU-funded projects, Stressgenes or Marine genomic Europe NoE (MGE): Thus, in trout, the new sequences generated within Aquafirst (mainly from cDNA SSH libraries) were fused to the Stressgenes EST collections. In sea bass, seabream and oyster, the Aquafirst sequences were fused to EST sequences produced by MGE NoE.
Bioinformatic analysis of these sequences in the 4 species allowed SIGENAE team to identify a list of contigs which could be spotted on the microarray. Thus, for each species, we obtained:

Trout: 3678 sequences representing the same number of contigs. From these sequences, design and synthesis of 3678 oligos has been carried out by a private company (Eurogentec). 63% of these sequences have a homology with at least one protein. The sequences chosen to be spotted were selected from different sources:

- All contigs represented on the microarray constructed during the Stressgenes project were selected after elimination of the redundancy for the annotated ones.
- Sequences obtained from SSH libraries constructed by Partner 1 (pathogen exposure) were selected after elimination of the redundancy with the Stressgenes collection.
- The same work was done with the sequences selected from SSH libraries constructed by Partner 2.
- A hybridization was performed to select genes differentially expressed between stressed and control fish from the AGENAE generic macroarray.
- A list of VHSV-induced genes selected from infected/non-infected genes was given by the VIM INRA team (Jouy en Josas) and added to the selection.
- Interesting genes selected from literature.

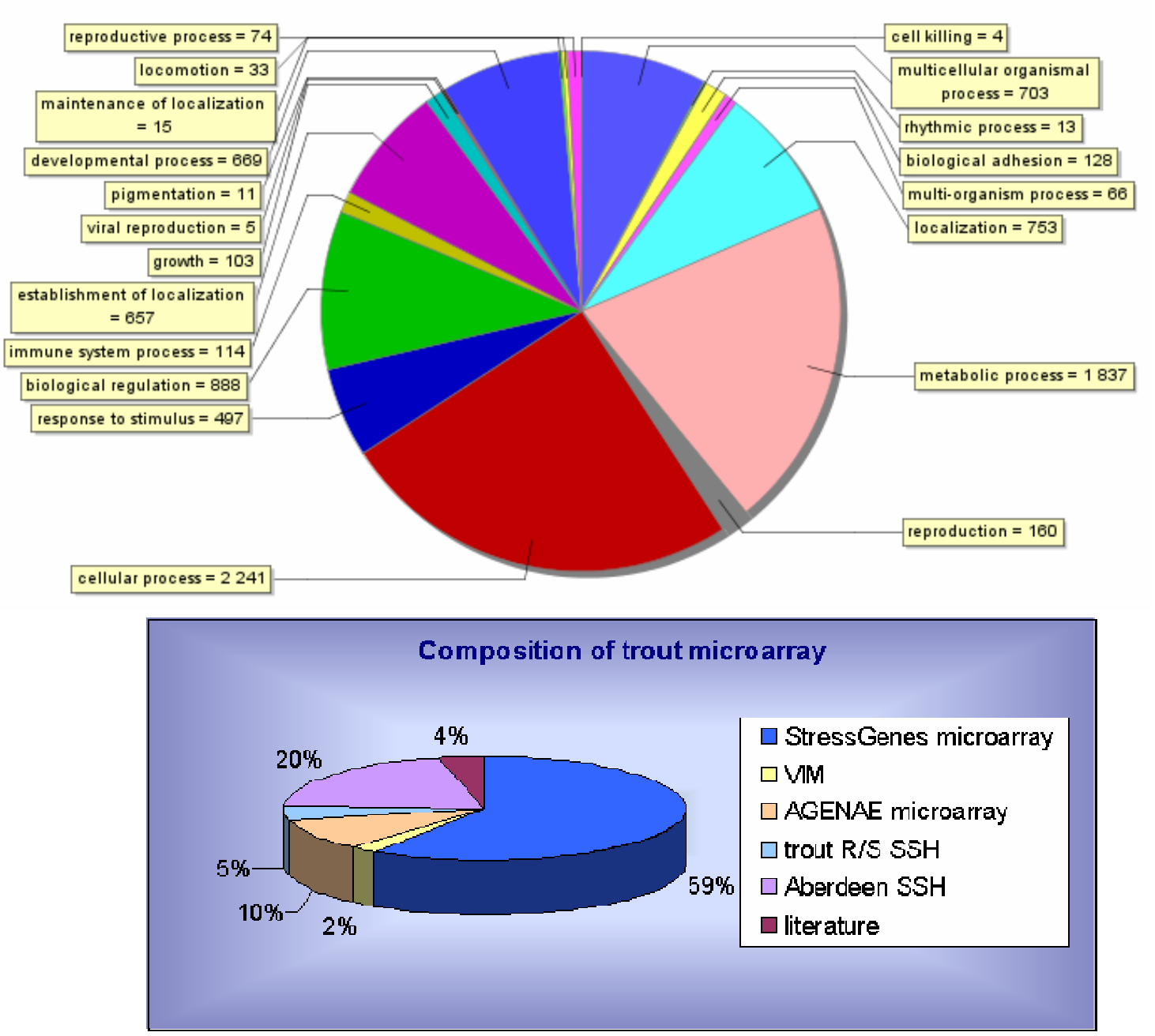


Figure 1. Gene ontology assignement (3rd level GO terms) of the x annotated ESTs for Biological Process GO categories. Each pie parts represents the distribution of the GO annotations into the major functional categories for each division. The value in each box represent the number of sequences associated with the gene ontology term.

Sea bass: 17,600 EST sequences have been identified.

Seabream: 19,000 EST sequences have been identified.

Oysters: 9,200 EST sesquences have been identified. In order to make an assessment for the putative identities of the sequences, all spotted ESTs were subjected to BlastX similarity searches on several databases. Of the analysed sequences, 46% had significant matches ($E<10^{-6}$) to the non redundant protein databases. More detailed functional annotation was performed on the annotated ESTs with the Gene Ontology (GO) Consortium structure. GO categories were assigned to 43% unique ESTs with BlastX using the Blast2GO software (see Annex 1). These analyses will be useful to determine if the metabolic pathways associated with the selected genes are over-represented in comparison to those associated to the whole spotted ESTs .

It is important to notice that this Aquafirst EST resource has been added to 2 others new EST collections: 1) a 47,889 EST collection generated by the Network of Excellence “Marine Genomics Europe” (MGE) 2) a 14,472 EST collection produced by the the french Genoscope project “EST sequencing from *Crassostrea gigas*. To maximize the utility of these collections, these ESTs together with those previously available from the MGE program and from public databases have been assembled in a publicly unique database wich include 26,724 sequences: the GigasDatabase (http://public-contigbrowser.sigenae.org:9090/Crassostrea_gigas/) containing 26,724 unique sequences.

Conclusion.

This technical part, which is crucial for the development of the functional genomic studies planned in this project, led to very significant progress in the construction of new genomic tools for aquaculture species:

1) Large collection of EST for oyster, seabream and sea bass which has not been otbtained before by any group. This was made possible by an efficient collaboration between MGE NoE and Aquafirst. It is important to notice that our colleagues from MGE fairly accepted to delay some fo their work and wait for Aquafirst EST production. This collaborative work allowed production of unique genomic tools.
2) For oyster, it is important to notice that EST collections generated by Aquafirst has been adequately added to the efforts carried out by MGE and Genoscope project and now have generated one of the most developped genomic database accessible among Lophotrochozoans.
3) In trout, we have been producing a unique collection of oligos corresponding to stress and disease-related genes which is very complementary to the AGENAE trout microarray produced by INRA. These EST collections have been added to the AGENAE trout EST database.
4) We have been developping an efficient functional genomic plate-form which rely on 2 major partners: The Max-Planck Institute in Berlin (for sequencing, clones management and amplification, spotting, database management) and SIGENAE team in Toulouse for all the bioinformatic work related to Aquafirst. It would be interesting to consider how this network could be maintained after the end of Aquafirst project. This question could be tackled within Aquagenome project of which one objective is to suggest new strategies for development of durable resources for fish and shellfish genomic.
5) These commun genomic resources have to be accessible to european research community: All these EST sequences are now under publication in public database. All oyster sequences have already been published. Publication of the trout, sea bass and sea bream sequences produced by Aquafrist is in progress and should be released before the end of 2008.

Development of the microarray technology within the project.

The main objectives of this work are (i) to accomplish construction of dedicated microarrays for each species (ii) to manage EST sequence data which have been previously generated (iii) to manage microarray data produced by the project.

1- Development of a microarray management structure

We have developped for Aquafirst project a microarray management structure based on 3 components:

- A microarray plateform devoted to the construction of the microarrays for the projects.
- A bioinformatic resource centre which will be helped by INRA/SIGENAE team (Toulouse, France).
- A coordination of this structure which is managed by partner 1 (INRA Rennes).

However, we have introduced several modifications in our functional plan of action:

- We have strengthen our collaboration with the Marine Genomic (MGE) NoE and particularly with his genomic resource center, the Max-Planck Institute –MPI- in Berlin (R. Reinhart): Initiated with the sequencing of EST produced by Aquafirst project, this collaboration lead us to also merge our EST sequence database for seabream, sea bass and oyster (trout was not involved in such process as MPI did not have any specific trout sequence information in their database). This merging leads to the construction of the largest oyster, seabream and sea bass databases presently available and which will be equally shared by the resource centers of the 2 European projects, MGE NoE and Aquafirst.
- Spotting of microarrays for Aquafirst: Trout oligos microarrays have been spotted by the West Genopole microarray plate-form in Nantes or in Rennes, according to availability. This was managed so that spotting could not be a confounding factor when analyzing the trout microarray data. For seabream, sea bass and oyster cDNA microarray, MGI in Berlin appeared to be fully organized for carrying out this task: Both MGE NoE and aquafirst agreed to construct for each species a unique type of cDNA microarray and MPI in Berlin accepted to carry out this task. Thus, there will be a unique place for spotting oyster, seabream and sea bass microarray.
- The cost fo these spotting have been consequently updated but still appeared reasonnable: For trout, 4 partners (Galway, Aberdeen, Uppsala and Rennes) bought the oligos collection (about 30,000 Euros) and have also been charged 30 Euros/slide for oligos spotting. For seabream, sea bass and oyster, slide have been charged 120 Euros which include cost of all steps fo the cDNA spotting technology, i.e. cherry-picking of the clones, PCR-amplification, filtration, spotting and quality control.
- Planning of the microaray hybridization and scanning have been organized: Some partners are keen to carry out this task in their lab (Galway, Aberdeen, Uppsala, Rennes for trout or seabream), other are interested to benefit the INRA microarray plate-form in Rennes (IFREMER/Plymouth for oyster, Canarie island for sea bass).

2- Training course for microarray analysis.

An important task developped within the porject is the development of training courses for microarray analysis. Partner1 (INRA/SCRIBE, Rennes) benefits from molecular and genomics facilities (IFR platform) and an INRA team dedicated to bioinformatics tools. This allowed him to develop in Rennes all experiences necessary for carrying out microarray analysis and to teach the interested partners the whole process, from microarray hybridization to analysis of the results.
One traning cession has been organized last November 2006 and has covered the following apsects:
Education for the use of BASE, a comprehensive web-based database for the management of data generated by microarrays.
Education for the use of TmeV, a verstatile microarray data analysis tool with algorithm for clustering, visualization, classification, statistical analysis and biological theme discovery.

Training to analysis of sequences, batches and libraires (SURF tool).
Training to aseembling processes and annotation (R-Icard tool).
Training to contig analysis and manipulation using the Sigenae contig browser (Biomart and Ensembl tools).
This session was organized during 3 days and allowed participants not only to have formal presentation of the tools but also to practise using microarray data and aquafirst database. 13 participants coming from 9 different laboratories were attending this cession (IFREMER Brest and Montpelliers, PLM, Galway, CSIC, CCMAR. Concerning the traning for technical pastects (labelling of the probes, hybridization and scanning of the slides...), this was organized with the partners who will carry out their hybridization in Rennes just before the start of their experiments.

3- Production of glass slide microarrays
As indicated above, aquafrist partners involved in the functional genomic studies decided to merge their genomic resources with that of MGE NoE, at least for seabream, sea bass and oyster. Moreover, we also decided to construct all arrays related to these 3 species in the MGE plate-form. This allow th eporject to produce the following glass slides:

- For trout: Based on the 3678 EST sequences previusoly identified, we have ordered 3678 oligos designed by the manufacture (Eurogentec). This design was validated by the company using the trout Agenae database (containing 250,000 EST sequences corresponding to ~40,000 contigs). These oligos were spotted on glass slides which were used for all transcriptome analysis of trout samples.
- For sea bream, sea bass and oyster: These cDNA glass slides will be produced in Berlin (MPI). All EST sequences obtained from MGE NoE or from Aquafirst have been merged by SIGENAE partner (Toulouse) and by this allowed selection of relevant clones to be spotted on the arrays. All arrays related to these species have been manufactured by the Max-Planck Institute in Berlin.

In conclusion, the Aquafirst project is able to produce trout oligos microarrays (corresponding to 3600 different contigs) and also cDNA microarrays for oyster (9000 different contigs), seabream (19,000 different contigs) and sea bass (17,600 different contigs). This was made possible owing to a large collaborative efforts between partners and the project was able to produce about 700 glass slides necessary for the transcriptome analysis developped in trout, sea bream, sea bass and oyster.
Beyond the tasks related to the Aquafirst project, it was also important to question about the sustainability of these genomic resources. Aquafirst project was a very good example of this problem owing to the importnace of the resources generated by the project and which could be abandonned at the end of the project. This question is also widely discussed within the on-going Aquagenome project (European Coordination Action FP6). Within Aquafirst, the next future of the genomic resources will be:
- For trout microarrays, it will be possible to manufacture on the West Genopole plate-form more slides containing the Aquafirst oligos in Rennes. Moreover, it is now possible to also buy new microarrays produced by Agilent (44,000 oligos).
- For oyster, sea bream and sea bass microarray, MPI in Berlin has decided to stop production of this type of glass slides. Presently, there will be the possibility to use sea bream slides produced by Agilent. Clearly, the next future for getting new microarrays in these species will be to buy them by private company where the quality might be better but 2 or 3 times more expensive compared to what we produced in Aquafirst.
- For the Aquafirst database which contains all EST information on the 4 species and considering the potential interest for the european scientist, the INRA-SIGENAE team (Toulouse, France) agrees to keep this database updated and alive during the 2 next years on his own means. However, such effort can only be maintained if other projects (European or national) accept to support this task and plan in their research budget the cost of such functioning. Considering the importance of such information for the genomic of the 4 major aquaculture species, it is important that such a solution could be found to make this resource really sustainable.

Part 2: Transcriptome analysis in trout, sea bass, sea bream and oyster exposed to stress conditions.

The major goal of this activity is to characterize stress and pathogen-responsive genes in the four species using analysis of their expression profiles. Those genes which will be differentially expressed will be relevant for our genetic analysis and according to our project strategy will be used for development of SNP. These studies have been carried out using teh microarrays developped in the previous part.
The main objective of the different gene profile analysis carried out on various tissues collected during these stress experiments were to provide list of stress-regulated genes as potential candidates involved in the trait resistance to stress or to pathogens. For trout and oyster, this information have been the starting material for characterizing new SNP markers.

Several in vivo stress experiments were carried out. This included:

i) **Trout:** Exposure of two selected lines of trout (high responsive –HR- and low responsive –LR- to stress) to confinement stress. These 2 lines were also exposed to pathogen (*Yersinia ruckeri*). Moreover, two clonal lines of trout expressing high sentivity or resistance to virus (HSV) were exposed to the pathogen (virus HSV). .

ii) **Oyster:** Selected resistant or sensitive lines (R or S lines) of oysters were exposed in situ to natural environemnt and tehse animals were studied before and during the summer mortality period.In addition, these 2 lines were exposed to hypoxia stress. Finally a bacterial challenge (a mixture of V. Splendidus and V. Easturianus) was done on the 2 lines (R and S).

iii) **Sea bass:** juvenile fish were exposed to confinement stress or to pathogen (bacteria *Pasterella piscicida*).

iv) **Seabream**: the same type of experiments (confinement or pathogen stress) were also carried out: The pathogen chosen for this species was a myxozoan parasite (*Entoromyxum leii*).

In all these experiments, physiological parameters relevant for characterizing the stress response in the animal were followed. In addition, gene expression profiles were analyzed in various tissues chosen as the most relevant for the stress factor. Thus, these analysis have been carried out in the following tissues:

Trout (HR and LR lines): Head-kidney (confinement, pathogen), gill, liver and brain (confinement), spleen (pathogen).
Trout (R and S clonal lines): fin (virus), spleen (virus).

Sea bass: Head kidney (confinement, pathogen) and liver (confinement).

Sea bream: Head kidney and intestin (pathogen), liver (confinement).

Oyster (R and S lines): gonads (field study), gill (field study and pathogen), muscle (field study), mantle (pathogen).

Overall, all these analysis led a very rich set of transcriptome data from which we have been producing short lists of candidate genes of which expressions were significantly modified after stress exposure. These genes are interesting candidates for further characterization of SNP markers which would be used for analysis of the QTL.
At this stage of progress of the transcriptome, we will present in the following pages more information on all these experiments and associated gene profile analysis. According to the species and stress application, the reader will find most of the important data. However, as indicated in the conclusion, we are fully aware of the richness of such ste of transcriptome data: An overall meta-analysis of these data (the same tissue between the 4 species or between tissues within the same species) has not already been carried out but will have to be developped in the future.

Rainbow trout: Genes expression analysis exposed to confinement stress or to pathogens.

a) Families selected for high- and low-responsiveness to acute confinement stress: Testing of F3 lines by standardised confinement stressor

Two strains of rainbow trout divergent for the plasma cortisol response to a standardized stressor (confinement) have been produced by selective breading (Pottinger and Carrik, 1999). These 2 lines are charcaterized as low responsive (LR) or high responsive (HR). An F4 generation of these strains have been generated and used for studies on the effects of confinement exposure on gene expression profiles.

Experimental protocol.

A large time-course study using HR and LR fish from the F3 generation have been carried out. LR and HR fish have been distributed between 11 1000 litre holding tanks (approx 100 fish/tank). After a 2 week period of acclimation, LR fish have been sampled at intervals following transfer to 50 litre confinement tanks, with undisturbed LR fish acting as controls. The start of HR time-course have been offset by 1 week because of the difficulty in running the time-course for both lines in parallel. At each time-point (0, 2, 4, 6, 8, 24, 48, 96, 168, 336, 504h) 10 control and 10 confined fish have been sampled, 5 each from replicate tanks, and samples of blood, liver, brain, pituitary, head kidney have been collected and processed immediately or frozen for subsequent analysis. Plasma cortisol and glucose levels have been measured to confirm that the expected differences between lines are evident.

For the present transcriptome analysis, we have been focusing on interrenal tissues where corticosteroid synthesis is localized and carried out a large scale gene expression profile analysis. Following quality analysis of interrenal RNA samples using the Agilent 2100 Bioanalyzer, the 90 remaining samples were labelled using the Chipshot Direct labelling Kit (Promega) with the Cy5 dye and the control samples were pooled and labelled with Cy3 to be sued as reference. Following hybridization, glass slides were scanned using the Axon Genepix 4000B scanner and signal was quantified with Genepix pro 5.1 software (Axon). Normalizations of the data were carried out with MADSCAN software available at the url:

http://cardioserve.nantes.inserm.fr/mad/madscan.

Physiological analysis:

The HR and LR populations were closely matched for size at the start of the study (LR weight: 181.3 ± 4.4 g; length: 24.7 ± 0.21 cm; HR weight: 184.7 ± 4.9 g ; length: 24.9 ± 0.2 cm). Only 1 maturing male was detected and no maturing females were found. Plasma cortisol levels in the confined and unconfined HR and LR fish are shown in Figure 1.

Statistical analysis of the plasma cortisol data (ANOVA, Genstat) with time, treatment (control or stressed) and line (HR or LR) as factors revealed that overall, there was a highly significant difference in plasma cortisol level during confinement between the HR and LR lines ($P<0.001$) and this was resolved on a time by time basis as shown in Fig. 2. The difference between lines was sustained (with the exception of the sample at 6h) to between 48 and 96h after which there was no significant difference evident, although the mean plasma cortisol levels in the LR fish were consistently higher than those in the HR fish. Significantly elevated plasma cortisol levels in both lines, relative to control unconfined fish, were evident until between 48 and 96h, after which there was no significant difference between confined and unconfined fish in either line. No significant difference in plasma cortisol level could be detected between the unconfined control fish of either line.

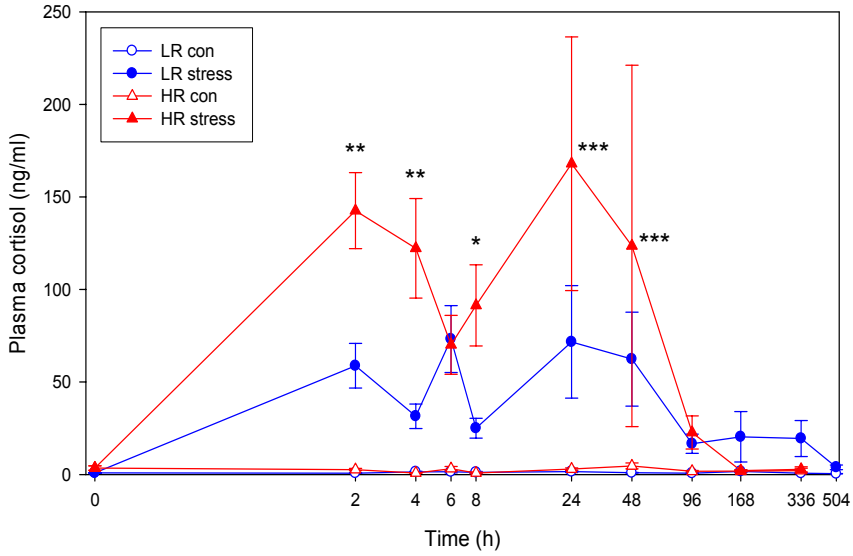


Figure 1. Plasma cortisol levels in 1+ HR and LR fish subject to confinement (stress) or sampled directly from undisturbed holding tanks (con). Each point is the mean ± SEM of 6 fish. Significant differences between HR and LR fish are denoted by * P<0.05; ** P<0.01; *** P<0.001.

Analysis of microarray data:

Selection of genes showing a significantly different expression have been carried out using the SAM statistical analysis. The strategy was to compare different conditions (HR/LR for control, HR/LR for stressed fish) at all time points. In case of missing values, we estimated it by using the K-nearest neighbours method. For this SAM analysis, the FDR was settled at 0 level in order to have in our list of genes the minimum false positive candidate.

This analysis allowed us to identify of list of 192 genes among which 90 are genes significantly different between HR and LR stressed fish, 59 between LR and HR control fish and 43 are common to these 2 comparaisons (see enclosed list of genes in annex 1). Among these 192 genes, 101 have an annotation whereas the rest of the genes appear as unknown.

When carrying out hierrachical clustering, we obtain 6 clear different clusters which included genes involved in a large variety of cellular and molecular functions (see list in annex 1). These clusters are now under more precise annotation and validation in order to investigate potential biological and molecular functions involved in the confinement stress responses.

This list of 192 genes have been given to partner INRA/Jouy for SNP analysis.

b) Characterisation of the effects of pathogen (VHSV virus) exposure on gene expression profiles in R and S trout (clonal lines).

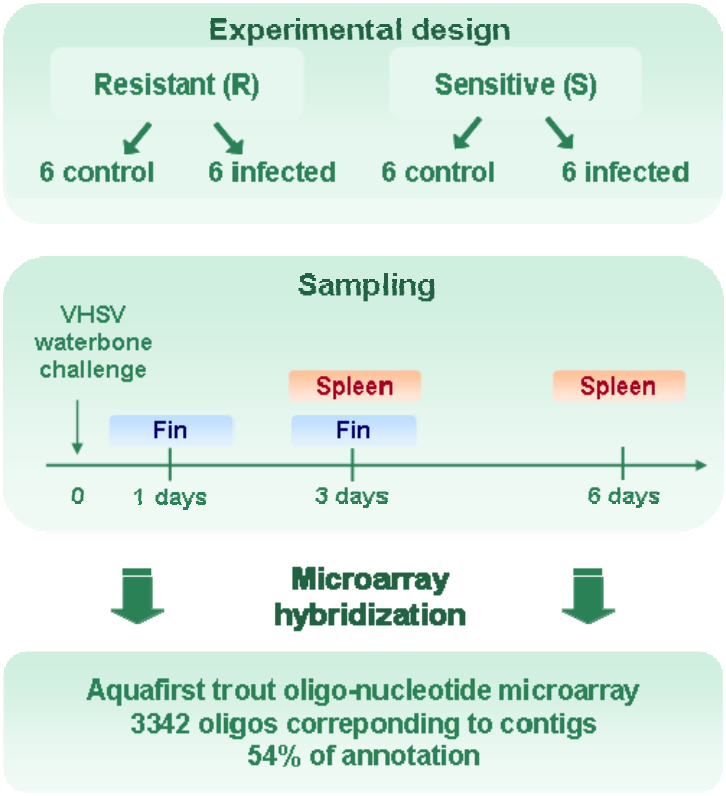
Two groups of trout selected for their resistance or susceptibility to waterborne VHSV challenge (R and S homozygous clones) were subjected to two replicated pathogen challenges. Gene expression analyses were performed on RNA samples prepared from spleen (a major site of immune response[1]) and fin (the site of viral entry in the fish[2]) obtained from the second VHSV challenge with R and S trout.

Experimental protocol:

The trout distributed according to their resistance (R) or susceptibility (S) to VHSV were challenged or not with this virus in 6 replicates per condition. Spleen and fin were sampled at 1, 3 and 6 days post challenge. For fin, gene expression analyses were performed at 1 and 3 days and for spleen at 3 and 6 days. In total 96 hybridizations were planned using the Aquafirst trout microarrays composed of 3342 oligonucleotides spotted in triplicates.

The quality of the 96 RNA samples was first checked using the Agilent 2100 Bioanalyzer. Five samples (3 for spleen and 2 for fin) were removed for the following steps because their quality was not sufficient. The 92 remaining RNA samples were labelled using the Chipshot Direct labelling kit (Promega) with the Cy5 dye and the control samples were pooled and labelled with Cy3 to be used as a reference. Each sample was then hybridized with a reference on the Aquafirst trout glass slides using the automatic hybridization platform (Ventana Discovery) available on the Rennes transcriptomic platform.

The slides were then scanned using the Axon Genepix 4000B scanner and the signal was quantified with GenePix Pro 5.1 software (Axon). The results files obtained were used for the normalization step with MADSCAN (**M**icro-**A**rray **D**ata **S**uite of **C**omputed **A**nalysis) available at the following url: http://cardioserve.nantes.inserm.fr/mad/madscan.



Experimental protocol

Statistical analysis:

SAM

The strategy chosen for gene expression analysis is a SAM comparison of the different conditions to select genes differentially expressed among the 2 lines and those differentially expressed in response to the pathogen challenge. When more than half of the values were missing for a condition, the gene was removed from the analysis. The other missing values were estimated using the K-nearest neighbours method. Eight SAM two-class analyses were so performed for each tissue with TMeV software (see Figure 1). The FDR was settled to 0 that is the lowest FDR in most of the cases.

In fin, SAM analysis identified a significant number of differentially expressed genes in S versus R clones. In contrast, very few genes were induced or repressed after pathogen exposure, whatever the resistant or susceptible status of fish (see Figure 1).

In spleen, similar results were observed as many genes were differentially expressed between R and S clones at the two time points studied. On the other hand, a significant number of genes differentially expressed in response to pathogen challenge was observed only in susceptible fish at day 6 (see Figure 1).

	SAM	Nb of significant genes	delta	median FDR (%)	nb of false positive (median)	Remark
FIN	RCont/SCont 1j	71	0,8017	0	0	
	RInf/SInf 1j	131	1,178	0	0	
	RCont/SCont 3j	92	0,646	0	0	
	RInf/SInf 3j	138	1,052	0	0	
	RCont/RInf 1j	2	0,435	0	0	following FDR > 13%
	SCont/SInf 1j	2	0,521	0	0	following FDR > 6%
	RCont/RInf 3j	1	0,233	0	0	following FDR > 33%
	SCont/SInf 3j	0	0,458	0	0	following FDR > 28%
SPLEEN	RCont/SCont 3j	70	1,037	0	0	
	RInf/SInf 3j	63	0,869	0	0	
	RCont/SCont 6j	159	1,249	0	0	
	RInf/SInf 6j	189	1,25	0	0	
	RCont/RInf 3j	0	0,004	0	0	following FDR > 99%g
	SCont/SInf 3j	4	0,518	0	0	following FDR = 20%
	RCont/RInf 6j	0	0,008	0	0	lowest FDR > 99%
	SCont/SInf 6j	207	1,109	0	0	

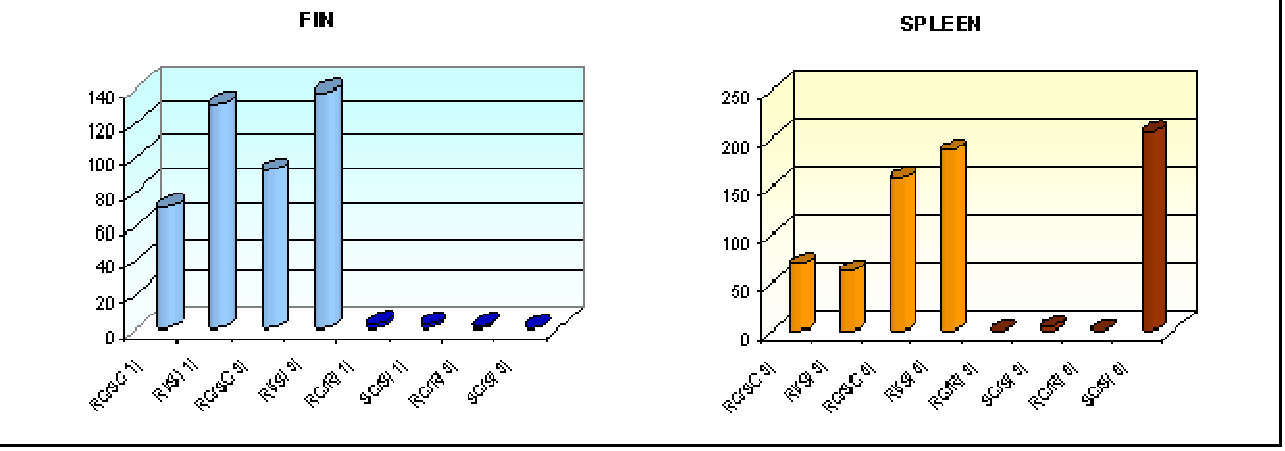


Figure 1: SAM analyses results

Hierarchical clustering

For each tissue, a hierarchical clustering was performed with the unique genes selected by SAM (269 for fin and 453 for spleen). For each condition, the median of the slides was used to have a better visibility of the clusters. This analysis was performed using the hierarchical clustering module of TMeV software. The genes were first median centred, then the hierarchical clustering was performed with the average linkage method and the Pearson uncentered metric.

In fin, 6 clusters were obtained of which 4 seemed interesting (2 of them had the same expression profile). The cluster 1.1 seems to group 143 genes overexpressed in S fish compared to R. The cluster 1.2 seems to group 8 genes overexpressed in S infected fish at day3 compared to the other groups. The cluster 2 corresponds to 112 genes underexperessed in S fish compared to R.

Results analysis

The main result observed in both fin and spleen is that R and S control fish showed pronounced differences of basic global expression profiles. Overall, 322 genes have been selected from the analysis in the 2 tissues for their responsiveness to pathogen exposure and transferred to INRA Génétique des Poissons (Jouy-en-Josas) for SNP analysis (see list in annex 1).

These differentially expressed genes may or may not participate to the susceptible status of the fish. Actually, some well-known genes involved in the antiviral responses are not strongly modulated by the infection but are constitutively expressed.

Some other results were rather unexpected. Indeed, the very low number of differentially expressed genes in fin of susceptible fish was surprising, since the fins are considered as the main site of the

virus entry and early replication. The interferon-responsive genes present on the array do not seem to be significantly induced at either day1 or day 3.
In spleen of R fish, exposure to VHSV did not modify significantly the gene expression profile, which was consistent with the very low level of virus production observed in these resistant animals[3]. On the other hand, a difference of response to infection was observed between R and S fish, mainly at day 6 where a lot of genes were selected in spleen and not in fin. Some of these genes were typical virus-induced genes confirming that this gene modulation was due to the infection (cf below the genes analysis). However it was somewhat surprising to observe so little modification of the spleen profile on day 3 post-infection.
T-test analyses were carried out with R software to check the absence of genes modulated by the infection in spleen and fin of S infected fish. The results of these complementary analyses confirmed those obtained with SAM.
In conclusion, this experiment allowed us to identify a list of genes responsive to VHVS exposure and then potentially interesting in the development of the trait 'resistance to VHSV exposure'. This list of genes (see annex 1) have been futher used used to identify new SNP markers.

c) Gene profile analysis in the brain of 2 selected trout lines (HR and LR) exposed to chronic confinement stress.

Rainbow trout selected for high (HR) or low (LR) cortisol responsiveness was exposed to standardized confinement stress. The brains were collected from fish exposed for 2h and 8h confinement stress together with brains from controls (0h). From each time-point five HR and five LR brains were used for extraction of RNA and the quality of the RNA was checked using an Agilent Bioanalyzer. The 30 RNA samples were labeled using the Chipshot Direct labeling kit (Promega). Also, RNA from all samples was pooled to be used as a reference in all the hybridizations. A dye swap approach was used, so a total of 60 hybridizations were made. The slides used for the hybridizations were the Aquafirst trout microarrays, which contained a set of 3342 oligonucleotides spotted in triplicates. The slides were scanned on an Axon Genepix 4000B scanner and quantified with GenePix Pro 5.1 software. The results were analysed in BASE (AGENAE Version) accessed from the Aquafirst web site. SAM was used for the statistical analysis of gene expression. For thoas analysis, we used FDR which is presented as a q value for each gene in the final list of significant genes. We choose a FDR value of 5% which is also what we choose for the 30 genes that were changing in the comparison of the HR/LR fish. For selcting confinement stress-responsive genes, 2 analysis were carried out, i.e. for each line (HR or LR) in LR comparaison between control and stressed fish.

The SAM analysis identified about 300 genes that were differentially expressed due to confinement stress and about half of them could be annotated (see Table1 and Table 2). When comparing HR with LR fish during confinement stress only 29 genes were found to be differentially expressed of which 15 could be annotated (see Tablle 3 in annex 1).
From these nalaysis, we have been Selecting genes that will be used for SNP analysis. This include the 10 most up or down regulated genes after confinement stress found in HR and LR trout (see figure 4 in annex 1) and those genes which showed significantly differently expressed genes in HR compared to LR (see figure 5 in annex 1).

d) Gene expression profiles in liver during to confinement stress in LR and HR selected trout lines.

Experimental design & microarray hybridisation:

In this study, we have been analyzing liver tissues collected in the same confinement stress experiment as the one previously described .For this study, two experimental designs were used - a loop design (Figure 1), where stress and control (pooled) RNAs from different time-points for both HR and LR families were compared directly, and a reference design where only stress HR and LR individuals at 168h were compared indirectly through a reference sample of pooled RNAs. Control samples (HR and LR) in the loop design were prepared from 3 individual fish and stress samples (HR and LR) were prepared from 5 individual fish. Only 6h, 24h and 168h time-points were considered for the loop analysis and on the basis of the analysed results the 168h time-point was chosen for the individual analysis. Both designs incorporate dye balance or dye swap. In total 41 hybridisations were performed (21 for the loop and 20 for the reference designs).

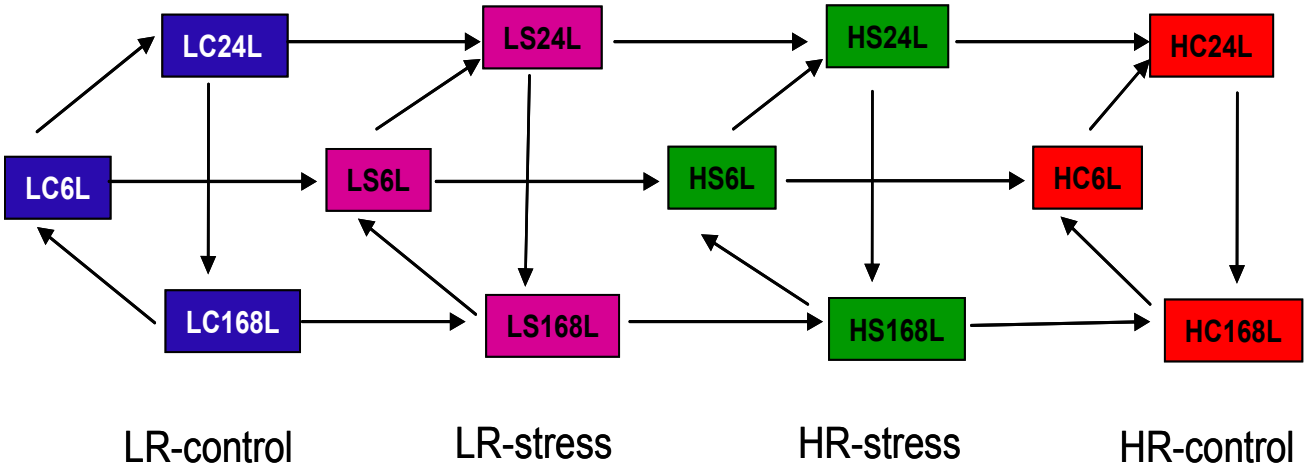


Figure 1: Loop design for the trout confinement time-course

Five micrograms of total RNA per individual fish or pool of fish (or reference) were directly labelled with Cy5 and Cy3 dyes using the Promega ChipShot Direct Labelling System. Hybridisations were performed in a Genetix hybidisation chamber as described in the protocol provided by Nantes. Washes used the Advawash automatic washing station. Slides were scanned as described above for the sea bream experiments.

Data acquisition & analysis

Two channel TIFF images were imported into the Genepix Pro 6 software program for feature (spot) finding and alignment using a batch alignment process. Features were flagged as present, absent or bad by this software program and pixel intensities for feature and background were quantified. Output GPR files from the loop design were imported into LIMMA (an R package). All resulting raw and normalised data is in process of submission to SIGENAE version of BioArray Software Environment (BASE) database. A list of genes differentially expressed between HR-stress and LR-stress (and any time) was determined (using various functions within LIMMA) and displayed with the TopTable function. LIMMA cannot average triplicates when, as in this array, the triplicates do not have a consistent displacement, therefore all spots are given in the TopTables. This analysis identified differences between HR and LR fish which might be due to individual differences between the original fish that spawned the HR and LR families. In further analysis, HR-stress and LR-stress fish were compared to the respective HR- and LR- control fish at all three time-points (6h, 24h and 168h). These analyses also suggested that the time-point where most changes were occurring was at 168h. Individual fish from this time-point were hybridised in the reference design and analysed in Genespring GX 7.3. Data was normalised for signal, dye swap and intensity (intensity dependent lowess normalisation) and log 10 ratios of signal (experimental) to control (reference) channel were calculated. A t-test identified those genes differentially expressed (P<0.05) between HR and LR fish (under stress at 168h). The data was further analysed using filters for fold change and by k-means clustering (by fish).

Real-time PCR confirmation:

A selection of genes potentially differentially expressed between the HR/LR families in control and stress conditions were chosen for real-time PCR confirmation. PCR primer details are provided in Table 1. Real-time PCR confirmation was carried out using an MX3000P Real-time PCR system. Reverse transcriptions of 2.5 µg aliquots of total RNA from each individual were performed and cDNA diluted to 10 ng/µl. Quantitative real-time PCR (qPCR) assays were performed using the Quantitech SYBR Green PCR kit. The threshold value (Ct) was fixed arbitrarily for all genes and expression was normalised to the housekeeping genes, □-actin and 18S RNA. Data was analysed using the Pfaffl method (Pfaffl,M.W. (2001) Nucleic Acids Res. 29, e45)

Results:

Trout confinement stress: A panel of 1015 spots (735 different genes are represented) were identified as being differentially expressed ($P < 0.05$) over the time course of the confinement (Appendix: List D) between HR-stress and LR-stress. However, only 304 of these genes were differentially expressed ($P < 0.05$) between HR-stress and HR-control (at 168h) or between LR-stress and LR-control (at 168h), i.e. their apparent differential expression was linked to the family (HR or LR) rather than to the stress.

A panel of 999 spots (660 genes are represented) were identified by LIMMA as being differentially expressed ($P < 0.05$) between **HR-stress** and **HR-control** (at 168h). Similarly a panel of 977 spots (648 genes are represented) were identified by LIMMA as being differentially expressed ($P < 0.05$) between **LR-stress** and **LR-control** (at 168h). These latter two lists (stress-related) have 186 genes in common, however, unique to the LR family are 462 genes and unique to the HR family are 474 genes. In Genespring these unique lists were analysed between the HR and LR 168h stress individuals: 98 of the 474 genes were significant in the HR group (filtered on 2-fold and only annotated genes illustrated in Table 2, annexe 1) and 34 of the 462 genes were significant in the LR group (annotated genes illustrated in Table 3, annexe 1). Therefore, for example, **this LR group represents genes which are stress-related in the LR family but behave significantly differently in the HR family**. This list was clustered by animal by k-means and illustrated by Gene Tree in figure 2 – this gives some idea of the variation between individual fish.

Real time PCR confirmation was performed for 4 candidate genes (Table 4, annexe 1). Most expression profiles determined from the microarray analysis were confirmed by real time RT-PCR. GRP 78 was a stress-related gene common to both LR and HR families. H-2 class II histocompatibility antigen (gamma chain) and ATP-binding cassette (sub-family G member 2) were not stress-related genes but differed between the HR and LR families. The unknown EST CR944591 was a stress-related gene associated with the HR family: individual PCR analysis for this EST is shown in figure 3.

e) Characterisation of the effects of pathogen exposure on gene expression profile in HR/LR trout.

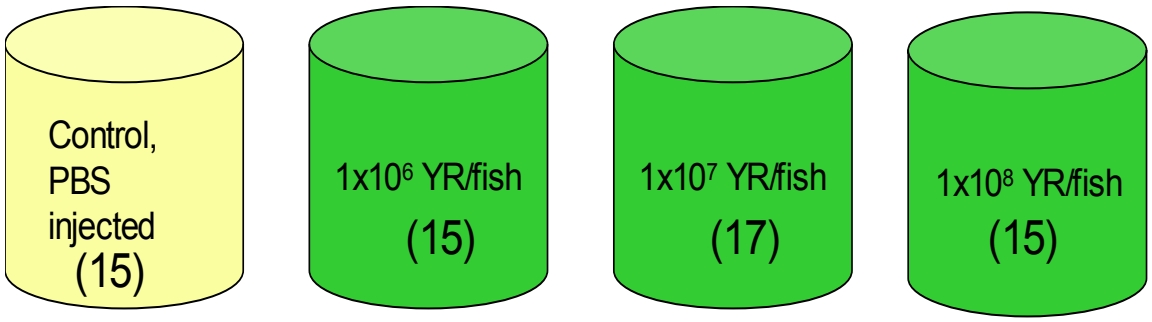
***Yersinia ruckeri* trial:**

HR/LR trout have previously been shown to have differential disease resistance to viral hemorrhagic septicemic virus (VHSV) and the gram negative fish pathogen *Yersinia ruckeri*, the causative agent of enteric red-mouth disease (ERM). The challenge experiment carried out at Aberdeen used *Y. ruckeri*. To prepare for the initial HR/LR expression profiling experiment and phenotyping of disease resistance, a pre-experimental trial was conducted using un-selected trout. The purpose of this experiment was (1) to find out if the bacterial strain available is pathogenic and suitable for QTL phenotyping, (2) to determine the optimal dose for the challenge, (3) to determine the optimal time points for sampling, and (4) to look at marker gene expression in selected tissues.

Material and Methods

Preparation of bacteria: *Y. ruckeri* MT3072 was obtained from Dr. Ellis, Marine Laboratory, Fishery Research Services, Aberdeen, Scotland. The bacteria were grown in TSB medium at 22 overnight, and harvested by centrifuging at 5000 rpm for 15 min. The bacteria were resuspended in PBS containing 20% glycerol, aliquoted and stored at –80 . The bacteria were titrated by plate counting on TSA plates.

Challenge and sampling: Normal un-selected trout were distributed over 4 tanks in the challenge facility of the School of Biological Sciences aquarium, for 4 weeks before intra peritoneal (i.p.) injection with three doses of bacteria in phosphate buffered saline (PBS) or with PBS alone as controls (Figure 1). The fish were fed with commercial trout pellets over the experimental period and the water temperature was controlled at 14.5 . Fish were killed at day 1, 2 and 3 post-challenge and the body weight and spleen weight were measured. Head kidney, liver and spleen were then collected for RNA isolation.



Fish : Normal trout

Bacteria: *Yersinia ruckeri* in PBS

Sample: Spleen, liver, and head kidney

Measuring: Body and spleen weight,

Figure 1. *Yersinia ruckeri* challenge trial: 4 groups of fish (Number of Fish in bracket) were injected i.p. with *Y. ruckeri* or PBS, and sampled at day 1, 2 and 3 post-challenge**.**

Q-PCR analysis of gene expression: The expression of a set of genes known to be related to infection was investigated by Q-PCR using cDNA prepared from liver and head kidney.

Results:

The fish injected with higher doses of *Y. ruckeri* ($1x10^7$ bacteria/fish and $1x10^8$ bacteria/fish) started to show disease symptoms 36 h after injection, whilst the fish injected with the lowest dose ($1x10^6$ bacteria/fish) started to die 3 days after injection. The diseased fish showed typical symptoms of ERM. The spleen size was generally increased over time in individuals without symptoms (Figure 2). The fish succumbing to infection also showed an increase of spleen size, but the size increase was negatively correlated to the bacterial number used for injection (Figure 3).

Total RNA samples were prepared from head kidney and liver tissues and reverse-transcribed to cDNA. Gene expression of some marker genes shown to be modulated by *Aeromonas salmonicida*, another fish bacterial pathogen, were detected by Q-PCR. CD209, IL-1β, and MCSF were shown to be up-regulated by *Yersinia ruckeri* infection in head kidney. In the liver samples, COX-1 was up-regulated by infection, whilst COX-2 and IL-1β were down-regulated.

From this experimental trial, it was concluded that the strain of *Y. ruckeri* used (MT3072) is pathogenic and could be used for QTL phenotyping. A lower dose ($1x10^6$ bacteria/fish or lower) should be used for injection and day 2 could be chosen for phenotyping. *Y. ruckeri* MT3072 exposure modulates expression of a number of immune genes.

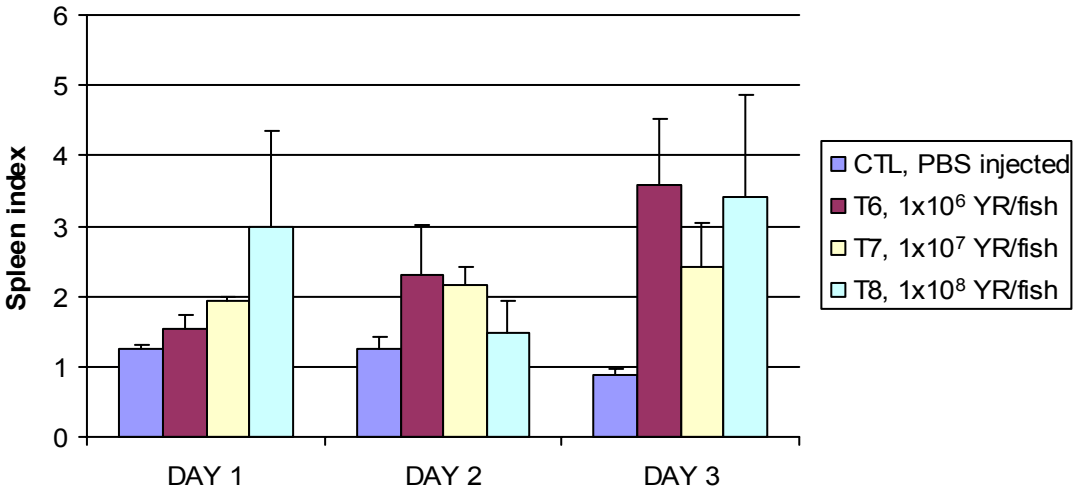


Figure 2. Spleen index of trout after *Yersinia ruckeri* infection. Spleen index = Weight of spleen/weight of body x 1000.

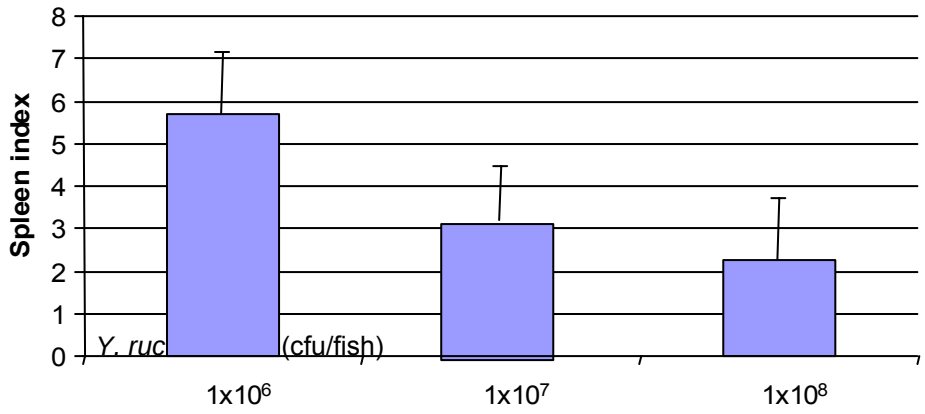


Figure 3. Spleen index of trout succumbing to *Yersinia ruckeri* infection. Spleen index = Weight of spleen/weight of body x 1000

HR/LR pathogen exposure on gene expression profile and phenotyping

HR/LR trout were transported to the Aberdeen facility on from Windmere. The fish were observed for one week for any sign of infection before being transferred to the challenge facility. The fish were acclimatised in the challenge room for 19 days before being challenged with $5x10^5$ cfu/fish of *Y. ruckeri*. Two sets of challenges, one for observation of mortality, and one for sampling of tissue samples for analysis, were carried out as described in Figure 4.

In the first challenge, the disease progression was recorded. The fish with disease symptoms were killed, and body and spleen weight measured.

In the second challenge, at 6 h, 1 day, 2 day and 3 day after injection, at least 6 fish from each group were sampled. Blood samples were collected for plasma glucose testing, and body and spleen weight were measured for each fish. Tissues (head kidney, liver and spleen) from selected fish were also collected in RNAlater for RNA preparation.

Spleen samples were collected ready for gene expression profiling using the trout oligo-microarray. Other tissue samples may also be used for Q-PCR analysis.

Plasma glucose levels were determined using a Glucose Assay Kit (GAHK-20, Sigma) with a variation in the protocol to allow for the small sample volumes available. Three microlitres of each test sample were pipetted, in duplicate, into the appropriate wells of a 96 well assay plate (Nunc) and 100 µL Glucose Assay Reagent (Sigma) added. The samples were incubated for 15 min at room temperature before being read at 340 nm on a plate spectrophotometer (Spectromax) against dH_2O. A standard curve was prepared for each assay plate using a standard glucose solution (1 mg/mL, Sigma) diluted in dH_2O to a final concentration of 0.083, 0.333, 0.5, 0.667, 0.833 and 1 mg/mL and a final sample volume of 3 µL tested as for the unknown samples. Sample blanks were prepared for each sample and standard by adding 3 µL of the appropriate sample/standard to 100µL dH_2O. A reagent blank was prepared, in duplicate, for each plate by substituting the 3 µL sample with 3 µL dH_2O. After the absorption reading for the reagent blank and the appropriate sample/standard blank had been made from each sample/standard a standard curve was constructed for each assay plate and a trend line fitted. The equation of the trend line was used to calculate the level of glucose in each of the samples on that plate.

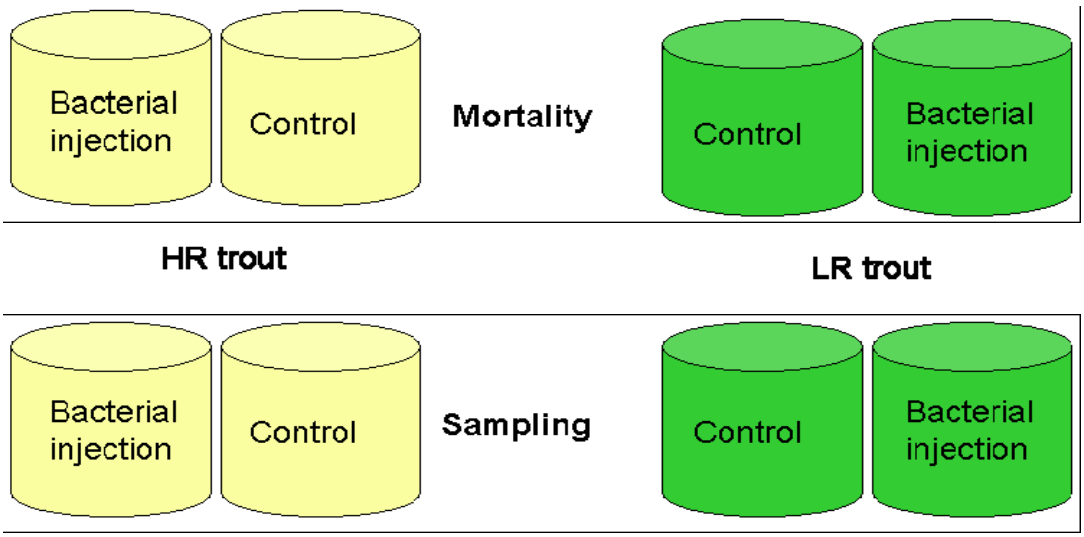


Figure 4. Challenge protocol for disease resistance and phenotyping, and microarray analysis.

Results

Mortality:
As shown in Figure 5, the HY fish were relatively more resistant to *Y. ruckeri* challenge than the LR fish. This difference might be affected by the challenge dose of bacteria used ($5x10^5$ CFU/fish) and the fish age. A large challenge experiments with more fish is needed to obtain a statistically reliable resistant difference to *Y. ruckeri* challenge between the two trout strains.

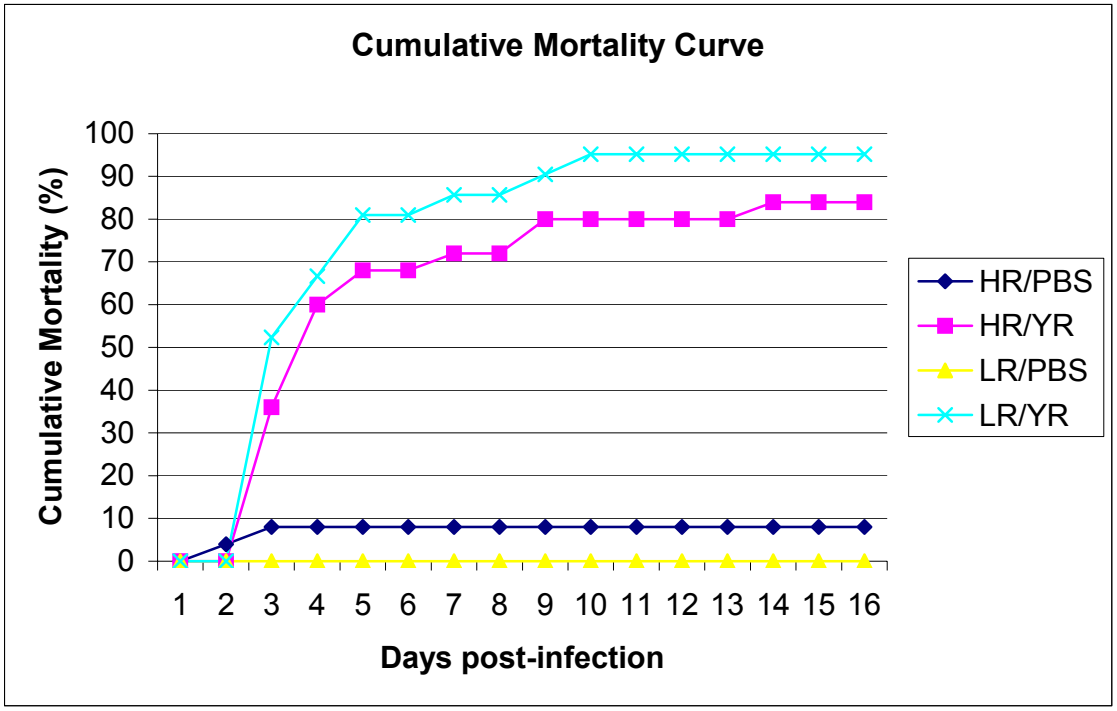


Figure 5. Cumulative mortality curve of HR/LR trout challenged with $5x10^5$ cfu /fish of *Y. ruckeri*. The fish size (g) +/- SE are: HR, 54.21+/-1.36, and LR, 55.74+/-1.67.

The spleen size:

The spleen size of both HR and LR trout was significantly increased at 48 hand 72 h post-challenge with *Y. ruckeri* (Figure 6). The spleen index of succumbing trout was also increased, but showed no significant difference between the two strains (Figure 7).

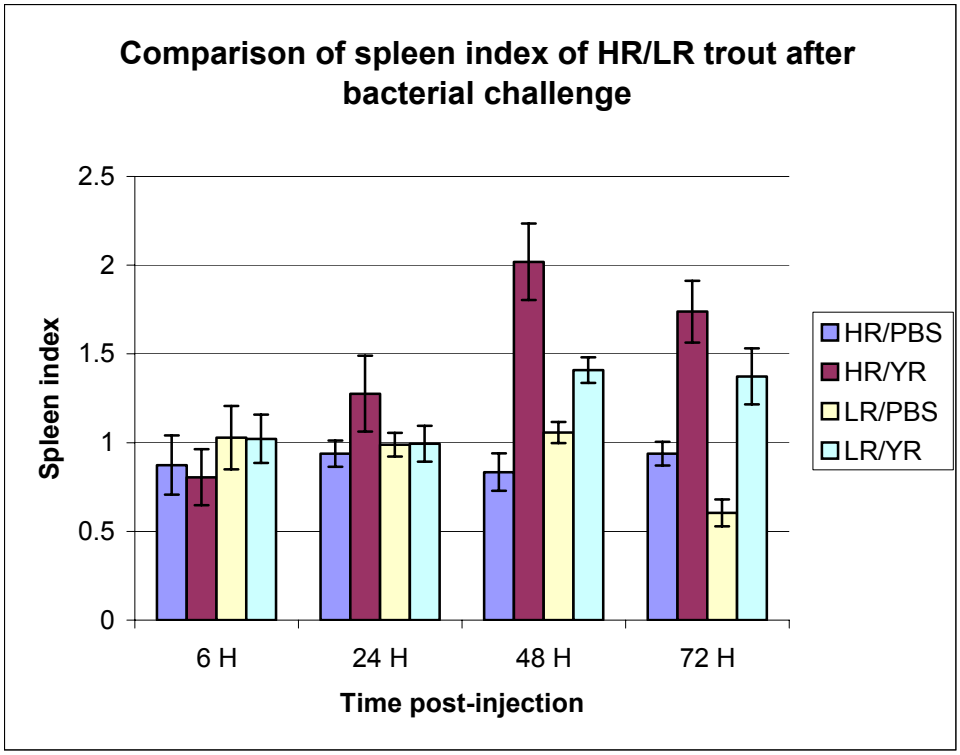


Figure 6. Spleen index of trout after *Y. ruckeri* infection, Spleen index = Weight of spleen/weight of body x 1000.

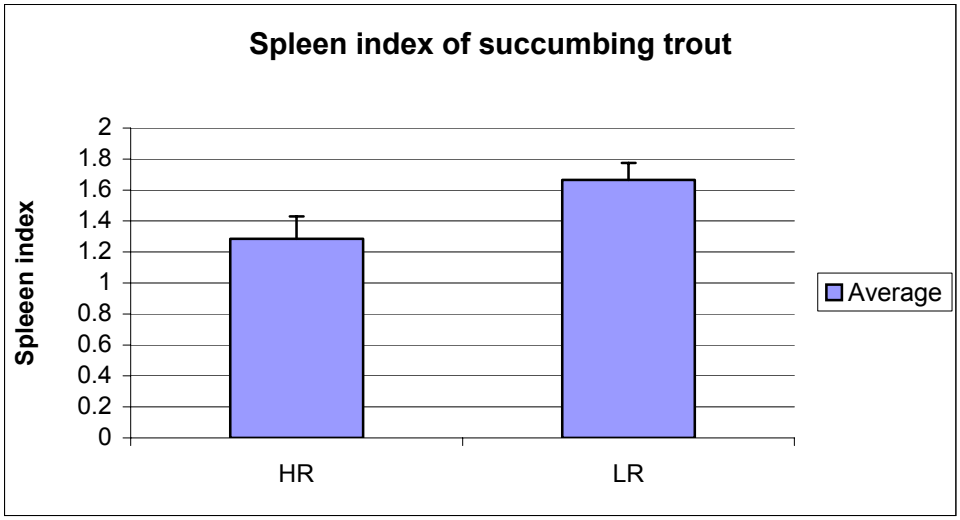


Figure 7. Spleen index of HR/LR trout succumbing to *Y. ruckeri* infection. Spleen index = Weight of spleen/weight of body x 1000

The plasma glucose concentration:

The HR trout in the PBS injected group maintained a higher plasma glucose concentration, with plasma glucose level decreased significantly at 48 h and 72 h after *Y. ruckeri* challenge. In contrast, the plasma glucose level increased at 48 h in the LR fish after challenge (Figure 8).

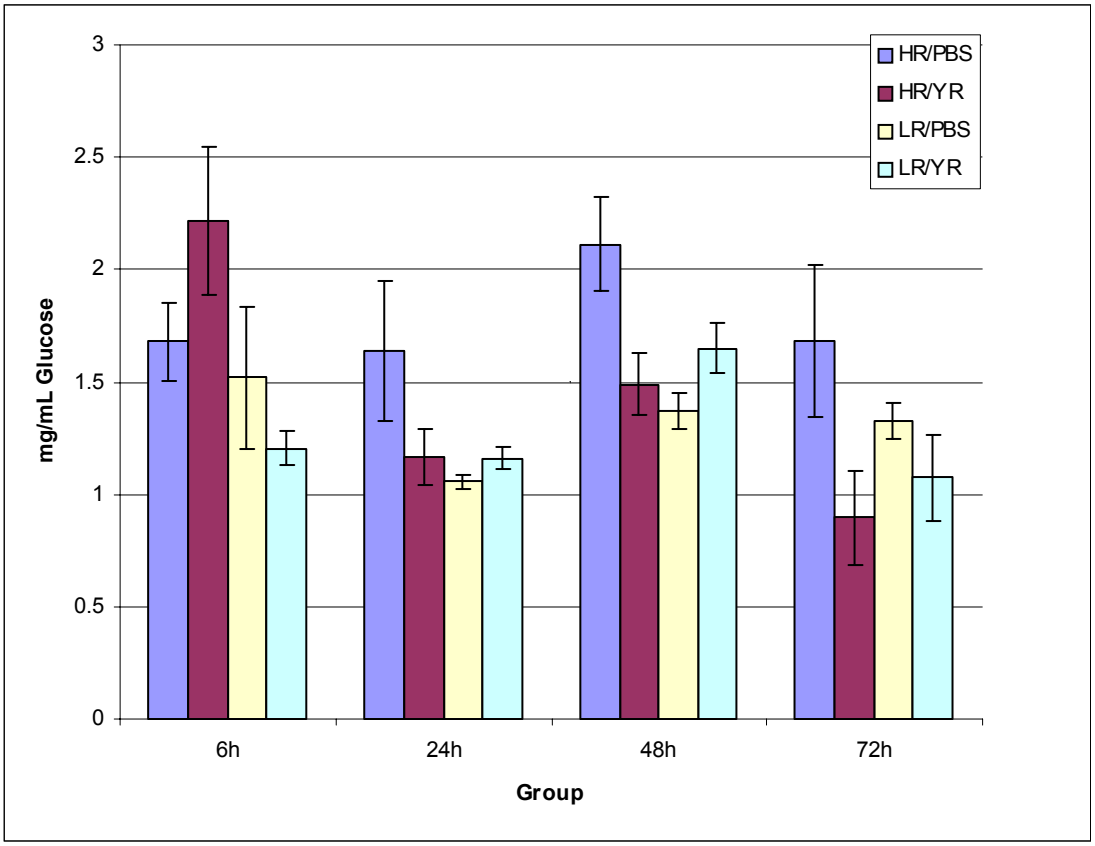


Figure 8. Plasma glucose of trout after *Y. ruckeri* infection.

Conclusion: Differential responses to *Y. ruckeri* challenge have been observed between the HR/LR strains. HR trout showed relatively higher resistance to *Y. ruckeri* challenge, although the difference in mortality is relatively small.

3.1.3 Microarray analysis:

Experimental design:. Spleen RNA samples prepared at 48 h after challenge were used for microarray analysis to investigate the differential gene expression profile between the HR and LR fish. Four groups of fish, HR/control and challenge, and LR/control and challenged, were analysed. Six individual samples (fish) from each control groups and twelve samples from each of the challenged groups were analysed, bring the total fish number to 36. A common reference was constructed by combining an aliquot of RNA from all the samples prepared. Each RNA was compared with the common reference with dye swaps. Thus 4 groups of samples brought the number of slides analysed to a total of 72.

Probe preparation: The total RNA was treated with RNase-free DNase (Ambion) before probe synthesis using a SuperScript Plus Direct cDNA labelling System (Invitrogen). Briefly, 20 µg of total RNA and 5 µg of anchored oligo$(dT)_{20}$ in 15 µl was heated at 70 ^{o}C for 5 min, and placed on ice. The denatured RNA was then mixed with 15 µl of labelling mix with a final volume of 30 µl containing 1x First-strand buffer, 10 mM DTT, 40 U RNaseOUT, 800 U SuperScript III reverse transcriptase, and 1 mM dNTP containing either Alexa Fluor 555 or Alexa Fluor 647. The labelling reaction was incubated at 46 ^{o}C in a PCR machine for 3 h. After the labelling reaction, the RNA templates were hydrolysed by addition of 15 µl of 0.1 M NaOH and incubation at 70 ^{o}C for 30 min, and neutralized by addition of 15 µl of 0.1 M HCl. Finally, the labelled probe was purified using the PureLink PCR purification system (Invitrogen).

The hybridisation was performed on an Amersham Lucidea hybridisation station. Prior to hybridisation, an equal amount of the Alexa Fluor-647-labelled probe and the Alexa Fluor 555-labelled common reference, or vice versa, was combined with 2 µl 100 mM oligo $(T)_{25}VN$, 110 µl formamide and 55 µl 4x hybridisation buffer (Amersham Biosciences) to make 220 µl. The probe mixture was heated to 95 ^{o}C for 2 min before being injected into the hybridisation chamber. The microarray slides were pretreated by heating in water at 95 ^{o}C for 5 min, then washed with water and air dried and kept at 42 ^{o}C before the probe was injected. After 16 h hybridisation at 42 ^{o}C, the slides were washed with 1x SSC for 10 min, then with 1x SSC+0.2% SDS for 10 min, and finally, with 0.1x SSC+0.2% SDS for 10 min, and air dried before being scanned.

Scanning and analysis: The slides processed as described above were scanned on an Axon 4200A scanner (Axon Instruments) at a resolution of 5 µm and saved as *.tif files. The images were initially analyzed using the GenePix (Version 5.1, Axon Instruments) program to flag out bad spots. The resulting data was inputted into Acuity (version 4.0, Axon Instruments) and normalized using Lowess normalization (Cleveland, 1981). Three methods, significance analysis of microarrays (SAM - Tusher et al, 2001), analysis of variance (ANOVA) and fold change, were used to identify the differentially expressed genes between groups. ANOVA analysis was performed with the Acuity program based on the normalized data. The significance cut off was set to $p<0.05$ for ANOVA.

Quantitative RT-PCR analysis

Verification of differential gene expression from microarray analysis was carried out by quantitative RT-PCR using a SuperScript III Platinum Two-Step qRT-PCR kit with SYBR Green (Invitrogen). All the primer pairs were verified that they could not amplify genomic DNA under the conditions used for Q-PCR.

Results:

Using microarray analysis, 195 genes were identified being differentially expressed ($p<0.05$, Table I) between groups of HR/LR trout at 48 h after *Yersinia ruckeri* challenge. This list of selected genes are indicated in table I, annex 1. This information has been sent to INRA- Génétique des Poissons (Jouy-en-Josas) for SNP analysis.

As only a limited immune genes presenting in the microarray and we have identified a number of immune genes potentially important for ERM pathology. We conducted a real-time PCR analysis on a selected list of genes with samples from a time course of ERM challenge. As shown in Figure 9, 10, 11 and 12, the expression of IL-2, two Interferon –gamma, IL-1 members, and some members of SOCS genes were up-regulated by *Yersina ruckeri* infection.

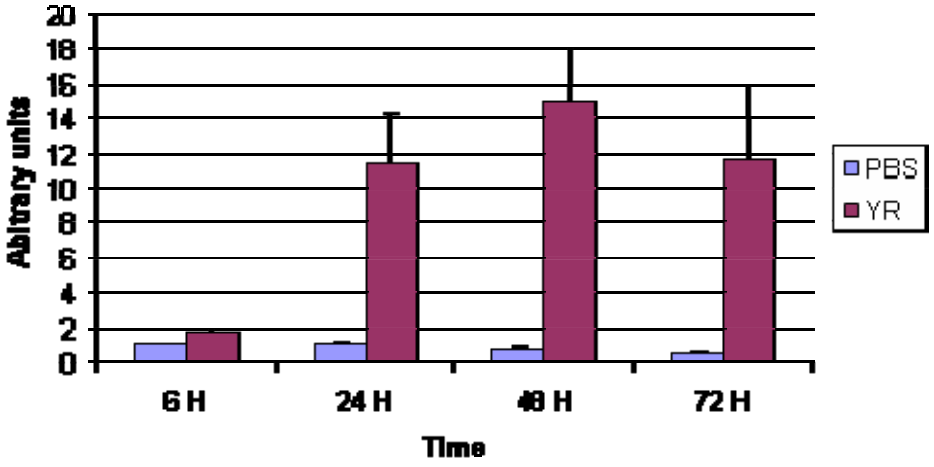


Figure 9. Up-regulation of IL-2 by *Yersinia ruckeri* challenge

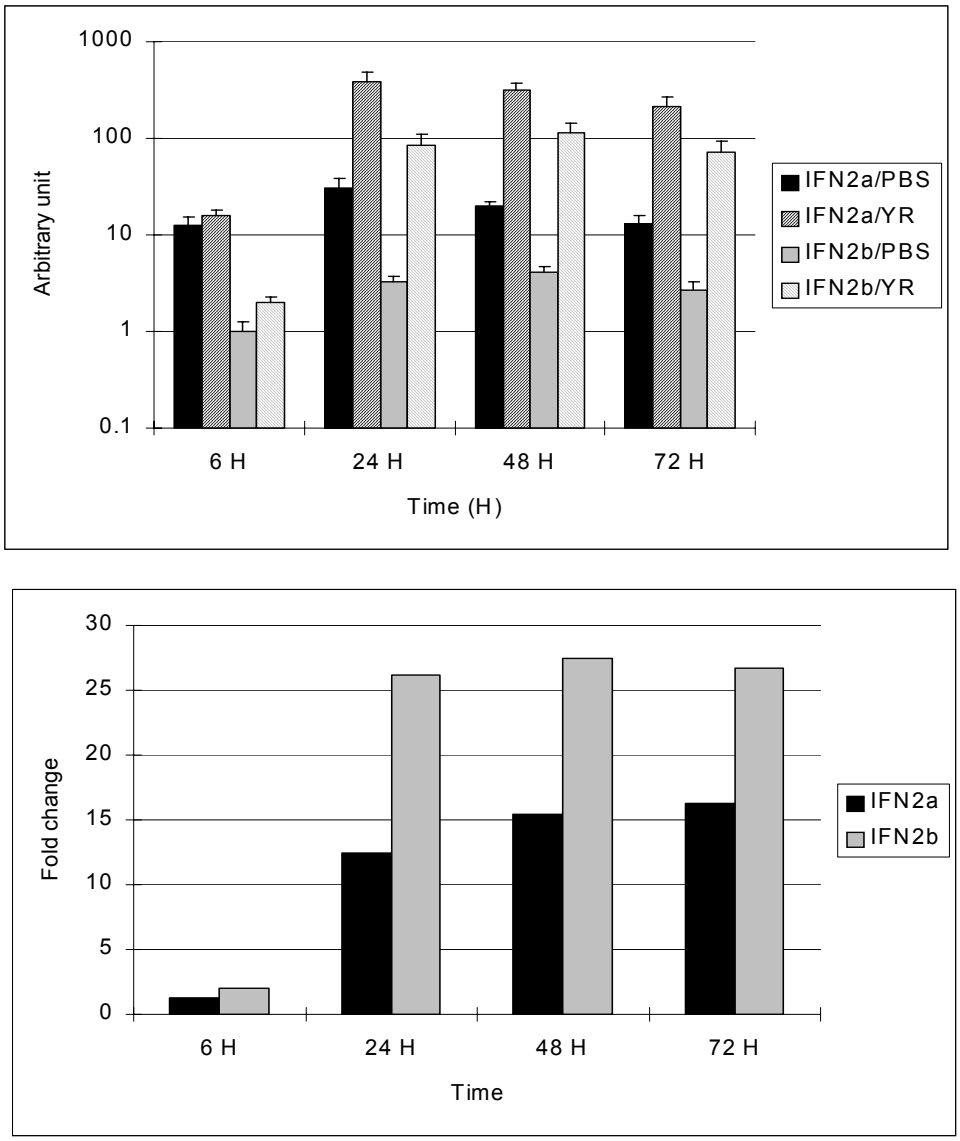


Figure 10. Modulation of the expression of the two interferon-γ genes in the spleen of *Yersinia ruckeri* challenged rainbow trout

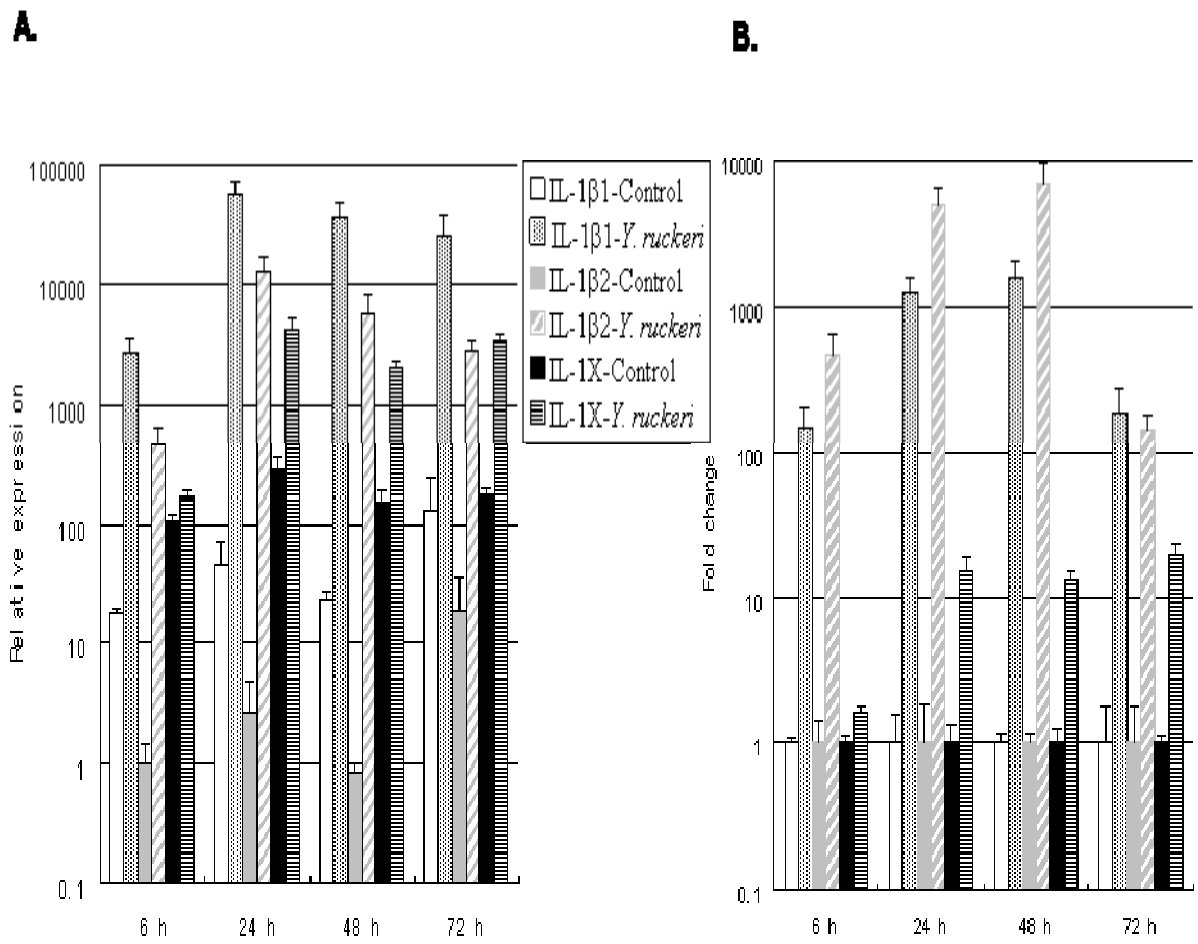
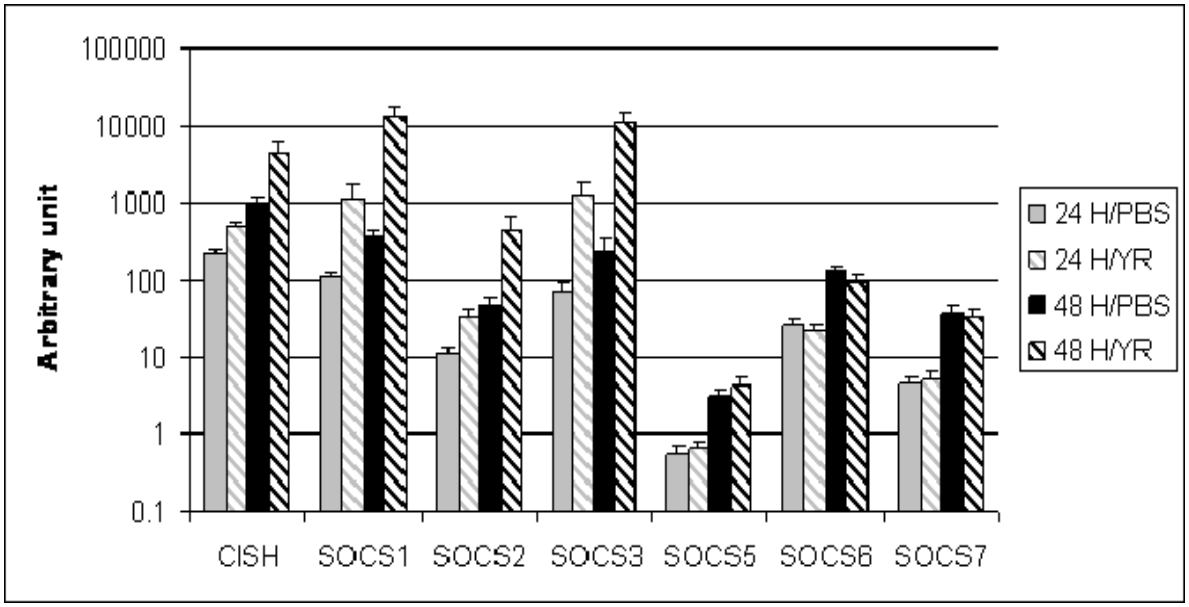


Figure 11. Modulation of the expression of the IL-1 family members in the spleen of *Yersinia ruckeri* challenged rainbow trout



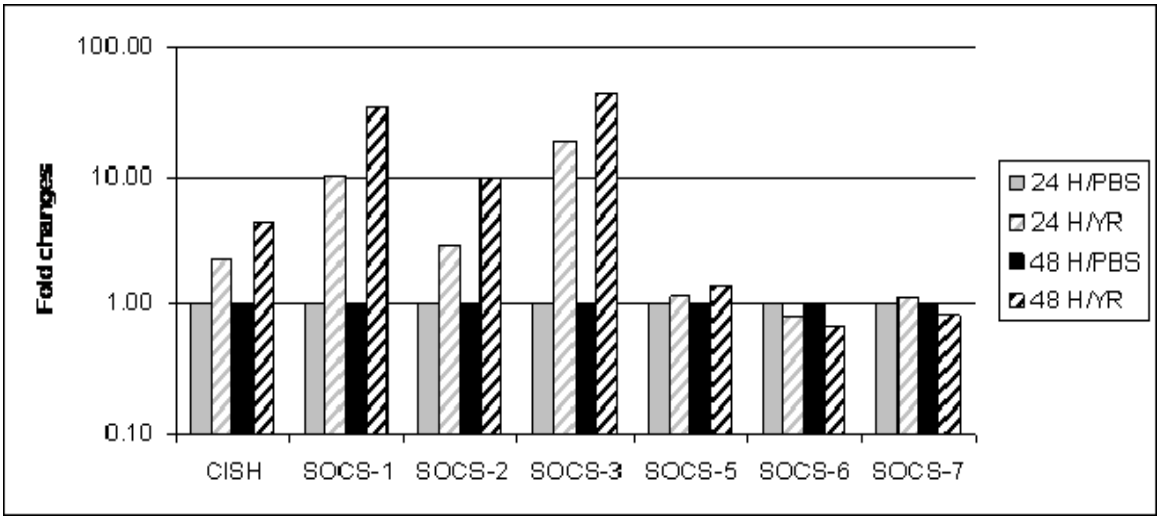


Figure 12. Modulation of SOCS gene expression By Yersinia ruckeri challenge

Sea bream gene expression profiles during confinement stress and chronic exposure to the enteric pathogen *E. Leei*.

Two experimental conditions were considered to analyze the gene expression profile in gilthead sea bream: confinement stress (i) and pathogen exposure (ii).

Material and methods

Juvenile gilthead sea bream were exposed to confinement stress and to a myxozoan parasite (*Enteromyxum leei*). Description of the protocols and data on a range of physiological variables [respiratory burst activity and circulating levels of cortisol, glucose, lactate, growth hormone and insulin-like growth factor) were reported in Annex 1 (part 1, Collections of biological material, Sea bream).

Experimental design & microarray hybridisation: For both confinement stress and pathogen microarray experiments a reference design was used, where both experimental and control RNA were compared to a common reference RNA for that experiment.

- Sea bream confinement stress: microarray analysis involved the comparison of liver gene expression profiles of control (10 kg/m^3) versus stress (120 kg/m^3) fish for the 6h, 24h, 72h and 120h time-points. The profiles of 5 individual fish were analysed for each time-point in a reference design with a Cy3/Cy5 dye-swap therefore in total 80 microarrays hybridisations were performed.

-Sea bream pathogen exposure: microarray analysis involved the comparison of gene expression profiles of control, non-infected (R-NonPAR; recipient fish exposed by water effluents to the parasite, but non parasitized), and infected fish (R-PAR; recipient fish exposed to the parasite and parasitized) for 2 tissues, head kidney and intestine. The profiles of 5 individual fish were analysed per condition in a reference design with a Cy3/Cy5 dye-swap therefore in total 60 microarray hybridisations were performed.

For Microarray hybridisations 10 micrograms of total RNA per individual fish (control/experimental) and ten micrograms of reference RNA per slide were indirectly labelled following an 'in house' protocol with Cy5 and Cy3 dyes. Hybridisations were performed in a Genetix hybridisation chamber and washes used the Advawash automatic washing station. Microarray slides were scanned in two channels (543 & 633nm) at 5 µm resolution using a confocal laser scanner (ScanArray Express, Perkin Elmer).

Data acquisition & analysis: Two channel TIFF images were imported into the Genepix Pro 6 software program for feature (spot) finding and alignment using a batch alignment process. Features were flagged as present, absent or bad by this software program and pixel intensities for feature and background were quantified. Output Genepix results (GPR) files were imported into Genespring GX 7.3. Data was normalised for signal, dye swap and intensity (intensity dependent lowess normalisation) and log 10 ratios of signal (experimental) to control (reference) channel were calculated. All resulting raw and normalised data is in process of submission to SIGENAE version of BioArray Software Environment (BASE) database.

For the confinement study normalised data were filtered on confidence, signal strength and flags present. A one-way Anova of time and condition was performed to select those genes differentially expressed (P<0.05) between control/stress over the time course. For the pathogen experiment, a one-way ANOVA was conducted to select genes differentially expressed among control, R-NonPAR and R-PAR.

Real-time PCR confirmation: A selection of genes potentially differentially expressed between control/stress (confinement) or control/R-NonPAR and R-NonPAR/R-PAR (pathogen) were chosen for real-time PCR confirmation. For confinement stress, genes were chosen representing the most interesting k-means clusters over the time course. For pathogen, genes were chosen representing infection, resistance (or both) and on the basis of the gene ontology analysis carried out by Partner CR06.

PCR primer details are provided in Table 1. Real-time PCR confirmation was carried out using an MX3000P Real-time PCR system. Reverse transcriptions of 2.5 µg aliquots of total RNA from each individual were performed and cDNA diluted to 10 ng/µl. Quantitative real-time PCR (qPCR) assays

were performed using the Quantitech SYBR Green PCR kit. The threshold value (Ct) was fixed arbitrarily for all genes and expression was normalised to the housekeeping gene, β-actin. Data was analysed using the DDCt method (Livak & Schmittgen, Methods 25, 402-408, 2001).

Table 1: Primer sequences for real-time PCR confirmation (sea bream confinement & pathogen exposure)

Gene name	Primer sequence	Study
L-2 hydroxyglutarate dehydrogenase	FOR AAGGTCTTCACAATGACAATGGCG REV TCCCTCGCCATCGCTGAAAT	Confinement Kmeans set 1
Lipoprotein lipase	FOR TTTACGCTCTGTGAGGTCTCCGG REV GGGACGTTGCCAAGTTTGTGAC	Confinement Kmeans set 1
94 KDa glucose related protein	FOR AGAACGTGGCAAAGGAGGGTGT REV TGTCCTTCAGGGCCTTGTCCTT	Confinement Kmeans set 2
Complement C1s subcomponent	FOR CCCACCCAGTGATGACTCCTGA REV GGCTTCCAGAACCGATCTGACTG	Confinement Kmeans set 3
Complement component C7	FOR TTGATTCCTGACAGACGGTCCCC REV CGGCTCAACTCCACCACGTTTACTT	Confinement Kmeans set 3
Ras-related protein Rab-1A	FOR GCTGAAATCAAGAAGAGGATGGGC REV GAGTAAGAGGGCGGGGTGTCAA	Confinement Kmeans set 3
Apolipoprotein Eb	FOR ACTGAACCACTAAAAGTGCCCTTCT REV TAGCCGCAGGACGTGCATTTA	Confinement Kmeans set 3
Alcohol dehydrogenase	FOR GTGCTGCAGTTTATGGGAACCAGTA REV TATTGACTGCTGCTCCGTATCCTGT	Confinement Kmeans set 4
Interferon regulatory factor 8	FOR TGGAGGCAGTGAACATGCGG REV GGGCATGTTGTCCTTGTAGCAGG	Confinement Kmeans set 4
Glutathione S transferase A	FOR GGAGGGTGATGATCGCTCTGGA REV GCGGCAAAGGACTCATTCAGGA	Confinement Kmeans set 4
Col1A1	FOR GAGGCACAGCCGCTTCCATACA REV GGAGCAATGTCAATGATGGGCA	Pathogen
DNase 1	FOR AGTCTCCCAGCAACACGATGTCG REV CGGGGAAGAACTTTGTTCTGATCC	Pathogen
Mannose binding lectin 2	FOR AAAGGAGCCCAACAACGCAGG REV CTTGATTTCTTCCTGGAGCAGACG	Pathogen
High choriolytic enzyme 21	FOR CTGGTGGCTGTGTCAGCTTTGG REV TCAGCGTTTCTCTCAGCCTCAGAA	Pathogen
Cofilin 1	FOR ACTTTCGCTGTCTCAGGAGCCC REV CCAAGCCTGACACCTGAAGAAGAA	Pathogen
Leukocyte cysteine proteinase inhibitor 1	FOR CCTCCATAACAGGGAAGTGCTTGG REV TGTGCAGGGAAAGACGAACAAGAT	Pathogen
Beta actin	FOR CTGGAGAAGAGCTATGAGCTGCCC REV GGTGGTCTCATGGATTCCGCAG	Housekeeping gene

Gene expression analyses

Sea bream confinement stress and analysis of the liver transcriptome: A panel of 660 genes (Appendix: List A) were differentially expressed over the time course of the confinement (Welch Anova $P<0.05$). The top 20 annotated candidate genes in this list are provided along with their fold changes between control/stress for the 4 time-points analysed (Table 2). This panel of candidate genes (those that were annotated) was further analysed by k-means clustering and grouped in 4 major clusters, each one showing a different expression pattern. Set 1 was composed of genes early (6 h) up-regulated after confinement. Genes grouped in set 2 displayed a strong positive peak stress response at 24 h. Genes up-regulated at 24 h and later on (24 h late-stress response) were grouped in set 3. Finally, set 4 was composed of down-regulated genes at 24 h and later on.

Ontology analysis of annotated genes in clusters 1-3 and 4 allowed a time series analysis of the physiological processes involved in the confinement stress response (for ther complete list of genes see Annexe 1, part 2, Seabream section: Table 3, up-regulated genes and table 4, down-regulated genes).

Set 1 encompassed a rapid metabolic readjustment with enhanced uptake and intracellular transport of fatty acids to cover the increased demand of metabolic fuels. Set 2 was associated with a wide variety of tissue repair and remodelling processes that were mostly mediated by the stress response of the endoplasmic reticulum. Sets 3 and 4 encompassed the reestablishment of cellular homeostasis with an increased intracellular trafficking and ROS (reactive oxygen species) scavenging, accompanied by a general decline of oxido-reductase activity and thereby ROS production. The dynamics of this stress response are summarized in Figure 2, and they serve to identify the most relevant processes and genes involved in the sea bream stress response after confinement exposure.

Real time PCR confirmation was performed for 10 candidate genes (Table 5 and Figure 3) that were selected to be representative of the functional processes involved in the k-means clustering. Ongoing data indicate that most expression profiles determined from the microarray analysis agree with real time RT-PCR results.

Table 2: Top 20 most significantly differentially expressed genes (annotated) from sea bream confinement.

Gene name	Uniprot accession	E value	Fold change (stress vs control)			
			6 h	24h	72h	120h
NADH-ubiquinone oxidoreductase subunit B14.5b	Q9CQ54	4.68E-10	1.37	0.45	0.52	1.74
Ras-related protein Rab-1A	Q6NYB7	9.69E-07	1.29	2.50	1.61	1.24
Transmembrane and coiled-coil domains protein 1	Q5I0H4	1.11E-05	1.03	2.56	1.14	0.97
Complement component C7 precursor	Q9TUQ3	2.09E-05	1.78	7.53	7.06	9.13
Cell death-inducing DFFA-like effector protein C	Q96AQ7	2.76E-05	3.03	5.39	3.90	1.10
Glycoprotein-processing glucosidase 1	Q80UM7	4.10E-05	1.20	2.67	1.53	1.65
ATP-binding cassette sub-family F member 2	Q9UG63	4.28E-05	1.44	3.27	1.35	1.0
Six-transmembrane epithelial antigen of prostate 2	Q8BWB6	0.000109	1.84	2.69	3.99	2.32
Heterogeneous nuclear ribonucleoprotein Q	Q7TMK9	0.000115	1.31	1.39	1.83	0.77
Complement factor I precursor	Q9WUW3	0.000152	1.72	3.16	2.3	4.16
Peptidyl-prolyl cis-trans isomerase	P54985	0.00019	0.94	2.63	1.84	1.35
Cytochrome P450 2J6	O54750	0.000195	1.34	0.55	0.50	0.40
Angiopoietin-related protein 4 precursor	Q9Z1P8	0.000214	1.14	1.72	2.0	0.68
Cytochrome P450 1A1	O42457	0.000214	0.58	0.23	0.50	0.64
Glutathione peroxidase 4	Q9N2J2	0.000214	1.02	1.34	1.2	1.7
Isocitrate dehydrogenase	Q9Z2K8	0.000214	1.25	2.51	2.20	1.60
Interferon-related developmental regulator 1	Q5S1U6	0.000214	1.55	1.67	1.06	1.71
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	O35286	0.000287	1.04	3.75	2.80	1.60
Peptidyl-prolyl cis-trans isomerase 1B	P68106	0.000437	1.42	0.38	0.46	0.78
Protein pelota homolog	Q7Z4	0.000451	1.17	3.08	1.67	0.15
Hemopexin precursor	P02790	0.000475	0.82	0.71	0.48	0.60

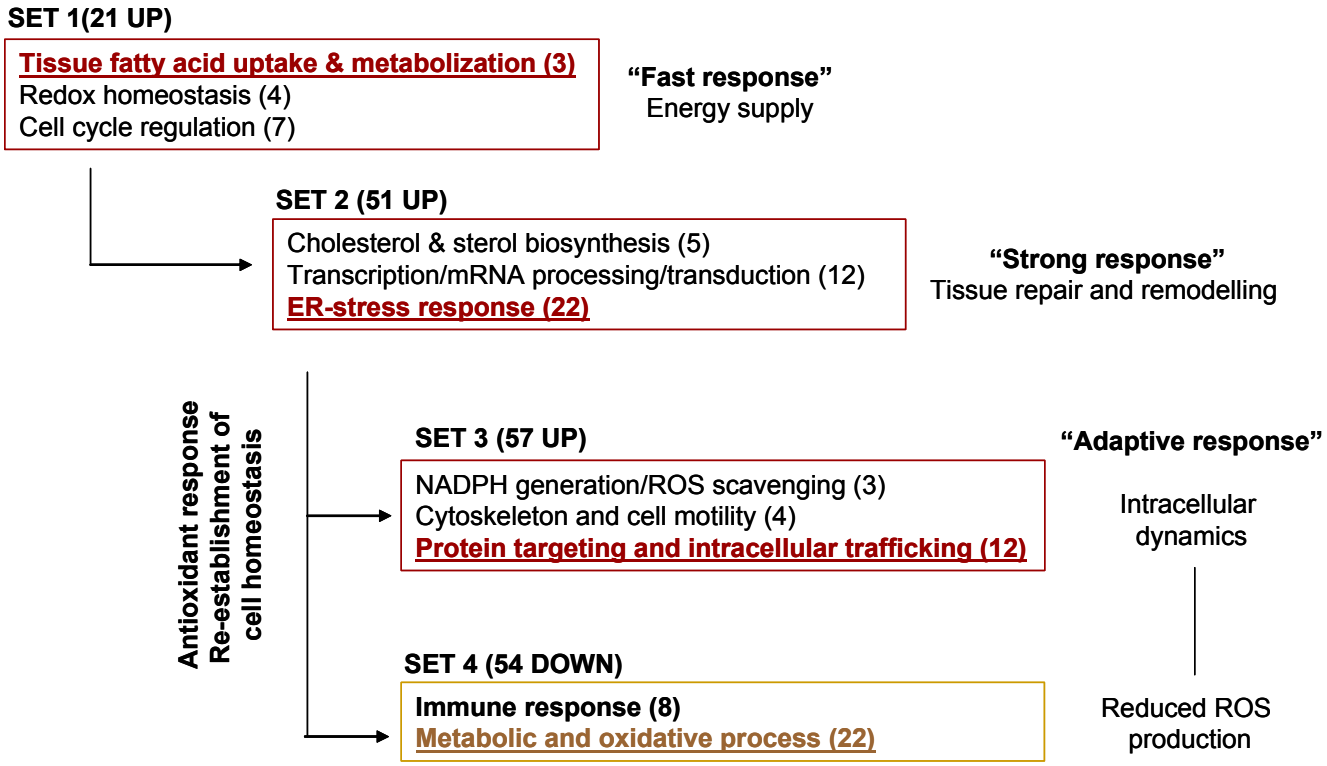


Figure 2: Dynamics of sea bream stress response after confinement exposure. Numbers in parentheses indicate the number of genes involved in each set of biological process.

Table 5: Sea bream confinement: fold changes comparison of real-time RT-PCR and microarray.

Gene name	Real-time PCR				Microarray[1]			
	6	24	72	120	6	24	72	120
Set 1[2]	**Up-regulated, 6 h early-stress response**							
L-2- hydroxyglutarate dehydrogenase	1.44	1.89	2.03	2.68	1.67	1.81	1.66	1.45
Lipoprotein lipase	1.66	1.57	0.94	2.29	2.07	1.72	1.08	1.72
Set 2	**Up-regulated, 24 h strong stress response**							
94 kDa glucose related protein	0.56	7.28	1.96	3.46	0.91	4.87	1.81	1.71
Set 3	**Up-regulated, 24 h late-stress response**							
Complement C1s subcomponent	0.93	3.37	6.04	11.78	1.72	2.86	3.3	3.73
Complement component C7	3.14	21.87	17.16	22.51	1.78	6.80	6.04	7.60
Ras-related protein Rab-1A	1.12	3.13	1.69	2.33	1.29	3.14	1.77	1.50
Apolipoprotein Eb	1.60	8.29	25.93	120.40	1.18	5.97	7.03	7.79
Set 4	**Down-regulated**							
Alcohol dehydrogenase	0.68	0.43	0.38	0.12	0.93	0.42	0.38	0.18
Interferon regulatory factor 8	0.51	0.86	0.95	1.06	1.01	0.60	0.34	1.28
Other								
Glutathione S transferase A	0.72	1.00	1.08	1.82	0.97	0.88	0.97	1.34

[1] Multiple values for the same gene were averaged

[2] These sets correspond to the major k-means expression profile sets.

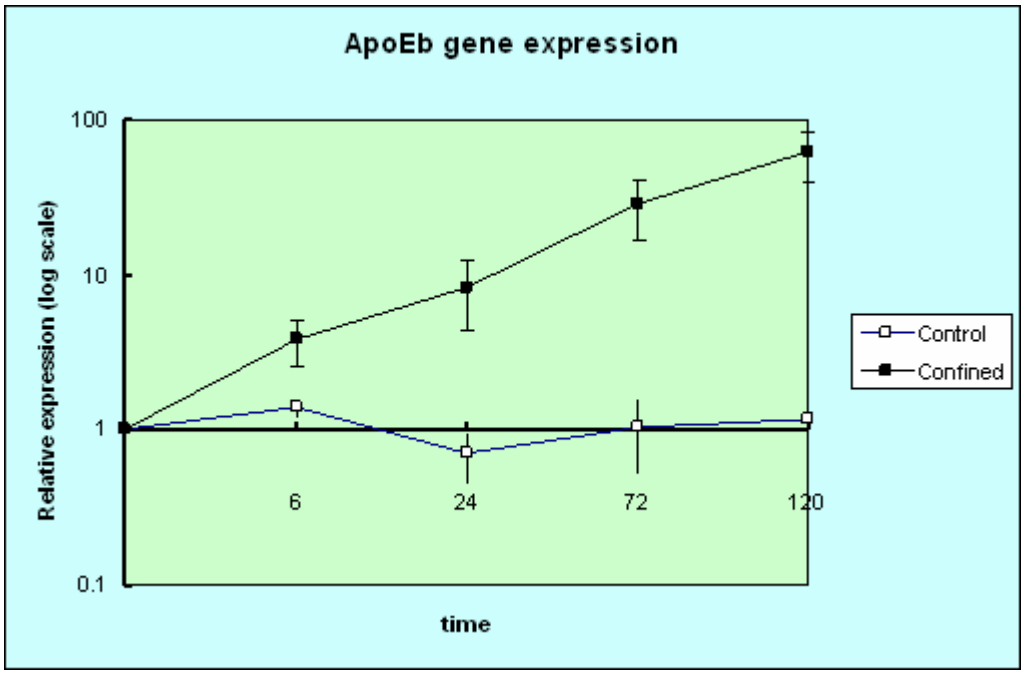


Figure 3: Expression of ApoEb in confinement stress by real time PCR

Sea bream pathogen exposure:

INTESTINE: At the local site of infection, a panel of 765 genes were differentially expressed (P<0.05) in control, R-NonPAR and R-PAR fish. This panel of candidate genes (those that were annotated) was further analysed by k-means clustering and grouped in 4 major sets that were filtered on fold change (1.5, -1.5, -2) **(see Annexe 1, part 2, seabream section, table 6).**

Set 1 was composed of genes that were down-regulated in R-PAR animals. This set had the most abundant number of genes, even with the more restrictive filter ratio value at -2.0. Amongst these genes, the down-regulation of lysosomal and digestive (lipases and proteases) enzymes was most notable. This could be provoked directly by the destruction of the normal intestinal tissue at the local site of infection, with the subsequent intestinal dysfunction. **Set 2** suggests activation of the processes related to cell adhesion and proliferation at the intestine of R-PAR animals in order to regenerate the injured intestinal epithelium. The expression of genes related to cell transcription, translation and mitochondrial respiration was also increased, as well as the expression of ROS scavengers to prevent concomitant oxidative damage. In **set 3**, many of the genes up-regulated in R-NonPAR animals are involved in the activation of the immune response via interferon signalling and antigen processing and presentation. In **set 4**, several genes with no clear physiological significance were down-regulated in R-NonPAR fish. As the pathogen was not present in the intestine of R-NonPAR fish after chronic exposure to it, genes of sets 3 and 4 could potentially be considered candidate genes for pathogen resistance. Nevertheless, whether the different expression profiles of R-NonPAR animals are due either to their innate resistance to the parasite or to the activation of their immune system to avoid the proliferation of the parasite remains to be established.

HEAD KIDNEY: A panel of 871 genes were differentially expressed (P<0.05) in the head kidney of R-PAR or R-NonPAR fish. This panel of candidate genes was filtered by fold change (higher than 1.5 or lower than -1.5) and a top list of stress, metabolic and immune-relevant genes was identified in the head kidney of R-NonPAR fish (see **Annexe 1, part 2, seabream section,Table 7)**.

A preliminary real time PCR confirmation of microarray results has been performed initially for 4 candidate genes: 2 for intestine and 2 for head kidney (Table 8, bellow). These candidate genes were selected on the basis of involvement in infection and/or resistance. Most expression profiles determined from the microarray analysis were confirmed by real time RT-PCR.

Table 8 : Seabream pathogen exposure: fold change comparison of real-time RT-PCR and microarray.

Gene name	Real-time PCR			[2]Microarray		
	Control	Infected	Resistant	Control	Infected	Resistant
Intestine						
DNase 1	1.00	0.29	0.70	1.00	0.33	0.62
[1]Mannose binding lectin 2	1.00	0.11	2.62	1.00 0.91	0.25 0.91	1.07 0.93
Collagen 1A1	Incomplete			1.00	0.68	1.18
Head Kidney						
High choriolytic enzyme 21	1.00	0.34	0.10	1.00	0.57	0.50
Cysteine proteinase inhibitor 1	1.00	86.28	63.67	1.00	7.15	3.06
[3]Cofilin 1	Primers failed			1.00	1.02 0.29	1.01 0.20

[1]MBL2 was represented by 3 clones on the array: 1 clone (upper figure) followed this pattern (differential expression) while 2 clones followed the lower pattern (no change). Interestingly, the MBL2 identifications were from different species as per expression patterns.
[2]Values are averages of all clones with this identification (except MBL2 and CFL1): there were 5 instances of COL 1A1, 4 of CPI1, 3 of HCE21 and 1 of DNase 1 on the array.
[3]CFL1 was represented by 2 clones on the array which followed quite different patterns.

Characterisation of the effects of stress and pathogen exposure on gene expression profiles in sea bass.

Non-selected sea bass were exposed to confinement stress or to pathogen. Three experiments were performed: a pathogen challenge, a confinement experiment with the study of acute and chronic phases and a chronic stress experiment with confined, fasted and fed fish. Genes expression profiles were studied in head kidney for the three experiments and an analysis of the response in liver was also performed for the chronic stress experiment.

Material and methods.

1. Confinement stress test

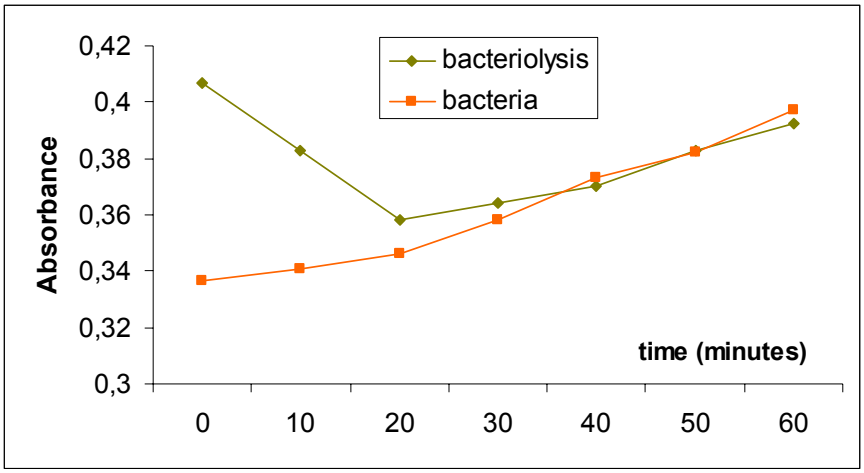
Sea bass of 110g mean body weight were held in 1000 l tanks at a density of 16 kg/m3, the density used in farm conditions. Fish had an acclimatization period of 1 month under standard culture conditions: Natural photoperiod (12L/12D), continuous air-flux, oxygen when necessary to keep oxygen levels no lower than 6.7 ppm, water provided at a rate of 17l/hour. Fish were fed twice a day, 6 days a week during the acclimatization period. After this period, sea bass were exposed to confinement stress, consisting on a net period of 7 days (triplicate group), consider as stressed group, whereas another group (triplicate group) held at standard conditions mentioned above (no net period), including feeding regime during those 7 days was consider as control group. Besides, a third group of fish were kept at the same conditions as control group, but these fish were fasted during the 7 days period (fasted control fish), in order to check the effects of starvation on the physiological parameters related to stress response.
Fish of stressed groups were placed into floating small cages (50x30x15 cm) at a number of 10 fish per net. Triplicates were used for both stress and control fish. At specific time intervals (1h, 4h, 1d, 2d, 5d and 7d) fish were sacrificed. Time intervals were chosen to encompass "early" and "late" responses to the stressor, based on primary and secondary stress responses of plasma cortisol, having then the acute stress period (from 0 to 1 day of stress) and chronic stress period (from 2 to 7 days of continuous stress).
In each sampling point, both stressed and control fish were sampled. Besides, at day 7, head kidney and liver from fasted control fish were also sampled for the chronic stress period. Fish handling time was less than 1 min per fish. Thus, total capture time was less than 8 min per tank to minimize capture stress effects on analyzed parameters (Sumpter, 1997).

Blood was obtained by caudal sinus puncture with a 1 ml plastic syringe. A first aliquota of blood samples were immediately transferred to an Eppendorf tube coated with lithium heparin as anticoagulant. The plasma was obtained by centrifugation at 3000 rpm for 10 min and stored at -80º C prior to cortisol determination. A second aliquota of blood was transferred to another Eppendorf tube without anticoagulant and was used for serum determinations. The second aliquot of blood was allowed to clot at 4 ºC for 2 h. Serum was then separated by centrifugation at 3000 rpm for 10 min, and stored at -80 ºC until analysis.

Plasma cortisol was determined by radioimmunoassay using the trypsin-antitrypsin method as described for gilthead seabream (Molinero and González 1995). Serum lysozyme activity was assayed by a turbidimetric assay as described by Anderson and Siwicki (1994), by measuring the lytic activity of the gilthead seabream serum against Micrococcus lysodeikticus, using hen egg white lysozyme as standard. Alternative complement pathway activity using rabbit RBC was determined as described by Sunyer and Tort (1995) for gilthead seabream. The reciprocal of the serum dilution causing 50% lysis of RBC is designated as the ACH50 and results are presented as ACH50 units/ml.

Serum bacteriolitical activity was also measured. Bacteriolitical assay is a “challenge” type of test, i.e., it measures a general response to a microorganism, pathogen, etc. It has the advantage of measuring the result of a set of mechanisms rather than just one of them. Therefore, some labs are changing or completing the immune tests with these types of tests since these are better overall indicators. The bacteriolytic assay is based in the ability of the fish serum to perform a lytic action

on bacteria. Bacteria (E.Coli; Listonella anguillarum...) were grown in a suitable medium, and then diluted in new medium, checking absorbance each 30 min
When absorbance reach 0.5-0.6 the bacteria was ready. Once the bacteria is ready was placed in an Elisa plate and samples were added. Absorbance were measured each 10 min during 60 min.
The most suitable time (depending on best lytic ability and bacterial growth) was selected and thus, the readings of the time chosen were also selected. Bacteriolytic values were calculated and the bacteriolytic power (in % respect to controls) was determined.



Head kidney was sampled for RNA extraction. Tissue was washed with sterile DEPC-treated water, pre-treated with RNA later and frozen on liquid nitrogen and stored at -80ºC to RNA extraction.

2. Response to pathogen infection: challenge test

The response of sea bass to pathogen exposure will also be investigated. Sea bass of 110g mean body weight were held in 500 l tank at a density of 9 kg/m3. The acclimatization period was 30 days under standard culture conditions described above. Fish were exposed to sub-lethal dosages of
Pasterella piscicida (Photobacterium sp. Piscicida) (108 CFU) by intraperitoneal injection. A sterile saline solution was injected to control fish. After 1h, 4h, 1d, 2d, 5d and 7d fish were sampled.
Blood samples were obtained as described above and plasma and serum determinations were conducted with the methodology described in the confinement stress test section. Head kidney was ²collected for RNA extraction following the same methodology described above.

Results

1. Confinement stress test

As expected, fish subjected to confinement stress into floating cages showed a significant increase ($P<0.05$) of plasma cortisol concentration (measured as ng of cortisol per ml of plasma) just after stress begun (Fig. 1.1). During the chronic stress period, stressed fish showed significantly ($p<0.05$) elevated values of plasma cortisol.

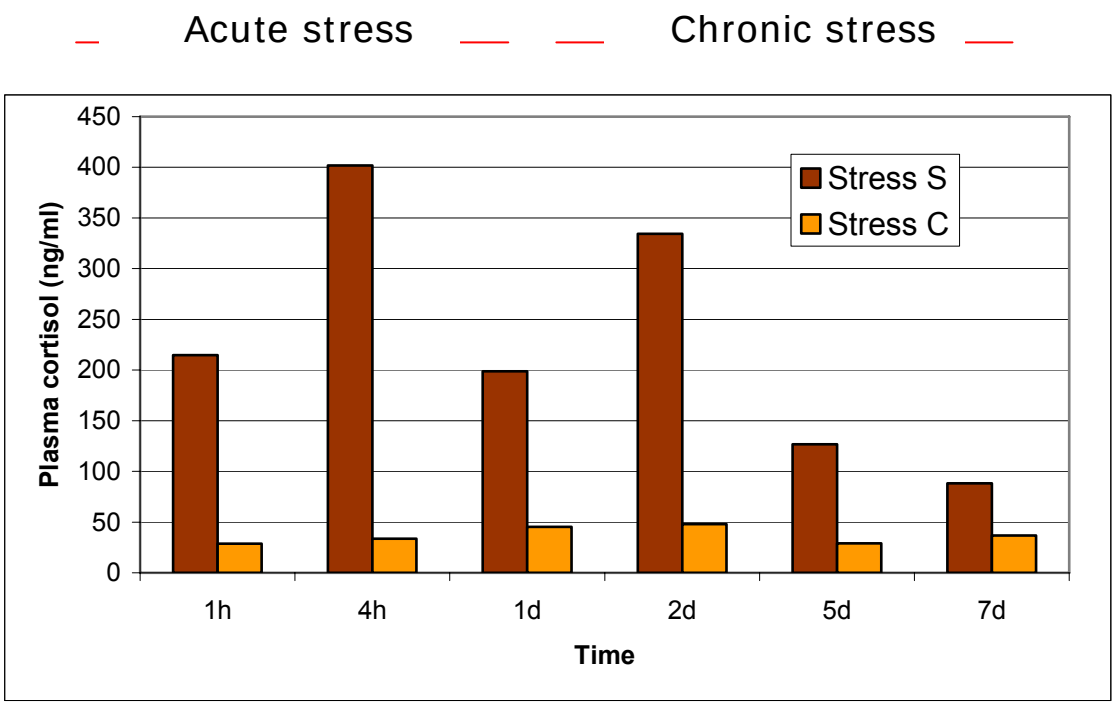


Fig. 1.1. Time-course study of plasma cortisol after confinement stress. Stress S: stressed group; Stress C: control group
*** denotes significant differences with control fish ($P<0.05$)**

Regarding serum analysis, fish subjected to stress showed a significant increase (P<0.05) of lytic activity just after stress, being this one a response to stress. After one day of stress, values were no significant different from control fish, except for day 7 after stress were a significant reduction can be observed, as a negative effect of chronic stress (Fig. 2.2)

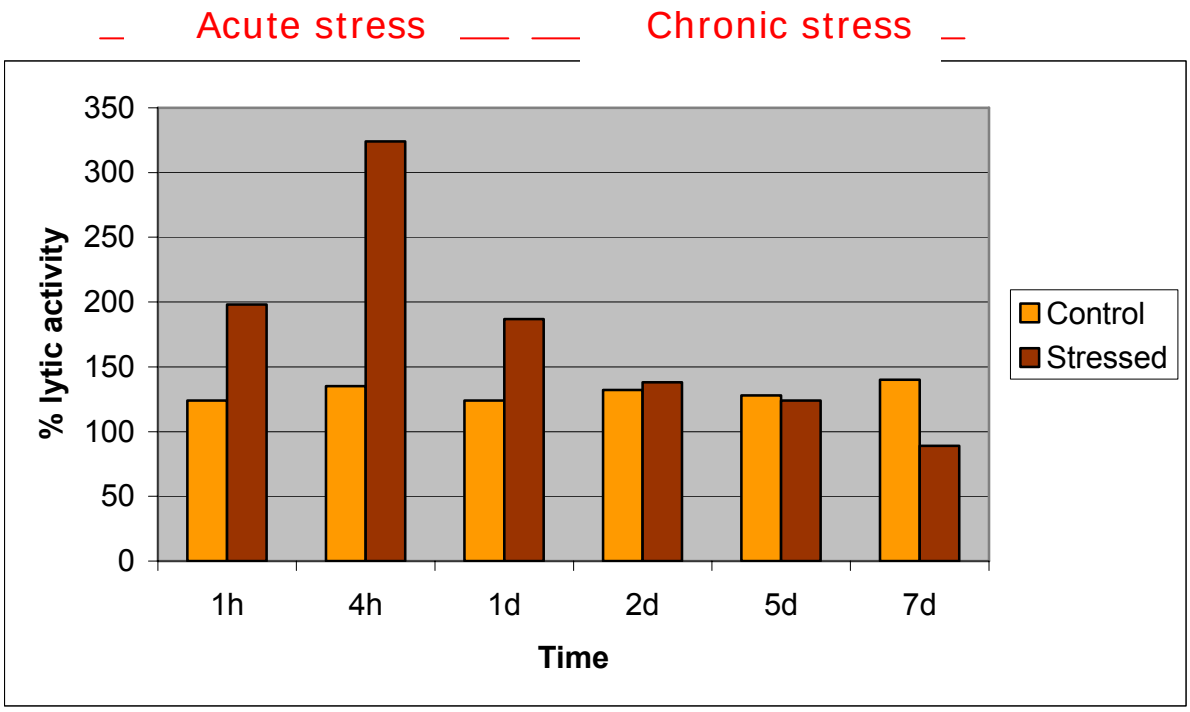


Fig. 1.2. Time-course study of serum bactericidal activity after stress test. * denotes significant differences with control fish (P<0.05)

Lysozyme activity were also significant different (P<0.05) for those fish subjected to stress and during the first hours after stress in response to stressful conditions. After one day of stress, values were no significant different from control fish (Fig. 1.3).

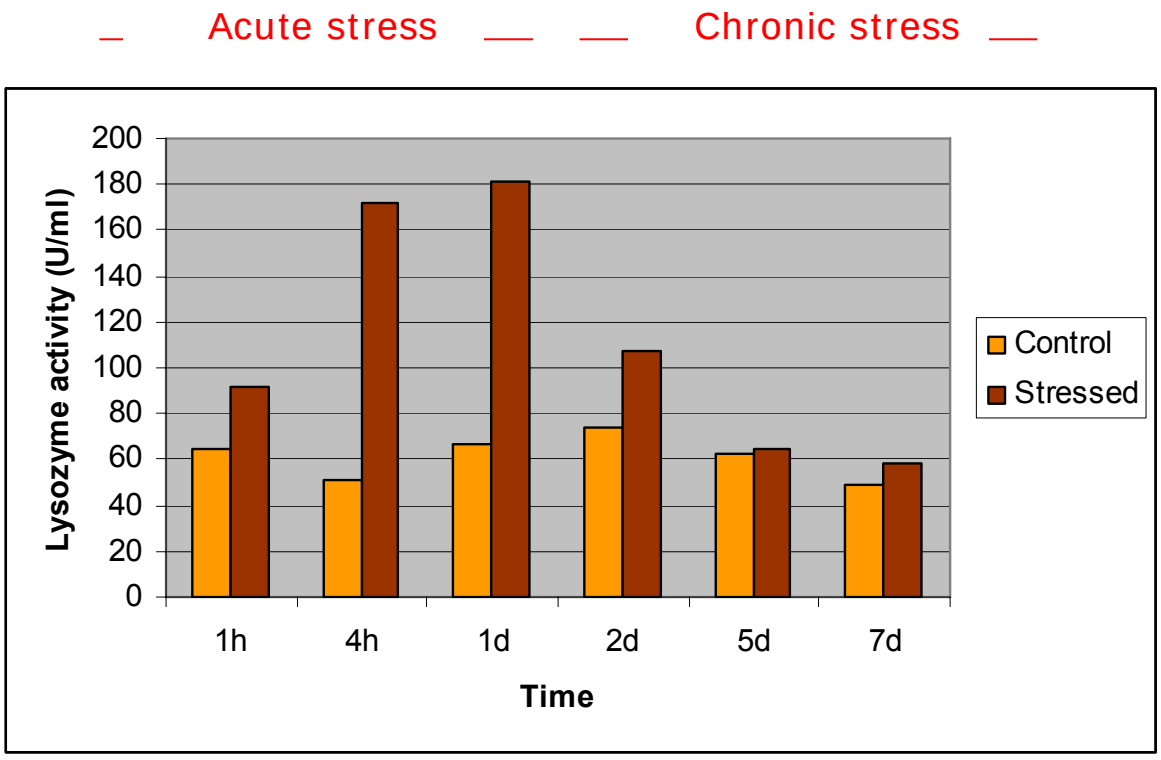


Fig. 1.3. Time-course study of serum lysozyme activity after confinement stress. * denotes significant differences with control fish

Alternative complement pathway activity significantly (P<0.05) decreased in stressed group un der chronic stress situation, being those values in agreement with the effect of stress on this parameter evolution (Fig. 1.4).

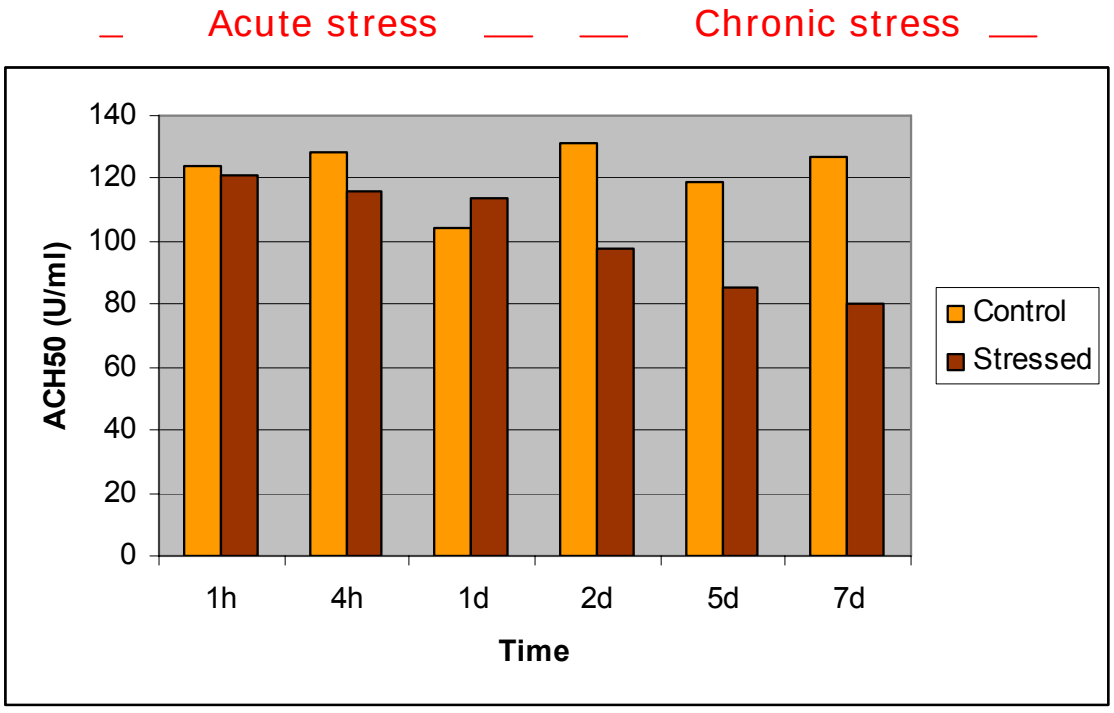


Fig. 1.4. Time-course study of serum alternative complement pathway activity after confinement stress. * denotes significant differences with control fish (P<0.05)

2. Fasted group.

In order to avoid the effect of fasting on stress-related physiological parameters, previously described for this species (Montero et al., in press), those fish held at standard culture conditions described in the material and methods section, but fasted during the experimental period of 1 week, was considered as a "fasted control".
Fasting induced an elevation of plasma cortisol, values being no significant different from control fish at day7 of experiment, but similar to those showed by stressed group after 7 days of confinemet stress (Table 1).
On the other hand, starvation of 1 week do not induced changes on either serum lysosome or lytic activities, with values close to those showed by the control group at 7 days of experiment. However, starvation induced a decrease of ACH50 giving values for starved fish lower, but not significant, than control group, being those values higher (but not significantly) that those showed by stressed group at day 7 (Table 1).

	Control	Stressed	Fasted
Plasma cortisol † (ng/ml)	39.81 ± 9.14	78.55 ± 10.45	76.23 ± 12.47
Bactericidal activity (%lytic activity)	146.22 ± 6.98^{a}	88.17 ± 11.33^{b}	147.04 ± 7.64^{a}
Serum lysozyme activity (U/ml)	48.73 ± 3.24	68.11 ± 16.18	53.21 ± 10.07
Alternative complement pathway activity (ACH50) (U/ml)	127.78 ± 14.64^{a}	80.21 ± 7.48^{b}	99.25 ± 12.45^{ab}

* Values within a row with different letter denote significant differences (P<0.05)

Table 1. Effect o starvation on the different physiological parameters studied after 7 days, comparing with both control un-fasted and stressed (confinement in net for 7 days) fish. Values expressed as mean ± SD, except for † (mean ± SE).

3. Microarray hybridization

.

For RNAs extraction, and quality checking, we used the same protocol as for trout. Some RNAs with an insufficient quality were removed for the following steps. The 101 remaining samples (22 for pathogen, 2*17 for chronic stress and 55 for confinement stress experiments) were labelled using the ChipShot Direct Labelling kit (Promega) with the Cy5 dye. The control samples were also pooled and labelled with Cy3 to be used as a reference. Each sample was then hybridized with the reference on the Aquafirst sea bass glass slides containing 16694 cDNAs spotted in duplicate. As for trout, the automatic hybridization platform (Ventana Discovery) of the Rennes transcriptomic platform was used for this step.

The slides were then scanned using the Axon Genepix 4000B scanner and the signal was quantified with Genepix Pro 5.1software (Axon). The results files obtained were used for the normalization step performed with R software. The lowess method was used for the normalization within array, and a normalization between arrays was then performed. The means of the duplicates was made and the genes with more than half of missing values in a condition were not used for the statistical analysis.

4. Statistical analysis of the microarray data on gene expression profiles in head-kidney:

a) ANOVA analysis.

Pathogen challenge

A 2 factors Anova was performed for each gene to analyse the effects due to the infection and the effects due to the time. (see Table 1)

Threshold	Time	Treatment (Inf/Cont)	Interaction
0,01	465 (1)	35 (0)	57 (0)
0,02	803 (1)	103 (0)	160 (0)
0,03	1120 (1)	176 (0)	251 (0)
0,04	1425 (1)	256 (0)	343 (0)
0,05	1684 (2)	357 (0)	448 (0)

AQUAFIRST – Pathogen
() : After Benjamini-Hochberg correction

Table 1: Anova results for pathogen challenge

The 1% type I error rate was chosen for the Fisher global test, so 465 genes were differentially expressed according to the time, 35 according to the control or infected state and 57 had a significant interaction.
These results suggested that the time factor had more effect on gene expression than the infection that is rather unexpected. The interesting genes to be used for SNPs detection and QTLs research are those differentially expressed in response to the pathogen challenge. Now, the Benjamini-Hochberg corrected p-values of the genes differentially expressed according to the treatment were higher than 99%. So we were rather interested in the genes that had a significant interaction and the lowest corrected p-value (See Annexe 1, part 2, sea bass section).
The 6 genes indicated in table 16 (Annexe 1, part 2, sea bass section) are probably the most interesting genes revealed by this experiment although their corrected p-value is high.

Confinement stress

A 2 factors Anova was performed for each gene to analyse the effects due to the stress and the effects due to the time. (see Table 2)

Threshold	Time	Treatment (Stress/Cont)	Interaction
0,01	136 (4)	115 (0)	172 (1)
0,02	277 (5)	207 (0)	374 (1)
0,03	419 (5)	303 (0)	593 (1)
0,04	559 (5)	392 (0)	784 (1)
0,05	703 (5)	512 (0)	1007 (1)

AQUAFIRST – confinement stress
() : After Benjamini-Hochberg correction

Table 2: Anova results for confinement stress experiment

The 1% type I error rate was chosen for the Fisher global test, so 136 genes were differentially expressed according to the time, 115 according to the stress factor and 172 had a significant interaction. In the objective to identify the genes involved in the stress response, we were interested in particular to the genes specifically differentially expressed according to the treatment independently or in interaction with the time effect.

The genes significantly differentially expressed according to the stress and independently of the time can be divided into 2 groups: those overexpressed in stressed fish and those overexpressed in control fish. The main genes in these groups with the lowest corrected p-value are described in **Erreur ! Source du renvoi introuvable.** and **Erreur ! Source du renvoi introuvable.** (see Annexe 1, part 2, section sea bass).

Chronic stress

Head kidney

A one-way Anova with 3 modalities (stressed/controls/fasted controls) was performed for each gene to analyse the effects due to the stress and the effects due to the time. (see Table 3)

Thres hold	Fisher test	Contrasts		
		Stress vs Fed Cont.	Stress vs fasted cont.	Fed cont vs Fasted cont.
0,01	**167** (0)	350 (0)	101 (0)	77 (0)
0,02	350 (0)	708 (3)	247 (0)	163 (0)
0,03	555 (0)	1068 (8)	414 (0)	256 (0)
0,04	767 (0)	1391 (11)	580 (0)	379 (0)
0,05	969 (3)	1685 (15)	773 (0)	494 (0)

AQUAFIRST – chronic stress

() : After Benjamini-Hochberg correction

Table 3: Anova results for chronic stress experiment in head kidney

The 1% type I error rate was chosen for the Fisher global test, we were so interested by the 167 differential genes obtained with this test. The details of the division of these genes between the different groups is presented in the Figure 2. Six groups can be highlighted in addition to a central gene that will be reclassified. The 6 main groups could be interpreted as follow:

The group 1 with 26 genes presents a significant difference between the stressed and control fish independently of the fasted or fed state. Indeed the contrasts between stressed/controls and stressed/fasted are significant whereas the other contrasts are not. The genes could be divided into two subgroups according to they were overexpressed in control or stressed fish.

The second group with 5 genes is characteristic of the difference between fasted and fed states independently of the stress. These genes could also be divided into two subgroups depending of they were overexpressed in fasted or fed controls.

The third group contains 30 genes representative of the fed control state according to the 2 other states. These genes had the same behaviour between stressed and fasted control fish.

It is more difficult to provide a clear interpretation for the three other groups of differentially expressed genes provided by Anova, their classification should be improved by hierarchical clustering.

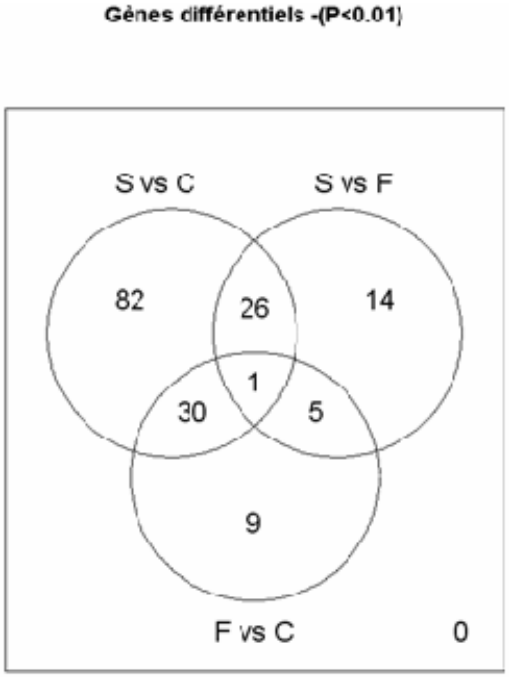


Figure 2: Groups of genes selected with Fisher test

In the framework of Aquafirst project we are more particularly interested in the first group of genes that are more specifically involved in the response to the chronic confinement stress. These genes could be divided into 2 subgroups: those overexpressed in stressed fish and those overexpressed in control fish. The list of these genes is detailed in Annex 1, part 2, sea bass section).

Liver

A one-way Anova with 3 modalities (stressed/controls/fasted controls) was performed for each gene to analyse the effects due to the stress and the effects due to the time. (see Table 4)

Seuil	Test de Fisher	Contrastes		
		Stress vs Contrôle	Stress vs Cont. à jeun	Cont. vs Cont. à jeun
0,01	**183** (0)	339 (0)	195 (0)	29 (1)
0,02	360 (0)	660 (1)	417 (0)	71 (1)
0,03	540 (1)	938 (1)	650 (0)	120 (1)
0,04	742 (1)	1198 (1)	894 (0)	159 (1)
0,05	939 (1)	1454 (1)	1126 (0)	205 (1)

AQUAFIRST – chronic stress
() : After Benjamini-Hochberg correction

Table 4: Anova results for chronic stress experiment in liver

The 1% type I error rate was chosen for the Fisher global test, we were so interested by the 181 differential genes obtained with this test (2 genes with too many missing values were removed). The details of the division of these genes between the different groups are presented in the Figure 3. As for the chronic stress experiment on head kidney, 6 groups can be highlighted that could be interpreted as follow:

The group 1 with 101 genes presents a significant difference between the stressed and control fish independently of the fasted or fed state. The genes could be divided into two subgroups according to they were overexpressed in control or stressed fish.

The second group with 3 genes is characteristic of the difference between fasted and fed states independently of the stress. These genes could also be divided into two subgroups depending of they were overexpressed in fasted or fed controls.
It is more difficult to provide a clear interpretation for the four other groups of differentially expressed genes provided by Anova, their classification should be improved by hierarchical clustering. Thus, the third group contains 7 genes representative of the fed control state according to the 2 other states. These genes have the same behaviour between stressed and fasted control fish. The other groups have not a clear interpretation and a hierarchical clustering was performed to try reclassifying them into the three first groups.

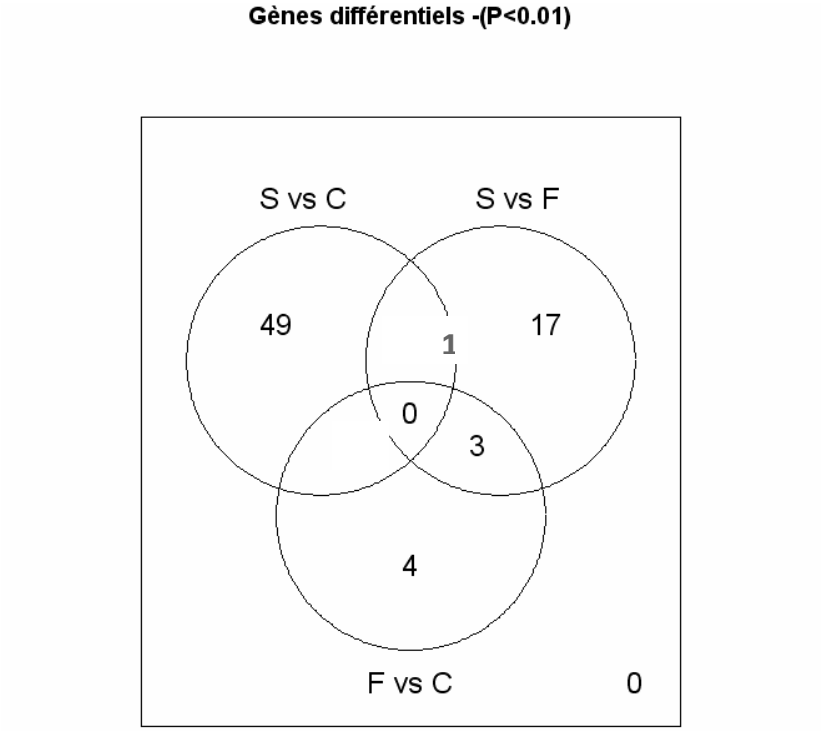


Figure 3: Groups of genes selected with Fisher test

To conclude on the effects of the stress factor, we could highlighted 2 groups of genes: those overexpressed in stressed fish and those overexpressed in control fish. The detail of these genes is presented in annexe 1, part 2, sea bass section.

Exposure of resistant or sensitive lines (R or S lines) of oysters to hypoxia stress or to pathogen.

A- Material and methods.

1) in situ survey during mortality event

In 2005, oysters reared at Fort Espagnol at 70 cm out off the sediment, were sampled *in situ* at seven dates from May to September. At each sampling date and for each group (R and S), 3 replicates of 8 oysters were collected per tissue (gonad, gills, muscle).

As the reproductive investment was a major characteristic differing between R and S families (Samain et al., 2007), we selected the gonad at five dates during the active gametogenesis surrounding the mortality event (figure 1) to do the hybridizations and perform the expression profiles. At 3 dates which surrounded the mortality, gills and muscle samples were also used and hybridised. These data will be compared to data from hypoxia and bacterial challenges obtained on the same tissues in order to do a meta-analysis and look for some common gene expression profiles between the field and experimental stresses.

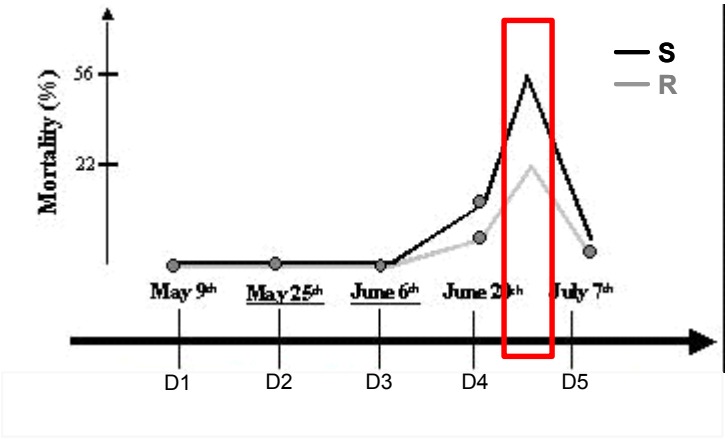


Figure 1 : Time course of mortality observed at Fort-Espagnol in 2005 for resistant R and susceptible (S) oysters. In abscissia are noticed sampling dates.

2) hypoxy experiment

During the hypox experiment performed in May 2005, R and S oysters in normoxia and hypoxia were sampled at day 2 and day 10, for 6 tissues (muscle, gills, gonad, digestive gland, mantle edge and hemolymph) in four replicates. We selected muscle and gill samples at day 10 for hybridization based on the clear effect of hypoxia observed in these tissues in previous studies (Le Moullac et al., 2007).

3) bacterial challenge

In 2006**,** oysters produced in March 2005 by divergent selection of inbred resistant (R) and susceptible (S) families, were challenged with a mixture of alive bacteria *V. splendidus* and *V. aesturianus* (Vibrio pathogen for oyster) in the conditions of summer mortality (temperature above 19°C and during the reproduction period). Samples from gills, mantle and hemolymph were collected to 12h, 24h and 48h post challenge in 4 replicates.

For hybridization of microarray slides, we used the samples from gills and mantle collected 24h post challenge. We chose these tissue samples because the total RNA from hemolymph tissue was insufficient and will therefore be only used in Q-PCR analysis. 24h post challenge samples were selected according to the preliminary results obtained in our laboratory ((Tirape et al. 2007). Indeed, we have shown that for the greatest majority of genes related to immunity a modulation of expression can be observed 24h after the challenge.

RNA extraction of **all** R and S samples (n=130) to be hybridized was realized in our laboratory (P5.1, done by the same manipulators) and their quality checked by using the Bioanalyser (Agilent). A unique total RNA sample was realized and used as control for all the slide

hybridizations. This control was done by mixing the whole tissues of 10 oysters, and was then, labelled 130 times by Cy3 and pooled before use.
Briefly, total RNA was extracted in large volume of "Extract-all" (5ml) to avoid DNA carryover and the use of DNase (cf troubles related in 2006 report). The concentration was controlled with the Nanodrop and the quality checked by Bioanalyser Agilent before the labelling.
The Direct Labelling by Reverse transcription from Promega (Chipshot Tm) was performed just before the hybridization step. The sample target was labeled with Cy5 (red) and the control with Cy3 (green). The label is introduced during the cDNA synthesis with reverse transcriptase and then purified. The label quantity and quality is then controlled using a Nanodrop.

B- Hybridization

The oyster microarrays were produced and send to partners by the MPI Berlin in May 2007. The microarray design corresponds to 22176 spots (48 blocks of 462 wells) of 9059 singletons spotted in duplicate (18118 spots), and controls (empty spots, water 4058 spots). The hybridizations were realized immediately at the West Genopole –IFR140 in Rennes in May and June 2007, using a hybridization station (Ventana).
By the whole, 130 microarray slides were hybridized with total RNA samples coming from three experiments with resistant and susceptible oysters to summer mortality: 66 slides for in situ survey, 32 for bacterial challenge and 32 for hypoxia challenge.
After hybridization the slides were scanned with Genepix 4000B (Axon).The spots were quantified, filtered (noise background) with Genepix Pro5.1 software, before normalization treatment by Madscan software. Finally, the statistical analysis were performed using different softwares : SAM (statistical analysis of microarrays), GeneAnova, R.
Here are only reported the bioinformatic and statistical analysis for seasonal cycle and bacterial challenge.

C- Statistical analysis of the field experiment

Hybridizations concerned 3 tissues: gonads (5 dates) and gills and muscle (3 dates). Here are only reported the analysis on the gonad. Different statistical analyses were performed; we report the most valuable results obtained by the SAM analysis.
SAM analysis assigns to each gene a relative difference score based on its change in gene expression relative to the standard deviation of replicate measurements for that gene.
After logarithmic normalization, the data were corrected by the mean of control values (green). SAM analysis has been performed to compare R and S gonad samples for each date of sampling. The main result is the characterization of 423 differentially expressed genes between both R and S progenies (see Annexe 1, part 2, section oyster, Figure 2).
The gene expression profiles for the 5 sampled dates has been generated for each of the 423 selected genes, allowing the selection of the most interesting candidates, according to the mortality phenomena observed in the field. This shorter list of genes (104) has been transmitted to geneticists (Ifremer La Tremblade, Partner 5) to look for polymorphim and QTLs linked with survival.
Gene ontology analysis was performed on the 423 selected genes, and could be determined for 115 candidates. Statistical analysis (with EASE software) of these GO allowed finding over-represented GO terms, in comparison to GO obtained for all spotted genes. Among them, specific terms were found associated with stress, nutrition/ energy and reproduction (and also others functions, still under analyses).
Hierarchical clustering (with Tmev software) has been performed on the 423 differentially expressed genes, in order to highlight principal groups of correlated genes. Even if the differential of gene expression between R and S for the selected genes didn't seem to be stable during the 5 studied dates, the clustering allowed the identification of 6 groups of correlated genes (see Figure 3). Studies are now running to check if the correlated genes are involved in the same biological process (by the analysis of the annotations and GO represented in each cluster).
The quantitative real-time PCR confirmation of some microarray results are under process and close to be finished: RNA prepared for microarray analysis are now used for confirmation of microarray expression patterns, on 25 selected differentially expressed genes, chosen for their differential between R and S, or for their integration in some particular GO group (see Figure 4). Specific primers of these genes were designed, and PCR efficiency was determined for each primer pair by calculating the slopes of standard curves obtained from serial dilution analysis of cDNA.

Among the differentially expressed genes between R and S families, we noticed some genes implicated in the reproduction and in energy metabolism, both being suspected to play a role in the resistance to summer mortality (Samain et al., 2007). Due to the high number of gametes produced, gametogenesis is a period of negative energy budget in oysters (Soletchnik et al., 1997; Heude-Berthelin, 2000) where most of the acquired energy is used for the production of gametes. This critical period has been shown to be detrimental for the defense mechanisms (Perdue et al., 1984; Myrand et al., 1995) and appears to be correlated with summer mortality phenomenon (Worrall and Widdows, 1984). Interestingly, lines selected for high (R) or low (S) resistance to summer mortality, a highly heritable trait in *C. gigas* (Huvet et al., 2004; Dégremont et al., 2007; Samain et al., 2007) show a differential investment for reproduction with R families displaying lower reproductive effort and spawning than S families thus suggesting that R lines survive summer mortality because they are not as reproductively active as S lines (Samain et al. 2007). One gene characterized for its differential expression between R and S families, a TGFb superfamily member, encodes a gonad specific growth and differentiation factor (Fleury et al, 2008). Other identified genes (NDK5, CD109, in figure 4) need to be comforted for their role in reproduction. Another interesting EST was observed, the neuro peptide Y receptor. Indeed, in vertebrates, intracellular signaling pathways that control energy balance, such as insulin, adipokines and NPY signaling, also play a pivotal role for gametogenesis (Wade and Jones, 2004). In bivalves, seasonal changes in the storage of energy components contribute to the gametogenetic effort (Berthelin et al., 2000) and some genes differentially expressed (NPY dopamine, retinoic acid receptors, calmodulin) comfort the idea that vertebrate-like signaling pathways exist in oyster, such as recently described in *Lymnaea Stagnalis* (De Jong-Brink *et al*., 2001). Following the results of the real-time PCR, some genes of the energy signaling pathway will be selected to pursue studies on links between energy and reproduction.

To conclude, these studies of functional genomics allow the characterization of genes that are differentially expressed between R and S progenies. To evaluate these genes as potential candidate markers of resistance to summer mortality, they will be next studied “one by one” in terms of the physiology of these oysters with functional promising tools and also in the time of this project terms of valuable markers in the search for Quantitative Trait Loci (QTL).

D- Statistical analysis for bacterial challenge

SAM analysis performed on gills and mantle, comparing both R and S families and the challenged and control oysters identified 132 responsive genes (False Discovery Ratio (FDR) <13%) (Table 1). One gene was differentially expressed in 3 conditions and 5 genes were differentially expressed in 2 conditions. A weak response to the bacterial challenge was noticed and the genes are not annotated. More differentially expressed genes were measured between genotypes (S-R challenged or S-R control). In gills 78 and 53 genes were respectively differentially expressed whereas in mantle only were 4, and 1 gene. Among these genes 21 are annotated and according to the analysis with Blast 2 Go (biological process level 2) these genes are distributed in response stimulus, metabolism process and cellular process mainly.

	Challenge		**Genotype**	
	S bact-control	**R bact-control**	**S-R challenge**	**S-R control**
gills	**0**	**0**	**53 (12 annoted)**	**78 (11 annoted)**
mantle	**2**	**1**	**1**	**4 (1 annoted)**

Table 1: Genes expressed differentially in oysters R and S after a bacterial challenge in gills and mantle.

These analyses show that with the environmental conditions of summer mortality, the bacterial challenge have a low effect on gene expression both for S and R, and more particularly on gene expression related to immune response previously studied in our laboratory. However, these results show a genotype effect (R vs S) suggesting that the gene differentially expressed could explain the resistance capacity of R *vs.* S to summer mortalities. It will be necessary to improve the potential role of these genes, particularly by their identification by Gene Ontology to elucidate their function. The validation of these results is in process by quantitative PCR. Actually, 38 genes with 0% FDR have been selected to research SNPs (Figure 5).

General Conclusion

All these microarrays analysis led to the production of lists of genes of which expression was significantly up or down regulated after exposure to stress. Moreover, for trout and oyster, comparaison of gene profiles obtained din divergent selected (HR and LR in trout, R nad S in oyster) lines allowed to focus on genes potentially involved in genetically-based expression differences. These lists of genes are presented in Annexe 1, part 2 see section sea bream, sea bass, oyster and trout. For oyster and trout selected genes, this information was transferred to other partners for SNP analysis.

In addition to this production of candidate genes for SNP which was the main objective of these studies, we have also started to analyze these transcriptome data in order to suggest some major biological processes (molecular or cellular) which could be associated with stress (confinement, pathogen exposure) responses in the four species. Although these gene expression profile studies were carried out on different species and using different microarrays, overall these analysis showed interesting similarities:

- Labelling of target and hybridization protocoles, scanning protocoles, quantification and filtering of the spots, normalization treatment of the scanning data were very similar in the different microarray experiments performed in this WP. Moreover, stress factors were in the 3 fish species or very similar (confinement stress) or of similar nature (exposure to pathogens (virus or bacteria). In this contexte, it was tempting to carry out a general analysis of the different sets of microarrays produced by each stress situation for each species. However, as these studies have been using different microarray (one for each species), on different transcriptome platforms and, of course, all sources of biais (labelling, hybridization, spotting of arrays...) have not been properly ramdomized which exclude the possibility to drive a general statistical analysis of all the microarray data obtained for all species. The only validated study is the one carried out on each stress experiment carried out on 1 species.

- When comparing the different statistical analysis for each stress/species combination, it was interesting to note that in all cases, strigent statistical criteria, i.e. SAM or ANOVA analysis with FDR = 0 led to selection of a limited number of genes (a few dozen) and in several case in selection of no genes. We have to use ANOVA with non-corrected p values to be able to select several hundreds of genes for which a significant number are false positive. This situation does not seem to be linked to the quality of the microarray analysis but more probably, to the physiological process analyzed, i.e. responses to stress exposure.

- For all collections of EST in sea bream, seabass, oyster or trout, we had to face a significant problem of gene annotation and, overall, only 50% of our EST collections were automatically annotated. This means that among statistically selected genes, about half cannot be used for identification of biological functions. However, those candidates are still interesting for genetic studies and SNP search.

- Based on the progress of these microarray data analysis, the only comparaison which can be proposed is a meta-analysis of the biological processes identified from the list of genes selected after statistical analysis. However, the present data suggest that many different biological and molecular processes would be involved in the responses to stress and it is still difficult to proposed a more consensual picture which could emerged from those data. Such objective will require to develop more synergestic efforts between the research teams involved in this topic and commmun analysis of the data produced by all groups.

In conclusion, such meta-analysis appears as an interesting objectives which, however, was not included in the initial tasks of the project. Presently, partners involved in this topics agree about the interest of such studies and are trying to find other possibility to carry out this collaborative work.

Part 3: Single Nucleotide polymorphism (SNP) characterization in trout, oyster and sea bass genes or EST (Expressed Sequences Tag).

Single nucleotide polymorphisms (SNPs) are the most common type of polymorphism in the human genome, with an approximate frequency of one every kilobase. These biallelic variants are relatively easy to genotype compared with VNTRs and microsatellites In marine bivalves, we expect to detect a high number of SNPs, considering that the allozyme diversity observed in these organisms is the highest recorded among eucaryotes generally. High genetic diversity has been also described in fish.
The objective of this part is to develop Single Nucleotide Polymorphisms (SNPs) in genes or ESTs (Expressed Sequences Tags), in trout, oyster and sea bass, to be mapped in the QTL analysis. The analysis of these SNPs are to be carried out on:
- the candidate genes previously characterised in oyster and sea bass. ,
- the new ESTs developed in the previous part and which correspond to up or down regulated genes. This approach ahs been developped in trout.

Material and methods

In oyster

DNA was extracted from samples of mantle tissues following the Promega DNA Wizard cleanup Kit recommendation. The quality of DNA was first checked on a 1.5% agarose gel (15V/cm; 40 min) and secondly by quantification using an Eppendorf biospectrophotometer. This allowed to get equal concentration of DNA (100µg/mL) prepared for the amplification step.
Primers were designed using the Primer3 software (Rozen and Skaletsky, 2000). PCR conditions were standardized by the use of a touchdown PCR protocol. The same reaction mix was used for each pair of primers and they were amplified with three range of annealing temperature (60 to 50°c, 65 to 55°c and 82 to 74°c). The mix was composed for each sample of 0.3U *Taq* polymérase (New England Biolabs), 10mM of provided buffer, 1mM $MgCl_2$, 2mM of dNTP (Eurogentec), 10µM of each primers and 100ng of genomic DNA in a final reaction volume of 48µL. Pairs of primers with the same optimal range of annealing temperature were grouped and ran on a Perkin-Elmer ABI 2700 PCR machine (Applied Biosystems) as following: initial denaturation step for 5 min at 94°C, then 2 cycles of 1 min denaturation at 94°c, for every subsequent 2 cycles, the annealing temperature is decreased by 1 degree Celsius during 1 min and an extension step at 72°c of 1 min. At the lowest annealing temperature (50, 55 or 74°c), 25 cycles at 94°c for 30s, 30s annealing and 1 min extension at 72°c are applied. A quantity of 5µL of PCR products was purified with 2µL of the ExoSAP-IT enzymes (Amersham Biosciences) to remove non-incorporated dNTPs and primers according to the manufacturer manual. Then, purified PCR products were sequenced in a single direction using one primer with the ABI Prism BigDye v3 Terminator Cycle sequencing Kit (Applied Biosystems) and the sequences were analysed on an ABI 3100 *Avant* genetic analyser (Applied Biosystems).
Sequences were edited, corrected by hand if needed and aligned using Seqscape v2.1 software (Applied Biosystems) using the KB basecaller algorithm. After a Blast homology search (http://www.ncbi.nlm.nih.gov/BLAST/), the sequences that didn't correspond, at least in part of the EST used to design primers, were removed. Because of the single pass sequencing, sequences of poor quality were removed from the data. We expected to find identical alleles by descent (i.e. the same sequence) several times because individuals of the sample are related. As a consequence, singleton mutations were inferred to be technical artefacts arisen during PCR and sequencing. Although SNPs were characterised by direct sequencing of PCR products, this sampling strategy allowed us to obtain a valid characterisation of SNPs. SNPs were identified as transitions or transvertions for both coding and non-coding regions. For SNPs occurring in coding sequences, variations were classified as synonymous or non-synonymous changes.

In Trout

Genomic DNA samples were prepared from ethanol preserved fin tissues following simplified phenol extraction and ethanol precipitation procedures (Estoup et al 1993).The quality of DNA was first checked on a 0.8 % agarose gel (15V/cm; 40 min) and secondly by quantification using an spectrophotometer (Multiskan spectrum thermo). This allowed to get equal concentration of DNA (20µg/mL) prepared for the amplification step.

Various bioinformatics processes were performed to ensure a proper location of primers.

- Tracking of repeats in the EST sequence (mreps software, *Kolpakov et al 2003*)
- Tracking of dispersed repeated sequences in the genome (salmonids repeat masker software, Cgrasp consortium)
- Screening of the contigs structure, with INRA bioinformatics platform (SIGENAE *http://www.sigenae.org/)*
- Comparison with zebra genome in order to gain information about likely position and size of introns (Ensembl software, *Birney et al 2004*)
- Use of primer3 software to design the primers taking into account the above constraints (*Rozen and Skaletsky, 2000)*.

PCR conditions were standardized. The same reaction mix was used for each pair of primers and they were amplified at two temperatures (60 or 50°c,). The mix was composed for each sample of 0.6U *Taq* polymerase (*Flexi Taq Promega*); provided buffer 1X; $MgCl_2$, 2mM; dNTP 2,5mM (*Amersham Biosciences*) ; each primers 4µM , and 40ng of genomic DNA in a final reaction volume of 20µL. PCR were performed either on MJresearch (Watertown, MA) or Hybaid MBS0.2G thermocyclers with the following protocol: initial denaturation step for 5 min at 95°C, then 5 cycles [95°C/ 30 sec denaturation, annealing temp/30sec, 72°C/1mn extension step], then 25 cycles [94°C/30sec denaturation, annealing temp/30sec, 72°C/1mn extension]. Depending on DNA concentration estimated by migration on minigel, between 1 to 5 µl of PCR products were purified with Exonuclease Phosphatase protocol to remove non-incorporated dNTPs and primers. Then, purified PCR products were sequenced with the ABI Prism Big Dye v3.1 Terminator Cycle sequencing Kit (*Applied Biosystems*) and the sequences were analysed on an ABI 3730 DNA Analyzer (*Applied Biosystems*).
SNP polymorphism in the parents of the 5 QTL families studied in the project and in the parents of the reference mapping families was studied. Sequences were edited, corrected by hand if needed and aligned using Seqscanner v1.0 software (Applied Biosystems) and multalin software (Corpet, 1988). SNPs and indel were listed. SNPs were identified as transitions or transvertions for both coding and non-coding regions. For SNPs occurring in coding sequences, variations were classified as synonymous or non-synonymous changes.

In sea bass

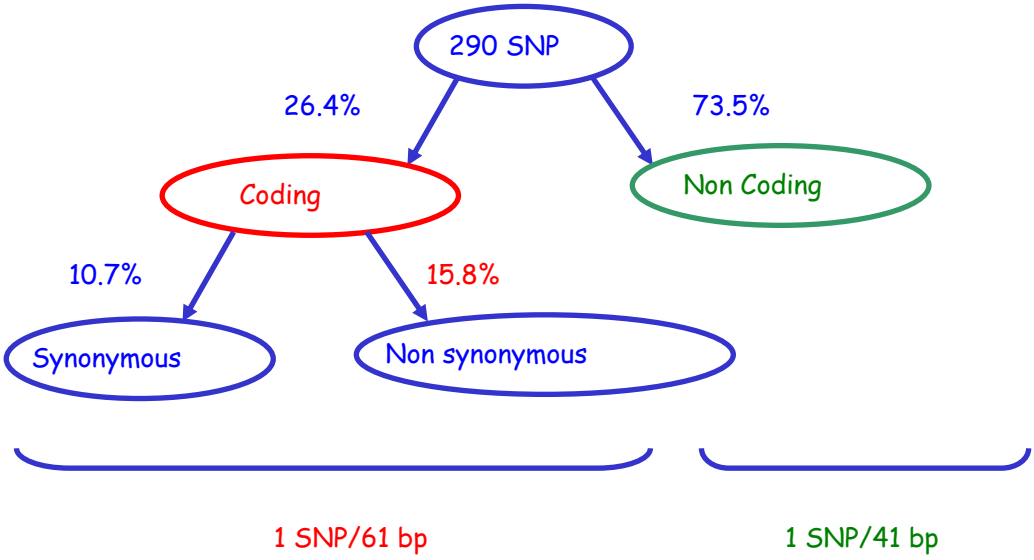
Single Nucleotide Polymorphisms (SNP) were data mined and developed from 30,000 ESTs available from 14 normalised cDNA tissue libraries (Pasos et al. in prep.). Redundance in the SNPs was exploited *in silico* as in Souche *et al.* (2007) through six software packages (MiraEEST, PolyFreq, PolyBayes, QualitySNP, ssahaSNP and SNPServer). Primer sets for a limited number of *in silico* detected SNPs were optimized and validated in the laboratory on a panel of 20 wild European sea bass.

Results section

In oyster

Although we worked with a total of 61 ESTs, the sequences of only 41 were completely corrected and taken into account as some ESTs turned out to be ribosomal genes, and others could be amplified but were never correctly sequenced. We focused on the DNA polymorphism and codon usage bias in this set of 41 nuclear loci for the 24 F0 parents of families. A total of 321 single nucleotide polymorphisms (SNPs) were detected in exons or in non-coding regions (for more details and sequences see Part 3, section oyster) . Our results revealed an astonishing level of DNA polymorphism in oysters. Non-synonymous mutations represent 10% of markers, and 16% are synonymous mutations in coding DNA. Average density of SNPs was estimated to be among the highest levels reported to date with one SNP every 61 bp in coding regions and one every 41 bp in

non-coding regions. As an example, in humans one SNP is found every 1kb and in soybean one SNP is found every 610bp. Non-synonymous mutations contributed substantially to the polymorphism observed in coding regions. The ratio of non-synonymous to synonymous polymorphism was 1.5 on average. We also examined codon usage bias. The table of optimal codons use in the Pacific oyster was deduced from the analysis of an EST dataset. As recently observed in a few other taxa, we found a significant negative relationship between codon bias and non-synonymous diversity suggesting correlated selective constraints on synonymous and non-synonymous mutations



In addition, more work on SNP analysis has been done on 27 genes, chosen for their role in protein turnover, metabolism or stress. The success rate of primers reached 33%, and the percentage of amplicons that were incorrect was 15%. Finally, 5 genes successfully amplified and the F0 generation amplified, sequenced and screened for SNPs. A total of 251 SNPs were characterized.
Altogether, these studies allowed us to characterize 572 SNP. Unfortunately none of them could be successfully genotyped and consequently were not included in the QTL analysis.

In trout
In trout, SNP polymorphisms was searched within candidate genes provided by pCandidates genes were selected from three stress experimentations carried out for gene profile expression analysis (cf. Previous part): a confinement stress challenge and two infectious challenges, a bacterial one (*Yersinia ruckeri*) and a viral one (VHS, viral hemorrhagic septicaemia). For the two first ones (stress and bacterial challenges), the set of genes we worked with included genes that were differentially expressed between the two lines previously selected for their differential cortisol response to confinement stress: HR (high responsive) and LR (low responsive). For the viral disease challenge, the genes were those that were differentially expressed in 2 doubled haploid lines provided by Partner 01.1 and exhibiting opposite resistance to the virus (R, resistant and S, susceptible).

Expression studies resulted in wide lists of candidate genes. The available list of corresponding EST included:

- Stress challenge : 192 EST from head kidney, provided by Partner CO.01
- Viral challenge : 322 EST, selected from 890 and provided by partner CO.01 (fin and spleen)
- Bacterial challenge : 192 EST provided by Partner CR02

After elimination of non appropriate sequences (too short sequences, presence of obvious repeated sequences such as microsatellites,..), primers were designed. Various bioinformatics processes were performed to ensure a proper location of primers.

- Tracking of repeats in the EST sequence (mreps software)
- Tracking of dispersed repeated sequences in the genome (salmonids repeat masker software)
- Screening of the contigs structure, with INRA bioinformatics plateform (SIGENAE)
- Comparison with zebra genome (Ensembl software) in order to gain information about likely position and size of introns (the assumption is that the size and the location of introns is partly conserved between these two genomes)
- And last, use of primer3 software to design the primers taking into account the above constraints

Due to the additional recent duplication in salmonids genome, it is expected that some primers may anchor at two different loci in the genome (the two homologous regions), which makes it difficult to make the difference between a true allelic polymorphism within a single locus and the superposition of two slightly different sequences. To overcome this problem, four doubled haploid individuals were systematically sequenced for each EST studied Then, any "variability" on the chromatogram observed could unambiguously be attributed to the presence of two or more locus (duplication, gene family,...). In a first approach, we did not work with genes showing multiple amplifications.
Next step was to study polymorphism in the parents of the 5 QTL families studied in the project and in the parents of the reference mapping families.

The table 4 below summarizes the success yield at the different steps.

Total number of EST studied	**555**		
Number of pair of primers designed	**609**		
No amplification	*127*	20.8%	*30%lost*
Non specific amplification	*59*	*9.7%*	
Successful amplification, a clear-cut band in agarose gel	423	69.5%	
Studied (27 still have to be studied)	**396**		
Multiple sequences	*139*	*35%*	*35% lost*
Ready for polymorphism studies	**238**		
Monomorphic loci	75	31.5%	*31.5% lost*

Table 4. Success rate in SNPs discovery

The initial plan was to screen 300 EST to get useful SNP polymorphism in around 150 genes. Considering the high lost rate (31.2% only of the examined loci provided informative polymorphism -no amplification, duplication and low level of polymorphism), we finally screened a a much higher number of loci.
Excluding loci with either lack of amplification or multiple sequences, we have sequenced 58178 base pairs accounting for 220 loci. Among those loci, 62 genes came from confinement stress experimentation, 149 from disease challenges, and 30 genes were shared by bacterial and viral challenges. They have been detailed in Part 3, section Trout.
Those 59239 base pairs are split up into 11459intronic base pairs (bp) and 47780 exonic base pairs.On the whole, 247 SNP were found, 185 in exons (i.e. 1 SNP every 255 bp), a
nd 62 in introns (i.e. 1 SNP every 183 bp, a slightly higher rate).

Below, we break down the data from each experimentation.

Challenge	N locus	bp in exons	SNP/ every N bp	bp in introns	SNP/ every N bp	% transversion	% transition	% loci with indel
Stress	62	11966	1/398	6221	1/239	50	50	28
Bacterial	47	11956	1/229	1510	1/137	48	52	36
Viral	111	23858	1/230	3728	1/150	56	44	30

As expected, frequency of SNP is higher in introns than in exons. A rather high level of locus with an indel was detected in both coding or non coding regions. Genes involved in the response to diseases exhibit more polymorphism than those involved in stress challenge. High level of polymorphism in immune gene is also reported in other species. Sequences, introns, exons, and all variability observed (SNP and indel) of polymorphic genes are detailed in annexe 1 (part 4, trout section). Whether the SNP is associated to a synonymous or non synonymous codon in the three open reading frames is also reported.

According to their polymorphism in the different families (see table below), SNP polymorphisms were genotyped in the reference mapping families of Partner CO01.1 and/or in QTL families. The objectives were to map them and to use them as additional markers in the genome scan for QTL detection.

Number of polymorphic loci in the set of QTL families only	27
Number of polymorphic loci in the set of reference mapping families	15
Number of polymorphic loci in both sets of families	102
Total	144

Genotyping was performed in collaboration with CNG (National Center for Genotyping) using the SNplex method (multiplex of possibly 48 loci). Additional technical constraints were generated by the SNplex method (no indel greater than 6 bp, no close SNP) requiring to provide complete information on more SNP than finally used in the SNPlex. Three SNPlex were successfully designed, standing for 119 genes. Two main reasons explain the difference between the theoretical (48x3= 144 genes) and actual number of genes. First, some loci with indel mutation could be used, but they had to be represented twice in the multiplex, with two complementary sets of primers (n=16). Secundly, some locus have been put twice, once for mapping in our reference map families and once for polymorphism analysis in QTL families (n=7). Successful rate of these SNPlex was around 91%. 5% of the loci exhibit pattern of duplicated loci even when they were not supposed to be from first data. For half of them, data are still usable as one of the locus is fixed and the other variable.

Loci that were not compatible with multiplexing but exhibited polymorphism in the reference map families (n=20) were mapped using individual sequencing (details in deliverable D04.02) or microsatellite polymorphism (n=8) .

Finally,101 genes were located on the reference map of Partner CO01.1 (details provided in deliverable D04.02). This will provide information on whether some of them can be positional candidates for QTLs detected in WP05. 132 were also available as additional markers for QTL detection in the Aquafirst families.

In sea bass

Several strategies were used in order to find SNPs in stress candidate genes.

(1) Thirty five SNPs were discovered in 15 publicly available ESTs. One EST from the beta actin gene containing one SNP was found to be homologous to Aquafirst ESTs.

(2) the European NoE *Marine Genomics Europe* (MGE) produced about 30,000 ESTs from 5 F1 offspring from wild Atlantic parents. 55.1% (16,117 ESTs) of the 29,260 processed sequences were redundant and clustered in 4,573 contigs; 44.9% (13,143 ESTs) remained single. A total of 974

(21.3%) contigs qualified for *in silico* SNP discovery, representing 5548 (19%) ESTs and 477,224 overlapping base pairs. Six tools were evaluated: SNPServer, PolyBayes, QualitySNP, PolyFreq, MiraEST and ssahaSNP. A pooled DNA sample containing the DNA of the 5 individuals used to produce the cDNA libraries was cloned and sequenced to evaluate the performance of the tools. 49.5% of ordered primer pairs were successfully optimised. MiraEST turned out to be the most efficient tool with both a high positive predictive values and a high sensitivity. The number of true positives detected by MiraEST is reasonable. A total of 139 validated SNPs have been found in 21 genes. Out of these, 30 SNPs are in amino acid sequences and 10 of them are predicted to be non-synonymous. It is not yet known whether these SNPs are related to stress.

(3) ESTs produced by Aquafirst in SSH libraries were compared to publicly available ESTs and the ESTs developed by the European NoE *Marine Genomics Europe* (MGE). Various *in silico* SNP discovery tools were evaluated on MGE ESTs. The most efficient tool (ssahaSNP) was used on a selection of Aquafirst ESTs which were homologous to stress candidate genes in rainbow trout and gilthead sea bream. A total number of 307 candidate SNPs putatively related to stress have been detected in 112 polymorphic sequences and in 100 unique genes. Of these, 240 SNPs should be true positive SNPs.

For the selected genes, 24 sequences are predicted to be polymorphic, containing 1 to 25 SNPs per sequence with an average of 5.08 candidate SNPs per polymorphic sequence. This number is quite high compared to the average number of candidate SNPs found in sea bass MGE ESTs (1.51 SNPs per polymorphic sequence), using the same software. For the other SSH genes, the average number of candidate SNPs per polymorphic sequence (2.07) is similar to those found for MGE ESTs. The same pattern is observed when considering the number of candidate SNPs detected by the number of overlapping base pairs. In the selected genes, one SNP occurs every 252.61 base pairs whereas one SNP occurs every 718.91 bp in the other SSH genes. The selected genes seem to be much more polymorphic than the other SSH genes.
In the same way, the proportions of transitions, transversions and insertions-deletions (indels) found in selected genes (81.10%, 13.39% and 5.51% respectively) are different from those obtained using MGE ESTs (59.65%, 30.00% and 10.35% respectively); whereas the proportions detected for MGE ESTs and other SSH genes are similar. Indeed, 56.11%, 31.11% and 12.78% of the candidate SNPs in other SSH genes are transitions, transversions and indels respectively. The values found for MGE ESTs and other SSH genes are in the same range as the proportions found in validated SNPs. Out of the 501 SNPs molecularly detected, 58.68%, 33.53% and 7.79% were transitions, transversions and indels respectively. The higher number of candidate indels may be due to the detection of polymorphic microsatellites.

Conclusion:

These studies allowed us to identify a large collection of new SNPs in 3 species, trout, oyster and sea bass. In trout, these SNP have been obtained from candidate EST slected from functionla genomic studies (cf. Part 2) and genotyped. They will also be used for characterization of the QTLs related to stress or pathogen resistance. For oyster, SNP have been generated from previuosly known genes and new SNP are currently developped from the EST generated by the functional genomic studies (see part 2). Genotyping of these SNP are in progress. In sea bass, SNP have been detected in silico in known genes and will have to be validated.

References

Rozen, S. and Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132, 365-86.

Estoup, A., Presa, P., Krieg, F., Vaiman, D., Guyomard, R., 1993. (CT)n and (GT)n microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity 71, 488-496.

Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res. 16 (22), 10881-10890.

Kolpakov, R., Bana, G. and Kucherovmreps, G., 2003. Efficient and flexible detection of tandem repeats in DNA. Nucleic Acids Res. 31, 3672-3678.

Birney, E., Andrews, T.D., Bevan, P. Caccamo, M., Chen, Y. et al., 2004. An Overview of Ensembl. Genome Res. 14(5):925-928.

Part 4: High resolution mapping of the sea bass and the Sea Bream genome.

This part has been carried out in order to bring more mapped genetic markers and new genomic tools which will be useful for the localisation of QTL on the genetic map in seabream and sea bass. This work was the continuation the work initiated within Bassmap and Bridgemap EC projects, i.e. improvement of the construction of a high resolution map for sea bass and sea ream. Moreover, this task also benefited from strong collaboration with Marine Genomic Europe which has been developping large EST collections for those 2 fish species. Such updated linkage maps for sea bream and sea bass is also the starting point of all QTL analyses. Finally, success of such work will rely on the maintenance of database integrating all these genomic and genetic data and accessible to the Consortium through the web. In the 2 species sea bass and seabream, the following tasks have been carried out:
To produce a high-resolution ordered linkage and gene map containing uniformly distributed type I and type II markers

To prepare a comparative genomic analysis with other fish and vertebrate maps.

Sea bass

Materials and methods

1. Isolation of DNA microsatellites
Microsatellite markers have been data mined, isolated and characterized from 30,000 Expressed Sequence Tags available from 14 normalised cDNA tissue libraries (Pasos et al. in prep.), candidate genes, public sources and agreements with ongoing projects (Chistiakov *et al.* in press). Newly developed microsatellites were tested for variability and reliability on 20 outbred wild sea bass.

2. Linkage analysis of microsatellites and SNPs
A well-characterized panel of European sea bass including 2 parents and 50 progeny (Venezia Fbis family) was genotyped using the newly developed microsatellite and SNP markers (cf. Previous part). Using the CRI-MAP program (Green *et al.* 1990), the analysis was performed as described by Chistiakov *et al.* (2005, in press). The microsatellite-based linkage map of Chistiakov *et al.* (2005) was used as a framework to incorporate new markers. Sex-average and sex-specific linkage distances were estimated for each linkage group assuming the Kosambi mapping function.

3. Construction of a radiation hybrid panel and radiation hybrid map
A panel of radiation hybrid cell lines was constructed using the methodology of Senger *et al.* (2006). Briefly, sea bass spleen cells were irradiated by a 3000-rad γ-ray exposure. After counting cells were fused with $HPRT^-$ derivative Chinese hamster ovary cells. Following selection in HAT medium and minimal cell culture expansion, DNA was extracted from individual clones. After verification of the presence of seabream DNA and analysis of the retention profile, a total of 92 radiation hybrid cell lines were selected for the panel. Amplification products were pooled, DNA quantified to produce a stock panel and a working panel compiled by pooling the DNA samples from four parallel amplification reactions. More than 1000 sequence tagged sites (STS) markers were scored on the radiation hybrid panel. They correspond to 150 genome-wide distributed polymorphic microsatellites (Chistiakov *et al.* 2005) and 915 ESTs selected from the sea bream map (Sarrapoulou *et al.* 2007) and unpublished contributions from the European Sea Bass Mapping Consortium. The microsatellite markers allowed for anchoring on the linkage map of sea bass (Chistiakov *et al*. Inpress).
To achieve this task we had to face two main difficulties. The first one, the most challenging one by far, was to prepare enough donor cells. We first thought to use a strategy similar to that used for the construction of the sea bream RH panel that we have done within the frame of a previous European program. Unfortunately this approach did not work. After testing many other strategies we found one that worked perfectly well. It consisted in starting from splenocytes prepared from fresh spleen instead of trying to grow fibroblast cells from dorsal fins. The second difficulty arose with the growing of the hybrid cells. It happened that the hybrid cell clones did not grow well in the selective medium and we

were very rapidly loosing them. This difficulty was circumvented with the help of an in vitro whole genome amplification.
With these two mains modifications to the protocol used with the seabream panel we finally obtained more than 350 hybrid cell lines from which a subset of 93 cell lines were selected through the typing of a number of microsatellite markers. These microsatellites markers were selected from the linkage map produced by Chistiakov et al 2005. Their mapping allowed to select out of the 350 hybrid cell lines the 93 hybrid cells having a high retention value of sea bass DNA while assuring the best coverage of the sea bass genome.

4. Physical map of BAC end sequences
A previously prepared high-quality BAC library (Whitaker *et al.* 2006) was exploited through end-sequencing of all clones. Clones were assembled into contigs and aligned chromosome by chromosome on the fully sequenced genome of the three-spined stickledback (Kuhl *et al.* in prep.). Microsatellite makers used in the linkage map of Chistiakov *et al.* (in press) were cross-mapped on the BAC end sequence map for comparability and ESTs detected in the end sequences were cross-mapped on the Radiation Hybrid map.

5. Comparative mapping
A suite of 186 microsatellite flanking regions of European sea bass were mapped comparatively on the genomes of five model teleosts (*Tetraodon nigroviridis*, *Takifugu rubripes*, *Oryzias latipes*, *Gasterosteus aculeatus* and *Danio rerio*) (Chistiakov *et al*. in press). First, sequences were aligned and BLAST search hits with e values $<10^{-5}$ were considered significant. The aligned region and the flanking sequence (± 1 kb) were then extracted from the genomes, searches for non repeat masked sequences and a Smith-Waterman alignment performed with sea bass microsatellite clones as queries, and microsatellites identified and characterized. Maps were graphically visualized.

Results.

Developing and mapping microsatellite markers and SNPs on the European sea bass linkage map:

Microsatellite markers have been isolated and characterized from ESTs and candidate genes as well as from public sources and agreements with ongoing projects (e.g. EU NoE Marine Genomics Europe). New microsatellites are tested for variability and reliability on 20 outbred wild sea bass. The genotypes of these microsatellites are scored on a well-characterized panel of sea bass including 2 parents and 50 progeny (Venezia Fbis) for mapping purposes.

Microsatellites: About 250 liver ESTs deposited to GenBank in April 2005 have been searched for microsatellites. Ten primer pairs have been designed; 4 of those showed strong homology with known genes and are valuated. Three of these 10 were polymorphic for the Venezia Fbis family (DLA0282e till DLA0284e) and have been submitted to the ResSpecies database (http://www.resspecies.org).
Partners of the EU Network of Excellence Marine Genomics Europe (http://www.marine-genomics-europe.org) developed about 30,000 ESTs from cDNA libraries prepared from 5 individuals progeny of Atlantic wild parents. Fourteen normalised cDNA libraries, corresponding to 14 tissues, have been developed. After aligning and assembling all sequences, they were screened for repeats. 953 clones were found containing repeats. The cDNA libraries were searched for microsatellites with special attention to tri- and di- repeats. From these 118 primer pairs were developed and 76 showed polymorphism for the Venezia panel (DLA0285e till DLA0360e). Sixty six (of which 7 were redundant) showed polymorphism in the Venezia Fbis family and are ready for mapping. All those markers and genotypes have been submitted to the ResSpecies database (See also Annex 1.1).
Additionally 11 loci were selected from the sea bass genomic sequences, based on a comparative approach with three-spined stickleback and using marker DLA0164 as anchor. Nine of these were polymorphic in the family and have been submitted to the ResSpecies database (DLA0402 till DLA0404, DLA0407, DLA0408 and DLA0410 till DLA0413).
Furthermore, gilthead sea bream (*Sparus aurata*) and European sea bass MGE-ESTs datasets were screened for di-, tri-, tetra-, penta and hexanucleotide repeated motifs (Annex 1.2). According to the search criteria adopted, 1604 ESTs that harbor 1748 simple sequence repeats (SSRs) were identified in both species. We used as template the non-redundant EST-SSRs; for sea bream and sea bass, 91.32% (i.e. 821 out of 899) and 73.9% (i.e. 521 out of 705) of the EST-SSRs, respectively. The

chance of finding a SSR-containing sequence in the non redundant EST database was thus 4.51 and 2.94%, with an average density of one microsatellite every 17.2 and 23.7 kb for sea bream and sea bass, respectively.
Primers were ordered for 260 sequences that approximately correspond to all those containing equal or greater than 15, 6, 6, 5 and 5 repeats for di-, tri-, tetra-, penta- and hexanucleotides, respectively. When these primers were PCR checked on wild sea bream individuals but also on sea bass, we observed that 44.2% of the gilthead sea bream primers gave PCR products of expected size also in European sea bass, whereas 39.3% did not amplify.
Using a significance threshold of $e<10^{-5}$, an initial *in silico* analysis (i.e. BLAST searches of sea bream versus sea bass EST-SSRs) had pointed out approximately 60 sequences with good homology (~7.43%). These matches were usually due to blocks of highly conserved sequences, and between the two species they had a length of 25 to 800 bp (average 212 bp) with sequence similarities between 73 and 100% (average 85%).

Mapping on the RH panel.
Using a panel of 93 independent radiation hybrid (RH) cell lines (WP08), more than 1000 markers have been mapped. The markers fall into two categories. The first category is made of 150 polymorphic markers (microsatellites) that have been selected from the first generation linkage map of European sea bass (Chistiakov et al., 2005) through the criterion of being well dispersed on the map. Placing on these two independent maps a common set of markers was one of the objectives of WP09, i.e. to anchor the RH map on the linkage map and produce a unique genomic resource.
Then 915 gene markers were selected by comparison of a number of ESTs retrieved from GenBank as being identified in Medaka (*Oryzias latipes*), Three-spined Stickleback (*Gasterosteus aculeatus*), Tetraodon (*Tetraodon nigroviridis*) and fugu *(Takifugu rubripes*), Zebrafish (*Danio rerio*) and Gilthead Seabream. Pairs of primers for specific sea bass genomic amplification were designed, synthesized and tested for faithful amplification on Sea bass DNA, Hamster DNA and Hybrid cell DNA. Finally 860 of them were typed on the 93 cell line RH panel. Presence, scored as 1, and absence, scored as 0, were recorded, tabulated and used to compute a sea bass RH map with the Travelling Salesman Problem (TSP) approach as specified by the CONCORDE algorithm (Agarwala et al., 2000).
These 860 markers, all with an ortholog precisely located in one to five of the species listed above, will allow a good anchorage of the sea bass genome onto the genomes of the above mentioned species allowing (1) to identify common blocks of synteny and (2) to derive a number of important genomic information from the sequences of Medaka, Tetraodon, Zebrafish and Stickelback that have been recently completed
We have now entered in the process of verification of the RH map and will start very soon to write a paper that will be submitted this summer for publication.

Database maintenance:
Entry of marker genotype data was done to the ResSpecies database via the web continuously, with partners 12 and 13 submitting the data as they were generated in the laboratory. Data are screened upon entry and those containing Mendelian errors are flagged and rejected by the database. This allows submitting scientists to rapidly identify errors in the data and correct these before final submission of the data. The database accepts microsatellite, AFLP and EST data, and has a commercial database management system. It harbours 408 microsatellite markers, 8 PCRs-RFLPs and 226 AFLP markers, contributing to 80846, 344 and 10168 genotypes respectively (as of 31.60.07). Tools are put in place to export the data in several formats, including the appropriate format for linkage analysis using Crimap. Access to data was restricted to the collaborating participants for the course of the project. In May 2007 the Sea Bass database has been added to a new implementation of ResSpecies.

Genotyping:
Microsatellite genotyping of the panel follows a semi-automated procedure of PCR amplification, electrophoresis, scoring and database submission. Data submission is organized and quality controlled through the limited access and purpose built database ResSpecies. SNP genotyping of the panel is well advanced; a first multiplex assay of 13 loci based on the Sequenom technology (http://www.sequenom.com) has been developed and three additional multiplex assays are being set up. Genotyping with the first multiplex assay of a panel of 20 outbred European sea bass was successful.

Genetic linkage map:
Microsatellite and SNP markers will be built into linear maps using linkage analysis and the maximum likelihood software package Crimap, which has proven to be very effective for analysis of marker data from many species including sea bass. The mapping of new markers (SNPs and microsatellites linked to ESTs) on an updated map, including cross-referencing to the Radiation Hybrid map (see WP08) and a BAC end sequence physical map (H. Kuhl and R. Reinhardt, MPI Molecular Genetics, Berlin) is in progress (partners 12, 13 and 16). It was decided at a special coordination meeting (Paris, 31.08.07) to await the preparation of a consolidated second generation linkage map (which most likely will cover tightly all 24 chromosomes) till publication of the BAC end sequence physical map (Kuhl et al., submitted) and the Radiation Hybrid physical map (publication scheduled in Summer 2008).

Comparative mapping:
Comparative mapping of microsatellite flanking regions with five model teleosts (spotted green pufferfish *Tetraodon nigroviridis*, tiger pufferfish or fugu *Takifugu rubripes*, medaka *Oryzias latipes*, three-spined stickleback *Gasterosteus aculeatus* and zebrafish *Danio rerio*), revealed a high percentage (35.5%) of evolutionary conserved regions with the three-spined stickleback. It highlights that functional analyses of European sea bass may have a high potential by combining sea bass genetics and functional information from model species. Our findings contribute to a better understanding of fish evolution and provide an excellent prospect for enhanced selection of aquacultured fish.

Sea Bream

Extension of the BRIDGE-MAP linkage map to up 300 microsatellites.

We used all gilthead sea bream (*Sparus aurata*) and European sea bass (*Dicentrarchus labrax*) ESTs which were unigenes downloaded from the MGE unigene site in February 2006. As of that date, there were a total of 18,196 sea bream unigene clusters containing 29,895 ESTs and a total of 17,716 sea bass unigene clusters containing 29,260 ESTs listed and annotated in the MGE database.

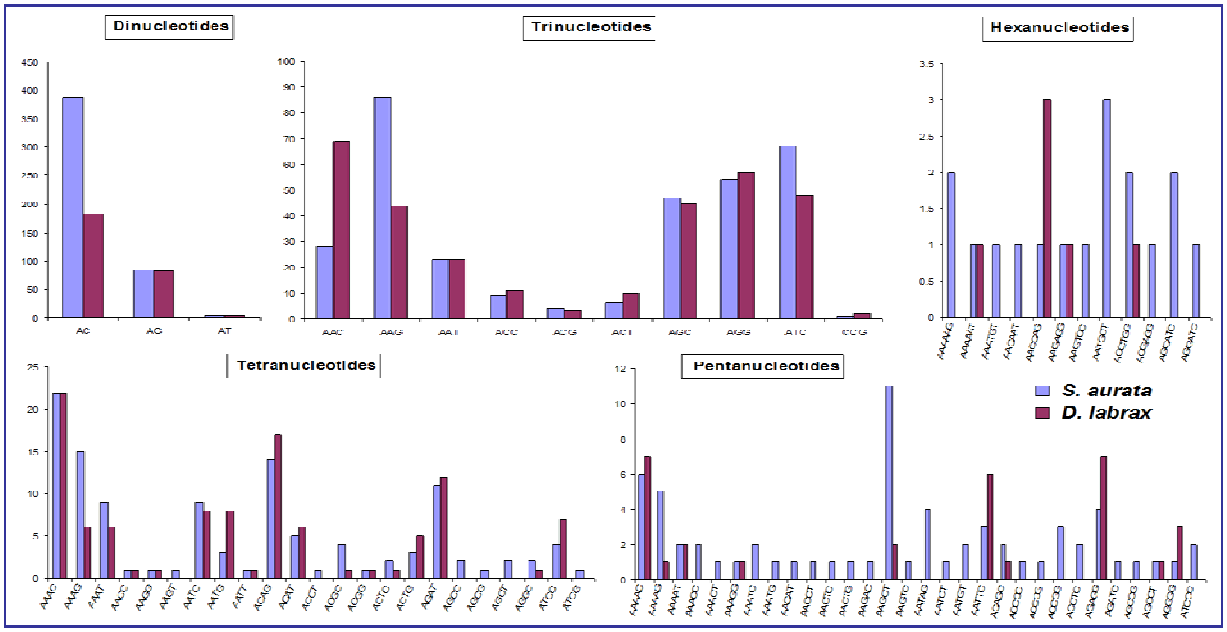
Both unigene databases were used to identify and characterize SSRs using a Perl script based on the algorithm of the MISA script (http://pgrc.ipk-gatersleben.de/misa). Perfect tandem repeats were defined in sequences with a minimum number of repeats of 9 for dinucleotide, 6 for trinucleotide, 5 for tetranucleotide, and 4 for penta- and hexanucleotide repeats. These results were added as a table in the data-set. SSR-ESTs analysis was performed with a set of SQL queries, and tables were automatically generated with Perl scripts; SSR-ESTs were analyzed for redundancy by Blast and/or Cap3 softwares.

Table 1: Detection of SSRs found in sea bream and sea bass ESTs (SSR-ESTs). Number of repeats is considered 9, 6, 5, 4 and 4 for di-, tri, tetra-, penta- and hexanucleotides, respectively. Data for other fish are from Ju Z *et al*. (2005, *In Silico Biology* 5, 0041).

Species	Unigenes	SSR-ESTs			Number of SSRs / motif					Total Nb SSRs
		Nb	%	Annotated	Di-	Tri-	Tetra-	Penta-	Hexa-	
Sea bream	18,196	899	4.94	284 (31.6%)	474 (47.6%)	325 (32.6%)	115 (11.5%)	66 (6.6%)	17 (1.7%)	997
Sea bass	17,716	705	3.97	321 (45.5%)	267 (37.6%)	312 (41.5%)	109 (14.5%)	47 (6.3%)	16 (2.1%)	751
Zebrafish	24,003	1,749	7.3	-	1,497 (64%)	579 (25%)	202 (8.6%)	55 (2.4%)	-	2,333
Medaka	8,158	209	2.6	-	105 (47%)	81 (36%)	27 (12%)	10 (4.5%)	-	223
Fundulus	16,726	369	2.2	-	237 (52%)	155 (34%)	43 (9.5%)	18 (4.0%)	-	453

Totally, we found 1,604 ESTs that harbor 1,748 SSRs in both species. The frequency of SSR-ESTs ranged from 3.98% for sea bass to 4.94% for sea bream and is in the range reported for model fish species which ranged from 1.5% in *Xiphophorus* to 7.3% in zebrafish [2.6% in medaka and 2.2% in *Fundulus*. In both species, the percentage of the annotated ESTs ranges from 31.6% in the sea bream to 45.5% in the sea bass.

Figure 1: Distribution of SSR motifs in gilthead sea bream and European sea bass



The dinucleotide repeat motifs are the most abundant SSRs in sea bream accounting for 47.5% of the SSRs [47%, 52%, and 64% for medaka, *Fundulus*, and zebrafish, respectively (Ju et al. 2005), followed by the tri-, tetra-, and penta- and hexanucleotide SSRs. Surprisingly, in the sea bass the most abundant repeats are the trinucleotides.

Among the dinucleotide motifs, AC/TG was the most abundant type in both species (Fig. 1) [also in all species in (Ju et al. 2005), except *Fundulus* in which AT/TA was the most abundant motif]. No CG/GC motif was found in sea bass and sea bream ESTs.

Ten only trinucleotide motifs were observed in sea bass and sea bream; the AAG and AAC motifs were the most abundant for sea bream and sea bass, respectively, whereas the other eight were approximately equal (Fig. 1).

We also evaluated SSR-ESTs distribution in the fourteen tissues used per species for cDNA library generation (Tab.2). Even though, in absolute numbers, SSR-ESTs are fewer in sea bass when compared to sea bream, in the former species 58% of the SSR-ESTs are tissue specific; this percentage is only 46% for sea bream.

Table 2: Total number, type (di-, tri-, etc) and number of tissue specific SSR-ESTs (in light blue) per tissue for sea bream (SA) and sea bass (DL).

Tissue	Total						Di-				Tri-				Tetra-				Penta-				Hexa-			
	SA		%	DL		%	SA		DL		SA		DL		SA		DL		SA		DL		SA		DL	
Liver	99	48	48.4	94	50	53.2	45	20	48	30	39	17	25	8	8	4	14	10	5	1	4	1	2	1	2	1
Ovary	81	37	45.7	80	54	67.5	33	15	20	13	34	13	45	29	8	3	9	9	4	4	2	1	2	1	3	2
Testis	92	48	52.2	92	55	59.8	36	18	28	18	37	19	40	24	9	5	12	6	6	5	8	6	4	1	4	1
Bone_Cartilage	122	58	47.5	84	48	57.1	61	29	26	15	40	22	43	25	9	5	12	7	6	1	2	0	3	1	1	0
Brain_Pit	123	68	55.3	94	39	60.9	58	36	29	15	45	29	21	16	10	6	10	6	6	3	2	2	4	1	2	0
Heart_Vessels	114	46	40.4	79	46	58.2	57	25	33	21	36	19	25	12	10	7	13	8	9	3	7	4	2	1	1	1
Adipose	101	44	43.6	86	48	55.8	39	17	27	15	39	12	41	26	15	11	12	5	6	3	4	2	2	1	2	0
Head_Kidney	124	54	43.5	62	33	53.2	64	34	19	8	40	13	30	17	13	5	9	6	5	1	4	2	2	1	0	0
Trunk_Kidney	96	35	36.5	81	40	49.4	40	15	26	12	35	8	32	17	10	5	13	8	9	7	6	3	2	0	4	0
Gill	105	45	42.5	81	47	58	46	23	33	20	41	12	25	14	8	4	10	6	9	5	7	3	2	1	5	4
Intestine	127	63	49.6	71	41	57.7	54	27	23	13	47	19	31	15	17	10	9	6	8	7	8	6	1	0	0	0
Spleen	123	68	55.7	80	45	56.3	51	33	26	14	45	19	37	20	14	8	11	7	9	5	5	4	4	1	1	0
Muscle	127	58	45.7	58	33	56.9	66	36	19	11	44	15	24	13	10	5	7	5	6	2	6	4	1	0	2	0
Skin	96	45	46.9	27	17	63	41	22	9	5	32	7	15	9	14	11	2	2	8	4	1	1	1	1	0	0

We used as template the non-redundant SSR-ESTs; for sea bream and sea bass, 91.32% (*i.e.* 821 out of 899) and 73.9% (*i.e.* 521 out of 705) of the SSR-ESTs, respectively, were non-redundant.

As compound microsatellites were considered those present in the same EST and distant by a maximum of 25bp. SSR-ESTs were used as target for the Primer3 software with a) primer length from 18 to 27 bases, b) primer's Tm ranging from 55 to 63°C, and c) product size in range of 100 to 250bp. The primers output of Primer3 was further analyzed in order to avoid or minimize the occurrence of tandem repeats in the primer sequence, the self and pair complementarity. This procedure was automated with a Perl script.

For the sea bream, there were finally 664 pairs of primers accepted (73.85% of the initial 899 SSR-ESTs). Primers were ordered for 260 sequences that approximately correspond to all those containing equal or greater than 15, 6, 6, 5 and 5 repeats for di-, tri, tetra-, penta- and hexanucleotides, respectively. When these primers were PCR checked on wild sea bream individuals, we saw that 78.1% of primer pairs gave PCR products of expected size in sea bream, 8.5% gave much bigger or non-unique products, and 13.4% did not amplify at all. Preliminary genotyping analysis with approximately 20 fluorescent primers, mainly for dinucleotides, indicated that more than half of them are polymorphic.

More specifically, data on 260 non-redundant EST-SSRs defined on sea bream's MGE database show that:

- 203 have a PCR band of expected size (78.08%),
- 22 gave a product of much different size (8.46%),
- 35 gave no product (13.46%)

Table 3: Type (di-, tri-, etc) and number of EST-SSRs found in sea bream, for which there was primer design and positive PCR results.

	Type of repeats in SSRs					
	Composite	2-nucleotides	3-nucleotides	4-nucleotides	5-nucleotides	6-nucleotides
Total	2	61	94	35	7	4
Annotated*	2	18	28	8	1	1

* In most cases annotation is like '*Tetraodon nigroviridis* (green puffer), chromosome undetermined scaffold XX, whole genome shotgun sequence...', i.e. those ESTs show significant Blast hit with sequences in other model fish but their function is yet unknown.

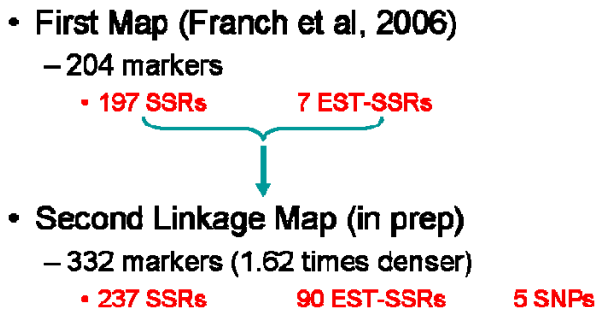
Additionally, in UNIPD (partner 14) seventy (70) new clones from a microsatellite-enriched genomic library were sequenced, and after sequence analysis 45 clone sequences contained a repeat greater than 18bp and were suitable for primer design. Those clones were used for "cold" primer design and 36 give good PCR products. When 36 fluorescent primers were ordered and synthesized for the above clones there were finally 22 informative loci that were informative and scored.

The first linkage map reported in gilthead sea bream by Franch *et al.* (2006, Genetics 174: 851-861) included 204 markers, of which 6 markers were not linked to any linkage group (LG) and 7 were linked but not surely positioned in some of the LGs; in total, 26 linkage groups were calculated with a map length of 1241.9 cM.

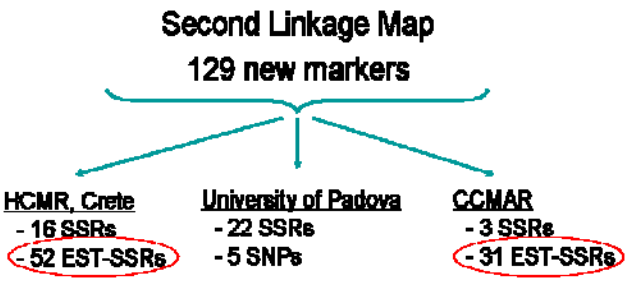
The outcome of this deliverable is the **second generation linkage map which will finally include 332 markers (Figure 3a)**; of these, 129 markers are new and 19 were reported in the 2nd year and 110 in the last year of the project (Figure 3b). [We finally took into account 203 markers from the Franch *et al.* (2006) map because marker SAGT41a was excluded since it was part of the same sequenced clone with marker SAGT41b]. In contrast to the first linkage map which mainly consisted of 'normal' microsatellites (SSRs) and only few EST-SSRs, the new version of the map is mostly enriched in EST-SSRs (Figure 3a) which will greatly facilitate comparative mapping with other teleosts and model organisms, since ESTs in general are markers that have 'sequence analogs' in other organisms.

Figure 3:

(a)



(b)

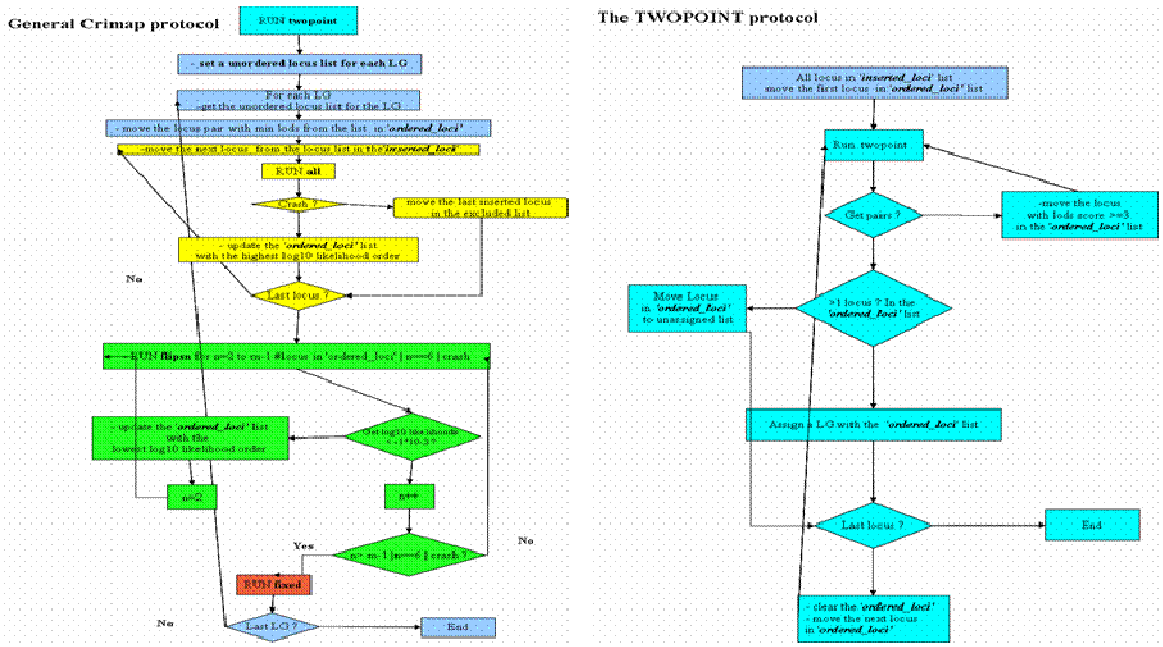


Map Development

For linkage analysis, the same HCMR's family was genotyped using the newly developed microsatellite and SNP markers. Using the CRI-MAP software v3.0, the analysis was performed as described by Franch et al. (2005). We did not use the previously reported microsatellite-based linkage map of Franch et al. (2005) as a framework to incorporate new markers and the analysis was done from the beginning. Sex-average and sex-specific linkage distances were estimated for each linkage group assuming the Kosambi mapping function (Kosambi 1944).

At this point, we have to mention that apart the ordinary use of the Crimap software, which demands many runs in order to achieve the best possible estimates, and therefore its use is not trivial, the HCMR's bioinformatics team developed a series of Perl scripts to render the above mentioned software more user-friendly, and flexible and able to produce a series of linkage maps using different criteria in short time intervals (see Fig. 2). The results of the newly developed scripts were always tested against results obtained using conventional use of the software.

Figure 4: Bioinformatic set-up of scripts for Linkage mapping using CriMap v.3.0



Linkage mapping resulted in a new version of the linkage map consisting of 332 genetic markers: 237 'normal' and anonymous microsatellites, 90 microsatellites located in expressed sequences (EST-SSRs) and 5 SNPs in 27 linkage groups. Ten markers (4 SSRs, 5 EST-SSRs and 1 SNP) remained unlinked: EId-10, Ct40, Saimbb16, G6, cDN03P05K21, CL428Cng1, CDN01P01G08, cDN10P06I02, CL848Cng1 and FIA07. The first two markers were also not mapped in the previous map version.

The total length of the sex-averaged map is 1769.7 cM (Fig. 3) and the number of markers in each LG ranged from 2 to 35 (LG9). Among linkage groups, LG9 is the longest having a size of 349.1 cM. In the map, the intermarker distance varies from 0 to 48.8 cM, with an average of 5.75 cM; 65.5% of the intermarker intervals are in the range 0 to 5 cM, 15.7% range from 5 to 10 cM, 12.8% range from 10 to 20 cM, and only 6.0% are larger then 20 cM.

Structural differences with the first generation Linkage Map

Two LGs are the result of a fusion of LGs of the first generation map. The LG9 comprises the old LG9 and parts of the old linkage groups LG12 and LG23. The new LG12 is made of the remaining part of LG12, the one not fused with the new LG9, and the old LG13. The new LG21 is having the old LG21 and part of LG14. There are also 3 new LGs (13, 23 and 27), having however only few linked markers (2 to 4).

We have also built sex-specific linkage maps (see annexe 1, part 4, sea bream section). The male map has a length of 1349.2 cM, an average distance between markers of 4.38 cM; the largest marker interval was observed in LG9, in which between markers CL2032 and CId-35 a distance of 100cM was observed. The length of the female map is 2172.1 cM, with an average intermarker distance of 7.05 cM; again in few linkage groups very large distances were observed, e.g. 100cM between Hd-33 and SauK140 in LG4, CId-35 and SaI-14 in LG9, cDN11 and cDN05 in LG12 etc. (Nevertheless, during map construction distances larger than 100cM were never allowed in both sexes since it is an indication for a linkage group splitting). The sex ratio of map lengths is thus 1.61-fold higher in the female compared to the male map. Most linkage groups are smaller in males than in females, with some exceptions, such as LG5, LG13, LG15 and LG26.

Extension of the BRIDGE-MAP RH up to 2000 markers (500 new) and a merged linkage & RH map reflecting recombination rates across the genome

In total 503 markers were genotyped on the RH panel constructed by Senger et al., 2005. Markers were designed based on sequences coming from four different approaches: i) expression study of different developmental stages and expression study of cortisol induced stress (Sarropoulou et al., 2005) ii) comparative study of seabream to medaka linkage map (Naruse et al., 2004) iii) comparative study of seabream to seabass based on homology found between Tetraodon and the seabass linkage map (Chistiakov et al. 2005) and finally based on sequences encoding for genes involved in the immune response.

All primer pairs were tested on genomic DNA of seabream prior to mapping on the RH panel and PCR products were observed for each designed primer pair. 289 (~58%) were successfully amplified. Markers for which no amplification were possible were at ~28 % and markers for which amplification were observed in every hybrid clone were observed at ~6% of the markers tested. Markers for which only seabream DNA gave a PCR product were found for 8% of the markers. Finally 141 (28%) markers were successfully mapped at strict conditions of map calculation. Example of four RH groups shown in fig.3.

One cR_{3000} on the seabream RH map corresponds approximately to 140 kb (Senger *et al*., 2006).

- BAC sequencing
- Comparative approach for fine mapping candidate regions

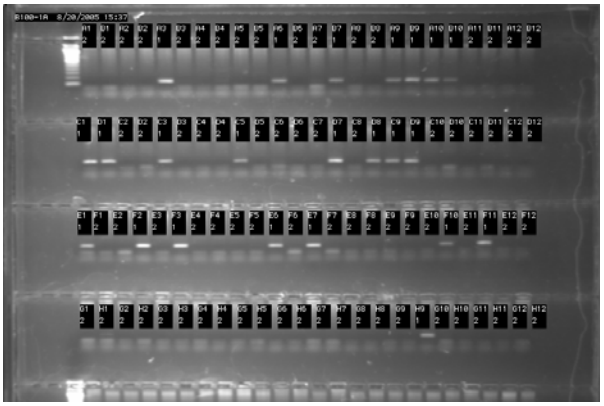


Fig.6a Successfully amplified marker

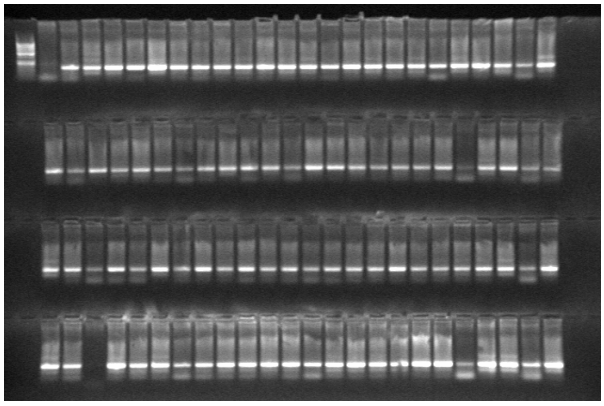


Fig.6b: Markers for which amplification were observed in all hybrid clones

40 genes) coming from the expression studies of seabream within AquaFirst were compared to the genes mapped on the RH panel (Table 5). Out of the already mapped genes related to stress and immune response only 3 were included in the list of differential expressed genes coming from functional analysis of seabream within AquaFirst (in the above table, in yellow are highlighted genes, among the co-expression data, that had already been mapped on the RH (3/40).

The additional relatively small scale mapping effort targeting 500 genes, demonstrates the need in the future, not for randomly increasing the density of the map in order to acgiev better coverage, but the need for targeting mapping. This approach is also becoming the only option, given the high potential of comparative mapping, presented in D10.03, and given ongoing BAC-end sequencing and partial full genome sequencing of seabream.
As intrachromosal rearrangements are expected to be found even between closely related species, although comparative mapping will be very helpful, confirmation at a fine genome scale will demand targeted mapping. BAC-end sequencing or/and low coverage sequencing will be the best way, but such an approach wll grately profit from RH mapping, in order to achieve closing some gaps in assembled maps.

Hereafter are given some examples of the interest of combining RH and linkage map.

Table 5. Fourty markers from seabream differentially co-expressed genes. Among them, 3/40 had been previously, indepdently mapped

Clone Name	Gene Name	RH-Position
cdn01p0002_K12	SAL_SILAS (Q9PVW8) Rhamnose-binding lectin precursor (SAL) (RBL) (Roe lectin)	
cdn02p0005_L16	CATION TRANSPORT REGULATOR TYPE-1	
cdn04p0002_P22	FA96B_MOUSE (Q9D187) Protein FAM96B	
cdn05p0002_P2	PPIA_BLAGE (P54985) Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8) (PPIase) (Rotamase) (Cyclophilin) (Cyclosporin A-binding protein)	
cdn05p0005_N10	?	
cdn07p0003_E6	PTPS_POERE (Q90W95) 6-pyruvoyl tetrahydrobiopterin synthase (EC 4.2.3.12) (PTPS) (PTP synthase)	
cdn07p0006_E24	CHAC1_BRARE (Q5SPB6) Cation transport regulator-like protein 1	
cdn08p0002_P12	?	
cdn09p0002_E8	RAB1A_RAT (Q6NYB7) Ras-related protein Rab-1A	13
cdn09p0004_P13	?	
cdn11p0003_C13	?	
cdn11p0003_I11	cdn11p0003i11	
cdn11p0006_H11	CIDEC_HUMAN (Q96AQ7) Cell death activator CIDE-3 (Cell death-inducing DFFA-like effector protein C) (Fat-specific protein FSP27 homolog)	
cdn13p0002_M17	?	
cdn14p0003_D3	APOEB_BRARE (O42364) Apolipoprotein Eb precursor (Apo-Eb)	14
gsgi08b06_F22	?	
gsgi08b25_C10	HVCM5_MOUSE (P84751) Ig heavy chain Mem5 (Fragment)	
gsgl01b15_D11	LECT2_MOUSE (O88803) Leukocyte cell-derived chemotaxin 2 precursor (Chondromodulin II) (ChM-II)	
gsgl01b15_D13	?	
gsgl01b15_D15	HMCS1_CHICK (P23228) Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase)	
gsgl02b16_B9	?	
gsgl03b02_A18	UNKNOWN EST	
gsgl03b02_A23	?	
gsgl03b02_C14	?	
gsgl03b02_E17	?	
gsgl03b02_E18	HEPC_MORCS (P82951) Hepcidin precursor	
gsgl03b02_G17	RIB1_HUMAN (P04843) Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 67 kDa subunit precursor (EC 2.4.1.119) (Ribophorin I) (RPN-I)	
gsgl03b02_G21	CGL_HUMAN (P32929) Cystathionine gamma-lyase (EC 4.4.1.1) (Gamma-cystathionase)	
gsgl03b02_G22	CFAB_MOUSE (P04186) Complement factor B precursor (EC 3.4.21.47) (C3/C5 convertase) [Contains: Complement factor B Ba fragment; Complement factor B Bb fragment]	
gsgl03b02_K19	CO7_PIG (Q9TUQ3) Complement component C7 precursor	3
gsgl03b02_M19	?	
gsgl03b17_G13	HEPC_MORCS (P82951) Hepcidin precursor	
gsgy01b19_N16	LECT2_BOVIN (O62644) Leukocyte cell-derived chemotaxin 2 precursor (bLECT2) (Chondromodulin II) (bChM-II)	
gsgy03b04_M5	HEPC_MORCS (P82951) Hepcidin precursor	
gsgy07b22_L23	?	
gsgy07b26_M18	?	
gsgy08b05_D9	?	
gsgy08b23_D18	XYLB_RAT (Q3MIF4) Xylulose kinase (EC 2.7.1.17) (Xylulokinase)	
gsgy08b23_J14	HYPOTHETICAL PROTEIN	
gsub01b02_I17	?	

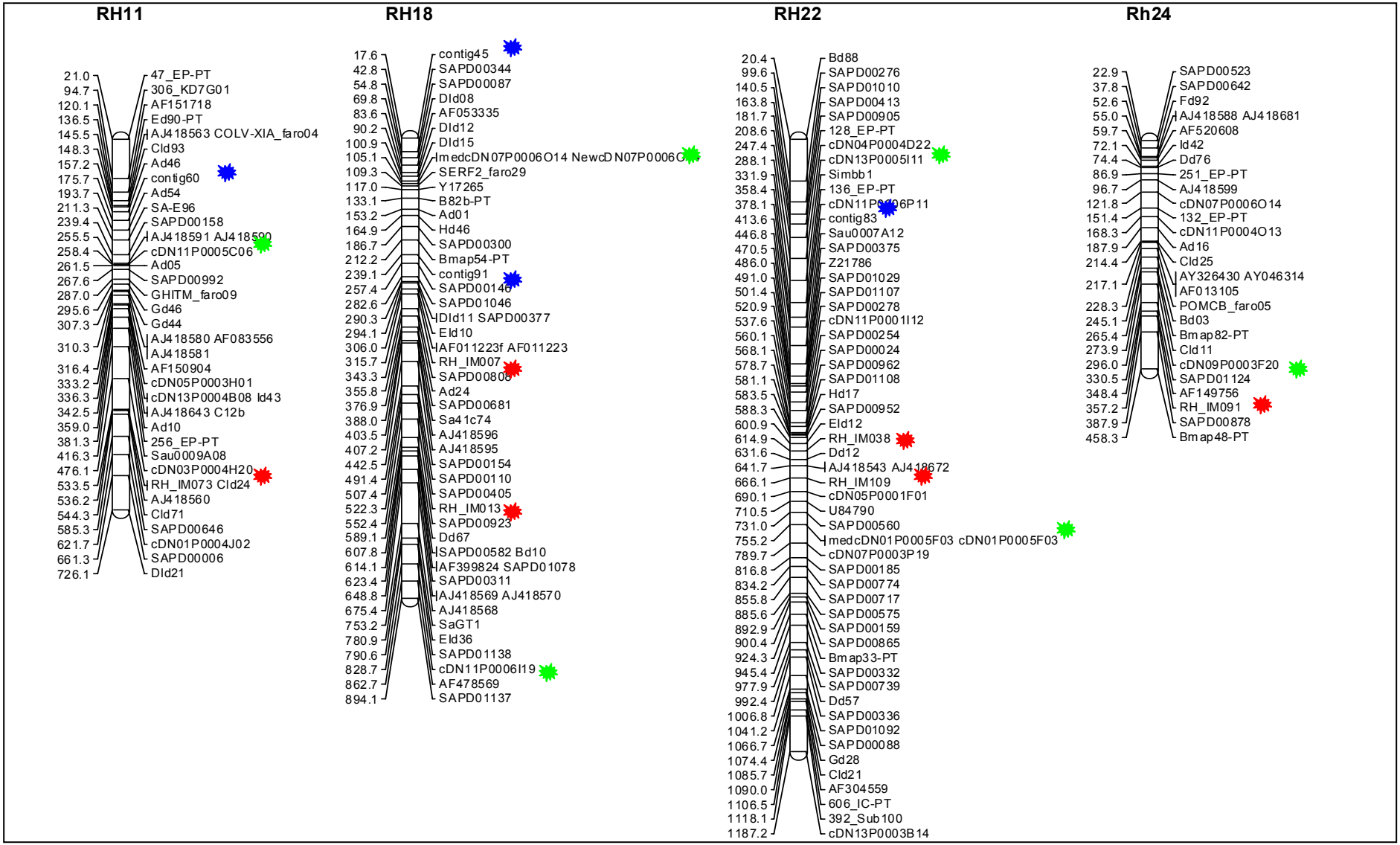


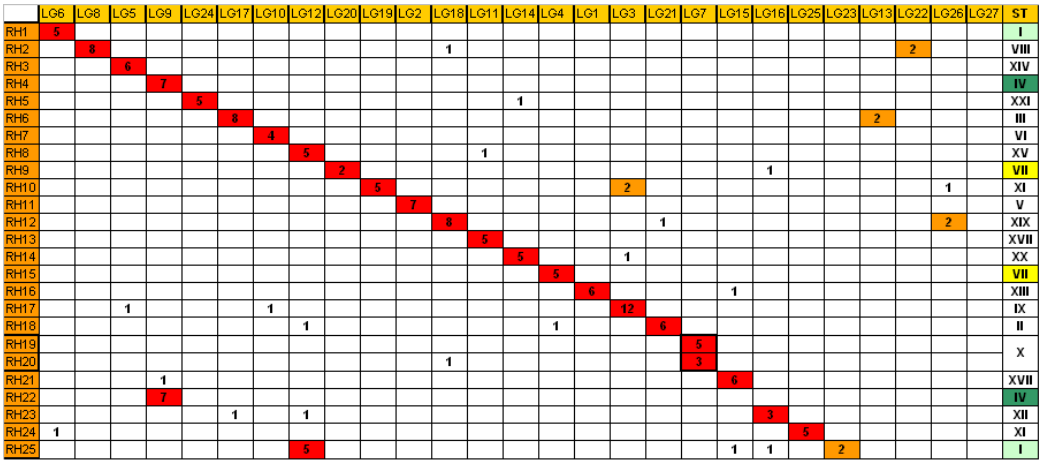
Fig.7: exemples of some RH groups (among the smallest)

- Primers designed based on differentially expressed genes (sets of co-expressed genes).
- Primers designed on seabream ESTs (MGE, NCBI, BRIDGEMAP) among the ones that showed high homology to genes mapped in medaka by linkage (Naruse et al. 2004**)**
- Primers designed based on sequences of immune related genes (NCBI & project WEALTH).

Merged linkage & RH map reflecting recombination rates across the genome.

The comparison with the RH map provides independent support for the observed LGs, as concordance is observed between RH groups and LGs (Fig.8). Karyotype analysis of *S. auratus* indicates a haploid set of 24 chromosomes (Cataudella *et al*. 1980; Vitturi *et al*. 1992).

Fig. 8. Linkage group 9 and linkage 12 of the genetic map split into two groups regarding to the radiation hybrid group. Looking at the comparative map LG9 (RH4 and22) is also splitting into two groups in stickleback but shows a 1:1 relationship in medaka (Med23 and Med10 respectively) as well as in tetraodon. Concerning Linkage group 12 splitting into RH 8 and 25 a 1:1 relationship is observed in stickleback as well as in medaka and tetraodon.



Comparing gene order between genetic linkage map and radiation hybrid map concordance is not given (fig. 9,10,11). However the order of markers can be changed in the genetic linkage group by only a minor difference in the lod score. The RH groups in fig. 9, 10 and 11 are symbolized by a vertical bar. The position of each marker is reported along the RH group and symbolized by a horizontal bar. The five RH maps automatically delivered by TSP/CONCORDE are compared. When a marker is present on all five maps at the same position, the horizontal bar has a maximum length, indicating high confidence. When a marker is present at different positions, the horizontal bar is shortened, reflecting a lower confidence level. Numbers in parentheses correspond to the markers as they appear in the consensus map. Asterisks are indicating a possible different location of labeled marker. Distances between RH markers are reported in cR3000. However as show in fig. 11 comparison of genetic linkage map to radiation hybrid reveal regions where resolution is insufficiently resolved in the genetic map and allows identification of regions with no recombination.

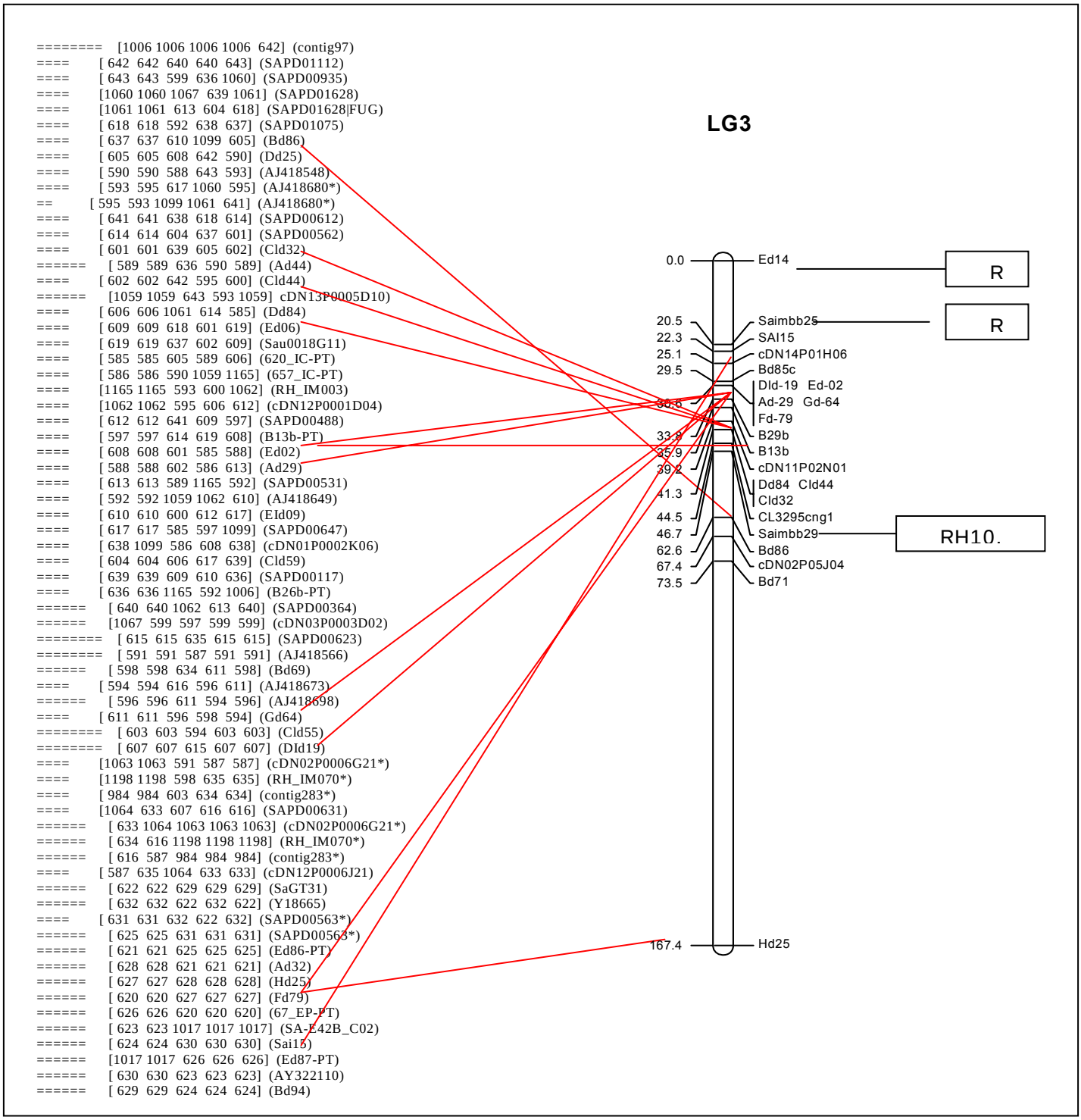

LG3
0.0
Ed14
20.5
Saimbb25
22.3
SAI15
25.1
cDN14P01H06
29.5
Bd85c
Dld-19 Ed-02
Ad-29 Gd-64
Fd-79
B29b
B13b
cDN11P02N01
Dd84 Cld44
Cld32
CL3295cng1
Saimbb29
Bd86
cDN02P05J04
Bd71
167.4
Hd25
R
R
RH10.
[1006 1006 1006 1006 642] (contig97)
[642 642 640 640 643] (SAPD01112)
[643 643 599 636 1060] (SAPD00935)
[1060 1060 1067 639 1061] (SAPD01628)
[1061 1061 613 604 618] (SAPD01628|FUG)
[618 618 592 638 637] (SAPD01075)
[637 637 610 1099 605] (Bd86)
[605 605 608 642 590] (Dd25)
[590 590 588 643 593] (AJ418548)
[593 595 617 1060 595] (AJ418680*)
[595 593 1099 1061 641] (AJ418680*)
[641 641 638 618 614] (SAPD00612)
[614 614 604 637 601] (SAPD00562)
[601 601 639 605 602] (Cld32)
[589 589 636 590 589] (Ad44)
[602 602 642 595 600] (Cld44)
[1059 1059 643 593 1059] cDN13P0005D10)
[606 606 1061 614 585] (Dd84)
[609 609 618 601 619] (Ed06)
[619 619 637 602 609] (Sau0018G11)
[585 585 605 589 606] (620_IC-PT)
[586 586 590 1059 1165] (657_IC-PT)
[1165 1165 593 600 1062] (RH_IM003)
[1062 1062 595 606 612] (cDN12P0001D04)
[612 612 641 609 597] (SAPD00488)
[597 597 614 619 608] (B13b-PT)
[608 608 601 585 588] (Ed02)
[588 588 602 586 613] (Ad29)
[613 613 589 1165 592] (SAPD00531)
[592 592 1059 1062 610] (AJ418649)
[610 610 600 612 617] (EId09)
[617 617 585 597 1099] (SAPD00647)
[638 1099 586 608 638] (cDN01P0002K06)
[604 604 606 617 639] (Cld59)
[639 639 609 610 636] (SAPD00117)
[636 636 1165 592 1006] (B26b-PT)
[640 640 1062 613 640] (SAPD00364)
[1067 599 597 599 599] (cDN03P0003D02)
[615 615 635 615 615] (SAPD00623)
[591 591 587 591 591] (AJ418566)
[598 598 634 611 598] (Bd69)
[594 594 616 596 611] (AJ418673)
[596 596 611 594 596] (AJ418698)
[611 611 596 598 594] (Gd64)
[603 603 594 603 603] (Cld55)
[607 607 615 607 607] (DId19)
[1063 1063 591 587 587] (cDN02P0006G21*)
[1198 1198 598 635 635] (RH_IM070*)
[984 984 603 634 634] (contig283*)
[1064 633 607 616 616] (SAPD00631)
[633 1064 1063 1063 1063] (cDN02P0006G21*)
[634 616 1198 1198 1198] (RH_IM070*)
[616 587 984 984 984] (contig283*)
[587 635 1064 633 633] (cDN12P0006J21)
[622 622 629 629 629] (SaGT31)
[632 632 622 632 622] (Y18665)
[631 631 632 622 632] (SAPD00563*)
[625 625 631 631 631] (SAPD00563*)
[621 621 625 625 625] (Ed86-PT)
[628 628 621 621 621] (Ad32)
[627 627 628 628 628] (Hd25)
[620 620 627 627 627] (Fd79)
[626 626 620 620 620] (67_EP-PT)
[623 623 1017 1017 1017] (SA-E42B_C02)
[624 624 630 630 630] (Sai15)
[1017 1017 626 626 626] (Ed87-PT)
[630 630 623 623 623] (AY322110)
[629 629 624 624 624] (Bd94)

Fig.9 RH vs. LG

(ST5)

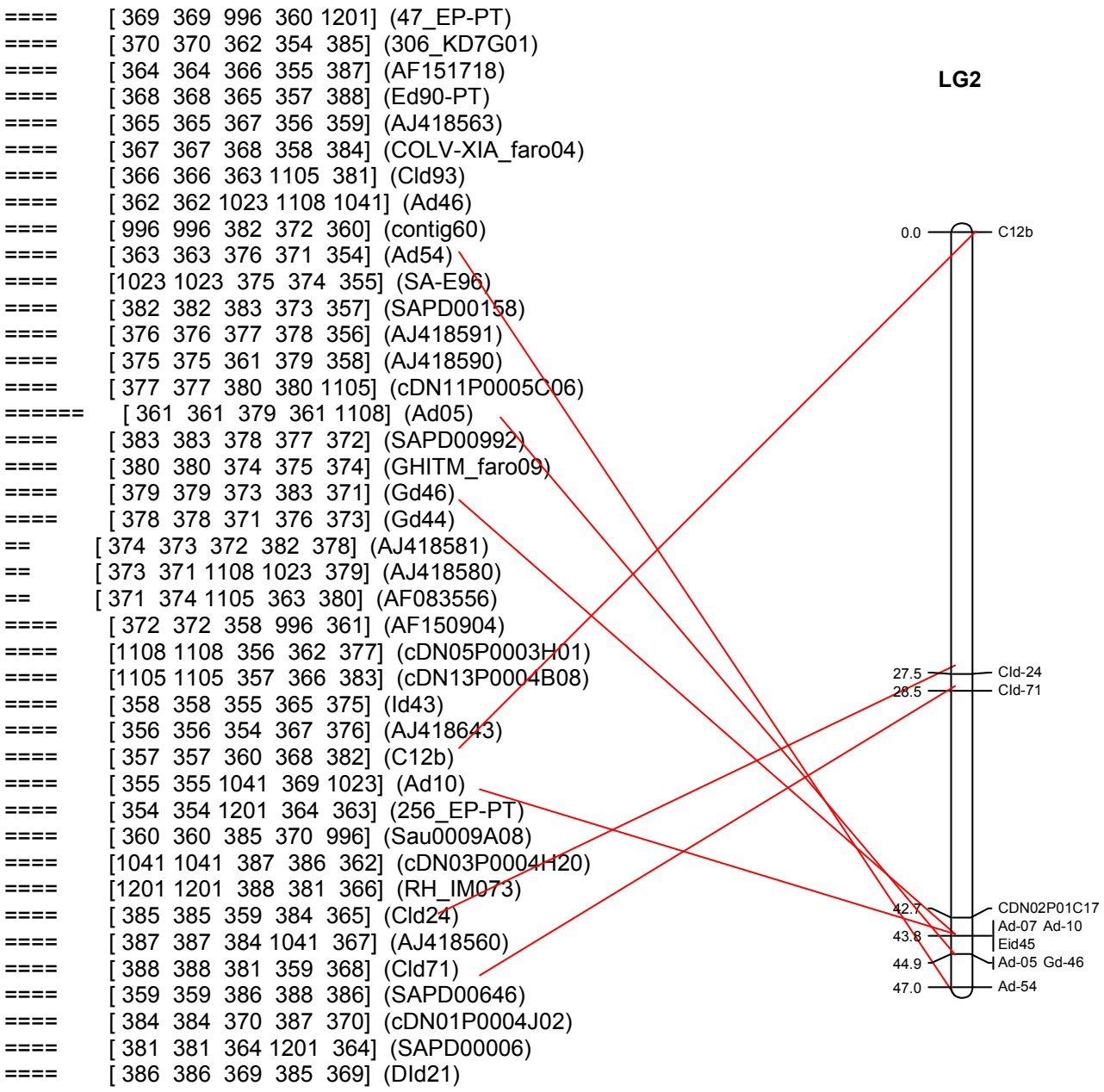


Fig. 10 RH vs. LG

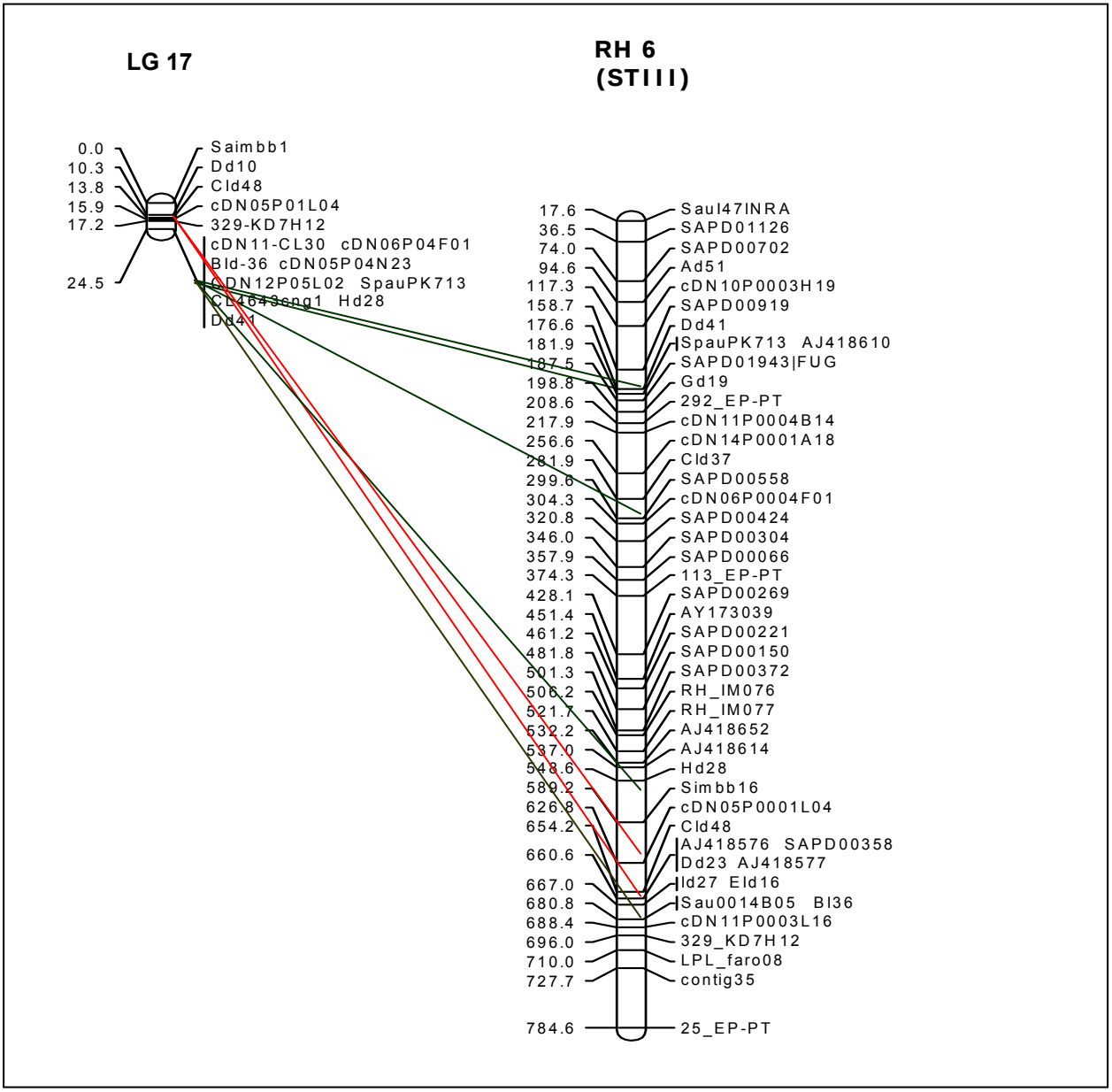
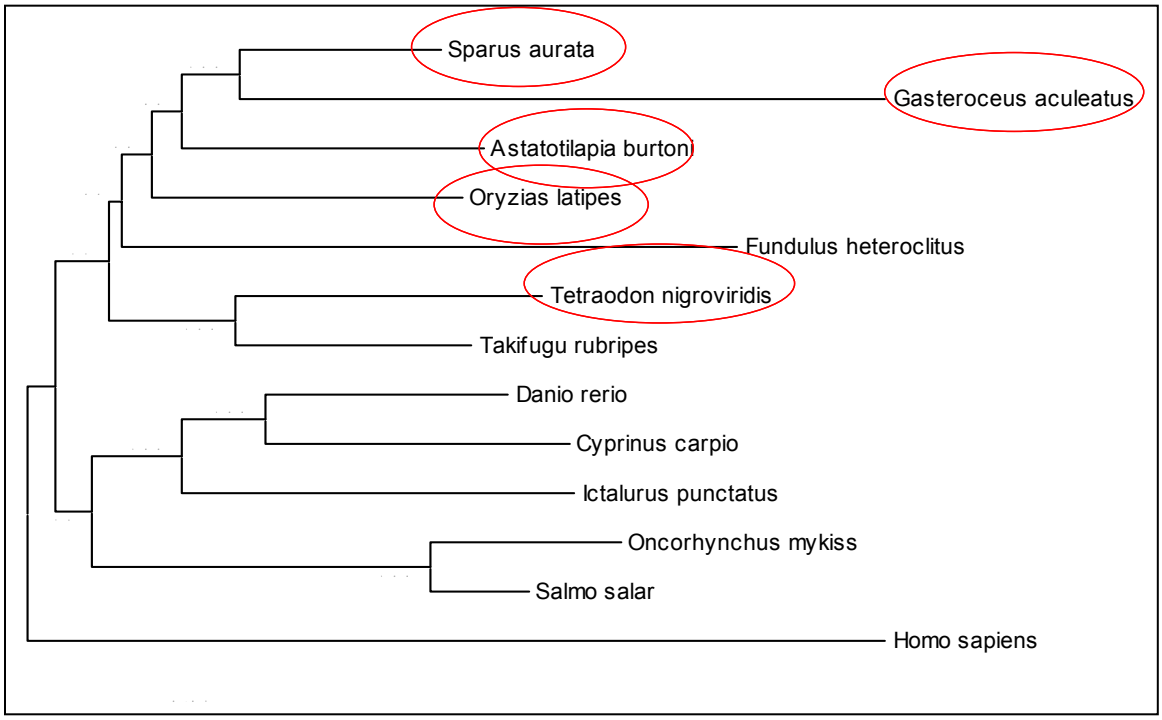


Fig.11 RH vs LG

Comparative genome maps between sea bream, sea bass, model fish species and human

Among teleost fish so far fully sequenced genomes are available for five species: zebrafish (*Danio rerio,* Cypriniformes), fugu (*Fugu rubripes*, Tetraodontiformes), Tetraodon (*Tetraodon nigroviridis*, Tetraodontiformes), medaka (*Oryzias latipes,* Beloniformes*)* and three-spined stickleback (*Gasterosteus aculeatus*, Gasterosteiformes).

Figure 12. Phylogeny based on a combined dataset of 22 genes modified after {Steinke, 2006 #113}. Maximum parsimony (MP) analyses of the combined amino acid alignement were performed with MEGA version 2.1 {Kumar, 2001 #114}. Neighbor-Joining tree was calculated with bootstrap test of phylogeny (1025 replicates) Mega Version 2.1 and PAUP Version 4 .0b10.



All of them are model fish species mostly used for addressing questions related to gene function or, in the case of three-spined stickleback also for addressing questions related to adaptation, speciation and evolution (e.g. Bell et al., 2001; Cresko et al. 2007). Recently, relatively rapid development of genomic resources has been made for economically important aquacultured fish species such as the channel catfish, *Ictalurus punctatus,* the tilapia of the genus *Oreochromis, such as Oreochromis niloticus*, the rainbow trout, *Oncorhynchus mykiss*, Atlantic salmon, *Salmo salar* in Northern Europe and North America, and the gilthead seabream (*Sparus aurata*) and the European seabass (*Dicentrarchus labrax*) in the Mediterranean and Northeast Atlantic.

These latest developments in genome resources of aquaculture fish species, although in their infancy, allow a comparative genome analysis of model fish with that of commercial fish species. Consequently, these efforts will bring valuable improvements to their genomics and functional genomics.

Here we studied the conservation of syntenies flanked by three Perciformes, the gilthead seabream *Sparus aurata,* representing the species rich family of Sparidae, the European seabass *Dicentrarchus labrax,* and the cichlid *Oreochromis niloticus*. The former two species are important species for the Mediterranean mariculture and could be considered as cultured model species, since they were extensively studied in the areas of physiology, behavior, reproduction and phenotypic diversity. Tilapia is an important species for developing countries around the tropics with ecological and evolutionary interest, due to its uniquely high adaptive radiation (Kocher et al. 1998). Comparative analysis in the present study is based on 800 sequences successfully mapped on the gilthead seabream radiation hybrid (RH) map (Sarropoulou et al., 2007). They were *in silico* mapped to the three fully sequenced genomes of Tetraodon (Sarropoulou et al. 2007), medaka and three-spined stickleback.

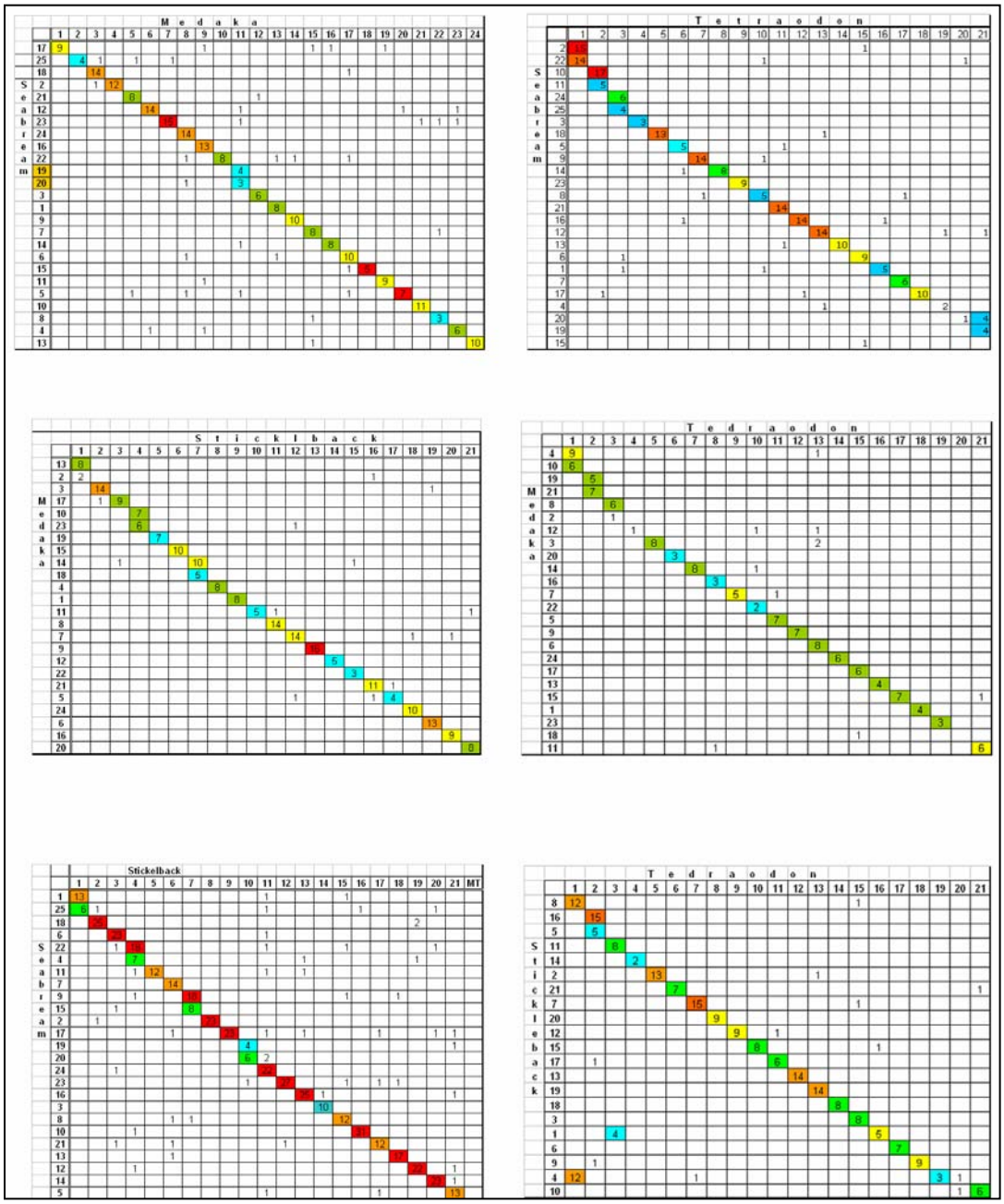


Figure 13. Syntenic relationship illustrated by six oxford grids between five teleost species constructed with 800 molecular markers mapped on the radiation hybrid panel of the gilthead seabream *Sparus aurata.* Number in each square is the number of matching genes between the two species. Colors refer to found matches: white 0-2, blue 3-5, green 6-8, yellow 9-11, orange 12-14, red >15.

We show that with a few markers reliable syntenic relationships can be built. Furthermore we demonstrate syntenic relationships between gilthead seabream and seabass as well as between gilthead seabream and tilapia Oreochromis spp. via a third or even a fourth species.

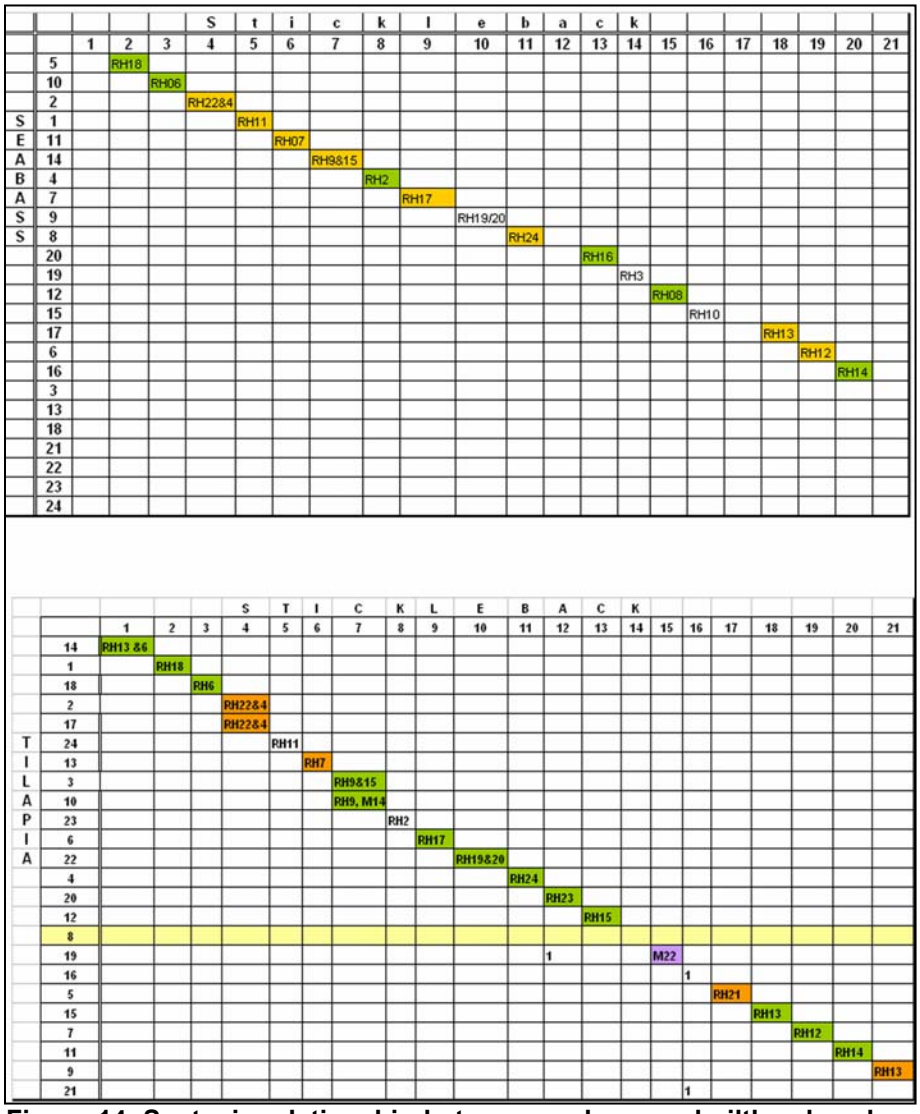


Figure 14. Syntenic relationship between seabass and gilthead seabream and between tilapia and gilthead seabream via three-spined stickleback is illustrated here by a three-dimensional Oxford grid format. X-axis: chromosomes of three-spined stickleback. Y-axis: chromosomes of seabass. Indication in each rectangle refers to the corresponding RH group in gilthead seabream.

Whenever sufficient genomic information for direct comparison is not available using a third of even a fourth species as stepping stone gives the possibility to link two species with poor genome information. The association between the three Perciformes gilthead seabream, seabass and tilapia is of great interest for economically aspect mainly in terms of QTL detection. In addition to the significance for the aquaculture industry, this study can encompass important ecological and evolutionary implications as it broadens the range species for studying chromosome evolution in fish by adding phylogenetically related non-model fish species.

More sequences mapped in the gilthead seabream and in seabass are required in order to obtain a better picture between the Perciformes as well as between the more distant species such as for example the catfish, and economically important fish species and the well-known model fish, the zebrafish. Recently, Xu et al., 2007 have shown that the comparison of catfish with Tetraodon produced only 7% of homologues hits, which has as a result that more sequence information or mapping information is needed to assess syntenic conservation between more distant fish species. Obviously comparison to human was not feasible at this stage. However BAC-end sequencing is in progress permitting comparison to the human genome as well as to other species of interest from which genome data is available.

The DBseabream can be found at :http://bioinfo.her.hcmr.gr

The database can be queered and data can be downloaded. As a demonstration is given the linkage map, and RH map will be integrated. Forthcoming is the integration of map position for each marker (both for Linkage and RH maps) **A User Instructions of the database are given in Annex1, part 4, sea bream section.**

Conclusion

We are extremely pleased with these genomic developments as they have put the European sea bass and sea bream among the most advanced models of fishes in term of genomic resources and genetic map available for the European scientific community. These significant advancements started with the Bridgemap and Bassmap projects and are finalized during the Aquafirst project.
These genetic maps will also strongly benefit from the sequencing of the full genome for these 2 species. In fact, such strategy is now in progress in sea bass in collaboration with the Max-Planck Institute in Berlin (Dr. R. Reinhart) and thank to the dynamism of a research group issued from Aquafirst and MGE projects (the European Sea Bass Mapping Consortium). The recent development of high-throughput sequencing technologies make possible such strategy which appears fully possible when a phyllogenetically close fish genome has been already sequenced. In the case of sea bass and sea bream, access to the stickelback genome is certainly a strong argument in favour of the success of using such high-throughput sequencing technlogies. In conclusion, it is not irealistic to suggest that within a few years, sea bass and seabream will benefit from a large panel of complementary genomic and genetic tools, putting these fish species among the most advanced fish species. The work carried out within Aquafirst project have certainly contributed to this evolving situation.

References:

Agarwala R., Applegate D.L., Maglott D., Schuler G.D., Schaffer A.A. (2000): A fast and scalable radiation hybrid map construction and integration strategy.*Genome Res*, 10:350-64.
Agarwala R., Applegate D.L., Maglott D., Schuler G.D., Schaffer A.A. (2000) A fast and scalable radiation hybrid map construction and integration strategy. Genome Research 10 : 350-364
Chistiakov D.A., Hellemans B., Haley C.A., et al. (2005) A microsatellite linkage map of the European seabass *Dicentrarchus labrax* L. Genetics 170:1821-1826
Chistiakov D.A., Hellemans B., Haley C.A., Law S.A., Tsigenopoulos C.S., Kotoulas G., Bertotto D., Volckaert F.A.M. (2005) A microsatellite linkage map of the European seabass *Dicentrarchus labrax* L. Genetics 170:1821-1826
Chistiakov D.A., Tsigenopoulos C.S., Lagnel J., Yuanmei G., Hellemans B., Haley C.S., Volckaert F.A.M., Kotoulas G. (in press) A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. Animal Genetics
Chistiakov D.A., Tsigenopoulos C.S., Lagnel J., et al. (in press) A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. Animal Genetics
Green P., Falls K., Crooks S. (1990) Documentation for CRI-MAP version 2.4 Washington University School of Medicine, St. Louis
Green P., Falls K., Crooks S. (1990). Documentation for CRIMAP, version 2.4. Washington University School of Medicine, St Louis, MO.
Kuhl H., Reinhardt R.R., et al. (submitted). High throughput paired BAC end sequencing and a comparative BAC clone map of European sea bass (*Dicentrarchus labrax*).
Marshall T.C., Slate J., Kruuk L.E.B., Pemberton J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7: 639-655
Sarrapoulou E., Franch R., Louro B., et al. (2007) A gene-based radiation hybrid map of the gilthead sea bream *Sparus aurata* refines and exploits conserved synteny with *Tetraodon nigroviridis*. BMC Genomics 8: 44, 1-14

Sarropoulou E, Franch R, Louro B, Power DM, Bargelloni L, Magoulas A, Senger F, Tsalavouta M, Patarnello T, Galibert F, Kotoulas G, Geisler R (2007). A gene-based radiation hybrid map of the gilthead sea bream Sparus aurata refines and exploits conserved synteny with Tetraodon nigroviridis. BMC Genomics 8: 44

Sarropoulou E, Nousdili D, Magoulas A, Kotoulas G (2008). Linking the genomes of nonmodel teleosts through comparative genomics. Mar Biotechnol (NY) 10: 227-233

Sarropoulou E., Kotoulas G., Power D.M. and Geisler R. (2005b). Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by the application of cDNA microarray technology. *Physiol. Genomics*, 23, 182-91.

Senger F., Priat C., Hitte C. et al. (2006) The first radiation hybrid map of a perch-like fish : the gilthead seabream (*Sparus aurata* L.). Genomics 87: 793-800

Senger F., Priat C., Hitte C., Sarropoulou E., Franch R., Geisler R., Bargelloni L., Power D.M., and Galibert F. (submitted in *Genomics*). The first radiation hybrid map of a perch-like fish sea bream (*Sparus aurata*).

Souche E.L., Hellemans B., Van Houdt J.K.J., et al. 2007) Mining for single nucleotide polymorphisms in expressed sequence tags of European sea bass. Journal of Integrative Bioinformatics 4, 73: 1-10

van Berloo R & Kraakman A. (2002). Quality analysis by means of new software package: MapInspect. Department of Plant Sciences, Wageningen University. (http://www.dpw.wau.nl/pv/pub)

Whitaker H.A., McAndrew B.J., Taggart J.B. (2006) Construction and characterization of a BAC library for the European sea bass *Dicentrarchus labrax*. Animal Genetics 37 : 526-527

Part 5: Characterization of QTL related to stress and disease responses in seabream, sea bass, oyster and trout.

QTL analysis is a powerful approach which will allow mapping loci that contains genes which affect quantitative traits. The strategy is to relate genetic markers (whatever their nature: neutral markers or known genes) to the phenotypic values of individuals in populations of adequate genetic structure. The 'marked' chromosomal segments are expected to contain genes with functional polymorphisms that affect the expression of the traits of interest. The access to biological material of known and wide phenotypic and genetic variability for the traits of interest is a key point to improve the power of QTL detection for a given experimental effort.
This approach have been developped for stress-resistance traits in the four aquaculture species, oyster, sea bream and sea bass. In the fish species, 2 different trais were considered, i.e. confinement stress and pathogen exposure stress (for each species, a specific pathogen was studied). In oyster, significant mortality have been reported to occur during the summer months among both juveniles and aduts and this is a major concern of oyster farmers. the causes underlying this phenomenon are complex and physiological, environmental and pathological causes have been suggested. In this context, resistance to summer mortality was the trait we have selected in oyster. Existence of such QTL is strengthen by the selection of resistant and sensitive oyster families recently carried our by Dégremont et al (2005).
Two different strategies have been developped in the present study:

1) For trout and oyster, the availability of divergent selected lines/families (oyster: resistant/senstive to summer mortality; trout: high responsive/low responsive to confinement stress) allows the production of segregating families devoted to QTL analysis. In both species, divergent original lines (F0) will be inter-crossed in order to produce F1 families. In both species Linkage analysis between traits and genetic markers will be carried out in F2 families. Ideally, F1 parents should be heterozygous for the QTL related to stress response and Phenotypic observations and genotypic analysis are carried out on the F2 individuals.

2) In sea bass and sea bream, F1 generation was produced from a broodstock consisting of 50-80 individuals (male and female). Association between genetic markers and performances (morphometrics and each disease or stress physiological responses) to detect QTL was carried out with analysis by animal model restricted maximum likehood (REML) direct estimations of genetic correlations between all conditions. These parameters have been measured in individuals which are relatives (full- and half-sibs). They have been reared under different environments which allow us to have an estimate of genotype by environment (GxE) interaction.

For the clarity of the presentation of the results, we will briefly summarize the QTL protocols developped in the 4 species and summarize the main results on the charcaterization of QTLs in these species. Detailed information on these experiments have been given in the 3rd activity report.

Material and methods section

Characterisation of QTLs related to stress and disease resistance in trout

Biological material

Original breeders were issued from two divergent lines selected for opposite response to stress, measured as the blood cortisol level after a standardized confinement stress (Pottinger and Carrick, 1999, 2001). Grand-parents of the QTL families were issued from two or three generations of selection for stress response. According to the mating design optimised during the first of the project, five large size F2 full-sib families were produced, and reared in the experimental facilities of Partner CR03. When fish were one year old (around 60g mean body weight), 215 rainbow trout from each QTL families were individually fin clipped for further DNA extraction, measured (length, weight) and implanted with passive integrated transponder (PIT; Trovan ID100A). Fish were further grown until the commencement of phenotyping.

Phenotyping of F2 offspring

Families were then exposed to several successive stressors. The main traits recorded were:

- response to confinement stress (two rounds of 1h confinement in 50-liter tanks, 5 fish/tank, with one month delay). The plasmatic cortisol and catecholamine levels were recorded. Cortisol analyses were conducted following the protocol of Pottinger and Carrick (2001) and the catecholamine concentrations, i.e. noradrenalin and adrenalin were analysed according to Montpetit et al. (2001) and Winberg et al 1997).

- response to salinity stress (two rounds of a 24h salinity challenge, 3 weeks interval). Exposure to salt water was carried out in 150 plastic tanks containing lake water plus 4500g Red Sea Coral Reef Salt (30 g/litre). Plasmatic sodium concentrations were measured using a flame photometer (VWR international, Fontenay-sous-bois, France) after diluting the samples to 1/400 for sodium.

- response to bacterial challenge with *Yersinia ruckeri*, the causative agent of enteric redmouth disease. Fish were challenged intraperitoneally ($0.5x10^6$ cfu/fish). To meet the animal welfare regulation, fish were sacrificed 48 h after challenge, and phenotyped for sex, spleen size and spleen bacterial burden. Bacterial burden of each fish was expressed as the measurement of the bacterial 16S RNA gene expressed in the spleen (qPCR) divided by the measurement of the MCSF host gene.

Genetic markers and genome scan

First, 138 microsatellites selected for polymorphism (Guyomard et al., 2006) were genotyped in the whole set of QTL families. Then, in each family, 44 to 54 additional markers were tested, aiming at improving genome coverage. Finally, 118 to 130 informative markers (13 to 14 duplicated) were genotyped in each family. SNP markers were developed during the course of the project, within genes differentially expressed after stress or immune challenges. Three SNPlex were designed and used for additional genotyping (corresponding to about 70 informative loci per family). Microsatellite linkage data from the QTL data set were compared to the reference map (Guyomard et al., 2006) and revisions were performed when necessary. The incorporation of SNP data is to be finalized, before consolidated information can be provided.

Characterisation of QTLs related to resistance to summer mortality in oyster.

Biological material

Power optimization of QTL mating design was performed in order to implement a design with the higher statistical power (type, number and size of families) given the total number of individual oysters to be genotyped and phenotyped. A design including 5 segregating F2 families of more than 2000 individuals was recommended.

On this basis, F1 hybrid parents (crosses between parents from lines susceptible or resistant to summer mortality) were mated to produced F2 progenies. Five different F2 families with 4 replicates

each were produced by bi-parental crosses. Two control crosses (wild oysters) were used as a reference cross to measure the intensity of the mortality event

The final choice of one of the four replicated cross in every F2 family was based on the number and average size of settled larvae, success rate at metamorphosis and the fact that oyster should not be injured in any mortality event during the growth period in the nursery.

Phenotyping of F2 offspring

Within a three day period, 1000 randomly chosen oysters of each selected F2 families were individually tagged and placed into a raceway for the phenotyping experiment in laboratory conditions. Between the 10th of July 2006 when summer mortality began to occur and the 31st of August when the experiment was stopped, mortality was monitored daily by removing dying oysters. Herpes virus had been detected in the dying oysters during this period. The 5 chosen F2 families had been differentially affected by this mortality event (from only 1.2% of tagged oysters to 67%). A complementary study was performed in 2007 (year 3) in order to compare the viral load detected by qPCR in surviving animals sampled during the phenotyping experiment in two families where high mortalities had occurred (two families with respectively 67% and 40% mortality). The results indicate that surviving animals in a highly dying family bear a low viral load and confirm that there is a strong association between the mortality event and the viral load. In the context of searching for link between genome regions (QTLs) and survival and also herpes virus load, only the three families exhibiting fairly high mortality have been chosen for further genotyping.

Genetic markers and genome scan

A total of 49 microsatellite markers were finally genotyped in the 20 F0, 10 F1, and 540 F2 animals (Selective genotyping of 90 early dying animals and 90 surviving animals at the end of the experiment for each of the three families). In parallel we developed 18 new microsatellite markers (called in silico microsatellites) from 9272 ESTs contigs of sequences produced in collaboration with the Marine Genomics Europe network of excellence. Those 18 microsatellites were also genotyped for the 20 F0, 10 F1, and 540 F2 animals.

A consensus map was built with the Crimap software. In total, 80 markers could be mapped: 38 out the 49 already published microsatellites, 13 out the 18 in silico microsatellites, and 29 out of the 33 SNPs. The number of linkage groups corresponds to the haploid number of chromosomes in this species. These groups have been labelled according to the previously published microsatellite map as 35 markers were common in both studies. The length of the map is 1062 cM. The mean interval mapping is comprised between 9.9 cM and 20.1 cM.

Characterisation of QTLs related to confinement stress and disease resistance in sea bream.

Biological material.

Disease resistance: The pathology selected for sea bream was fish pasteurellosis (causative agent: Photobacterium damselae). A total of 3577 animals were sampled in challenge 4, divided in two tanks (A and B). Disease trends were similar between the two tanks, which represent independent replicates. After consulting with partner CR15, it was decided to analyse at least 1500 animals from tank A (in total 2144 fish) and keep animals from tank B (1433 fish) as a back-up. For each animal, body length was measured based on digitalized pictures of collected specimens.

Confinement stress: The F_1 generation was produced and reared at the Nireus Kanatadika – Evoia hatchery, from a broodstock consisting of 75 individuals from which 2/3 are estimated to be females and 1/3 to be males. After fecondation and hatching, larvae were reared at 19°C and later transferred to tank. The feeding regimes were adjusted to their age and size. The fish remained ungraded for the whole period of their life. Later, fish were divided in 4 groups of approximately 500 each and placed in net cages (1.1x1.1x1.12m) inside a concrete raceway where they were left to acclimatize until the confinement stress experiment was conducted 1 month later.

Phenotyping of F1 offspring

Disease resistance: A. phenotypic information (survival time, body length) were recorded and further used for QTL analysis.

Confinement stress: fish were divided in 4 groups of approximately 500 each and placed in net cages (1.1x1.1x1.12m) inside a concrete raceway where they were left to acclimatize until the confinement stress experiment was conducted 1 month later. For the confinement stress the net was raised so that the water level inside the cage was approximately 20 cm. After 4 h confinement, the cage was lifted and emptied into a tank of icy water and the fish were stunned within 3 minutes. Fish were immediately bled, weighed and digitally photographed. Each group of 500 fish was sampled in between 1h 50m and 2h 10m.

Genetic markers and genome scan.

Disease resistance:A total of 1735 fish from tank A was analysed for parentage assignment. . The total number of reconstructed Full Sib (FS) families was 245. Due to the high number of FS families it was suggested to focus onto the five largest FS or Half-Sib (HS) families. A total of 508 animals were selected to proceed with genome-wide scan using a panel of 151 microsatellite loci. Genotypes were provided for 140 markers. There is a reference map for Sea Bream but many of the typed markers were unassigned. While the groups were quite consistent the marker order changed quite a bit compared to the reference map. Linkage maps were constructed using Crimap software using the published map as a starting point for linkage groups. Three markers were omitted because they had too few meioses. Otherwise ALL markers were mapped to at least on other marker (only 1 LG with 2 markers!). One marker could be assigned to a LG but not mapped.

Confinement stress: All progeny (2200 individuals) and potential parents (177 fish) of sea bream were genotyped for a multiplex of 9 loci. Assignment was performed using the FAP software (Taggart, 2006) which assigns family of origin to progeny from known parental and progeny genotypic data. From the initial 177 breeders, for the 873 individuals of the single match progeny we find that 33 males and 113 females have participated in various degrees. QTL scan of 467 offspring coming from 7 males and 75 females have been carried out. These were the males that had mostly participated in the offspring. More than 10 multiplex PCR reactions which in total included approximately 80 loci, some of them being normal (or nuclear) and others expressed SSRs. Finally, we obtained good and trustworthy results for 50 loci and these data were used for QTL analysis.

Characterisation of QTLs related to confinement stress in sea bass.

Biological material.

The F_1 generation was produced from a broodstock consisting of 57 individuals from which 34 were estimated to be females and 23 to be males. The broodstock was fed on fresh and pelleted feed.

Phenotyping of F1 offspring.

The sea bass were divided in 4 different tanks of 45 m^3, each one with a net covering its inside surface, where they remained to acclimatize until the confinement stress experiments were conducted. The sea bass were divided in 4 different tanks of 45 m^3, each one with a net covering its inside surface, where they remained to acclimatize until the confinement stress experiments were conducted. For the confinement stress the inside net of each tank was slowly pulled to one end so as to concentrate the fish in a small volume of water. As much as possible identical conditions were set for each tank. After 4 hours of confinement the net was lifted and emptied into a tank of icy water a process which stunned the fish within 3 minutes.

Genetic markers and genome scan.

An updated linkage map of European sea bass (Chistiakov et al., Animal Genetics, in press) is the starting point of the QTL analyses. Based on the current estimates of the genome size (15-20 Morgans), 116 markers (covering 25 linkage groups) are used in a moderate resolution genome scan. Blood samples of sea bass F1 progeny were used for parentage and genome genotyping. Assignment has been completed on 56 parents and 2122 progeny genotyped at 31 loci covering 3 multiplex PCRs. Five large full-sib families (n=576) have been identified for QTL analysis. All 576 offspring and 7 parents were genotyped at 109 loci in 13 multiplex reactions. Genotyping data were futher used for QTL analysis.

Statistical analyses

Analyses of the data generally comprised of up to three stages as appropriate for the population analysed. Firstly, analyses of the phenotypic data utilising information on pedigree and potential cofactors and confounding effects. This involves cleaning the data (of outliers, etc) and allows estimation of trait heritabilities and identification of appropriate factors to include in later analyses. Secondly were the analyses of the marker data to build a genetic map and to do some secondary cleaning of the data to remove clear genotyping errors. The third stage is combining the trait and marker data for the analyses to detect quantitative trait loci.

Phenotypic data analyses

These are required both to detect and correct or eliminate potential recording errors or other errors and biases and to provide background information and context for the QTL analyses. Basic analyses require that the data for each trait recorded are examined and the distribution of the data are explored. This may identify outliers which could represent recording errors and are likely to impact substantially on the results if not removed. In addition substantive non-normality in the trait distribution may suggest the need to transform the data. Once the data distributions have been explored and modified if necessary, trait statistics and correlations between traits within a data set can be estimated. Finally, for some data sets from the outbred population structures it is possible to estimate trait heritabilities. These provide an estimate of the total genetic variation for each trait and hence an upper bound and context for any detected QTL effects.

Initial exploratory analyses of the data to detect outliers, look at trait distributions, estimate was performed using Excel. Phenotypic correlations were estimated with the Genstat statistical package.

Heritability estimation is appropriate for the random mating populations of sea bream and sea bass but not for the selected trout and oyster populations and so was restricted to the former two species. Heritabilities were estimated using ASReml fitting an animal model. Fixed effects included: sample set, day, tube number and assay number.

Marker data analyses

Optimisation of the QTL analyses requires high quality marker data where inevitable errors are identified and removed or corrected. The markers then need to be integrated into a linkage map relevant to the population under study for the QTL analysis. Some marker errors may be evident and corrected at the time the data is initially collected, however many only become apparent in the analyses to generate the linkage map. Thus this stage requires a continuing dialogue between data analysts and the laboratory where the data were generated in order to correct errors.

For the trout, where the linkage map is relatively well developed, the mapping exercise is largely revision of known maps with re-estimation of distances between markers so that they are relevant to the populations under study, although so revision of marker order is inevitable at this stage. The data used for QTL mapping were obtained as residuals from a linear model including the relevant fixed effects (sex, rearing tanks, testing tanks or date of challenge, depending on the trait) and covariates (usually body weight). The analyses were performed using the GLM procedure of the SAS software (SAS Institute, 1999). QTL detections were performed using the QTLMAP software described in Le Roy et al. (1998) and Elsen et al. (1999) with a full-sib family design. Further tests (multi-traits and multi QTL analyses) are necessary to further hone the results.

For oysters, with the least well developed map the process consists of generating the map *de novo* utilising the markers developed in the project. The situation for sea bream and sea bass it between trout and oysters, requiring some revision of already established maps.

Marker map development was undertaken using the Crimap software (http://compgen.rutgers.edu/multimap/crimap). For sea bass we built a genetic map specific to our experimental population using the 90 genotyped markers based on the previous published map (Chistiakov *et al.* 2005) which was used as a reference to assign the markers to different linkage groups. For both sets ofsea bream data linkage maps were developed *de novo* and compared back to the published map. A number of changes were noted compared to the reference map and because of

the large amount of data in our analyses were prefered to use the population specific maps generated here. For oysters where prior information was more limited the linkage map of the population was built *de novo* using Crimap to first assign markers to linkage groups (using pair-wise linkage analysis) and then to order markers within linkage groups.

QTL analyses

These bring the information on pedigrees, markers and traits together in order to detect regions of the genome contributing to trait variation (i.e. quantitative trait loci or QTL). The analysis utilised depends on the population structure of the species under study. For the sea bass and sea bream the study sample is essentially a sample from an outbred population. Where there are large half-sib families these can be analysed considering segregation for a single parent. Both sea bass and sea bream samples (the latter species generally having smaller families) can be analysed utilising the variance component approach. The oyster and trout study samples were based on a cross between selected lines and so there may be some benefit from taking this into account in the analyses, however the selection was limited to a few generations and so the lines will not be extremely different and analyses that treat the sample as a single outbred population are also appropriate.

The QTL mapping was done using interval mapping, where the presence for a QTL is tested at each position of the genome. A test statistic is calculated for each position to create a profile, and then a QTL is positioned at the position that has the best test statistic, provided it passed a significant threshold (which depends on the method used to calculate the test statistic).

For the non-selected outbred populations of sea bream and sea bass, two different analysis approaches where used to perform interval mapping: (1) half-sib regression analysis, (Knott, Elsen, and Haley 1996) and implemented in QTL express (Seaton *et al.*, 2002; http://QTL.cap.ed.ac.uk) and (2) variance component analysis (George *et al.,* 2000) also implemented in QTL express (http://qtl.cap.ed.ac.uk/puccinoservlets/hkloaderLoki).

The half-sib regression analysis regress the phenotype of each offspring on their probability of having inherited a given allele (calculated using the marker information) from their common parent (Haley, Knott, and Elsen 1994). The test statistic for this approach is the F-test obtained from the analysis. One of the characteristic from this methodology is that it can only consider QTL allele inherited from one of the parent of each family at a time.

In the variance component approach (VCA), the QTL is included into the model as a random effect with its covariance structure proportional to the IBD matrix (George *et al.*, 2000). Because the IBD matrix includes the degree of relatedness between sibs due to the allele inherited from both sire and dam, the VCA takes into account the segregation of the QTL from both parents, unlike the half-sib regression method. This approach contrasts a model including a randomly fitted QTL (full model) against a model without a QTL (reduced model). The test statistic assumed to be 2 * (Lf – Lr), where Lf and Lr are log-likelihood of the full and the reduced model, respectively.

For the sea bass data we performed both sire and dam half-sib regression analysis. Similar procedures were followed for sea bream. We used the VCA to confirm significant QTLs found with the half-sib regression analysis.

The primary analyses of the oyster data treated the population as an F_2 population derived from a cross between tow lines (i.e. susceptible and resistant) using the analytical approach described in Haley *et al.* (2004) as implemented in QTL Express (Seaton *et al.*, 2002). This analysis assumes the two lines are fixed for alternative alleles of the QTL (but not necessarily markers) and proves estimates of the additive and dominance effects in the resulting three genotypes.

Results section

1) Characterisation of QTLs related to stress and disease resistance in trout

Data collection and genotyping

In summary, around 200 progeny of 5 QTL families were individually tagged, genotyped for a number of informative microsatellite (120-130) and SNP (around 60) markers and included in the phenotyping process. Main traits recorded were:

- response to confinement stress. Fish were stressed during two similar rounds of confinement performed at one month interval. The recorded traits were weight, plasmatic cortisol level (CORT1 and CORT2 respectively), and the plasmatic adrenalin (E1 and E2) and noradrenalin (NE1 and NE2) levels.
- response to salinity stress. Fish were stressed during two similar rounds of exposure to salinity, at one month interval. Traits recorded were weight and plasmatic sodium concentration (SOD1 and SOD2).
- response to bacterial challenge with *Yersinia ruckeri*, the causative agent of enteric redmouth disease (ERM). Because of limited experimental facilities, fish were randomly distributed into two groups that were challenged at one month interval. Fish were sacrificed before mortality occurred, and resistance was estimated through tow indirect criteria, spleen weight corrected for body weight (SCW) 48h post-injection and spleen bacterial burden (SBB), measured by qPCR. Sex was also recorded.

Genotyping effort was put on microsatellite markers (number increased by about 30% from initial schedule) in order to anticipate with the delay in the identification of candidate genes required to design SNP markers. In the end, it was finally possible to design 3 SNPlex as initially scheduled and to use them to genotype SNP polymorphism inQTL families. Yet, data were available late in course of the project, and analyses are still continuing. Results below have been extracted from microsatellite data only, and are being refined with SNP data.

Analysis of marker data and linkage map

Linkage data of the population under study (F2 families) were compared to the linkage groups of the map previously published by Partner CO.01, using Carthagen software, and revisions were performed when necessary. They included re-estimation of distances and re-ordering of markers. In few markers, the polymorphism in the new set of families also revealed duplication that had not been previously identified. A more complicated case (possible re-association between chromosomal arms due Robertsonian polymorphism) requires additional investigations. It is anticipated that he incorporation of SNP data will imply new revisions.

Analysis of phenotypic data

Mean values of recorded traits are summarized in Table 1 for each of the 5 QTL families.

Table 1. Mean family value of major traits recorded in the F2 trout offspring.

Trait	Cross	X3	X4	X8	X14	X17	Unit
Confinement stress	CORT 1 CORT 2	151 197	109 158	108 172	134 168	88 155	ng/ ml plasma
	NE1 NE 2	10.4 2.9	14.5 3.1	8.8 2.9	14.6 3.2	12.9 2.7	ng/ ml plasma
	E1 E2	9.5 4.3	12.2 5.8	23.5 6.8	3.6 4.8	3.4 5.7	ng/ ml plasma
Salinity stress	SOD1 SOD2	185 202	195 206	209 221	173 164	200 197	mM /L plasma
Immune challenge	SCW	2.00	2.53	2.01	2.06	1.83	mg, body weight as covariable
	SBB (log)	1.03	1.46	1.34	1.17	1.24	Expression ratio to a host gene (16s RNA gene)

During challenge or date of challenge) as well as covariables (usually body weight) were tested using ANOVA and/or ANCOVA (Proc GLM, SAS software). Residuals after correction for significant effects and covariables were used for further QTL analyses.

Main results were as follows:

- Confinement stress: sex and cross had a significant effect on CORT1 and CORT 2. The second challenge resulted in higher cortisol levels than the first one (more than 40% relative increase). Correlation between CORT1 and CORT2 was low (Pearson coefficient of correlation of 0.10 to 0.39, according to families), as often recorded in such experiments. Surprisingly, post-stress E and NE values were extremely low (80% of individuals with value less than 10 ng/ml), and typical of unstressed fish. There was no explanation for such low values, and data were not considered in QTL analyses.
- Salinity stress: there was a significant effect of cross on SOD1 and SOD2, no effect of sex nor body weight. Correlation between SOD1 and SOD2 was good (within family Pearson coefficient of correlation from 0.65 to 0.83). The second challenge resulted in higher values (+ 7% relative increase).
- Immune challenge: both SCW and SBB differed among families. There was a positive relationship between the two traits (within family correlation from 0.26 to 0.42). Fish tested during the second challenge exhibited lower response for both traits, maybe because of the larger size of fish (445g *vs* 378g). Both traits were log transformed before analyses.

QTL analyses

Analyses were performed with the QTLMAP software, described in Le Roy et al (1998) and Elsen et al (1999) using a full-sib family design. Results still need validation (additional markers and further analyses).

Using single-trait tests, 14 QTLs with chromosome wide significance ($P<0.05$) were discovered on 10 different linkage groups as detailed in Table 2.

- Confinement stress: 6 QTLs were found for cortisol level after confinement stress. Such a result was expected owing to the origin of F2 families (F2 cross between two lines selected for high/low blood cortisol level after confinement stress). Four QTLs are responsible for CORT1 variability (one with a genome wide significance, LG6) while the two others explain CORT2 variability. Surprisingly, none of those QTL is common to CORT1 and CORT2, suggesting that different mechanisms may have been involved in the two successive challenges.
- Salinity stress: No significant QTL was detected for SOD1, though two suggestive ones were identified. Three significant QTL were mapped for SOD2.
- Immune challenge: QTLs were detected for both SCW and SBB, with no common QTL between the two traits.

Table 2. QTLs detected for confinement, salinity and immune stress in trout

Trait	LG	Position of maximum [a]	95% [CI] [b]	Number of heterozygous parents [c]	Average effect [d]
CORT1	6	32 **	22 - 41	6	0.27
	15	7*	0 -16	6	0.31
	23	59*	38 - 88	6	0.29
	30	24**	19 - nd	3	0.22
CORT2	21	22**	19-33	3	0.32
	31	35**	24-64	6	0.28
SOD2	14	49*	43 - 55	2	0.26
	19	93*	82 - 99	4	0.21
	23	67*	52-74	4	0.22
SCW	17	69*	43-Nd	6	0.19
SBB	6	16 **	Nd-113	2	0.25
	7	49**	Nd-58	3	0.44
	23	23*	0-44	4	0.34
	31	39*	21-50	4	0.20

[a]: chromosome-wide significance at P<0.05 (*) and P< 0.01 (**) respectively
[b]: 95% confidence interval of the QTL location estimated by lod drop-off
[c]: parents for which the effect of within family allelic substitution is significant (T test, P<0.05)
[d]: effects averaged on the heterozygous parents (σ_P units)

Summary

A population of more than 1000 individual (5 full-sib families) was individually tagged and individuals were challenged successively for confinement and salinity stress (for each, two repeated challenges with several weeks delay) and finally an immune stress (bacterial infection with *Y. ruckeri*). Genotyping with more than 120 informative microsatellite markers per family provided good coverage across the genome. QTLs were discovered for all trait groups, all together contributing to a substantial part of the variability of the traits under study. These results provide original data concerning stress and disease resistance in trout.

2) QTL analysis of summer mortality in oysters

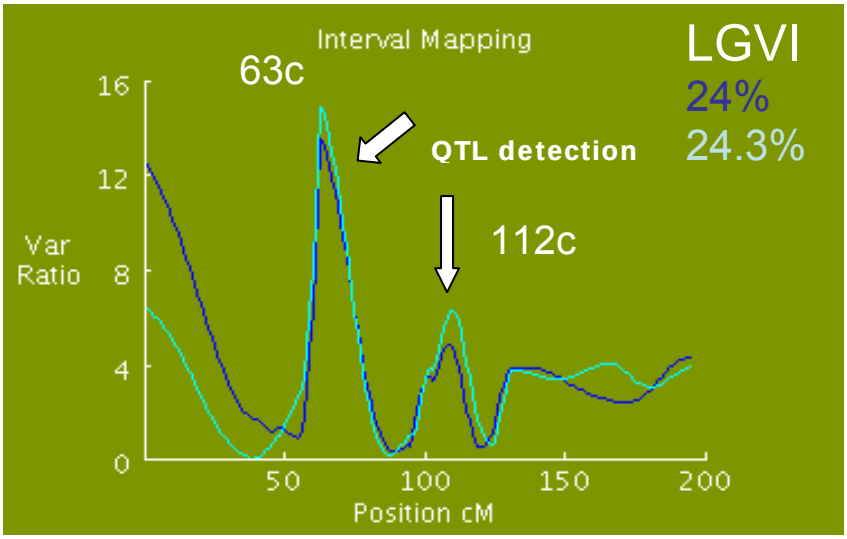


Figure 1. Example of 2 QTLs detected for both traits in Linkage Group VI.

Two traits were used in QTL mapping: survival (after the mortality event recorded in 2006) and viral load (measured individually during the same experiment). The analyses detected 5 QTLs located in four linkage groups V, VI, VII, and IX (see example in Figure 1) which explained 49% and 43% respectively for the survival and viral load characters. Most of the QTL co-localized for the two traits which supports the hypothesis that the genetic architecture of the two traits is very similar in our experiment. These results are very promising for potential Marker-Assisted Selection, but need to be confirmed by QTL experiments in the field, as well as experiments to link the selection to summer mortality and Herpes Virus. Furthermore, this approach will be complemented by the localization of genes differentially expressed in resistant and sensitive lines.

3) QTL analysis of disease resistance in seabream.

Challenge and parentage. A disease challenge was set up for pasteurellosis, with 1735 fish for 20 days, coming from a brood stock of 253 parents. Each day, the dead fish were recorded. The length in millimetre was also phenotyped to check on growth.
7 microsatellites were genotyped in all 1735 offspring and the 253 parents for parentage assignment using PAPA software. A total of 1256 fish were allocated 2 parents.

Quantitative Genetic Analyses.

The raw data on this challenge contains the fish identity, its parent's identity, its length in millimetre and the day it has been found dead.

Heritability results.

Heritability for growth: The heritability of the model using the log-transformed data is preferred here, as the data are more normally distributed. The heritability is therefore 0.3790 with a standard deviation of ± 0.0759. This heritability is sufficiently accurate as the standard error is very low.
Survival heritability: Results for heritability using a binary trait is shown in table 3. The log-likelihood, the heritability, standard deviation and log-likelihood ratio are reported here.
The heritability is not estimated for survival above 10 days. The additive component is too small at this point (survival is really small). The heritability goes from 0.4395 to 0. The standard error is fairly large but for early mortality the heritability is significant.

QTL Results

Body length. The Analyses showed a genome-wide significant QTL on LG6. This QTL was significant in all analyses. The VCA showed a LOD score of 5.4 with an estimated QTL heritability of 0.40 (i.e. the QTL was estimated to explain 40% of variation in body length). A second genome-wide significant QTL was detected with HSMain, segregating in a single family. Under VCA this QTL was suggestive, explaining 6% of the variation in length. Other chromosome-wide and suggestive QTL were detected on LG11, 12, 19, 24 (LG13_12_old) and 5.

Disease resistance. The results for disease resistance are summarised in Table 4. The analyses showed a highly significant QTL for DPC on LG1 under all analyses. Surprising the estimated overall heritability of DPC was 0.19 while the earlier results of the larger data suggested that the heritability was negligible. Figure 6 shows excellent agreement between the three analyses for this QTL. Further genome-wide QTL related to resistance were detected on LG3 (Day10), LG12(Day10), LG17 (Day5) and LG22 (all traits). Unfortunately, the QTL on LG22 and 12 only appear to segregate in a single parent (147). This offers limited prospect of fine-mapping this QTL via a linkage disequilibrium analysis across families. Figure 7 shows the QTL curve for family 147 on LG1 and LG12.

In conclusion, several QTL related to disease resistance have been identified and these explain a considerable part of the variation in mortality post-infection.

Table 4. Overview of QTL results for disease resistance

Trait	Linkage group	HSMain	HS147	VCA LOD	QTL heritability
DPC Day10 Day20	1	** NS *	** ** NS	2.47	0.11
Day15	2	*	*	NA	
Day10	3	**	NS	NA	
Day10	5	*	NS	NA	
Day15	7	NS	*	NA	
DPC Day10	12	* NS	NS **	NS	
Day15	15	NS	*	NA	
DPC Day5 Day 10 Day15	17	* ** * *	NS	1.6	0.17 NA
Day5 Day10	19	* *	NS	NA	
DPC Day5 Day10 Day15 Day20	22	NS NS NS NS NS	** ** ** ** **	2.5	0.05
Day15 Day20	24	NS NS	* *	NA	

* denotes chromosome-wide significance, ** genome-wide significant

Figure 6. QTL results for days post challenge on linkage group1 using HSMain (top left), HS147 (top right) and VCA (bottom)

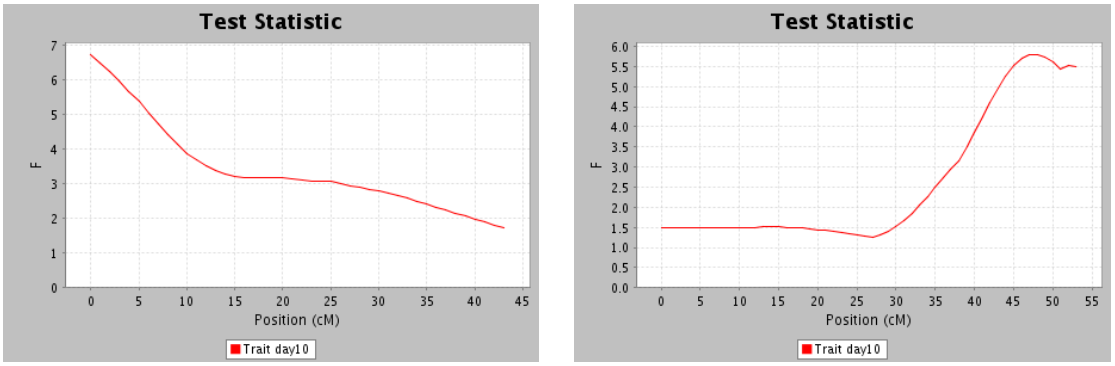


Figure 7. QTL results for survival at day 10 post-challenge for HS147 on LG1 (left) and LG12 (right)

4) QTL analysis of stress related traits in sea bass and seabream

Sea bass.

A large population of 2000 fish resulting from mass spawning of the brood stock were measured for cortisol response to stress and a number of linear measures of growth. Five large full-sib families from the stress response experiment were selected for full phenotyping and genotyping, including a total of 568 progeny. Although offspring originated from five different sires mated to two dams, this adding half-sib relationship the full-sib structure. All animal were kept in the same tank throughout the experiment and therefore parentage assignment is required. Table 6 describes full-sib families used for QTL mapping.

Table 5. Traits measured

Trait	Abbreviation	measured in
Body weight	BW	gram
Body depth	BD	centimetre
Body length	BL	centimetre
Head depth	HD	centimetre
Head length	HL	centimetre
Abdominal length	AL	centimetre
Post-anal length	POAL	centimetre
Pre-anal length	PRAL	centimetre
Standard length	SL	centimetre
Cortisol	CORT	milligram per litre

Heritability of traits: All traits related to size or morphology were found to be highly heritable ranging from 0.73 to 0.90 with exception of standard length (0.00) and CORT (post-stress plasma cortisol levels: 0.09). Similarly, the phenotypic correlations were positive and very high for all traits (0.88 – 0.99) but SL and CORT. The correlation between CORT and the other traits tended to be slightly negative, no overall trend was found in the correlations between SL and the other traits.

QTL mapping: – QTL mapping analyses were performed using both the half-sib analyses and the variance component approaches described in the section above on disease in sea bream. However here we only report results from the variance component approach for the sake of brevity.
Eight QTLs have a LOD score above 3 corresponding to a likelihood ratio test of 13.8 chi squared distributed, and are therefore highly significant. 18 other QTLs have a LOD score between two and three, so they are considered as putative QTLs. Table 8 shows the eight highly significant QTLs with their position, their likelihood ratio, LOD score and heritability of both the polygenic component and the QTL. The QTLs are located on LG4 and LG6 (4 on each chromosome) for growth: PRAL, HD, HL and AL QTLs on LG6 and AL, BL, BD and HL QTLS on LG4. The QTLs on LG4 corresponds to those found with dam analysis, while the ones on LG6 are at the same location than in sire analysis. QTLs heritabilities on LG4 show that they are major QTLs, responsible for 36% to 62% of phenotypic variation according to the trait. On the other hand, QTLs on LG6 seems to have less influence on the phenotypic variation (only 6% to 12%). Figure 9 and 10 represent the variance component results for LG4 and LG6.
For the cortisol response to stress no significant QTL were found.

Table 6. Major QTLs identified in the variance component analysis

Trait	LG	Position	LR	LOD	Polygenic h2	QTL h2
PRAL	6	19	32.416	7.039	0.887	0.114
HD	6	19	29.19	6.32	0.752	0.122
HL	6	19	25.48	5.533	0.897	0.089
AL	4	2	17.026	3.697	0.883	0.312
AL	6	19	16.504	3.584	0.907	0.06
BL	4	1	15.898	3.452	0.787	0.488
BD	4	2	15.728	3.415	0.825	0.525
HL	4	2	15.702	3.41	0.877	0.279

Conclusion:
– Sea bass stress – A population of 2000 fish derived for a commercial mass-spawning was stress challenged and the cortisol response measured. Linear measures of fish morphology were also recorded. Marker information was used to reconstruct pedigree relationships and revealed a number of large families with a good structure for QTL mapping. A group of 568 fish from 5 large families were selected for full genotyping. Genotyping of 90 microsatellite markers provided good coverage across the genome. Heritability was estimated to be moderate to high for morphometric traits, but was low for cortisol response. QTL with genome wide significance were discovered for several morphological traits on two linkage groups but no significant QTLs were found for cortisol response.

Sea bream.

A large population of 2000 fish resulting from mass spawning of the brood stock were measured for cortisol response to stress and a number of linear measures of growth. Fish were allocated to families using marker information. This revealed a structure with a relatively large number of small families which was not ideal for QTL mapping. Complete genotyping focused on the largest half-sib families. There were 468 recorded and genotyped individuals in 166 full-sib families. These comprised 7 sire half-sib families (of size 49, 50, 52, 59, 79, 85, 94) and 74 dam half-sib families.

Heritability: Estimated trait heritabilities were generally moderate for morphological traits but were zero for cortisol response to stress. (Table 7).

Table 7. Sea bream traits recorded

Trait	Meaning	Heritability
OPAN	Distance between operculum and anal fin	0.21
OPCA	Body length	0.29
OPDO	Length between operculum and dorsal fin	0.02
SNOP	Head length	0.30
BW	Body weight	0.52
SL	Standard length	0.45
SNAN	Length between snout and anal fin	0.37
DOAN	Length between dorsal and anal fins	0.41
DOCA	Length between dorsal and caudal fins	0.41
CORT	Cortisol response to stress challenge	0.0
LOGCORT	Log of corstisol response	0.0

The selected animals were genotyped for 51 microsatellite markers and the linkage maps were generated *de novo*.

QTL analyses were performed with the variance component approach described above and the significant QTLs from the initial studies are indicated in Table 8. These results need validation in further analyses of these data.

Table 8. QTLs detected in sea bream stress study.

D Trait	LG	Position	LR	LOD	Polygenic h2	QTL h2
OPAN	7&15	24	54.372	11.807	$8.66\ 10^{-9}$	0.878
OPCA	7&15	24	19.368	4.206	$1.189\ 10^{-8}$	0.760
OPDO	10	1	8.606	1.869	$1.54\ 10^{-8}$	0.767
OPAN	10	18	63.531	13.79	$9.657\ 10^{-9}$	0.867
OPCA	10	0	27.59	8.992	$9.26\ 10^{-9}$	0.844
SNOP	A	5	9.174	1.992	5.6410^{-8}	0.370
OPDO	A	5	12.866	2.794	7.5510^{-8}	0.489
OPAN	A	5	58.130	12.62	1.4010^{-8}	0.816
OPCA	A	5	32.548	7.068	1.6710^{-8}	0.645
OPAN	F	22	50.74	11.02	1.20510^{-8}	0.824
OPCA	G	22	17.396	3.778	1.81910^{-8}	0.663
BW	I	1	13.120	2.849	0.456	0.110
SL	I	1	6.422	1.395	0.379	0.090
SNAN	I	2	5.272	1.145	0.290	0.096
OPDO	I	20	28.26	6.226	$9.66\ 10^{-9}$	0.867
OPAN	I	10	71.1682	15.45	$9.95\ 10^{-9}$	0.864
OPCA	I	10	20.106	4.366	$1.297\ 10^{-8}$	0.701
DOAN	I	0	8.56	1.859	0.347	0.103
DOCA	I	0	5.97	1.296	0.333	0.077
DOCA	J	3	5.063	1.099	0.333	0.089

No QTLs were detected for cortisol response, but a number of significant QTLs were detected for the morphological traits.

Summary – Sea bream stress study – The population structure for this study was less than ideal with few large families. In addition the marker information only covered a proportion of the genome. The estimated heritabilities for cortisol response to stress were zero and so not unexpectedly no QTLs were detected for this trait. Moderate heritabilities were found for morphological traits and a number of significant QTL were detected.

Outline operational genetic protocols incorporating identified QTL and traditional breeding approaches.

Work in this area has focussed on general designs for marker-assisted selection in fish populations. The application of marker assisted selection (MAS) and gene assisted selection (GAS) techniques in aquaculture is becoming possible as genomic technology becomes established and new QTLs ar found. Hence the development of customised breeding progrqmmes using MAS or GAS is becoming important. A number of strategies for MAS and GAS have been developed, but these require the evaluation before implementation in real breeding programmes. The particular characteristics of fish populations, in pqrticular the potentially high reproductive rate, provide opportunities for high genetic progress utilising MAS or GAS, but also may lead to high rates of inbreeding.

First of all, using the genetic progress prediction software SelAction, we simulated mass selection in a range of scenarios with and without a QTL and with various heritabilities for the trait and various QTL effects.

Secondly, 3 MAS scenarios have been stochastically simulated using R to evaluate the effects of within family selection on inbreeding rates for heritability of 0.1, 0.5 and 0.9.

SelAction results shows that genetic gain and accuracy were higher in MAS scenario than in phenotypic selection as expected, especially for low heritability (15% increase of genetic gain and 16% increase for accuracy for a heritability of the breeding goal trait of 0.1). The unexpected results were for the increase of inbreeding per generation. For MAS scenarios, for a specified QTL size, the inbreeding first increases as the heritability increases and then reaches a maximum at an medium heritability and decreases as the heritability as heritability approaches unity.

The within-family selection results coming out from the stochastic simulation confirms the results from deterministic prediction.The strange results obtained for inbreeding are likely due to the high number of progeny produced per family and the assumption of random mating (used in selAction). Those results in inbreeding have been higlithed by Woolliams (1997) with the same assumptions. MAS performs better than own performance but with potential worrying consequences for inbreeding rates.

The results for genetic gain are consistent with the results from other species. However, **the behaviour of inbreeding is quite interesting and alerts us that we must be careful when dealing with aquaculture, where it may be especially important to utilise designed mating strategies to avoid inbreeding.**

Conclusion:

These studies allowed us to identify several QTL related to stress or disease resistance in trout, sea bream and resistance to summer mortality in oyster. Integration of these genetic information in traditional breeding programs for aquacuture species highlights both benefits but also worrying consequences for inbreeding rates.

References

Chistiakov D.A., Tsigenopoulos C.S., Lagnel J., et al. (in press) A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. Animal Genetics

Chistiakov DA, Hellemans B, Haley CS, Law AS, Tsigenopoulos CS, Kotoulas G, Bertotto D, Libertini A, Volckaert FA. (2005) A microsatellite linkage map of the European sea bass Dicentrarchus labrax L. Genetics, 170, 1821-1826.

Elsen JM., Mangin B., Goffinet B., Boichard D., Le Roy D., 1999. Alternative models for QTL detection in livestock. I. General introduction, Genet. Sel. Evol. 31, 213-224.

George, A. W., Visscher, P. M. & Haley, C.S. (2000) Mapping quantitative trait loci in complex pedigrees: A two step variance component approach. Genetics **156**:2081-2092.

Green P., Falls K., Crooks S. (1990) Documentation for CRI-MAP version 2.4 Washington University School of Medicine, St. Louis

Guyomard, R., Mauger, S., Tabet-Canale, K., Martineau, S., Genet, C., Krieg, F., Quillet, E., 2006. A type I and II microsatellite linkage map of rainbow trout (Oncorhynchus mykiss) with presumptive coverage of all chromosome arms. BMC Genomics, 7: art. n° 302.

Haley CS, Knott SA, Elsen JM. (1994) Mapping quantitative trait loci in crosses between outbred lines using least-squares. Genetics, 136, 1195-1207.

Knott SA, Elsen JM, Haley CS. (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. Theoretical and Applied Genetics, 93, 71-80.

Le Roy P., Elsen JM., Boichard D., Mangin M., Bidanel JP., Goffinet B., 1998. An algorithm for QTL detection in mixture of full and half sib families, in 6th World Congress of Genetic Applied to Livestock production, 11-16 January 1998, Vol. 26, University of New England, Armidale, pp 257-260.

Marshall T.C., Slate J., Kruuk L.E.B., Pemberton J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7: 639-655

Montpetit C. J., McKendry J. and Perry S. F. 2001. The effects of c-type natriuretic peptide on catecholamine release in the pacific spiny dogfish (Squalus acanthias), Gen. Comp. Endocrinol. 123, pp. 210-221.

Pottinger, T. G. & Carrick, T. R. (1999). Modification of the plasma cortisol response to stress in rainbow trout by selective breeding. General and Comparative Endocrinology 116, 122-132.

Pottinger, T. G. & Carrick, T. R. (2001). Stress responsiveness affects dominant-subordinate relationships in rainbow trout. Hormones and Behavior 40, 419-427.

Sarrapoulou E., Franch R., Louro B., et al. (2007) A gene-based radiation hybrid map of the gilthead sea bream *Sparus aurata* refines and exploits conserved synteny with *Tetraodon nigroviridis*. BMC Genomics 8: 44, 1-14

Seaton, G., Haley, C.S., Knott, S. A., Kearsey M. & Visscher, P. M. (2002) QTL Express: mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics **18**: 339-340

Senger F., Priat C., Hitte C. et al. (2006) The first radiation hybrid map of a perch-like fish : the gilthead seabream (*Sparus aurata* L.). Genomics 87: 793-800

Souche E.L., Hellemans B., Van Houdt J.K.J., et al. 2007) Mining for single nucleotide polymorphisms in expressed sequence tags of European sea bass. Journal of Integrative Bioinformatics 4, 73: 1-10

Whitaker H.A., McAndrew B.J., Taggart J.B. (2006) Construction and characterization of a BAC library for the European sea bass *Dicentrarchus labrax*. Animal Genetics 37 : 526-527

Winberg S., Winberg Y. and Fernald R. D. 1997. Effect of social rank on brain monoaminergic activity in a cichlid fish. Brain Behav Evol.;49(4):230-6.

General Conclusion

This report describes the activity which has been carried out during the last 4 years by the Aquafirst project. Developped in the context and in the continuity of previous EC-funded reserach projects (Stressgenes, Bassmap and Bridgemap), This project led to large development of the the genomic resources and new information related to genetic in aquaculture species. An overall analysis of these research activities highlighted the following major outcomes:

- **A large development of genomic tools**: This include not only functional genomic tools (EST collections, microarrays, database for management of the sequences and the microarray data) but also genomic tools (SNP, RH panel, new microsatellite markers, high-resolution linkage and gene map). This success was observed in the 4 species, seabass, sea bream, oyster and trout. On these issues, the project also beneficited from the strong support and scientific imput from one of NoE MGE partners (Max-Planck Institute, Berlin): These links have also stimulated gathering of researchers for futher research efforts on sequencing of the full genome of the sea bass European Sea bass mapping Consortium).

- **A need for sustainable resources:** This effort raises an important question related to the sustainability of these resources after the end of Aquafirst project. Efforts have already been made: For example, SIGENAE team (INRA, Toulouse, France) has proposed to keep on managing on his own resources the Aquafrist genomic database for 2 more years. This is an important question (cf. Conclusion from Aquagenome project) which would deserve full consideration if the European scientific community does not want to loose all the benefits from efforts developped during the last 8 years.

- **A large set of new scientific information**: Beyond the development of new genomic tools, this project has also generated several major new information related to the cluster of genes involved in the stress responses or in the first identification of QTL related to disease resistance or stress resistance traits in fish. In oyster, a QTL related to summer mortality have been identified. These information will be disseminated through publication in international peer-review journal. The list of presentations and publications (accepted or in progress) of these results (cf. Dissemination plan) fully confirm the scientific success of the present project and we can anticipate that more publication will come out.

- **A large scientific exchanges between physiologist using functional genomic tools and geneticians using genomic tools:** The present project have been a successful attempt to stimulate these discussions and more benefits will come later as such exchanges should be carried on. Development of genomic approaches to answer specific questions related to aquaculture species will lead to such an integration of these genetic and physiological answers.

- **The future of the Aquafirst project will be also very important:** This will include finalization of data analysis in order to publish them in high-ranked scientific journal. It will also include continuation of scientific discussion between partners on specific topics such as: (1) metagenomic analysis of the transcriptome data obtained in different tissues in the same species or for the same stress situation in different species (2) precise localisation an identification of the QTL using gene expression information (3) integration of the different genetic maps including BAC end sequencing map and RH map.

Such questions will mainly rely on scientific discussions and the possibility to maintain such exchange activity between Aquafirst partners will be very productive for the outcomes of the project. During the coming months, we willl try to stimulate such discussion and continue the dynamic which has been engaged within Aquafrist project.

Appendix 1 – Plan for using and disseminating the knowledge

Section 1 - Exploitable knowledge and its use

Not relevant to the project

Section 2 - Dissemination of knowledge

Planned/ actual dates	type	Type of audience	Partner responsible /involved
November 2006 (Bruxelles)	'Research on fish farming technologies for SMEs: Technological challenges and market opprotunities' Workshop organized by the European Commission DG Research	Research + Industry	COO1
February 2006 (Faro)	Presentation of the Aquafirst project ate the AQUAFUNC project meeting	Research	CO01
December 2007 (Bruxelles)	Seminar on Aquaculture Research	DG MARE	CO01
June 2008 (Calgary, Canada)	Symposium on Fish Endocrinology	Research	CO01
October 2008 (Tours, France)	Journées des Sciences du Muscle et Technologie de la Viande Abstract + Poster	Research	CO1
August 2008-Cork (Ireland)	Agricultural Biotechnology International Conference	Research and industry	CO01
May 2005 Rennes, France	Presentation of the Aquafirst results for the oysters West Genopole	Research + French genomic platforms	CR05.1
September 2005, Hinckley, UK	7th international meeting on Single Nucleotide Polymorphism and Complex Genome Analysis	Research	CR05
January 2006, San Diego, USA,	XIV Plant and Animal Genome conference	Research	CR05
June 2006, Montpellier France	International symposium of Genetics in Aquaculture	Research	CR05
October 2006 (Busan – Korea)	International Workshop on summer mortality	Research	CR05.1

October 2006 Sorrento Italia	Marine Genomics Europe conference	Researrch + EU representative	CR05 and CR05.1
April 2007 Roscoff, France	Marine genomics Europe	Researrch + EU representative	CR05 and CR05.1
March 2007, San Antonio, USA	Annual conference of the National Shellfisheries Association	Research	CR05 and CR05.1
Nov.2007 Hangzhou, China	International Oyster Symposium	Research	CR05 and CR05.1
February 2008, Paris, France	Aquagenome workshop,	Research + Producers	CR05 and CR05.1
October 2008, Paris, France	Aquabreeding workshop	Research + Producers + EU representative	CR05.1
January 2008, San Diego, USA	XV Plant and Animal Genome conference	Research	CR05
September 2008 Brest, France	Physiomar conference	Research	CR05 and CR05.1
July 2008 Amsterdam The Netherlands	ISAG conference	Research	CR05 and CR05.1
August 2008 Brisbane, Australia	International Conference “Australasian Aquaculture 2008”	Research + Producers	CR05.1
June 2007 Stirling (UK)	7th Nordic Symposium on Fish Immunology	Research	CR06
September 2007 Cádiz (Spain)	6th Congress of the Iberian Society of Comparative Endocrinology	Research	CR06
August 2008 Paris (France)	10th European Multicolloquium of Parasitology (EMOP 10), from satellites to microsatellites	Research	CR06

September 2008 Ravenna (Italy)	25th Congress of the European Society of Comparative Biochemistry and Physiology	Research	CR06 / CO01, CR09
October 2007 NUIG Ireland	Marie-Curie (TOK) meeting in Ireland	Research + EU representative	CR09
September 2008 Ravenna (Italy)	25th Congress of the European Society of Comparative Biochemistry and Physiology	Research	CR06 / CO01, CR09
October 2006 Sorrento, Italy	Marine Genomics Europe conference	Research + EU representative	CR13
March 2007, Eilat, Israel	8th International Marine Biotechnology Conference	Research	CR13
April 2007 Roscoff	Marine genomics Europe	Research + EU representative	CR13
June 2007, Alesnund, Norway	Fish Breeders Round Table	Research + stakeholders + EU representative	
October 2007 Chania, Greece	Marine Genomics Europe conference	Research + EU representative	CR13

Section 3 – Publishable results

Publications

Bermejo-Nogales A., Benedito-Palos L., Saera-Vila A., Calduch-Giner J.A., Sitjà-Bobadilla A., Pérez-Sánchez J. (2008) Confinement exposure induces glucose regulated protein 75 (GRP75/mortalin/mtHsp70/PBP74/HSPA9B) in the hepatic tissue of gilthead sea bream (*Sparus aurata* L.). Comparative Biochemistry and Physiology B: Biochemistry & Molecular Biology 149: 428-438.

Cairns, M.T., Johnson, M.C., Talbot, A. T., Pemmasani, J. K., McNeill, R.E., Houeix, B., Sangrador-Vegas, A., Pottinger, T.G. (2008). A cDNA microarray assessment of gene expression in the liver of rainbow trout (*Oncorhynchus mykiss*) in response to a handling and confinement stressor. Comparative Biochemistry and Physiology Part D Genomics and Proteomics, 3 (1). 51-66

Calduch-Giner J.A., Saera-Vila A., Cairns M., Davey G., Prunet P., Pérez-Sánchez J. Use of microarray technology to assess the liver stress response of gilthead sea bream (*Sparus aurata*) after confinement exposure. In progress.

Davey G, Houeix B, Calduch-Giner J.A, Talbot A, Pérez-Sánchez J, Prunet P and Cairns M. Molecular profiling of the Gilthead Seabream (*Sparus aurata*) response to chronic exposure to the myxosporean parasite *Enteromyxum leei.* In progress.

Fleury E, Fabioux C, Lelong C, Favrel P, Huvet A (2008) Characterization of a novel gonad-specific transforming growth factor-β superfamily member differentially expressed during the reproductive cycle of the oyster Crassostrea gigas. Gene, 410, 187-196.

Fleury E, Huvet A, Lelong C, de Lorgeril J, Boulo V, Gueguen Y, Bachère E, Tanguy A, Moraga D, Fabioux C, Lindeque P, Shaw J, Davey G, Reinhardt R, Prunet P, Lapègue S, Sauvage C, Moal J, Corporeau C, Boudry P, Favrel P. Generation and analysis of a large scale repertoire of ESTs from the Pacific oyster Crassostrea gigas assembled in a unique clearing-house. BMC Genomics, submitted.

Fleury E, Moal J, Boulo V, Daniel JY, Favrel P, Mazurais D, Huvet A. Identification of new genes associated with summer mortality in the oyster Crassostrea gigas using cDNA microarrays. In progress.

Le Borgne, G., S. Mativet, Pottinger, T.W., Valotaire, C., Auperin, B., Prunet, P. Expression profile ij head kidney of rainbow trout selected for high responsiveness to coinfinement stress. In progress.

Prunet P., Cairns M. T., Winberg S., Pottinger T.G.. Functional genomics of stress responses in fish. Rev. Fish. Sci. In press.

Prunet P., Auperin, B., Mativet, S., Le Borgne, G., Lepage, O. 2007. Gene expression profile analysis to detect confinement and salinty stress-related genes in rainbow trout (Oncorhynchus mykiss). Aquaculture 272 (supll. 1): S302.

Quillet E, Boudry P, Lapègue S (2007). Variabilité génétique de la réponse aux organismes pathogènes : un outil pour améliorer la santé des mollusques et poissons d'élevage. INRA Productions Animales 20(3): 239-252.

Saera-Vila A., Calduch-Giner J.A., Prunet P., Pérez-Sánchez J. Dynamics of liver GH/IGF axis and stress markers in juvenile gilthead sea bream (*Sparus aurata*) exposed to acute confinement. Lessons on the stress-specific regulation of growth hormone receptors. General and Comparative Endocrinology. Submitted.

Sauvage C., J.-F. Pépin, S. Varsamos, S. Lapègue, P. Boudry, T. Renault. OsHV-1 infection in Pacific oyster juveniles, *Crassostrea gigas*, during a summer mortality outbreak: differences in viral DNA detection and quantification related to oyster families. In progress.

Sauvage C., P. Boudry, S. Lapègue, 2008. Identification and characterization of 18 novel polymorphic microsatellite markers derived from expressed sequence tags in the Pacific oyster *Crassostrea gigas*. In press in Molecular Ecology Resources.

Sauvage C., P. Boudry, S. Lapègue, N. Bierne, 2007. Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific Oyster *Crassostrea gigas*. Gene, 406, 13-22.

Sauvage C, Boudry P, De Koning DJ, Haley C, Heurtebise S, Lapègue S*. Mapping QTL for resistance to summer mortality in the Pacific oyster. In progress.

Sitjà-Bobadilla A., Calduch-Giner J.A., Saera-Vila A., Palenzuela O., Alvarez-Pellitero P., Pérez-Sánchez J. (2008) Chronic exposure to the parasite *Enteromyxum leei* (Myxozoa: Myxosporea) modulates the immune response and the expression of growth, redox and immune relevant genes in gilthead sea bream, *Sparus aurata* L. Fish & Shellfish Immunology 24: 610-619.

Tanguy, A., Boutet, I. Moraga, D., Jollivet, D., Bachère, E., Boulo, V., Samain, JF, Moal, J., Huvet, A., Lapegue, S., Boudry, P., Favrel., P. Dossat, C., Saavedra, C., Pina, B., Reinhard, R., Canario, A., Bonhomme, F., Bierne, N., Klages, S.(2008). Increasing genomic information in bivalves through new EST collections in four species: Development of new genetic markers for environmental studies and genome evolution. Gene, 408, 27-36.
Wenne R, Boudry P, Hemmer-Hansen J, Lubieniecki KP, Was A, Kause A. What role for genomics in fisheries management and aquaculture? Aquatic Living Resources 20: 241-255.

Vogiatzi E., Louro B., T Chatziplis D., Magoulas A., C. C. Mylonas, Patarnello T., Power D. M., Kotoulas G., Canario A., Bargelloni L. & Tsigenopoulos C. S., A second generation genetic linkage map for *Sparus aurata* In progress.

Costas S. Tsigenopoulos, Jacques Lagnel, Emmanuella Vogiatzi, Victoria Pakaki, Bruno Louro, Adelino Canario, F.A.M. Volckaert, Richard Reinhardt, Georgios Kotoulas & Antonios Magoulas. *In silico* mining and characterization of simple sequence repeats from gilthead sea bream (*Sparus aurata*) and European sea bass (*Dicentrarchus labrax*) expressed sequence tags (EST-SSRs); PCR amplification, cross-species assays and polymorphism evaluation in sea bream. In progress.

Communications orales

Boudry P, 1. Gaffney P., Fleury E, Huvet A, Lelong C, Favrel P, Hedgecock D. Recent advances in the development of genomic resources in Crassostrea gigas: towards the sequencing of the Pacific oyster genome. Talk at the 31th International Society for Animal Genetics, Amsterdam, The Netherlands, July 20-24, 2008

Boudry P. Current status and prospects of genetic improvement of bivalve shellfish, with special reference to oysters. Talk at the workshop “The future prospects for aquaculture breeding in Europe”, co-organized by the the EU concerted actions “ReproFich” and “AquaBreeding”, Paris, France. October 1-3, 2008.

Boudry P. Current status of genetic and genomic research on the Pacific oyster in France and prospects for future research. Talk at the International Conference “Australasian Aquaculture 2008”, Brisbane, Australia, August 3-6, 2008.

Boudry P. Genetics and oyster mortality in France. Oyster Mortality Workshop, CSIRO, Hobart, Tasmania. August 15th, 2008.

Boudry P. Mortalité estivale et coût de la reproduction chez l’huître creuse : des approches génétiques aux approches génomiques. Talk at the workshop « Génomique et chimie bleue », organized by Europole Mer axe 1, Sea Tech Week, Brest, October 13, 2008.

Boudry P. QTL mapping in molluscan shellfish: progress, prospects and constraints. Talk at the applied training workshop for the aquaculture industry “From genomics to applications in European temperate and Mediterranean Aquaculture: Basis and prospects”. Aquagenome concerted action EU project., Paris, France. February 7-8, 2008.

Boulo V., Huvet A., Avazeri J., Bacca H., Moal J., Moraga D., Boutet I., David I., Burgeot T., Van Wormhoudt A., Gagnaire B., Renault T., Samain J.F. 2006. Functional genomic in the oyster Crassostrea gigas: contribution to high throughput genomic to explain the summer mortality phenomenon. International Workshop on summer mortality of marine shellfish. October 11-14, Busan, Korea.

Fleury E, Daniel JY, Boulo V, Lindeque P, Favrel P, Reinhart R, Mazurais D, Boudry P, Moal J, Huvet A. Talk at the 2nd Marine Mollusc Physiology Conference (‘Physiomar’), Brest, France, September 1-4, 2008.

Fleury E, Fabioux C, Lelong C, Samain JF, Favrel P, Huvet A. Identification of a novel transforming growth factor-b in the oyster Crassostrea gigas that is differentially expressed in progeny exhibiting opposed susceptibility to summer mortality. Annual conference of the National Shellfisheries Association, San Antonio, USA Feb 26 – March 2, 2007.

Huvet A, Avazeri J, Boulo V, Moal J, Bachère E, Samain JF. Identification de gènes impliqués dans la survie estivale de l'huître par la technique de macroarray: vers la compréhension du phénomène de mortalité estivale. Poster à la Journée OuestChips de Ouest Génopole, Faculté de Médecine de Rennes, France, 31 mai 2005.

Huvet A, Moraga D, David E, Meistertzheim AL, Tanguy A, Van Wormhoudt, Moal J, Jeffroy F, Sauvage C, Lapègue S, Boudry P. Polymorphisme de genes impliqués dans l'acquisition et la gestion de l'énergie chez l'huître creuse Crassostrea gigas. Talk at the 7th « Colloque national ressources génétiques », organized by the Fondation française de Recherche sur la Biodiversité, Strasbourg, october13-15, 2008.

Lapègue S, Lallias D, Sauvage C, de Koning DJ, Haley C, Beaumont A, Boudry P*. Mapping QTL for disease or stress resistance in oysters. Talk at the 31th International Society for Animal Genetics, Amsterdam, The Netherlands, July 20-24, 2008.
Lapègue S., 2008. Genome mapping in shellfish aquaculture species. Workshop "From genomics to applications in European temperate and Mediterranean Aquaculture: Basis and prospects", AQUAGENOME, Paris, 7-8 February 2008.
Saera-Vila A., Calduch-Giner J.A., Pérez-Sánchez J. Stress-mediated effect on the liver GH/IGF axis of gilthead sea bream. Functional partitioning of GH receptors. 6 Congreso de la Sociedad Ibérica de Endocrinología Comparada, Cádiz (Spain), September 10-13, 2007.

Prunet, P, Lepage, O., Le Borgne, G., Mativet, S., Krieg, F., Quillet, E. From functionla genomic to gentic: Search for genetic factors involved in stress resistance in rainbow trout. 6th International Symposium on Fish Endocrinology, Calgary (Canada), June 22-27, 2008.

Sauvage C, Boudry P, De Koning DJ, Haley C, Heurtebise S, Lapègue S*. Mapping QTL for resistance to summer mortality in the Pacific oyster. Talk at the 2nd Marine Mollusc Physiology Conference ('Physiomar'), Brest, France, September 1-4, 2008.

Sauvage C, Heurtebise S, Haley C, Boudry P, Lapègue S. Mapping QTL for resistance to summer mortality in the Pacific oyster. Talk at the Plant & Animal Genome XVI Conference, San Diego, California, USA, January 12-16, 2008.

Sitjà-Bobadilla, A. Fish immune response to myxozoan parasites: mini-review. 10th European Multicolloquium of Parasitology (EMOP 10), from satellites to microsatellites, Cité Internationale Universitaire de Paris (France), August 24-28, 2008.

Sitjà-Bobadilla, A. Living in a fish: A trade-off between parasites and the immune system. 7th Nordic Symposium on Fish Immunology, Stirling (UK), June 17-22, 2007

Lagnel J., Tsigenopoulos C. S., Kotoulas G, Canario A. and A. Magoulas Characterization of simple sequence repeats derived from expressed sequence tags (EST-SSRs) of gilthead sea bream and European sea bass; PCR assays and polymorphism evaluation. 8th International Marine Biotechnology Conference – IMBC 2007, March 11- 16 2007, Eilat, Israel, pp. 287.

Kotoulas G. Genome sequencing: scale dependent choices made and interrogation about future actions. Marine Genomics Europe Exploratory Workshop: High throughput sequencing applied to Marine organisms, a European perspective, Roscoff, France, April 15-17, 2007.

Kotoulas G. Aspects of genomics contribution to fish breeding. Fish Breeders Round Table, Alesnund, Norway, June 11-12, 2007.

Kotoulas G. Advances in marine genomics at HCMR. Marine Genomics, Workshop: An Ocean of techniques, Orthodox Academy of Crete, Chania, Greece, October 8-11, 2007.

Posters

Bermejo-Nogales A., Benedito-Palos L., Saera-Vila A., Calduch-Giner J.A., Sitjà-Bobadilla A., Pérez-Sánchez J. Confinement exposure induces glucose regulated protein 75 in the hepatic tissue of gilthead seabream (*Sparus aurata* L.). 6 Congreso de la Sociedad Ibérica de Endocrinología Comparada, September 10-13, 2007.

Bermejo-Nogales A., Saera-Vila A., Benedito-Palos L., Calduch-Giner J.A., Pérez-Sánchez J. Differential tissue expression of uncoupling proteins in gilthead sea bream (*Sparus aurata* L.): Lessons from UCP1 and UCP3 in fish. 25th Congress of the European Society of Comparative Biochemistry and Physiology, Ravenna (Italy), September 7-11, 2008.

Boudry P, Huvet A, Lapègue S, Sauvage C, Tanguy A, Moraga D, Goostrey A, Prunet P, Boulo V, Gueguen Y, Moal J, Fleury E, Samain JF, Mathieu M, Favrel P. Recent advances in EST sequencing in Crassostrea gigas: towards the sequencing of the Pacific oyster genome. Marine Genomics Europe Exploratory Workshop: High-throughput sequencing applied to marine organisms, a European perspective. Roscoff, France, April 15-17, 2007.

Boulo V, Huvet A, Lapègue S, Tanguy A, Moraga D, Goostrey A, Davey G, Le Moullac G, Samain JF, Prunet P, Moal J. Functional genomic in the oyster Crassostrea gigas : contribution of high throughput genomic to explain the summer mortality phenomenon. International Marine Genomics Conference, Sorrento, Italy, October 28 - November 1, 2006.

Calduch-Giner J.A., Saera-Vila A., Cairns M., Davey G., Prunet P., Pérez-Sánchez J. Time series analyses of sea bream (*Sparus aurata* L.) stress response after confinement exposure. 25th Congress of the European Society of Comparative Biochemistry and Physiology, Ravenna (Italy), September 7-11, 2008.

Fleury E, Fabioux C, Lelong C, Favrel P, Huvet A. Identification of a gonad-specific transforming growth factor-beta superfamily member differentially expressed during the reproductive cycle of the oyster Crassostrea gigas. 2nd International Oyster Symposium, Hangzhou, China, November 26-29, 2007.

Mativet S., P. Prunet, S. Le Guillou, F. Tiquet, E. Letellier, P. Boudinot And E. Quillet. Genetic resistance and innate response to VHSV: A transcriptome analysis in rainbow trout homozygous clones. Animal Genomics for Animal Heath symposiuym (1ABS), October 25-27, 2007, Paris (France).

Pemmasani, K., Cairns, M. T., Kane, M. Study of gene expression profiles in the liver of Rainbow trout (*Oncorhynchus mykiss*): low and high responders under
confinement stress using microarray analysis. Marie-Curie (TOK) meeting in Ireland, NUI, Galway, Ireland, October 2nd 2007.

Sauvage C, N. Bierne, S. Lapègue, P. Boudry, 2006. Development of SNPs and their use for the identification of summer mortality resistance QTLs in the Pacific oyster (*Crassostrea gigas*). International symposium of Genetics in Aquaculture, Montpellier, June 26-30, 2006.

Sauvage C, S. Heurtebise, D.-J. De Koning, C.S. Haley, P. Boudry, S. Lapègue, 2008. Mapping QTL for Resistance to Summer Mortality in the Pacific Oyster, *Crassostrea gigas*. XV Plant and Animal Genome conference, San Diego, USA, 10-18 January 2008.

Sauvage C, S. Lapègue, P. Boudry, 2005. Development of SNPs and their use for the identification of summer mortality QTLs in the Pacific oyster (*Crassostrea gigas*). 7th international meeting on Single Nucleotide Polymorphism and Complex Genome Analysis, Hinckley, Grande-Bretagne, September 22-24, 2005.

Sauvage C, S. Lapègue, P. Boudry, 2006. Development of SNPs and their use for the identification of summer mortality resistance QTLs in the Pacific oyster (*Crassostrea gigas*), 2006. XIV Plant and Animal Genome conference, San Diego, USA, January 14-18, 2006.

Sauvage C, S. Lapègue, P. Boudry, 2006. Identification of SNPs and their mapping for the identification of QTLs of resistance to summer mortality in the Pacific cupped Oyster *Crassostrea gigas* (Thunberg, 1793). International conference on Marine Genomics, Sorrento, October 28 – November 1, 2006.

Tsigenopoulos C. S., Lagnel JA, Kotoulas G, Hellemans, B., VanHoudt J.,Volckaert FAM, R. Reinhardt, Canario A. and A. Magoulas. 2006. *In silico* mining, PCR assays and polymorphism evaluation for simple sequence repeats from MGE generated expressed sequence tags (SSR-ESTs) of sea bass and sea bream. Marine Genomics International Conference, Oct.28th – Nov.1st 2006, Sorrento, Italy, pp. 118.

Pakaki V., Tsigenopoulos C.S., Kasapidis P., Lagnel J.A., Kotoulas G., Canario A. and A. Magoulas Cross-species amplification of expressed sequence tag -based microsatellites (EST-SSRs) from gilthead sea bream (*S. aurata*) to other teleosts. 8th International Marine Biotechnology Conference – IMBC 2007, March 11- 16 2007, Eilat, Israel, pp. 210.