Project n° 044481

Genomic in fish and shellfish: from research

AQUAGENOME

Instrument	Coordination actions
Thematic Priority	8.1 – Policy-oriented research – scientific support to policies

Final Activity Report

Period covered:	2007-01-01 to 2008-31-12
Date of preparation:	2009-04-31
Contract Start Date :	2007-01-01
Duration:	24 months
Project coordinator name:	**Patrick Prunet**
Project coordinator organisation name:	**INRA Rennes**

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

Table of contents

Publishable executive summary

Type of contract: Coordination actions		**Total project cost** (in euro): **893305**
Contract number **044481**	**Duration** (in months) 24 Months	**EU contribution** (in euro) **794040**
Start date 01 January 2007		**Period covered by the project** 01 January 2007 – 31 December 2008

PROJECT COORDINATOR

Name :
Dr Patrick PRUNET

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

Telephone : **(33)223-48-50-02**	**Telefax :** **(33)223-48- 50-20**	**E-mail address :** Patrick.Prunet@rennes.inra.fr

World wide web address: http://genomics.aquaculture-europe.org/index.php?id=41

List of participants

INRA (Coordinator) : Institut National de la Recherche Agronomique
Responsible scientist : Patrick PRUNET, director,INRA/SCRIBE research unit, campus de Beaulieu, 35042 RENNES Cedex, France.
Tel : 33-(0)-223-48-50-02
Mobile : 33-(0)-631-96-59-23.
Mail : prunet@rennes.inra.fr

UOS : University of Stirling
Responsible scientist : Prof. Brendan McAndrew Head of Fish Genetics and Reproduction Section. Institute of Aquaculture.
Mail: b.j.mcandrew@stir.ac.uk

GFP: Genesis Faraday
Responsible scientist : Christopher Warkup, Chief Executive, Roslin BioCentre, Midlothian, Scotland, UK.
Mail: chris.warkup@genesis-faraday.org

HCMR: Hellenic Centre for Marine Research
Responsible scientist : Giorgios KOTOULAS, Hellenic Centre for Marine Research Thalassocosmos, Former US base at Gournes, Gournes Pediados, 71500 Heraklion, Crete, Greece.
Mail: kotoulas@her.hcmr.gr

SYSAAF: Syndicat des Selectionneurs Avicoles et Aquacoles Français
Responsible scientist : Pierrick Haffray, SYSAAF, SCRIBE, campus de Beaulieu, 35042 Rennes cedex, France.
Mail : pierrick.haffray@rennes.inra.fr

UGOT : Zoological Institute of Göteborg University
Responsible scientist : Kristina Sundell, Fish Endocrinology Laboratory, Department of Zoology, Zoophysiology, Göteborg University, Göteborg, Sweden.
Mail: k.sundell@zool.gu.se

CCMAR : Algarve University

Responsible scientist : Deborah Power, Comparative and Moleculare Endocrinology, Universidade do Algarve, Centro de Ciências do Mar, 8000-117, Faro, Portugal.
Mail: dpower@ualg.pt

IMR : Institute of Marine Research

Responsible scientist : Gier-Lasse Taranger, Head of Research Group/Research Program Manager
Institute of Marine Research, PO Box 1870 Nordnes, 5817 Bergen, Norway.
Mail: geir.lasse.taranger@imr.no

CSIC: Insituto Investigaciones Marinas

Responsible scientist :Antonio Figueras, nstituto Investigaciones Marinas. CSIC. Spanish National Reference Laboratory for Mollusc Diseases , Eduardo Cabello 6, 36208 Vigo, Spain.
Mail: rodaballo@teleline.es

VNIRO: Russian Federal Institute of Fisheries and Oceanography

Responsible scientist : Nikolai Mugue, Russian Federal Research Institute of Fisheries and Oceanography, Department of Molecular Genetics, 107140, 17, V. Krasnoselskaya str., Moscow, The Russian Federation.
Mail: mugue@mail.ru

Publishable Executive Summary

In recent years, a relatively large number of projects investigating the biology of aquaculture species through genomic approaches have been funded both at European levels and national levels. These have generated an extensive core of genomic tools and expertise that is spread among European research laboratories. However, this scientific resource is not coordinated within research and transfer to the European aquaculture industry is still in its infancy. In this context, the objectives of AQUAGENOME are (i) to improve coordination of the ongoing and future national and international research projects in the field of genomics in fish and shellfish European aquaculture and identify bottlenecks in the diffusion of genomic approaches within research laboratories (ii) to enhance the transfer of information and knowledge towards the aquaculture industry.

In order to reach these objectives, the Aquagenome project has carried out the following tasks:

- Inventory existing genomic resources and identify and evaluate new bioinformatic resources, genomic information and tools that are vital for the development of European aquaculture.
- Development of new bioinformatic tools aiming to improve a major problem faced by genomic approaches, i.e. poor quality of the annotation of the fish and shellfish genomic resources.
- Achievement of a benchmarking exercise on a genomic approach which is developping in many fish and shellfish species, i.e. transcriptome analysis using conventional microarray and new high-throughput sequencing technology.
- To identify specific research domains in which genomic approaches should be developed in order to support the European aquaculture industry.
- To support exchanges of scientists and genomic resources between members of the AQUAGENOME Associated Partner Network through the distribution of grants.
- To improve tranfer of genomic information toward aquaculture industry:
- establishment of a network of contact with aquaculture stakeholders (ii) achievement of two workshop associated with site visits for aquaculture stakeholder on ‘best practice’ and ‘best technologies’ for genomic approahces (iii) achievement woth aquaculture stakeholders of a consensus Technology Road-Map for the sector that prioritises short, medium and long term research and knowledge transfer needs.

These achievements required the use of a number of instruments, e.g. establishement of network of contact with research laboratories working on genomic and aquaculture industry, development of a project website for providing database and information on fish and shellfish genomic but also for external communication of the project, working groups to gather experts in specific domains, organisation of workshops.

In addition to all the achievements described above, the Aquagenome project allowed us to suggest a limited number of recommendation which should improve coordination and transfer of genomic research toward aquaculture industry. Dissemination of these conclusion toward the European community will be engaged through active participation of Aquagenome partners to EATIP and FABRE technology platforms in close link with Aquabreeding and Reprofish EU-funded projects.

All information related to Aquagenome project can be found at the project website: www.aquaculture-europe.org

Overview of general objectives.

Background.

In recent years major investments have been made in understanding the biology of aquaculture species through genomic approaches. The approaches proposed are diverse and cover a large variety of questions related to fish and shellfish aquaculture production. Thus, important issues related to safe and healthy food for optimal growth, selection for a desirable trait, fish welfare in aquaculture production, vaccination against major disease threats, reduction of the negative effects of aquaculture on environment have been raised. These issues need to be addressed within a sustainability perspective, i.e. through interactions with societal development, implications for biodiversity and environment and through contribution to economy growth.

Initiated 15 years ago with the Human Genome project, the development of genomics has more recently reached fish and shellfish species. The major benefits of genomic approaches are linked to the capability to drive a much broader and systematic analysis of the genome and its expression products. Such perspectives are profoundly modifying our ability to understand biological processes related to the issues raised above and to propose solutions. As such, many research projects have started to look at using genomic approaches to solve industrially relevant problems.

Research projects incorporating genomic tools require considerable resources and funding which have been made available through the European Commission, national agencies and on a few occasions through industry. This has led to the development of a large number of genomic tools and has increased associated research expertise. Although overlap between these projects is limited, it has become clear that the coordination of scientific activities related to fish and shellfish genomics and future actions among them is now necessary. For example, various projects aim to use similar genomic tools but their exchangeability has not been considered. The main drawback will be the development of redundant or under-used resources and a lack of efficiency of these funded research projects. In order that Europe can compete with other regions that are developing genomic approaches in aquaculture, it is vital that the work carried out here is coordinated to a high level.

At the same time, implementation of results from fish genomics within the European aquaculture industry is still at its infancy. Several reasons may explain this situation, e.g. complexity, lack of knowledge, and lack of applicability and usefulness, have often been perceived as stumbling blocks. However, this coordinated action would put effort into transferring the genomic results of aquaculture research to the industrial producers. A recently EC-funded project, AQUAFUNC, has started to tackle this issue. The overall aim of the AQUAFUNC project is to integrate the outcome of an EC-funded project through a common publication of the outcomes with the creation of a common web-page presenting projects and providing links to respective instituitons and research. This SSA project was a first step for acting on this integration but an overview of what is presently under development at a international level on animal genomics clearly indicates that we have to advance more in the coordination of these research projects within aquaculture.

Scientific and technological objectives.

A relatively large number of projects investigating the biology of aquaculture species through genomic approaches have been funded both at European levels and national levels. They have generated an extensive core of genomic tools and expertise that is spread among European research laboratories. In the European context previously described, the objectives of the AQUAGENOME are two fold:

- to coordinate the ongoing and future national and international research projects in the field of genomics in fish and shellfish European aquaculture and support diffusion of genomic approaches within research laboratories.

- to enhance the transfer of information and knowledge accumulated in research laboratoies towards the aquaculture industry to assist in solving their specific problems and strengthen the transfer of knowledge and technology.

In order to reach these objectives, the following aims will be achieved:

- To gather an inventory of the genomic resources, tools and expertise developed by European research groups and the needs of the research community that are working on fish and shellfish (to complement the work of the AQUAFUNC project).
- To identify new bioinfomatic resources, genomic information and tools that are vital for the development of European aquaculture and test a strategy for optimizing access of researchers to genomic resources.
- To identify specific research domains in which genomic approaches should be developed in order to support the European aquaculture industry.
- To support exchanges of scientists and genomic resources between partners associated with the AQUAGENOME project.
- To determine the actions to carry out in order to apply genomic tools and knowledge effectively and sustainably in aquaculture.

To reach these objectives, the AQUAGENOME project addresses these issues starting from the point of view of the research. Thus the project develops its activities according to a general strategy which will start from the coordination of research activities and logically advance to finally end with dissemination activity of genomic technology towards aquaculture industry. Along this line, the project is developped within 4 parts:

- Part 1 analyze the available genomic resources and expertise in order to optimize its accessibility and its use by researchers. This task which includes setting up and testing of common functional genomic database.
- Part 2 brings together research groups working on fish and shellfish genomic with the aim to identify further research needs to support aquaculture industry.
- Part 3 is an exchange program to extend the analysis and actions carried out in WP1 and WP2. This WP provides grants for visiting young scientists and for the exchange of genomic resources.
- Part 4 fosters the application of basic research results by the aquaculture industry sector. Within this part, we determined what actions will be carried out for effective transfer of knwoledge.

The instrument used:

These achievements require the use of a number of instruments, e.g. working groups to gather experts in specific domains, organisation of workshops where partners associated with the project can contribute to the project objectives, writing of technical reports, creation of a website to share information and provide links, development of a database and software to broaden access to genomic resources, distribution of grants to support exchanges. Another important aspect in this project is the ambition to involve a large representation of the European community interested in fish and shellfish genomic research. This cover all the research groups involved totally or partially in this kind of approach but also structures associated with aquaculture production. This representation has important roles during the life of the project: (1) They are consulted during the tasks carried out by WP1 (2) they are invited to the workshop organised by WP2 and be important actors in the identification of research needs (3) they will apply for grants within the exchange program (3) they nominate representatives who will participate in the various actions developped in WP4. As such, the project has created a network called **Aquagenome Associated Partner Network (A^2PN)** which gather as many research groups involved in fish and shellfish genomics as possible.

Part 1: Existing genomic resources and future developments

During these last years, many European and national programs examining various aspects of fish and shellfish genomics have been funded. Most of the time, these projects were aiming to answer/solve a specific question using genomic approaches but national projects have also supported programs aiming to develop genomic resources. Although these resources do exist, they are numerous and this has complicated the integration and comparison of results generated by different projects. In this context, it appears fundamental that access to these bioinformatic resources should be shared within the scientific community working on fish and shellfish genomics. The objectives of this first part was: (1)To bring together European research groups including national projects which have been developing genomic resources and expertise in order to give access to the whole community by exchange or interconnection of this knowledge and tools (2) to analyze the needs in term of genomic resources of users of fish and shellfish genomics which would require the combined effort of the community (3) to propose new developments of genomic resources with an appropriate strategy. This will include implementation of improved accessibility to bioinformatic solutions for common analysis of genomic data.

1. Inventory of existing aquaculture related genomic resources.

Introduction: Many European and national research projects have funded genomic researches and resources in many aquaculture species. As a direct consequence these numerous genomic tools are now dispersed over many laboratories in Europe. The objective of this project action is to gather as much as possible the information on these existing resources, on one website and to make it available to everyone. This information was also used in a second action to assess the needs on the most important genomics tools that should be prioritized for the development of research on European aquaculture species.

Methods: For this apporach, the first step was to identify all genomic tools that might be developed, regardless to the species. We then identify **22 types** of genomic resources: *complete genome sequence, BAC library, cDNA library, radiation Hybrid Panel, SNP marker, microsatellite marker, AFLP marker, candidate genes marker, microarray gene expression, microarray genotyping (SNIP), segregating families, special genetic background, transgenic line, physical map, genetic or linkage map, physical/genetic maps integration, QTL map, QTL detection, radiation hybrid map, bioinformatics database, bioinformatic service, bioinformatics tool*.
The second step was to select which species to be focus on. This choice was made with the consultation of all project partners and **13 species** of European importance for aquaculture were selected: *Carp, Clam, Cod, Crassostrea gigas, Crassostrea virginica, Mussel, Ostrea edulis, Salmon, Seabass, Seabream, Senegalese sole, Trout, Turbot.*
The final step was to setup a table with the possible resources for each species, begin to fill in with information available over the internet, and afterwards contact directly (by phone and email) a referent scientist for each species. **20 scientists** were reached.

Results: All results are stored and made available in a relational database in the Aquagenome Website (http://www.aquaculture-europe.org/, login Aquagen/ password Aquagen), which can be queried using the website. We identified on the 13 species investigated a total of 376 genomic resources, 45 projects referring to these resources, and 223 scientists from 124 different laboratories over 28 countries related to these resources. All species have at least some resources in genomics. Trout is the most represented in our database with 60 different resources, but this is also one of the few

species where we have been able to reach numerous referent scientists. One group gathers species with quite a lot of genomic resources: Trout, Salmon, Seabream, Seabass and *Crassostrea gigas*. And another one containing species with only a few numbers of genomic resources: Clam, Carp, Cod, *Crassotrea virginica*, Mussel, *Ostrea edulis*, Senegalese sole and Turbot.
Based on the list of the 22 possible resource types, we analyzed how many types of genomic resources are available for each species over the 22 possible.

A more detailed view is offered in table 1 (extracted from the White Paper, pp.11) which represents the existing type of resource (in blue) or to be planned (in light blue). From that analysis it is obvious that not only one species has its complete genome sequenced. This resource is however underway for seabass, and planed for Salmonids. The most common resource is cDNA libraries. Working on EST is the first step to study a genome and the new sequencing technology development will further make easier such EST sequencing approach in a number of additional species important for the aquaculture sector. Genetic maps are also developed in numerous species, but their quality depends on the number of markers available, and this might be variable from one species to another (data not collected now in the resource information gather within the Aquagenome Website).

Table 1. State of the art of the genomic resources available in important European aquaculture species. Non existing resources: open squares and coefficient 0; Resources ongoing or planned: blue squares and coefficient 1; Existing resources: black squares and coefficient 2.

SPECIES / **RESOURCES**	Crassostrea virginica	Clam	Ostrea edulis	Mussel	Senegalese sole	Crassostrea gigas	Cod	Carp	Seabream	Turbot	Salmon	Seabass	Trout
cDNA library, ESTs sequences	2	2	0	2	2	2	2	2	2	2	2	2	2
Microarray, gene expression	2	0	0	2	2	2	2	2	2	2	2	2	2
Genetic, linkage map	2	0	2	2	2	2	2	1	2	2	2	2	2
Bioinformatics database	0	0	0	2	2	2	2	2	2	2	2	2	2
BAC library	2	0	0	0	0	2	2	2	2	1	2	2	2
SNP marker	2	0	0	0	0	2	2	1	2	2	2	2	2
Microsatellite marker	2	2	2	0	0	0	2	2	2	2	2	2	2
Bioinformatics tool	0	0	0	2	2	0	0	2	2	0	2	2	2
Bioinformatic service	0	0	0	2	2	2	0	2	0	0	2	2	2
QTL detection	0	0	0	0	0	0	0	0	0	2	2	2	2
QTL map	0	0	0	0	0	0	0	0	0	2	2	2	2
AFLP marker	0	0	2	0	0	0	0	1	0	2	0	2	2
Candidate genes marker	0	0	0	2	0	0	0	1	0	2	0	2	2
Physical map	0	0	0	0	0	0	0	0	0	0	2	2	2
Special genetic background	0	0	0	0	0	2	0	0	0	0	0	2	2
Radiation Hybrid Panel	0	0	0	0	0	0	0	0	2	0	0	2	1
Segregating families	0	0	0	0	0	0	0	0	2	2	0	0	2
Radiation hybrid map	0	0	0	0	0	0	0	0	2	0	0	0	1
Transgenic lines	0	0	0	0	0	0	0	0	0	0	0	0	2
Complete genome sequence	0	0	0	0	0	1	1	0	0	0	1	1	1
Physical/genetic maps integration	0	0	0	0	0	0	0	0	0	0	1	0	1
Microarray, genotyping (SNP)	0	0	0	0	0	0	0	0	0	0	0	0	0

Conclusions: This first inventory of existing genomic resources in European important Aquaculture species reveals that these resources are growing rapidly as none of these important species is devoid of any genomic resources. The main lack in all these species is probably the availability of complete genome sequences, bother genomic resources are still currently lacking in important aquaculture species such as the integration of physical and genetic maps and the development of microarrays for SNP analysis. These microarrays are however directly link with the availability of whole genome sequences. The most common genomic resources are cDNA libraries and the availability of expressed sequence tags (ESTs) sequences. One important lack resulting from this resource is the current lack of common gene annotation in all these species. This is of main importance as these cross-species annotations would allow comparative genomic analysis and potential transfer of result from one species to another (mostly microarray results). All these information are available on the project website.
Unfortunately, the benefits of such inventory rely on the ability to maintain these tables updated: With amazingly rapid devlopment of new technologies around genomic approaches, that information become quickly exceeded and then loose much of its interest. It would be necessary to find a solution for keeping this inventory and the information of the whole website updated through another European initiative.

2- Improvement of genomic resources.

In agreement with the main conclusions of the Bergen Meeting devoted to the preparation of the White Paper (see Part 2), we have decided to focus our activity on the exploration of new approaches to improve gene annotation for aquaculture relevant species. This was a major issue for improving the quality of the fish and shellfish genomic resources. This has been done by two parallel approaches. We first investigated if high throughput phylogenomics could be used to functionally and phylogenetically annotate fish genes. The second approach, has investigated the gene annotation quality in the common carp and the impact of massively parallel new sequencing approaches on this annotation

2.1. Phylogenomics and gene annotation

Introduction In nearly all aquaculture species this simple homology search approach suffers from i) the lack of well annotated reference genome sequences especially in the context of the huge evolutionary distance between some of these species, ii) the quite few number of available DNA sequences in these species and iii) the extra whole genome duplication that is observed in some species (in the case of teleost fishes for instance). Due to these reasons blast search in aquaculture species often returns very bad quality or even false information in term of annotation. Therefore most of these homology annotations should only be considered as a best homolog bet.

Results: To try to overcome this problem we tested a phylogenomic approach to annotate ESTs in teleost fishes. For that we first used an already available pipeline called Figenix that has been developed by the phylogenomics laboratory (University of Provence, Marseille) by P. Pontarotti and colleagues. This pipeline is available online at http://sites.univ-provence.fr/evol/figenix/index.html and it retrieved sequences, provided multiple sequence alignments, performed phylogenetic reconstruction, and deduced orthology and paralogy relationships (for a detailed description of the pipeline and models used, see Gouret *et al*., 2005. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 6:198). In order to get a better representation of fish sequences we *de novo* created a new database made by the 6 frames translation of all ESTs contigs in all fish species (3.3 10^6 ESTs available in dbest release of July 12, 2007). This database was merge with the Ensembl database of deduced proteins from fish genomes (available in medaka, zebrafish, fugu, tetraodon and stickleback) and the Swissprot protein database on all species. Using these merged databases we run 100 phylogeny analysis using as sequence bait proteins deduced from

rainbow trout EST contigs. These analysis revealed that this strategy was in most case good enough to provide an accurate annotation of the bait proteins.

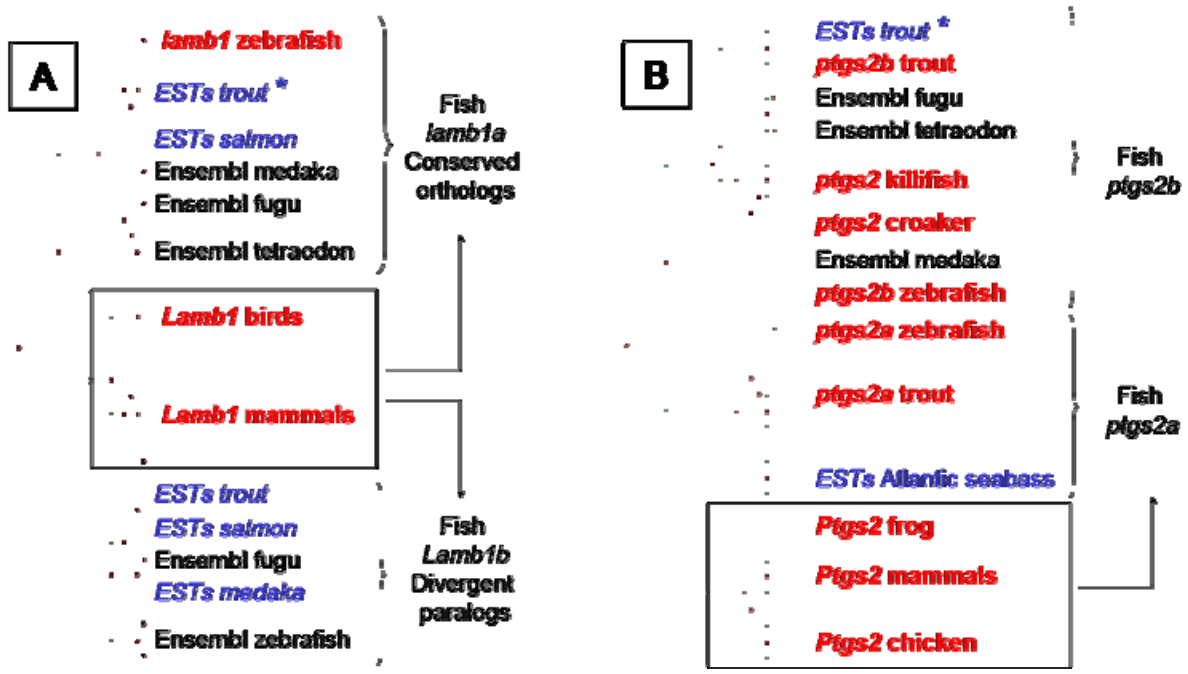


Figure 2. Two characteristic examples of gene annotation using a phylogeny approach. These two cases illustrate two different tree topology with in each case a specific fish gene duplication. In red bold types are represented proteins with an annotation in databases (mainly Swissprot), in black underlined are represented proteins deduced from fish genome sequences (Ensembl) and in blue proteins deduced from the 6 frame translation of ESTs contig assembly. The unknown protein used as a bait for running the analysis is marked with asterisk. In panel A the phylogenetic analysis returns an annotation with the laminin, beta 1 protein (lamb1) and in panel B with prostaglandin-endoperoxide synthase 2 protein (ptgs2). See text for details.

In the figure above (Fig. 2) we illustrated two characteristic examples of the resulting phylogenetic trees that we obtained. In these two cases protein annotations can be deduced from the analysis of the tree topology. In panel A, the unknown protein deduced from the EST contig used as a bait clearly belongs to the laminin beta 1 (lamb1) branch with at least one fish protein annotated accordingly (zebrafish lamb1) and birds and mammals lamb1 proteins branching at the root of the fish node in agreement with their position in the tree of life. By inference, we can then also annotate all non annotated deduced proteins within the same branch including proteins deduced from genome sequences. Interestingly, another branch roots at the basal node of this vertebrate lamb1 branch (including fish, birds and mammals lamb1). This branch gathers proteins with an unknown annotation, but with very clear lamb1 protein annotations when blast homology search are carried out. This topology actually reflects a gene duplication (in that case most probably due to the whole genome teleost specific duplication) followed by dramatic divergence of these sequences. Such a high rate of molecular evolution appears to be quite frequent for many genes resulting from fish genome duplications and these duplicated genes are often associated with functional shifts through either sub- or neo-functionalization. We then could now refine gene annotation for that lamb1 group in fish by naming the first branch as lamb1a corresponding to so-called “lamb1 conserved

orthologs", and the second branch as lamb1b corresponding to so called "lamb1 divergent paralogs".
This clearly show that this approach is much more precise for gene annotation than the simple blast search homology approach. An additional gain of this strategy is to be able to identify the potential functional shifts resulting from genome duplications. In panel B, the fish specific duplication in branched in agreement with the tree of life (i.e., the two prostaglandin-endoperoxide synthase 2 proteins (ptgs2a and ptgs2b) in fish branch together and are rooted by the higher vertebrate ptgs2). In that case the phylogenomic approach also allows a much precise annotation and could also be used to correct bad or incomplete gene annotation (in that example for instance annotation of the killifish and croaker ptgs2 are incomplete and should be precise as being ptgs2b instead of ptgs2).

Conclusion: This first approach was made to investigate the power of this phylogenomic annotation approach. The results we obtained are quite promising and show that this approach could be used at least for fish and would be much more precise especially in the correct assignment of gene annotation for duplicated genes. However it is still quite difficult to implement this strategy on a large scale dataset especially due to the current need of very demanding resources in term of computational time.

2.2. Phylosig: a first development towards high throughput EST annotation using phylogenomics.

Introduction: As presented above gene annotation is not a simple task when dealing with species having both no closely related good genome annotation and a addition whole genome duplication. The usual alignment strategy is often not accurate enough to separate hits coming from the duplicated genes or even the different members of a gene family. The exploratory work presented below aimed to study the feasibility of another strategy based on a high throughput phylogenomic approach. The phylogenomic approach is usually labor intensive on the computational side especially for the tree generation as much as on the human curation side. The idea is to use a robust pipeline based on several methods for the tree building steps in order to produce the final consensus tree. The most computational intense part of the pipeline are using a PC cluster in order to parallelize the processes. This part will present in details the different steps of the pipeline as well as some preliminary results obtained.

Results: The first step of the pipeline aims to retrieve known orthologous proteins for a given EST contig. The pipeline has two main inputs, the first is the set of sequences usually EST contigs to be annotated the second is the database containing all the known sequences build to find orthologous proteins in different closely or not so closely related species. This second input is more often than not build in order to maximize the number of possible orthologous proteins. In the example presented below the protein database included Swissprot and all the Ensembl fish proteins.

> The second step performs a multiple alignment of the retrieved sequences.

> The third step of the pipeline simplifies the multiple alignment to gain robustness for the tree building step. The simplification of the alignment is performed using Gblocks.

> The fourth step performs a multiple alignment of the simplified sequences

> The fifth step calculates a consensus tree using the three trees generated before and checks also the correspondence between the consensus tree and the tree of life in order to determine if the sequence comes from a speciation or a duplication event.

The pipeline is able to run on the local PC cluster for the most computer intensive steps such as the alignment to retrieve the sequences, the multiple alignment and the tree building steps. The source code and install documentation can be found following this link:

http://mulcyber.toulouse.inra.fr/gf/project/phylosig/docman/?subdir=84

The test results on two sets of contigs shown below (Fig. 3)indicate that this approach is able to provide new information to the biologists when compared to the classical simple alignment best hit annotation. It seems also that the gain of information is more important on lately duplicated species.

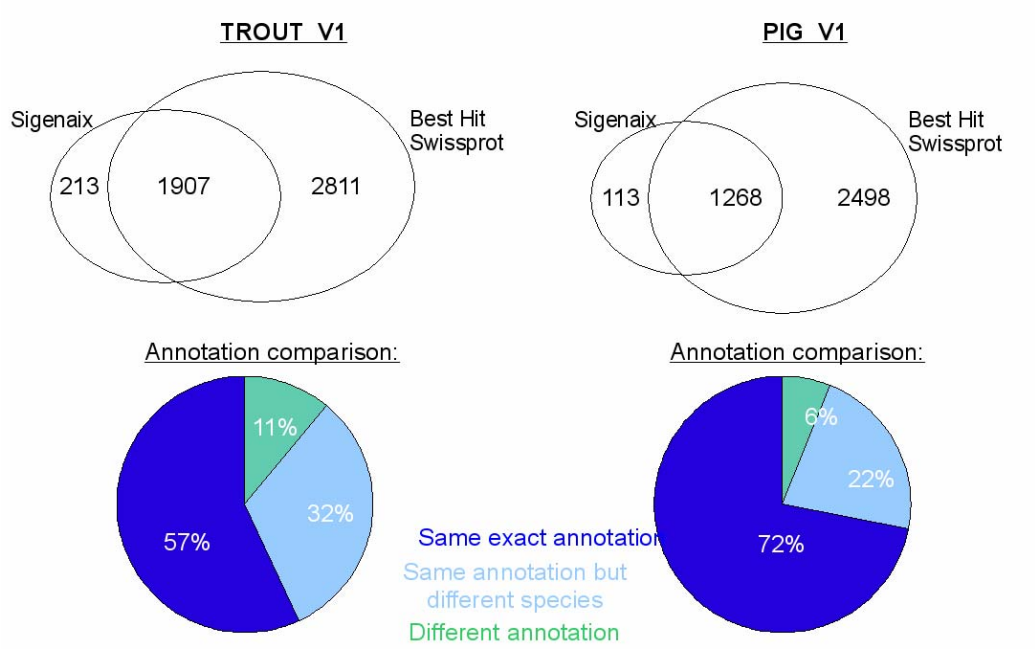


Figure 3. This figure presents results comparing the traditional best hit annotation strategy versus the phylogeny annotation strategy. Results for two ESTs contig sets are presented here in order to compare a well annotated species (Pig) with a not so well annotated species (trout). With the phylogeny annotation strategy less EST contigs are annotated but the annotation is different from the one provided by the best hit annotation strategy. The difference seems to be more important for not so well annotated species..

Conclusion: The first test carried out using that pipeline indicated this new tool is efficient in providing a high throughput phylogenic analysis, but the analysis is less robust especially in term of the number of genes well annotated than with the Figenix pipeline. A combination of the two pipelines would probably be wise with the massively parallel approach of Phylosig combined with the robust annotation of Figenix.

2.3. Annotation of a Carp gene database and impact of new sequencing approaches.

Introduction: As a part of inventory of aquaculture genomic recourses and identification of existing bottlenecks, we have taken as a case example one existing microarray datasets that has been developed for carp by University of Liverpool genomic center. The approach proposed in that part was to dissect quality of microarray annotation to reveal factors which cause insufficient annotation of existing genomic platforms in aquaculture.

Results:

Gene annotation of common carp with focus on duplicated genes (paralogs).

Common carp (*Cyprinus carpio*) is a recent tetraploid species (2n=100) and the result of this recent whole genome duplication is seen in carp by a high number of closely related duplicated genes (paralogs) with or without a known diverged function. In carp that is at an early stage a re-diploidization this process of "gene speciation" is quite under-explored.

We have manually analyzed by multiple homology alignments in carp 165 pairs of paralogous genes, represented by more than 10 EST sequences each. The zebrafish homologous sequence was used as an outgroup for this analysis. Tissue origin of EST sequences and Kn/Ks ratio (synonymous/nonsynonymous substitutions) have been studied when both paralogs presented as an EST sequences.

This analysis revealed that:

1. 75% of duplicated carp genes are expressed from both paralogs. Average divergence between paralog sequence is only about 3%, which indicates the recent allotetraploid origin of carp.
2. About one-quarter of all EST clusters do not reveal expression of one of the paralogs. This could indicate a rapid silencing of one copy, which can be explained either by coping with "dose-effect" or by relaxed selection acting on one copy of duplicated gene and transition of one paralog into pseudogene.
3. Some paralogs indicate uneven tissue-specificity in expression (one paralogous gene expressed in one tissue while the other paralog is expressed in different tissues.
4. Some paralogs indicate high dN/dS ratio, indicating that these genes are under selection.

Our results indicate an important identified gap in microarray annotations of this tetraploid fish species. Clones from duplicated genes have high sequence similarity, and therefore are often merged into the same group (cluster) during automated gene annotation procedure. However, differential gene expression regulation of duplicated genes can possibly be an important factor in adaptation of tetraploid organisms, and should be identified and studied as separate genes. High Kn/Ks ratio, as well as tissue-specific expression of paralogs indicates evolutionary plasticity of duplicated genomes.

Automatic annotation was implemented based on a BlastX algorithm (translated nucleotide sequences against protein databases). Sequencing errorscan significantly impact the deduced proteins from the translated nucleotide sequence, often causing failure in correct gene assignment. To avoid such methodological pitfalls, direct comparison of mRNA sequence against reference mRNA sequences of other organisms could be implemented. In order to improve the quality of gene annotation, we have performed comparative analysis with current release of NCBI mRNA reference sequences databases for three best annotated teleost species – Danio, Fugu and Tetraodon. BLAST analysis of 32000 carp EST sequences against *Danio rerio* RefSeq database revealed that 26088 carp ESTs (81%) have highly similar sequence among protein-coding genes known in zebrafish. Number of zebrafish genes identified as carp homolog is 16034. However, additional Blast analysis with other teleost species (Fugu and Tetraodon) did not improve significantly carp EST database annotation: number of carp cDNA clones which have
high sequence similarity to Fugu or Tetraodon genes, but do not have counterpart in Danio is rather low (355).

Impact of new 454 sequencing technology for the improvement of Carp gene annotation.

This study have been developped using data and experiments carried out within the EUROCARP project funded by EC. Within this project, the new version of Carpbase became

available in 2008. In 2008 the new version of Carpbase became available. This version (Carpbase 7) differs from previous versions as it contains not only contigs from Sanger sequencing but also hybrid contigs (containing data sequenced using 454 sequencing technology and Sanger clones). A total of 388,514 reads from 454 and 19,065 Sanger sequences have been deposited in GenBank. They now form 71051 clusters. The addition of 454 sequence data significantly improved the robustness of existing carp genome annotation. Hybrid contigs that have no match to Sanger sequences may be used to provide a resource for improved annotation for sequences from Carp and to be integrated into reconstruction of metabolic and regulatory networks for *Cyprinus carpio*. Despite that 454 reads are generally shorter that Sanger sequences (50-400bp with average 200 bp), the high number of sequences from the same gene allowed a better quality of assembly and in turn generate longer contigs sequences.

Within this analysis we found that 454 reads after proper trimming give much clear reads with high fidelity. To illustrate this, we manually analyzed all reads, originated from mitochondrial DNA (mtDNA) genes. Because complete carp mtDNA sequence was added to the assembly, all sequences of mtDNA genes are supposed to join the same cluster, containing mtDNA reference sequence. If assembled sequences have sequence errors, it could prevent them from clustering with the mtDNA cluster. 92,15% of the sequences are correctly assign with mtDNA cluster using the 454 sequences and only 7,24% of the sequences are correctly assign with mtDNA cluster using the Sanger sequences. This very low number of Sanger sequences that are correctly assigned to the carp mtDNA cluster is explained by the insufficient trimming and/or the low quality of end-reads of Sanger sequences.

The high quality of the resulting contigs based on large number of 454 reads and long Sanger clones allowed to reveal carp paralogous genes by pairwise alignment of cluster consensus sequences. Manual analysis of these sequence pairs revealed that 1) in most cases both sequences match the same single copy gene in *Danio rerio*, and b) sequence divergence is approximately 3%, which correspond to average divergence in carp paralogs.

Conclusion: From this carp gene annotation we can say that a large proportion (up to 50%) of genes do not have a proper gene annotation and this is representative of most aquaculture species, therefore not allowing a proper functional genomic annotation. The development of teleost-specific gene orthology clusters (cross-linked all model fish species and all EST databases developed for aquaculture species) would be useful genomic tool for improvement of gene annotation and for the better understanding of teleost system biology. A significant improvement of existing gene annotation is however very easily possible by the incorporation of a large number of sequences (short read of 200 pb) generated by 454 technology into existing EST databases based on Sanger EST sequencing data.

2.4. Conclusion on the imporvement of genomic resources and annotation issues.

Based on the previous studies described above, we suggest two very complementary approaches that would probably solve the major problem of gene annotation in aquaculture species. First, taking advantage of the new technical developments in terms of large-scale sequencing projects using the relatively affordable and rapid next-generation DNA sequencing methods it is judged high priority to support the production of large transcriptome sequencing projects in a large array of aquaculture species (fish and shellfish). Second, in parallel and using these extended transcriptome databases we suggest to develop research in the field of comparative phylogenomics approach in order to annotate genes in aquaculture species taking into account their genome complexity and evolution distance.

Part 2: State-of-the-art and future of genomics in aquaculture research.

1- Introduction: Many research areas in the aquaculture domain have benefited during the last few years from genomic approaches. An overview of the out-put from EU-funded and national projects on fish and shellfish sustainable aquaculture is presently being carryed out by the AQUAFUNC project. However, the state-of-the-art of the application of genomic approaches in various key research-domains of European Aquaculture is lacking, as well as a critical appreciation on the opportunities for each of these domains in using genomic resources to improve the knowledge base and so solve key industrial problems. Moreover, both knowledge are fragmented among research groups, and across different scientific domains within aquaculture – and further networking activities are needed to make these knowledge available across research groups and research domains in aquaculture and genomic research.
In order to answer these issues, we have build on - and extend - an initial analysis by the AQUAFUNC project surveying the state-of-the-art of functional genomics in sustainable aquaculture as developed by recent European research projects. In order to get a broader picture on the state-of-the-art of genomic applications in European aquaculture research, we have started from an a priori list of research topics/domains that are highly relevant for aquaculture and to which genomic approaches may bring substantial new information/answers and we have ended with a prospective analysis based on a compilation of recent developments and state-of-the-art for the key research domains in aquaculture, and suggest direction for future research to solve aquaculture problems with genomic approaches.

This work aims at the identification of further specific research needs in order to support industrial exploitation and at providing recommendations of future research at the European level to the Commission. This analysis includes the major research domains of fish and shellfish biology of high relevance to the key challenges in the aquaculture industry, but will not be an exhaustive list of all research carried out on these species.

2- Method : In order to carry out this task, the following strategy has been developped:

- Integration adn extension of the initial analysis developped by Aquafunc project which was aiming to survey the state-of-the-art of functional genomic in sustainable aquaculture as developped by recent EC-funded resaerch projects.
- Integration of the survey on the available genomic resources in fish and shellfish presented in Part 1.
- Identification and invitation of key-note speakers and participants to a two-days meeting organized in Bergen. This task was mainly developped by advertising and exchanging with the AQUAGENOME associated partners network (A2PN) initially set up when constructing the proposal. Since that time, several new members have been added in this A2PN list in order to provide a large representation of the European research community working on aquaculture species.

Organization of a two-day workshop was held in Bergen, Norway, in cooperation with AQUAFUNC on September 20-21 2007 with 63 participants from 11 countries. At day one 11 key-note presentations was given by invited leading scientist in each field covering:

Aquafunc review	Kristina Sundell (S)
Status genomic tools in aquaculture species	Luca Bargelloni (I)
Genomics in Host-patogen interaction	Simon Mackenzie (ES)
Genomics in Breeding	Sigbjørn Lien (NO)
Environmental adaptions, stress and genomics	Lluis Tort (ES)
Reprodutive genomics	Jean Jacques Lareyre (F)

Genomics of Growth and nutrition	Michael J. Leaver (UK)
Toxiogenomics, safety and product quality	Pål Olsvik (NO)
Genomics in mollusc Aquaculture	Beatriz Novoa (ES)
Genomics in monitoring environmental impact	Geir Dahle (NO)
The use of Bioinformatics in Aquaculture	Aleksei Krasnov (NO)

On day two of the work-shop, five working groups were established resulting in a short review of the state-of-the-art and suggested future directions and priorities within each of the topics:

1. Genomics of host-pathogen interaction (Moderator Antonio Figueras)
2. Genomics to study ecotoxicological, environmental and husbandry effects on stress, welfare and product safety & and environmental impact (Moderator Patrick Prunet)
3. Genomics in reproduction and breeding programmes (Moderators Pierrick Haffray and Geir Lasse Taranger)
4. Genomics of Growth and Nutrition (Moderator Brendan McAndrew)
5. Genomic tools and bioinformatic resources (Moderators Yann Guigen and Frank Nilsen)

The conclusions from the 5 working groups were presented and discussed in a final plenum session and adopted into the reports of each working group by the moderators and published on the project web site for comments.

- Publication of the conclusions of these 5 working groups were published on the AQUAGENOME website, advertised and submitted to the critical comments of the A2PN members.
- Final synthesis of these conclusions including critical comments brought by A2PN carried out by AQUAGENOME partners and presented as a White Paper on “Future of genomic in aquaculture research”.

3- Results:

The final product of this work was a synthesis document called “White Paper on Genomics in European Aquaculture Research” which review in short the state-of-the-art of genomic applications in European aquaculture research, and provide recommendations of future research and actions at the European level to the Commission. The complte document (17 pages) is presented in annex 1 of this report. From this document, we have extracted the following main recommendations:

a- General recommendations:

- The review reveal that substantial efforts have been carried out by a range of European research laboratories related to genomics in aquaculture species, and this have already involved considerable resources based on European and national funding. However, **the review also illustrates a rather low level of interactions between these projects.** It is also demonstrated that the large-scale nature of genomic approaches leads to very rich sources of information, but it is very demanding to fully exploit and analyse the vast amount of data that is generated. Hence, **there is a potential for new European actions to more fully exploit these data**.

Similarly to what has been observed in higher vertebrates, the major benefits of the genomic approaches are closely associated with the ability to carry out a much broader and systematic analysis of the genome and its expression products in close relation to a large set of biological processes. **Genomic approaches can profoundly extend our ability to understand biological processes**, **but the approaches are also very demanding both in terms of experimental set-ups and bioinformatic processing**. As an example, currently genomic approaches such as micro array analysis serve mainly

to generate hypotheses about affected processes and mechanisms in aquaculture species (i.e. to find groups of affected genes), but we foresee that these tools in the near future also should be used in hypothesis driven approaches in well defined experimental set-ups that will more fully exploit their potential.

- The direct adoption of genomic techniques and knowledge has so far been relative slow in the European Aquaculture industry. This may in part be a reflection of the characteristic of the industry which includes: 1) its diversity both in terms of species, environment and size of producers; 2) the rather unstructured nature of the sector which makes uptake of new methods difficult; 3) limited domestication/genetic selection of aquaculture stocks, 4) the lack of standard production parameters which leads to variable performance and 5) issues related to economic sustainability. The diversity of species cultured in aquaculture and their very divergent physiology compared to the animal production sector does not favour models which try to transfer technology and approaches between sectors. In summary, 1) there is a need to intensify studies of established species for which generally there is a deficit of basic biological knowledge, and 2) studies are required for new species to ensure establishment of sustainable production methods. **The high information content of well conducted biological trials analysed using a range of molecular resources represent a cost effective way of generating knowledge with high information content.**

- A substantial part of the European genomic research for aquaculture has been funded by EU research programs. However, EU-funded project are normally developed for a limited period of time (usually 3 years) and without clear policy in term of project exchanges and maintenance of genomic resources and databases after the end of the project. **This does not favour cross-talk between projects and various European research laboratories**, both at the level of approaches, methodologies and the level of new data for meta-analysis. Considering the large investment of the European Commission on the development of these genomic approaches, **there is certainly large potential in further developing such cross-talk** to explore the full potential of the genomic tools and resources that are available and currently being further developed. Measures should therefore be found to **secure the afterlife and availability of genomic resources and tools** after the completion of individual projects and to more fully explore results derived from individual projects developing and using genomic resources.

b- Specific recommendations:

Fish health
Genomics, have a vast potential in studies on host-pathogen interaction, e.g. by the ability to sequence whole genomes of both host and pathogen, to understand the immunological spectrum (innate and acquired immunity) in the host and the way of infection by the pathogen. In addition transcriptomic surveys can explain how the relation between host and pathogen change with life-stages, environments and treatments.

Reproduction
High priority should be given to the application of the available tools for functional studies on key reproductive issues such as control of sex differentiation, puberty, spawning and gamete quality in a range of European aquaculture species - both in terms of improving the knowledge base and in development of applied solutions to industrial problems, including reproductive isolation of farmed animals.

Breeding
Research to support the implementation in selective breeding programmes and to reduce their costs should have high priority in genetic research. These approaches should be adapted to the status of the breeding programme for each species, the different breeding goals (e.g. growth, yield, flesh quality, feed efficiency and disease resistance), and take into consideration potential side-effects of breeding on normal development, welfare,

disease transmission and diversity in disease resistance, as well as minimising risk of negative impact on wild stocks and natural ecosystems.

Nutrition and growth
In the area of nutrition and growth, and particularly in trying to understand basic fish physiology, global gene expressional surveys (microarray, cDNA libraries) are considered to provide a very promising tool. This would include the effects of varying nutrition regimes on energy homeostasis and metabolic control, disease resistance, fish welfare, larval development, muscle and skeletal development and growth and broodstock performance. Identification of QTLs linked to nutrition and metabolism should be a priority as it will offer the means to select strains of fish with improved performance or adaptability to modified diets.

Aquaculture-environment interactions
High priority should be given to the use of genomic approaches to provide a better understanding of the complex interactions between aquaculture and the environment. Such development will have major impacts on various aspects of aquaculture sustainability including fish welfare and health, product quality and safety, genetic impact on wild stocks. Key contributions for these topics should include genomic studies on basic biological information on complex traits related to physiological and behavioural responses to environment in relation to welfare, genetic impact of aquaculture species on wild stocks, basic information on genetic of plasticity and variability in relation to adaptation to challenging and fluctuating environment, new screening methods for effects of xenobiotics with the long term goal to improve traceability of aquaculture products.

c- Recommendations for resource development and maintenance

Whole genome sequences
The sequencing of whole genomes of selected aquaculture species is an ultimate addition to understand genome architecture and functioning. New sequencing technologies are emerging very quickly and facilitate already a much higher throughput at a lower cost than conventional Sanger sequencing technology, which is presently still essential for genome finishing. It is expected that the sequence of a single genome will cost less than 100,000 € in the near future. However target species should be chosen carefully; and we propose the following guidelines to make a selection (list without priority):

- ✓ The availability of background genomic resources.
- ✓ Economic significance (i.e., aquaculture species with a high economic value at the European scale).
- ✓ Support from a significant European and international scientific community to exploit the data.
- ✓ Genetic structure
- ✓ Phylogenetic relevance with regards to available sequenced genomes (new species should take phylogenetically unique positions).

As example of a whole genome sequencing project which is prioritized is Atlantic salmon. Much work has already been done in this species with about 400000 EST's available which will facilitate the assembly. In addition salmon is an important economic species. This species is also genetically interesting since it has undergone a recent genome duplication, which makes it partly tetraploid. Not much is known about genetic adjustment to genome duplication but by using this species many of the basic questions regarding gene silencing and evolution could be answered. In addition Atlantic salmon could fill out phylogenetically as a model for primitive teleost.

In parallel for less prioritized or newly emerging species there will be a need to develop physical maps to facilitate comparative approaches. The quality of the anchoring of these physical maps to full genome sequences will depend on the evolutionary distance. These

maps are not necessarily classic linkage or radiation hybrid maps, but should be based on characterized / end-sequenced BAC libraries. Additionally, BAC-libraries are ideal carrier for the exchange of genetic material and genomic information between groups and resource centres

Gene annotation

Gene annotation for aquaculture relevant species should be improved and standardized to characterize unique sets of sequences, to provide unified gene names and do functional annotation linked to model fish. This could be developed by a global approach that would cross-link genes (or ESTs) from between aquaculture species based on their phylogenetic affinity. A specific task within WP1 of the Aquagenome project has explored this approach to annotate teleost fish genes and ESTs. Other global approaches could complement and enrich this evolutionary annotation such as synteny relationships based either on comparisons of gene location on physical maps or the comparative analysis of full genome sequences. No specific new database bioinformatics environments needs to be developed for aquaculture species, but the implementation of already developed and existing databases has to be preferred (e.g., EMBL-EBI bioinformatics tools).

Resource maintenance and sharing

There is need for a European action to ensure that existing and new biological resources for aquaculture species are maintained and remain available for use by the wider scientific community. This could be achieved by assigning support and promoting networks to link research groups to maintain and share the biological resources (e.g. cDNA and BAC libraries, Radiation Hybrid panels, cDNA and BAC clones, mapping panels, QTL families), this process could be a combination of physical and virtual centres. Moreover, database maintenance and exploitation is important due to the considerable increase in sequence, mapping, and gene expression datasets in numerous aquatic farmed species. Data management at this scale requires sufficient resources and close cooperation between the involved parties, hence such topics may also be suited for new European actions.

4- Conclusion of part 2:

This approach allow us to carry out a rather broad discussion with many eurpean groups working on genomic and establish a comprehensive prospective review of the priorities needs for genomic research in aquaculture species. Dissemination of this White Paper will be described in the last section of this report.

Part 3: Development of benchmarks for genomic activities.

1- Introduction:

The Aquagenome project highlighted the need for a more systematic and extensive approach to gathering and integrating data about the application of functional genomics to aquaculture as well as about the application of genetic management of brook stock and its inplementation in aquaculture industry. This will be done by bringing together small working groups of experts in thematic meetings and by initiating important but as yet uneralized benchmarking activities among research laboratories for common research based procedure and data analysis. The generation of standardized procedures and data analysis will be an important step in the integration of functional genomic tools in aquaculture research and will facilitate future penetration in the production sector.
After ample discussion and evaluation of the outcome of the Aquagenome meeting held in Bergen, Norway, in several Aquagenome project meetings and finally during the Aquagenome thematic workshop on Microarrays, held in Göteborg, March of 2008, benchmarking activities with potential for implementation and standardisation were identified. The priorities identified arose from a "bottom up approach" in which members of the aquaculture community using functional genomics indicated the importance of establishing standardised procedures for microarray. The general confusion surrounding what "constitutes" benchmarking, a metric which is uncommon in aquaculture makes it important to define clearly the application of the term.
The absence of existing benchmarks in microarray applied to aquaculture and the existence of different microarray technologies and recently available competing methods (eg. transcriptome sequencing) makes the first benchmark to be established the determination of the best method. This is applied in the present project as outlined identify the reason for performance differences. below.
Benchmarking is the search for best practices that will lead to superior performance. This term is normally applied to an organisation or group of organisations that wish to compare for example, a given manufacturing process to establish the best performers and then
In the case of AQUAGENOME instead of applying benchmarking/standardization to a group of organisations a technique, microarray, was chosen. The aim was to compare the performance of cDNA microarrays, oligomicroarrays and new sequencing technologies. The exercise is rather modest when compared to the ongoing MicroArray Quality Control (MAQC) study being carried out with mammalian microarrays (Tong et al., 2006 *Nature Biotechnology* - **24**, 1132 – 1139; Shi et al., 2006 *Nature Biotechnology* - **24**, 1151 – 1161), however, significantly as far as we are aware it represents the first time that such an approach has been taken in aquaculture research.

2- Methods:

Several phases were identified in the planning of the benchmark and standardization exercise.

Phase One: Development of the benchmarking plan.

The key questions addressed:

1. What is to be benchmarked? – microarray versus sequencing technology (Shendure 2008, *Nature* **453**, 1197-1198).
2. What will be the best practice method? – cDNA array, oligoarray or 454 sequencing.
3. What is the method of data collection? – optical measureemnt or base sequence.

Phase Two: Data analysis

In this phase the data collected in the benchmarking study was analysed to provide a basis for comparison. The key questions in this phase were:

1. What is the performance of the three methods (cDNA array vs oligoarray vs 454 seqeuncing)?
2. What is the performance of the best method?
3. Why are they better?
4. How can the information be applied to aquaculture research?

Phase Three: Integration

The objective of this phase is to develop goals for the integration of the benchmarked process into standard methods to ensure significant performance improvements. The key questions in this phase are:

1. How will the results of the benchmark exercise be divulged to the aquaculture field?
2. Should recommendations about transcriptome analysis be made to the aquaculture community ?
3. Have the goals been clearly communicated to all involved parties (eg. project partners, Aquagenome participants, the scientific community and the European Commission?

Establish the "benchmark" activity: The microarray analysis was restricted to a single species and using a platform with the necessary resources for hybridisation and analysis of the cDNA and oligomicroarray in order to maintain the costs low and permit robust evaluation of the methods being compared. It was decided that trout (*Oncorhynchus mykiss*) would be the species chosen since 1) a cDNA array exists, ii) an Agilent oligo array is available from NIDDK (USA), iii) an appropriate platform for hybridisation and microarray analysis exists at the Université de Rennes 1/INRA-SCRIBE Laboratory, Rennes, France which is readily accessed at cost by partner 1 and iv) bioinformatics support and resources which are present in Toulouse, where the existing INRA trout database is hosted can be readily accessed by partner 1 (INRA, Rennes, France).

Consider the new technological developments: The development of new sequencing technologies associated with the drop in their cost has led to a switch in many fields to direct transcriptome sequencing (Rothberg & Leamon, 2008 *Nature Biotechnology* **26**, 1117 – 1124). The possibility that new sequencing technologies might substitute microarray technology was vigorously debated during the thematic workshop. It was concluded that although the application of new molecular technologies in aquaculture has lagged behind that in farm animal production this might not constitute a disadvantage. Moreover, **this conclusion was in agreement with the work carried out in Part 1 of the project which indicated that such methodology would be very beneficial for annotation of the fish and shellfish EST**.
In the case of aquaculture it was considered that microarrays could be compared with transcriptome sequencing and a benchmark established for this sector, before widespread indiscriminate use and establishment of costly microarray resources. The benchmark to be established is particularly pertinent for the aquaculture sector which is characterised by the diversity of species currently exploited and the emphasis on the diversification of aquaculture species for production which mean that common resources cannot be generated but have to be established for each species. The multi-species character of the aquaculture sector represents a significant technical and

economic challenge for the development of microarrays the establishment of which requires that significant numbers of cDNA sequences (either full length or short expressed sequence tags, EST) are available. For the preceding reasons 454 sequencing was also included in the comparison as it may in the future represent a cost effective way of obtaining transcriptome data from "resource poor" aquaculture species.

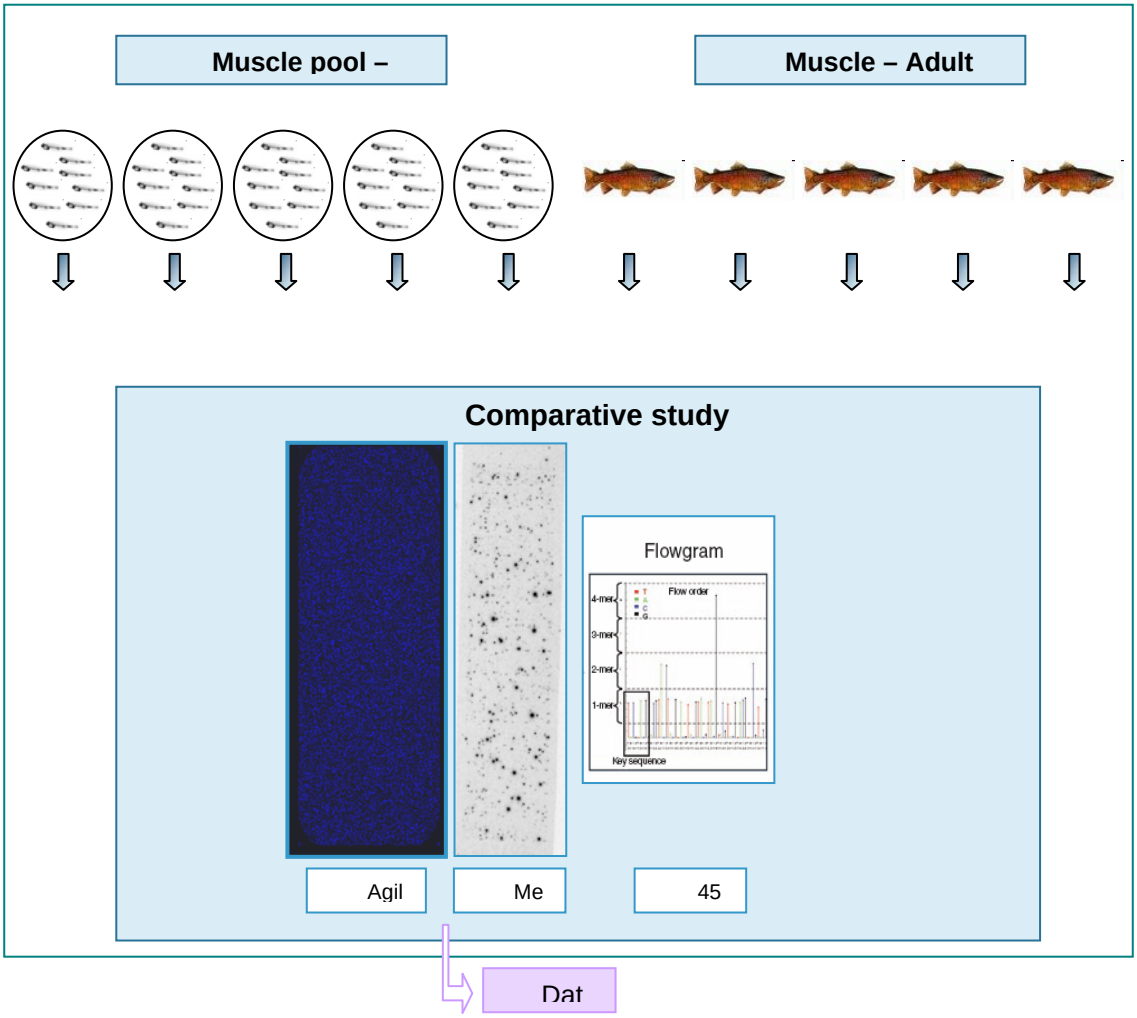
3- Results of the benchmark activity:

The goal of the experiments was to compare 3 technologies that allow the analysis of the transcriptome. Those 3 methods are:

- membrane microarrays (9026 cDNA)
- Agilent microarrays (37394 uniques oligos)
- High-throughput second generation sequencing (long read using titanium chemistry)

Procedure was as follows:
The biological material utilised was common for all three methods and was composed of RNA extracted from red and white muscle collected from five/six biological replicates at two development stages (larva and adult) from rainbow trout (*Oncorhynchus mykiss*). The exercise permitted, i) comparison of the muscle transcriptome elucidated with each of the 3 technologies selected and also ii) age specific comparisons in order to test the if biologically meaningful data is obtained with the 3 technologies selected.

Figure 1. Fluxogram showing the experimental design and demonstrating the comparison carried out.



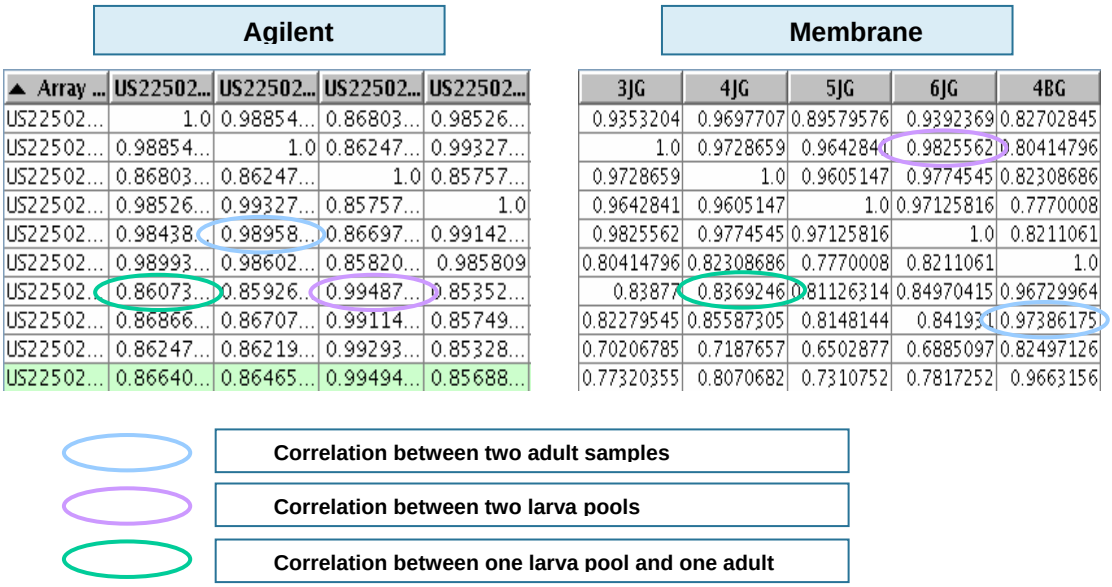
Microarray analysis

RNA was extracted from tissues collected from individuals for the adult stage and from a pool of several individuals for the larva stage. Five biological replicates were made for each condition. The same pools of total RNA were hybridized to the 2 microarrays and also used for synthesis of cDNA libraries which were subsequently sequenced using second generation technology to generate a digital transcriptome. Hybridizations were done by hybridizing one sample per array. The membrane microarray was spotted with 9216 unique trout cDNA. A rainbow trout high-density oligonucleotide microarray which contained 60-mer oligonucleotide probes representing 37 394 unique trout consensus (TC) sequences and 1417 control spots and 6409 randomly chosen duplicates (4 × 44 format) was purchased from Agilent Technologies (design number 016320). The custom microarrays are manufactured using a proprietary non-contact industrial inkjet printing process, in which oligo monomers are deposited uniformly onto specially-prepared glass slides. This in situ synthesis process prints 60-mer length oligonucleotide probes, base-by-base, from digital sequence files. The slides were designed by Salem *et al.*, 2008 (*J Fish Biol. 72, 2187-2206*) and the technical performance of the array has been characterised. The experiments were conducted by partner 1 at the microarray platforms at the Université de Rennes 1/INRA-SCRIBE Laboratory, Rennes, France.

Results of the comparative analysis carried out for the microarray methods take into consideration key measures of microarray performance which include *sensitivity*, *accuracy* and *reproducibility* and are presented below:

1) Reproducibility. Tables below show a part of the correlation matrix (figure 2).

Figure 2. Comparison of the results obtained for the total transcriptome using the Agilent microarray and the Membrane array. The table shows a part of the correlation matrix between different sample transcriptomes analysed using the Agilent oligoarray and cDNA membrane array. Each row and each column represent one sample (pool of larvae or adult muscle).

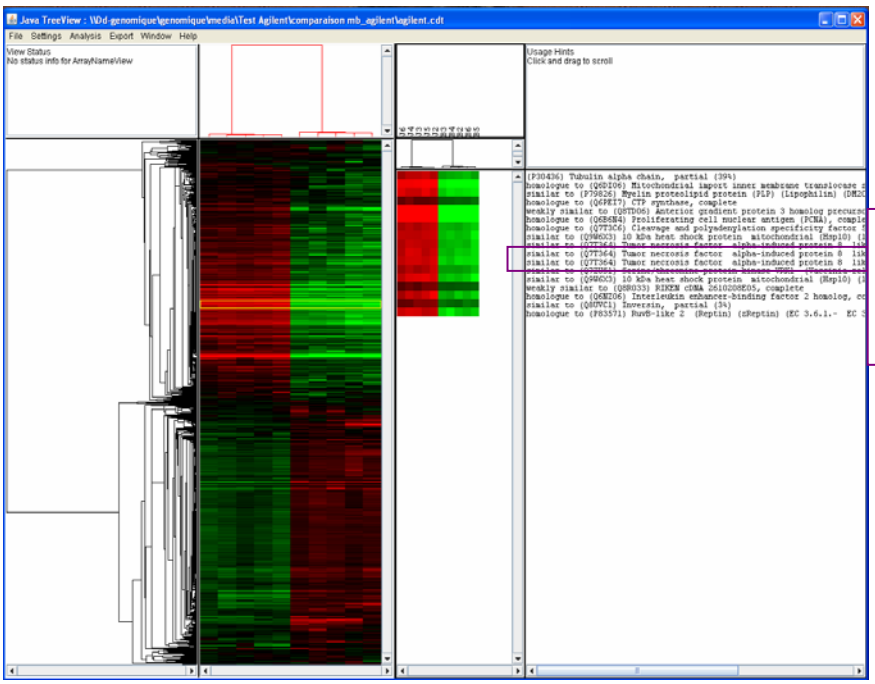


Array ...	US22502...	US22502...	US22502...	US22502...
US22502...	1.0	0.98854...	0.86803...	0.98526...
US22502...	0.98854...	1.0	0.86247...	0.99327...
US22502...	0.86803...	0.86247...	1.0	0.85757...
US22502...	0.98526...	0.99327...	0.85757...	1.0
US22502...	0.98438...	0.98958...	0.86697...	0.99142...
US22502...	0.98993...	0.98602...	0.85820...	0.985809
US22502...	0.86073...	0.85926...	0.99487...	0.85352...
US22502...	0.86866...	0.86707...	0.99114...	0.85749...
US22502...	0.86247...	0.86219...	0.99293...	0.85328...
US22502...	0.86640...	0.86465...	0.99494...	0.85688...

3JG	4JG	5JG	6JG	4BG
0.9353204	0.9697707	0.89579576	0.9392369	0.82702845
1.0	0.9728659	0.9642841	0.9825562	0.80414796
0.9728659	1.0	0.9605147	0.9774545	0.82308686
0.9642841	0.9605147	1.0	0.97125816	0.7770008
0.9825562	0.9774545	0.97125816	1.0	0.8211061
0.80414796	0.82308686	0.7770008	0.8211061	1.0
0.83877	0.8369246	0.81126314	0.84970415	0.96729964
0.82279545	0.85587305	0.8148144	0.84193	0.97386175
0.70206785	0.7187657	0.6502877	0.6885097	0.82497126
0.77320355	0.8070682	0.7310752	0.7817252	0.9663156

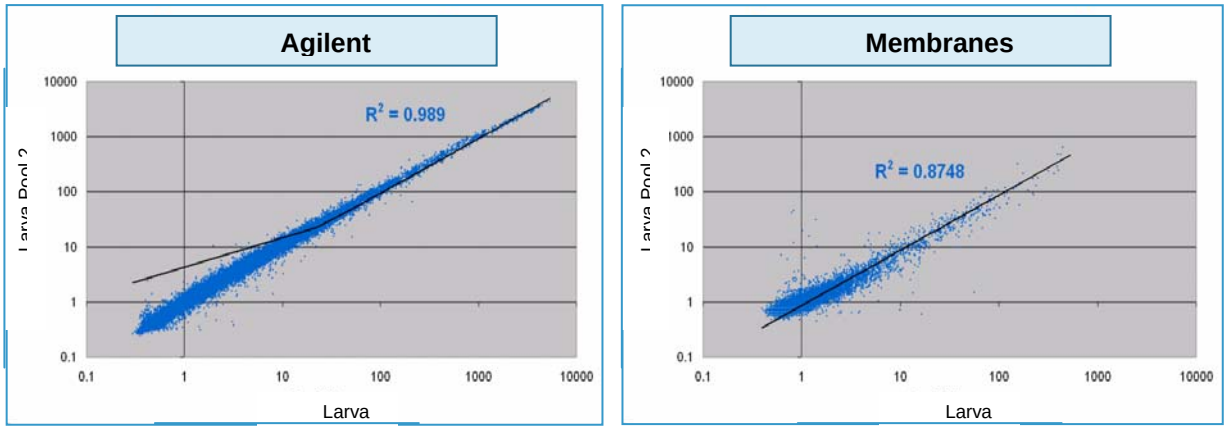
The perfect regression, = 1, is obtained when the same sample is compared to itself in the correlation analysis. The samples highlighted with different coloured rings correspond to the results of the analysis for three different samples (green ring, blue ring, lilac ring) using the oligoarray of cDNA array. For example, the sample highlighted by a blue ring has a correlation cooefficeint of 0.9895 with the Agilent oligarray and 0.9738 with the cDNA array. High correlations were observed between adult muscle RNA using the two microarrays, between two larval samples using the two microarrays and also between one pool of larva and muscle RNA extracted from one adult. In general higher correlation coefficients were obtained with the Agilent oligoarray compared to the cDNA array. Indeed, several oligos which were replicated several times on the Agilent microarray slide had the same measured value of expression (figure 3). Hierarchical clustering of the Agilent data, revealed the values quantified for the replicated oligos were highly correlated. Figure 3 illustrates the results of data clustering and the expression of tumour necrosis factor 8 (TNF8) which was represented by 3 different oligos on the Agilent array is highlighted.

Figure 3. Cluster analysis showing a similar expression profile for the oligos on the Agilent array which have a similar identity, as defined by their "e value" using the Blast algorithm



2) Dynamic range is the range in which a linear relationship exists between the detected signal (fluorescent or radioactive) and the concentration of the fluorescent dye (Agilent) or radioisotope (membrane) that is incorporated into the RNA from the test sample. Data falling within the linear dynamic range is most reliable.
- The benchmark exercise revealed that the linear dynamic range is higher with the oligomicroarray using fluorescence compared to the cDNA array (figure 4).

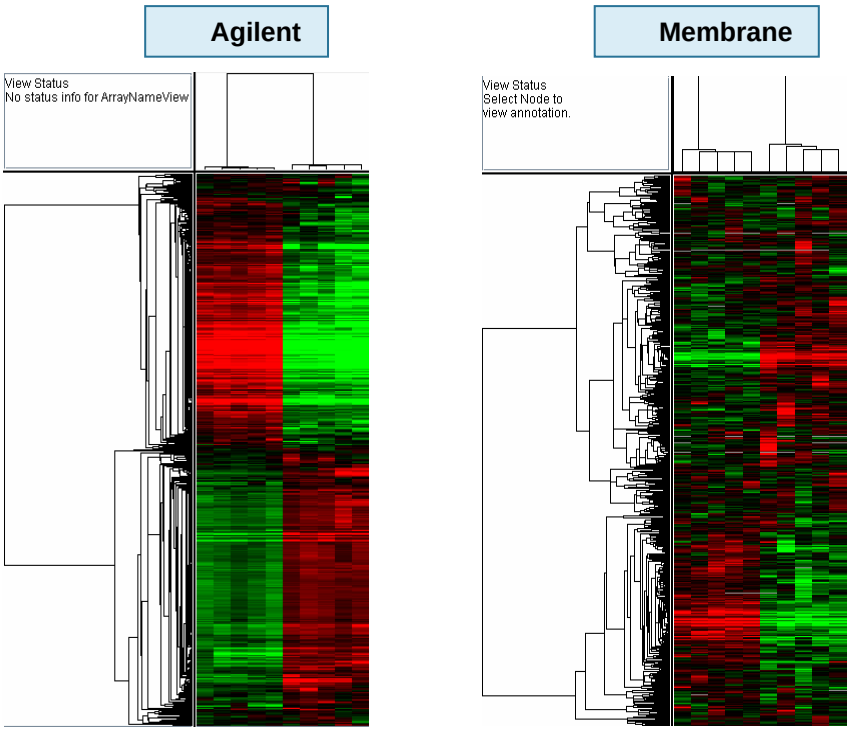
Figure 4. Comparison of the linear dynamic range of the oligoarray and the cDNA microarray using as the sample larval pool 2



Differential expression analysis

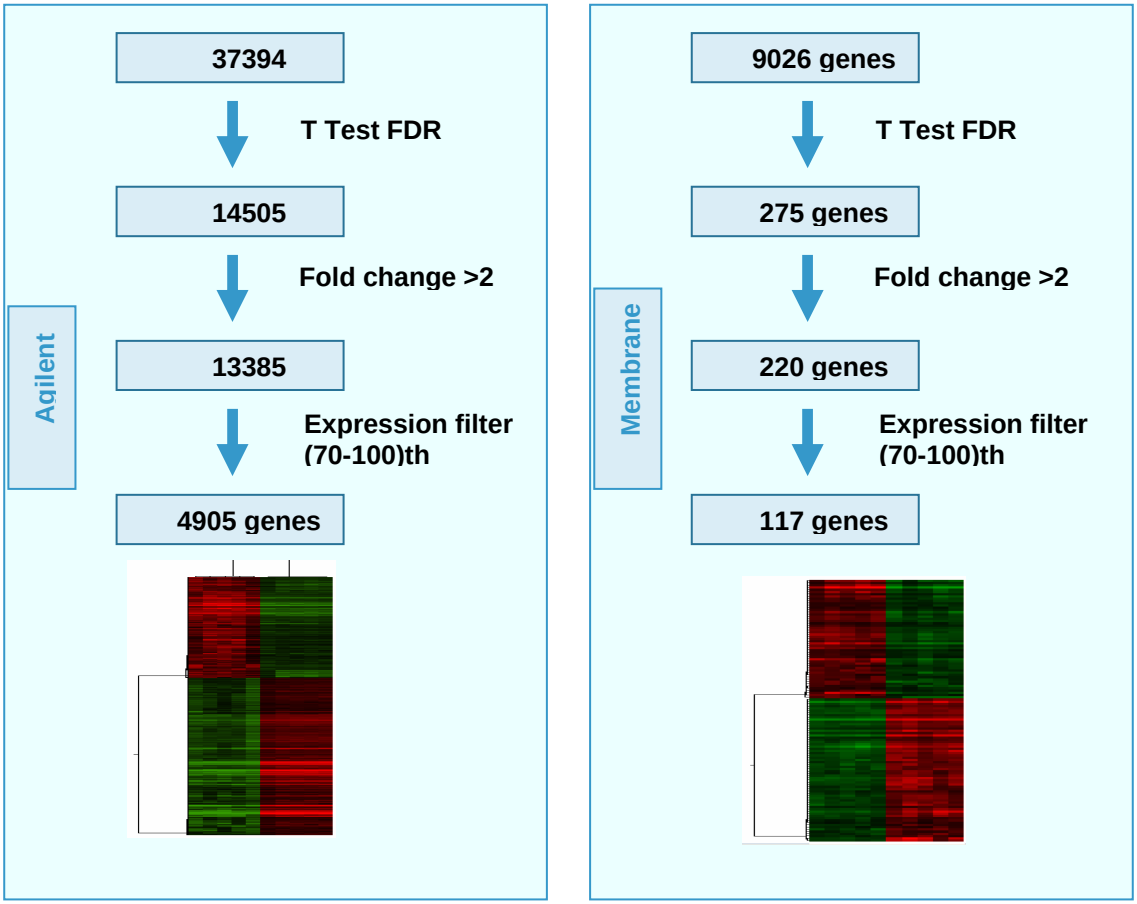
Genes with the same SwissProt hit on both microarrays were identified (approximately 700 genes) and hierarchical clustering performed and the results for the two microarrays compared (figure 5). This analysis revealed higher correlations between normalised expression of genes between individual samples using the oligoarray compared to the cDNA microarray. In other words the correlation obtained for multiple oligos in the microarray which corresponded to the same genes was much higher than observed with the cDNA array.

Figure 5. Heat map summarising the data obtained from microarray analysis and comparing differential gene expression between the oligo and cDNA array used. The green colour represents highly expressed transcripts and red represents transcripts with a low expressed transcripts.



Finally, differential analysis of the expression of all the genes present in each array was conducted. With the Agilent oligoarray, 4905 differentially expressed genes (or 13% of the totality of the chip) were identified between larva and adult muscle RNA while with the cDNA membranes array, only 117 genes were found to be differentially expressed (less than 1% of the array). The following diagram summarizes the steps performed (figure 6)

Figure 6. Scheme representing the relative number of differentially expressed genes using the Agilent oligoarray and the cDNA array.



In conclusion, while both oligoarrays and cDNA arrays have a number of advantages there is no question about the superior performance in relation to *sensitivity*, *accuracy* and *reproducibility* of the oligoarray compared to the cDNA array. The main drawback of the oligoarray in comparison to the cDNA array is the higher cost associated with its utilisation.

Table 2 summarises the relative advantages and disadvantages of the two microarray methods.

	AGILENT	MEMBRANES
Advantage	- Experimentation time shorter (ex: 60 samples⇨7 days) - Standardization (calibration of spotting) - Excellent Reproducibility between experimentation - Density - Very high dynamic range	- Low cost (77€) - Good Reproducibility between experimentation
Disad-Vantages	- High cost	- Experimentation time longer (ex: 60 samples⇨2 weeks) - Quality controls - Restricted dynamic range - Radioactive Waste Management - Radioactive risk

Second generation sequencing

Global transcriptome analysis is a key target of second generation sequencing. The significant reduction in cost of sequencing and increased throughput means that it is now possible to sequence a human genome in 1 week. Global transcriptome analysis allows a genome wide analysis of differential RNA expression and allows biological insight into a range of processes. The most widely used methods currently for whole transcriptome analysis is microarray. However the limitations of microarray is that only genes on the array can be analysed, they have limited dynamic range (sensitivity) and cannot reveal alternative splicing, single nucleotide polymorphisms (SNPs), coding and none-coding RNAs etc. It appears that far more information can be attained using second generation sequencing and it was considered by all scientists present at the microarray workshop of importance to conduct a “benchmark” exercise to assess the utility of sequencing approaches versus microarray approaches for global transcriptome analysis. Scientists at the microarray workshop were divided about the feasibility of second generation sequencing approaches for aquaculture as a consequence of questions related to; 1) the length of sequences attained with second generation sequencing which might not lead to transcript identification and ii) interpretation of data as a results of existing problems with gene ontology.
The questions raised were addressed by establishing a benchmark for microarray, cDNA versus oligoarray and also microarray versus second generation sequencing for global transcriptome analysis.

Procedure
The announced Roche 454 Titanium platform by summer 2008 was expected to be a good candidate for this comparison, as it provides about 500 Mbases of data information with sequence read length of about 500 bases. These technical features would make the titanium platform a suitable method of choice to compare sequence derived data with microarray data sets. The opportunity of sequencing RNA samples instead of comparing them to each others by microarray techniques, is quite easy to demonstrate, as microarray tests are limited to the information content once store on the glass-array in form of selected biological samples. In contrast, sequencing present the information of the actual sample composition (actual genes expressed, gene number, mutation etc), even if the sequences cannot immediately be compared (identified) to known genes. The advantage of the Roche titanium platform compared to the ultra short sequences produced by Solexa or SOLiD systems is the less complicated bioinformatic processing of the longer sequence reads.

Technical aspects:
In the second generation sequencing approach, RNA is isolated from target tissue and a cDNA library synthesised using specific adaptors and then amplified using standard emulsion PCR approaches. The PCR product is then deposited on a glass slide for sequencing and the relative expression levels are based on the number of tags. The 12 trout samples utilised for microarray analysis were also used for sequencing and MPI-MG in Berlin was chosen as the partner to perform the sequencing task of the test as they have well establish sequencing platform including the next generation sequencing systems.

Library construction: Total trout RNA was supplied as ethanol precipitated samples and mRNA purifications and concentration were performed using the Dynal's oligo-dT magnetic beads. Twelve cDNA libraries from trout samples were constructed using either the SMART PCR cDNA synthesis kit (Clontech) or the Mint universal kit (Evrogen) to generate twelve amplified cDNA libraries.
Unfortunately the titanium platform was delayed until late October 2008 and even worse, at the same time information was received that contrary to what was initially envisaged, amplicon sequencing was not supported by the Roche system at this time, and only genomic shot-gun protocols and kits were supplied. This meant that it was not possible to use the libraries produced in preparation for the Titanium sequencing as initially planned when Roche first provided the information about the system. It was therefore decided to modify the procedure and protocols received and perform the sequencing following the shotgun kit and protocol of the titanium platform, as the constructed cDNA libraries were optimized for an insert size of about 1400 bases.
From each cDNA library, a corresponding sequencing library was generated by sample ultra sonification (one minute) following the protocol and kit supplied by Roche, except that coded primers were introduced to allow multiplexed samples within the subsequent sequencing run. All samples were controlled by Agilent's bio-analyzer using DNA and RNA detection chips. Sample enrichment were done by emPCR with the Roche titanium propriety kits and bead recovery of sonicated cDNA after separation was calculated to be between 9% and 14%, all within the expected range, pronounced by the protocol.
As an alternative, to the first approach a second sequencing library set was constructed as described above, except that rigorous sample purification by spin columns for the cDNA libraries and gel purification combined with micro column separation for the sequencing libraries, were introduced. This was necessary to get a better separation of short nucleotide components, which became the absolute major components in the sequencing process.

454-sequencing: Two separate sequencing runs were performed on the Roche GS-FLX Titanium system using again Roche's titanium propriety kits and protocols. Primary data analyses were done on an separated LINUX cluster using Roche's data processing software package.
Results: the first run was a complete disaster, as all sequence data were discarded after processing with the qualifier: too short sequences.
Therefore a second sequencing library set was produced, introducing the purification of shorter fragments of about 400 nucleotides, because short fragments are extremely high processed by the following amplification steps.
Nevertheless the second run was very similar to the first one and further sequencing was stopped until the reasons for these results have been clarified.
Further analysis of the few data processed by the software showed two effects for the bad results. One was as expected the high amplification rate of very short fragments and secondly the high amounts of polyT sequences derived from the 3' part of the cDNA constructs, which are normally not accessible by amplicon sequencing. It was therefore decided not to continue with additional optimization of the shot-gun protocol, but to invest our efforts in establishing an amplicon based protocol compatible with the

kit components available from Roche. The time line is expected to be in the range of four to five month, thus being present not before May 2009. The costs of these developments will be covered by MPI budget, as RNA sequencing is a predominant sequencing request not only of the institute for the Roche platform.

4- Conclusion of benchmarking activity.

A benchmark has been successfully established which indicates that oligoarrays are the best performing method of global transcriptome analysis for Aquaculture research (of 27-02-2009).

The comparison of cDNA arrays versus oligoarrays revealed without a doubt that the oligoarray is superior in all respects. This benchmark has already had a significant impact and INRA have now switched to using oligoarrays for their microarray work. Moreover, the salmon community (UK and Norway) have initiated work on the design of an oligoarray and will stop using cDNA microarrays. The failure to successfully implement the second generation sequencing technology indicates clearly that currently the most relevant method for analysis of global transcriptomes in Aquaculture is oligonucleotide microarrays. Nonetheless, work is ongoing to adapt Titanium sequencing methods to amplicon sequencing and it is hoped in the near future to establish the benchmark initially identified by comparison of microarray data and sequencing data and establish if oligoarrays continue to be the best procedure.

It was considered important that the results of the benchmarking exercise in transcriptome analysis should be divulged to the aquaculture community and also to all the Aquagenome partners and workshop participants. Several steps have been taken to ensure adequate divulgation of the results of the benchmarking exercise.

1. The outcome of the benchmarking exercise will be divulged in a state-of the art lecture at the Genetics in Aquaculture meeting (July 5-7th 2009) in Bode, Norway.
2. A technical paper is being written to describe the outcome of the benchmarking activity and will be divulged to workshop participants, project partners and submitted for publication in an appropriate journal to ensure maximum visibility to the Aquaculture industry in Europe, US and Asia.

Part 4: Exchange programs.

1- Introduction:

The AQUAGENOME project addresses the need developping exchanges between research laboratories developping complementary expertise in fish and shellfish genomic. Moreover, fostering the application of fundamental research results in new aquaculture techiques by the industry will certainly benefit from development of training in genomic for young researchers (PhD students, post-doc) who should become active difuser of teh genomic approchaes towards industry. In order to carry out these tasks, we have established i) a program for the exchange and dissemination of resources, materials and information ii) a program for training grants particularly for young scientists to promote exchange of information, analyses, and dissemination of good practice in functional genomics in aquaculture.

2- Methods:

Mobility grant for individual training activities – the exchange of personnel between European laboratories with expertise/material or specialised equipment was developped in order to facilitate technology transfer and dissemination of expertise and information between European laboratories. AquaGenome funded the mobility of PhD students and young scientists among the laboratories within the APN. The grants covered travel and subsistence costs up to a maximum of 2000 euros for a period of two to six weeks.

Applications for participation in training activities were accepted continuously and evaluated after 6 fixed dead-lines: month 6, 9, 12, 15, 18 and 21. At each fixed dead-line, the partners responsible for this WP judged the application for qualification. The requirements included: a letter of invitation and engagement from the host institute, a letter of recommendation from the supervisor or principal investigator of the lab were the young scientist currently belongs, a research plan and a letter of motivation explaining how this planned research visit would increase the exchange of knowledge and resources on a genomics approach to sustainable aquaculture in Europe. The criteria for evaluating the applications that should be met/the information needed to be included by each application in order to pass the threshold for further ranking was:

1. *Curriculum vitae* of the applicant
2. Budget with specifications
3. Research plan - extended (max one page)
4. Motivation to how the planned research visit will increase transfer of knowledge and resources within genomics approaches to sustainable aquaculture in Europe
5. Motivation to how the planned research has relevance to the theme of Aquagenome
6. Motivation to how the planned research visit will contribute to the applicant's research training
7. Letter of recommendation from supervisor/principal investigator of the lab were the applicant currently belongs
8. Letter of invitation and engagement from the host University/Institute
9. Information on scientific carrier and gender

The core group of AquaGenome decided on the final ranking based on the scientific evaluation and on strategic considerations such as geographic distribution and transfer of technology to give the best complementation of competences. After completion of the mobility period, the trainee prepared and send an activity report and financial statement to the core group.

Grant for exchange of genomic resources - Through projects in functional genomics in FP5 & FP6 and the integrating activity of the SSA; AQUAFUNC, it is clear that numerous molecular resources exist in different European laboratories which could contribute in a cost effective way to ongoing European and National projects. A resource exchange and

methodology/results dissemination program will be established. A budget have been earmarked in order to cover the cost of generating resources such as copies of cDNA and genomic libraries, filters, microarrays and other materials. This budget was used to grant applications for utilisation of resources at any of the other laboratories within the APN. The resource utilisation grants were mainly directed towards the host-laboratories for providing manpower and consumables to cover the costs of production plus packing and distribution of the resources.
The procedure for the resource grants was that the applicant and the host will together request the grant but the costs will cover the cost of manipulation and material at the host institute. Each grant can maximally amount to 1000 euros and the purpose of the grants is to cover cost for handling of the resources, shipping (if applicable) and material for handling and shipping. The resource exchange grants could combined with Mobility grants – example: an applicant could go to another laboratory to utilise a special genomic resource and then apply for a Mobility grant – to cover travel and subsistence as well as apply for resource grants that can cover the costs for actually utilising the resources at the host laboratory. The resource grants could also be used to cover costs of handling and shipping of a resource from one laboratory (the host) to another (the user). The resource grants had the same deadlines as the mobility grants and were advertised together with the mobility grants.

3- Results:

Mobility grants: In 2008, the applicants and awardees from 3 rounds of applications, August 1st, 2007, November 1st, 2007 and February 1st, were reported. In 2008, an additional three applications rounds have been open. Thus, in total applications for participation in training activities have been accepted continuously and evaluated after the 6 fixed dead-lines: month 6, 9, 12, 15, 18 and 21, as specified in the Annex1.
For the first deadline of applications during the second reporting period, May1st, 2008, 4 applications were submitted, all very relevant to the main aim and objectives of Agenome and thus all four applicants were granted support. Next deadline of applications – August 1st, 2008, attracted only one applicant who without problems fulfilled the criteria and was of interest for the Aquagenome project and was thus granted. Finally, for the last round of Mobility grants, deadline November 1st, 5 applications were submitted. Unfortunately, only one of these applications were planned for and were able to fulfil the main part of the research visits within the time-frame of the project (ie. before February 15th). After enquiry of the scientific and financial officers of the project we were unable to support four of the applicants even though, scientifically, all four projects were of great relevance and clearly within the aim of Aquagenome. Thus, in the end only one application could be granted for the last application round of Aquagenome mobility grants.
For summary of all mobility applications granted see table 1 bellow.

Table 1 Mobility grants – applied, granted and reimbursed within Aquagenome

Applicants – August -07	**Host**	**Granted**	**Report accepted**	**Reimbursed**
Antonio Ibarz, University Barcelona	Canario & Power, Universidad Algarve	2500	YES	YES
Christopher Sauvage, IFREMER	C. Haley Roslin Institute	2500	YES	YES
Applicants – November -07	**Host**	**Granted**	**Report accepted**	**Reimbursed**
Ingun Stubhaug, NIFES, Bergen, Norway	Douglas Tocher, Institute of Aquaculture, University of Stirling	2500	NO – the applicant has not requested the grant	NO
Lars Niklasson Göteborg university, Sweden l.niklasson@zool.gu.se	Bertrand Collet, FRS Marine Lab, Aberdeen, Scottland B.Collet@MARLAB.AC.UK	2500	YES	YES
Dimitrios Lukovitis, Department of animal production, Alexander technological Educational dloukovi@mail.gr	Georgios Kotoulas, HCMR, Greece kotoulas@her.hcmr.gr	2500	YES	YES
Isabel Morgado, CCMAR, Faro, Portugal imorgado@ualg.pt	Catherine Boyen, Département des Sciences de la Vie, Roscoss boyen@sb-roscoff.fr	2500	YES	YES
Bruno Louro CCMAR Faro Portugal blouro@ualg.pt	Filip Volckaert Katholieke University, Leuven filip.volckaert@bio.kuleuven.be	2500	YES	YES

Applicant – February -08	Host	Granted	Report accepted	Reimbursed
Alexander Triantafyllidis, Aristotelse University of Thessaloniki, Greece	Filip Volkaert, K University, Leuven, Belgium	2500	YES	YES
Mbaye Tine, Université Montpellier, France	Richard Reinhardt, Max Planck Institute, Berlin, Germany	2500	YES	YES– in advance
Pedro Guerreiro, CCMAR, FARO. Portugal pmgg@ualg.pt	Shuttan Sundell, Göteborg University, Sweden	2500	YES	YES– partly in advance
Applicant – May -08	**Host**	**Granted**	**Report accepted**	**Reimbursed**
Sílvia Gregório, Centre of Marine Sciences, Universidade do Algarve, Faro, Portugal	Luca Bargenolli Departmenf of Public Health, Comparative Pathology,and Veterinary Hygiene, University of Padova	2500	YES	YES
Serena Ferraresso, Dep. of Public Health, Comparative Pathology and Veterinary, University of Padova	Chris Secombes, Immunology Research Centre School of Biological Sciences, University of Aberdeen	2500	YES	YES
Liliana Anjos Guerreiro, CCMAR, FARO. Portugal pmgg@ualg.pt	Richard Reinhardt, Max Planck Institute, Berlin, Germany	2500	YES	YES
Sofia Morais Institute of Aquaculture, University of Stirling	Pedro Rodrigues Functional and Comparative Proteomics, CCMAR, University of Algarve	2500	YES	YES– partly in advance
Applicant – August -08	**Host**	**Granted**	**Report accepted**	**Reimbursed**
Oscar Monroig, Institute of Aquaculture,	J. M. Cerda Reverter, Instituto de Acuicultura,	2500	YES	YES

Applicant – November -08	Host	Granted	Report accepted	Reimbursed
Massimo Milan Department of Public health, Comparatiive Pathology and Veterinary Hygiene. University of Padova, Italy	Dr. Pierre Boudry, Department of Aquaculture, IFREMER de Brest, France	2500	YES	YES– partly in advance
Alberto Falco Mol Cell Biol Institute, Miguel Hernandez Univerity, Elche, Spain	Dr. Victor Mulero, University of Murcia	**No**		
Laura Martin Gomez, Pathology department, CIMA, Cororn, Spain	Leonor Cancela, CCMAR, Algarve, Portugal	**No**		
Elena Viciano, CSIC, Catellon, Spain	Dr. Douglas Tocher, Institute of Aquaculture, University of Stirling	**No**		
Trine Haugen, IMR, Norway	Dr. Yann Guigien, INRA, France	**No**		

As indicated in the table, all granted applicants have provided the Aquagenome with completed Scientific activity reports, letter of confirmation from the Host principal scientist and itinerary along with receipts of tickets. The scientific activity reports have been evaluated and accepted by the WP-leaders of this activiy and all the reports are can be found in appendix 2.
In time for each new application round the Aquagenome mobility grants have been advertised through the Aquagenome Web-page, the Aquagenome A^2PN and through the AquaTT send list.

Grants for exchange of resources: The awardees from the first 3 rounds of applications, August 1st, November 1st, 2007 and February 1st, 2008 led us to support 5 exchange grants. During 2008 an additional three applications rounds have been open. The resource grants have had the same deadlines as the mobility grants and have been advertised together with the mobility grants through the different pathways.
In summary, of the 15 mobility grant applications awarded so far within Aquagenome, 9 have also applied for and been granted Resource Exchange grants. See summary table appendix 2.

Part 5: Knowledge transfer to the aquaculture sector.

1- Introduction:

EU funding of research in large scale genomic based science in any sector is aimed at improving the socioeconomic well being of the European community through job creation, training of scientists, better tools or products and economic development. However, the link between the basic research and its application and commercialisation has always proved difficult to achieve in many sectors, including genomic for aquaculture. Academics are notoriously reticent about commercialisation of their research findings because it will involve them in areas for which they are ill-equipped such as IPR protection, company creation, product promotion, pursuit of investors and partnership with industry.

The recent developments in genetic and genomic technologies, particularly those applicable to animal breeding, disease and nutrition, offer the means to formulate a rigorous generic framework for the domestication and management of any potential aquaculture species. This includes the establishment of techniques for genetic mapping, identification of QTLs and parentage assignment have particular application to broodstock improvement and management in aquaculture. Indeed the application of these techniques to fish species may be particularly effective due to the high fecundity of fish and the high degree of genetic variation available in wild stocks. But we should also consider genomic tools which are now powerful generic methods for understanding and diagnosing nutritional requirements, disease states and environmental impacts. Most of these new technologies revolve around the ability to measure the expression of very large numbers of genes simultaneously in individual fish tissues, often called transcriptomics. Indeed transcriptomic and other technologies for large scale gene expression analysis, offer enormous power when combined with marker assisted breeding programmes to enable the breeding of multiple complex traits.

In this context, Aquagenome project aims fostering the application of fundamental research by aquaculture industry, an objective which has to be associated with genomic knowledge transfer to the aquaculture industry. This task corresponds to one of the major objective of Aquagenome project and has implied large contact and mobilisation of aquaculture stakeholders on fish and shellfish genomics.

We have been developping our our activities on the following tasks:

1. To improve awareness of ‘Best Practice’ and ‘Best Available Technology’ by workshops attached to visits to leading players in aquaculture.
2. To engage industry in consideration of the potential for aquaculture genomics in a broad context of business, ethical, legal and social issues.
3. To develop a consensus Technology Road-Map for the sector that prioritises short, medium and long term research and knowledge transfer needs.
4. To provide practical advice to interested partners.

2- Results:

improvement of awareness on ‘Best practise’ and ‘Best available technology’ of genomic tools.

The accomplishment of such task rely on the establishment a network of contacts: Such database which invaluable in order to disseminate information. The contacts distribution list has been constructed and populated. In order to increase the number of contacts, presentations were given at Fish Breeders Roundtable (2007) and communication was sent by email and through publications in Fish Farmers International and at conferences. This list is constantly added to in order to ensure that a wide range of stakeholders are informed of Aquagenome activities.

In order to improve awareness on 'Best practise' and 'Best available technology' of genomic tools for leading players in aquaculture, 2 workshops associated with site visits were organized. These visits were aiming to present existing possibilities of fish and shellfish genomic and bring together scientists and industrialists in order to discuss future area of development in ther context of business, ethical, legal and socal issues. All presentations made during the 2 site visits have been put on the project website.
The first site visit workshop was held in February 2008 at the IFREMER Central Office in Paris "From genomics to applications in European temperate and Mediterranean Aquaculture: Basis and prospects". This workshop co-organised by SYSAAF, INRA, IFREMER, focused on the application of genomic technologies in trout, bass, bream, turbot, oyster and carp.. Twenty one presentations covered biological basis of reproduction and genetics, examples of genome sequencing and genetic maps (blue mussel, flat oyster, Pacific oyster, rainbow trout, sea bass, sea bream), first applications of functional genomics and future use of genomic tools in breeding was illustrated to trace pedigree (fingerprint) or to improve breeding programs (QTL, major genes, SNPs, genome-wide selection) as exemplified in the French dairy cattle breeding program. The workshop concluded with visits to facilities involved with genomic research (robot, PCR, sequencing) in the INRA Jouy en Josas research center (Bioressource Center CRB GADIE, Bioinformatic Resource Center SIGENAE) and the private LABOGENA laboratory involved in the use of genomic tools (DNA parentage assignement, QTL, genetic mutations, …) for livestock and aquaculture species in France, such as rainbow trout, sea bass, sea beam or turbot. The audience was 24 aquaculture professionals from Southern and Eastern European and Mediterranean countries involved in rainbow trout, sea bass, sea bream and oyster production. In total 37 delegates attended the workshop, including 24 representatives from aquaculture companies. Presentations can be download on the AQUAGENOME web site : http://www.aquaculture-europe.org.

The second site visit workshop was held in december 2008 at the Institute of Aquaculture in Stirling. The workshop began with a half day session on 'Genomics for Beginners' which is a course designed and run by GFP to introduce the topics and terminology in the study of genetics and genomics. The format includes a series of presentations, interactive activities and question and answer sessions. The 'Genomics for Beginners' course was attended by 19 delegates, including 10 representatives from 6 different fish breeding companies.
The following full day programme comprised a series of presentations. Initially, delegates were provided with presentations demonstrating how genetic and genomic techniques had been applied in other farm animal species. This included presentations on Marker-Assisted Selection in pigs, the application of genomics in chicken breeding programmes and Genomic Selection in dairy cattle. Delegates then heard a series of talks regarding the current state-of-the-art in northern aquaculture species with regard to genetics and genomics, including a description of how a national cod breeding programme was established in Norway. Finally, delegates heard about the experiences at Landcatch Natural Selection during a detailed presentation about their breeding programmes. In total 27 delegates attended the workshop, including 15 representatives from 10 aquaculture companies.

The Aquaculture Roadmapping workshop

This workshop was held in June at Brunel University in the UK, running over 2 days. This unique approach to Roadmapping used specific technology to capture the thoughts and inputs from all delegates, through the use of small laptop computers for each delegate to enter their input directly into the discussion. The roadmap involved 33 delegates from 27 organisations. A total of 11 delegates came directly from industry, representing a variety of backgrounds – from small breeding organisations, to large breeding organisations. Prior to attending the workshop, all delegates received a pack of 'pre-work' for them to read over in preparation for the event, so that they would have a chance to think through some

ideas for discussion. This allowed each participant to arrive at the workshop with an overview of what would be considered and discussed during the event.
The purpose of the workshop was to develop an Industry-led Roadmap to strengthen European Aquaculture through the application of genomics. The roadmap aimed to:

- be a strategic tool to focus the funding and implementation of research
- recommend research priorities based on a comprehensive analysis of industry needs and opportunities arising from advances in genomics
- develop shared approaches and specific plans for knowledge transfer
- Through the process of the roadmap, we aimed to:
- engage stakeholders in collaborative working
- promote the exchange of ideas and tools between different disciplines and different parts of the industry
- integrate information and ideas from other relevant programmes

Throughout the workshop, many areas were explored and discussed. Using the unique 'Brain-Pool' technology all delegates were able to add contributions and discuss topics raised by providing their own input into individual computer work stations, allowing everyone to contribute to the work of the group.
The workshop began with some short presentations to give an overview of the science, technology and opportunities available. Time was also spent looking at:

- Drivers influencing the future of the industry.
- Priority traits for genetic improvement.
- Research priorities to take forward the desired traits.
- Knowledge Transfer Priorities to ensure effective exchange of know-how with industry.
- Promoters to overcome obstacles to progress on research and Knowledge Transfer.

As each of these areas were addressed, delegates were invited to comment, discuss and prioritise the outcomes, allowing full and open dialogue that could be built on throughout the workshop.

After the workshop, a report was prepared by New Game Plan showing the main findings of the workshop and providing the full output of the two day event (see apendix 3). This report has been used by the Aquagenome partners to develop a succinct report for further distribution to funders, delegates, and national bodies to inform them of the work carried out and the priorities identified (see apendix 4).

Overall, the report confirm the potential for genetics and genomics in aquaculture and identified prioritized traits, research and knowledge transfer. In order to be successful in these tasks, this report also indicate that several important points which should be considered in parrallel and that a unique set of actions will not be able to cover all aspaect of the question. These various actions are presented in appendix 4.

From this exercise, the following main conclusions can be proposed:

- The Roadmapping exercise has clearly shown that the industrial participants believe that the application of genetic and genomic technology will advance our understanding of the biology of aquatic organism. The application of this knowledge by the different sectors that make up the global aquaculture supplies and production industry is presently constrained by a number of factors.

- There has been a lack of investment by the industry in the development of scientifically based breeding programmes. Having access to pedigreed populations of fish is an essential resource for researchers to identify the traits important to the industry. It is therefore not surprising that the first applications of genomics in aquaculture are being seen in collaborations between commercial salmon breeding companies and researchers. These early success can only be transferred when

companies producing other species begin to apply genetic management and improvement practice.

- In nearly every case these advanced studies are a collaboration between the breeding company and research scientists supported by national funds from governmental or Research council sources. The aquaculture industry still does not have the financial resources available to breeding companies in the livestock or poultry sectors to undertake in-house research. Companies will benefit from being early adopters of this technology but will continue to need support until the whole aquaculture sector becomes financially more stable.

- It is clear that both the industry and the academics want to continue a dialogue. There are misconceptions on both sides that can only be overcome by better communication. It appears that the industry would like its Associations to act as the conduit and lobby for more structured workshops and meetings. Genomics and the possibilities it offers for the future needs to be better presented to the industry and the general public to avoid confusion with other more contentious technologies.

- Many of the issues regarding networking between academia and the different industrial sectors will hopefully be dealt with by the links being forged between the different concerted actions and the Technology platforms for Aquaculture and animal breeding. Actions that encourage more contact between researchers and industry at the Knowledge transfer and collaborative working levels that help inform these higher level discussions are needed.

Practical advice to interested partners

At both the site-visit workshops and the roadmap, a number of potential issues were brought to surface, where individual aquaculture companies identified areas or projects in genetics and genomics that would benefit from further investigation. The project partners involved in this WP were able to conduct individual technology translation discussions with these interested parties to discuss development of ideas. In total, 5 discussions took place, involving 5 aquaculture companies in the UK, Italy and Spain.

Part 6: Dissemination of knowledge.

1- Introduction:

The strong mobilization and contribution of European research laboratories to the development of genomics in fish and shellfish biology is illustrated by the number of projects funded by EC. Thus, more than 13 projects contributing to the implementation of the common fisheries policy within the 5th and 6th Framework Programs have used genomic approaches to answer specific questions/problems in different species under the same general philisophy of improving the sustainability of European aquaculture. Under their specific objectives, these projects have generated various genomic resources, which included the EST collections, microarrays and genetic linkage map. In parrallel, through national funding, some projects have also been supported with the ambition to develop generic resources in genomics for aquaculture fish species. Overall, during the 5 last years, significant efforts have been made in Europe in the development of genomic approaches in European aquaculture species, such as salmon, trout, seabream, sea bass, cod, flounder, oyster and clam. A recently EC-funded project, AQUAFUNC, has started to coordinate these projects and to integrate the outcome of an EC-funded project through a common publication of the outcomes with the creation of a common web-page presenting projects and providing links to respective instituitons and research. This SSA project was a first step for acting on this integration.

An overview of what is presently under development at a international level on animal genomics clearly indicates that we have to advance more in the coordination of these research projects within aquaculture. Until now, very few opportunities to integrate within aquaculture production the outcomes of these projects at the level of the genomic approaches have existed. Implementation of results from fish genomics within the European aquaculture industry is still at its infancy. Several reasons may explain this situation, e.g. complexity, lack of knowledge, and lack of applicability and usefulness, have often been perceived as stumbling blocks. However,

In this context, it was an important task of Aquagenome project to put effort into transferring the genomic results of aquaculture research to the industrial producers. The major anticipated impacts should be on the relationship between research in genomic and aquaculture sector. Through a proper communication plan supported by a large database of contacts within aquaculture domain, our objectives werere to (i) demonstrate the current state of the art and future development of genomic in aquaculture, (ii) promote the potentialities of such approaches for solving aquaculture problems, (iii) give aquaculture producers technology translational advice to stimulate development of genomic within interested industrial partners.

Achievement of these objectives relies on a proper dissemination fo the outputs of the project. For this task, the following actions have been carried out:

- Development of a project website which centralize all information related to the Aquagenome activities.

- Development of a network of project associated partners (A2PN) which gather both aquaculture producers, stakeholders and researchers.

- Development of communication actions towards aquaculture industry.

2- Results:

a- The Aquagenome Website:

The Aquagenome Website (http://www.aquaculture-europe.org/) has been developed to support the objective of a coordination of future aquaculture related genomic projects and

to transfer the knowledge towards the aquaculture industry. On one hand, and in order to create a community of aquaculture related scientists, academic researchers will be able to find within the website some contacts and an updated and curated list of genomic tools in European Aquaculture species. This website gathers information about:

- The Aquagenome project description and a list of partners for communication and contact.
- The Aquagenome Newsletter to inform partners and contacts about Aquagenome project news.
- The inventory of existing aquaculture genomic resources by specie to forecast the future requirement in term of genomic tools.
- The application forms to intend for grant for resources exchange of for mobility.
- The organization of Aquagenome meetings and workshops
- The documents resulting from all Aquagenome meetings and workshops

On the other hand, this website also offers an opportunity for aquaculture stakeholders to survey the development of genomic research in aquaculture species, and to find contacts and information.

The Website structure:

Considering the complemntarity between Aquahunc and Aquagenome projects, for the aquagenome website, we have decided to use the same website as for aquafunc. This website entitled "www.aquaculture-europe.org" is hosted on a server ran by SIGENAE at INRA Toulouse, France (Head of Sigenae team: Christophe Klopp), and was created using the content management system typo3. The graphic theme was design thanks to the help of E. Bonnet from INRA, Rennes. The website is divided in 2 parts (Fig1), one dedicated to the former project AQUAFUNC, and the other on the AQUAGENOME project.

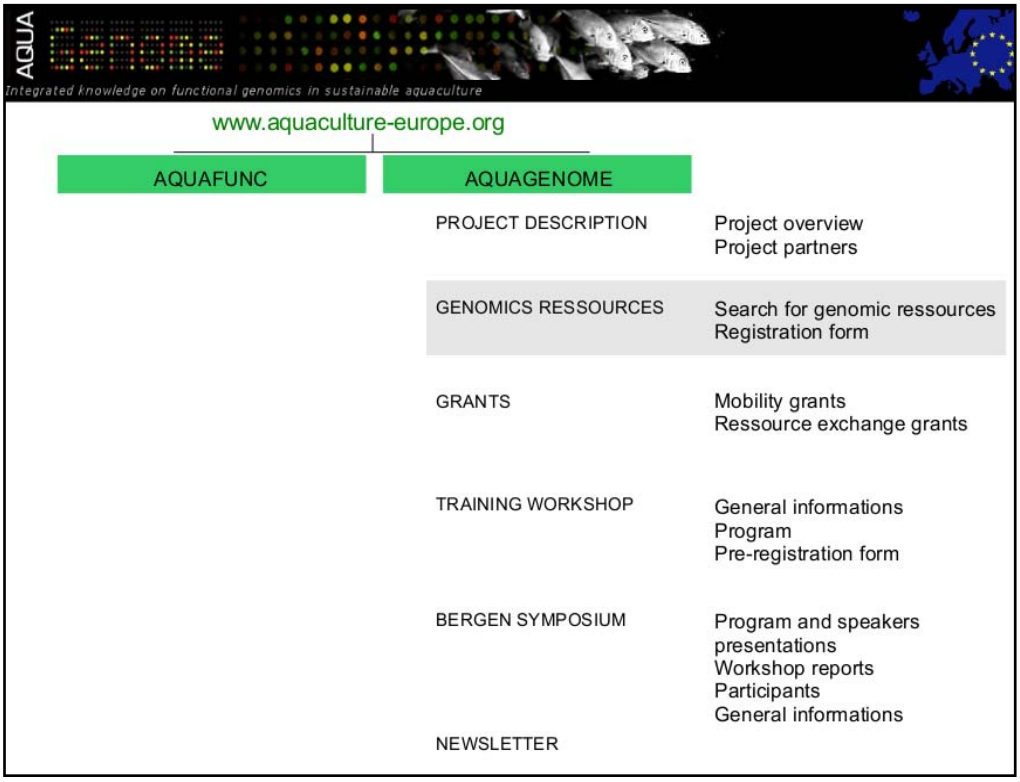


Figure 1: www.aquaculture-europe.org structure

The Aquagenome website gathers general information on the Aquagenome project (**Project overview:** description of the project), on Aquagenome partners (**Project partners:** list all partners with their contact details), on the project actions (**Project actions:** Actions achieved for the project). Within Project actions the website provided information on the achievements of different Aquagenome activities. This includes:

Existing genomic resources and future development.

Inventory of Genomic resources

This inventory contains:

- A list of the numbers of ESTs available in public databases for fish and shellfish (From NCBI dbEST summary August, 10 2007).

- A list of the genomic resources for the main European Aquaculture fish and shellfish species. This list can be queried using a mysql database hosted by the SIGENAE team, either with queries by species to get an overview of all genomic resources in that species (All to select all species, or you can select one specific specie) and/or with queries by specific genomic resources (All to select all type of genomic resource, or you can select a specific type of resource). For each resource a contact person or a published reference manuscript is given with an email link in order to contact this person if needed. This part is right now a restricted area protected with a login and password (invited login and password is: aquagen / aquagen). However, everyone is welcome to register, under the acceptance of the project coordinator by filling a registration form available on the website. Currently, **30 persons** are registered, mostly European researchers.

Analysis of genomic research to be developed for aquaculture.

Mini-Symposium, Bergen, Sept. 2007:"State-of-the-art and future of genomics in aquaculture research"

- Program & key Speakers presentations: Program and presentations in pdf format.
- Workshop reports: The second day of the Bergen meeting was dedicated to working group discussion about different topics, the endind reports are available under this section.
- Participants: List of all participants to the meeting with detailed contact information.
- General information: General information about the meeting and link to the registration form.

The final version of the “White Paper on Genomics in European Aquaculture Research” has been also, posted on the website.

Exchange program.

Mobility Grants.

In this section a description of one of the objective of the project is given such as the funds of researcher mobility between laboratories.

- Guidelines for submission: a pdf document that explains how to submit an application.
- An electronic application form, with a lot of field and attached document required. All information is send to Mrs Kristina Sundell, UGOT Sweden.

- After completion of the granted research visit, one can find the reimbursement form

Resource Exchange Grants:

In this section a description of one of the objective of the project is given on funds for genomic resource exchange.

- Guideline for Submission. a pdf document with all explanation
- Application form: an electronic form which send the information to Mrs Kristina Sundell, UGOT Sweden.

Knowledge transfer to the aquaculture sector.

In this section the following information have been posted:

Applied Training Workshop for the aquaculture industry 7th-8th February 2008 Paris:"From genomics to applications in European temperate and Mediterranean Aquaculture:Basis and Prospects"

- General information: link to the workshop presentation, objective, audience, the contact, the link through the pre-registration form. A pdf version is also available
- Program: Preliminary program of the workshop. A pdf version is also available
- Pre-registration form. The final registration is submitted to selection by the organizing comity according to the description of the job role. Information is send to Mr Pierrick Haffray, SYSAAF, France.
- All pdf files presentations made during this training workshop.

Applied Training workshop for Genomic and North European Aquaculture which was help in December 2008 in Stirling.

- General information and program on the workshop with all forms for registration were posted on ther website.

Aquagenome Newsletters.

Newsletters summarize the main activities of the porject during a 3-5 months period. All newsletter issues are available under this section and a frame named « Latest News » is always in evidence at the right side of the page. Those links refers to the latest news with a link to a brief explanation, and a link to the detailed page.

b- Aquagenome Associated Partner Network (A2PN):

To extend impact of the project to a wide network of stakeholders and researchers, we have develop such network. During the life of the project, A2PN included 26 membres coming from public/academic research groups and 9 aquaculture stakehoders (see appendix 5 for the complete list of A2PN members). Members of the A2PN will have access to resources and results of the project and have also to access the exchange grants offered by the project. They have been invited to participate to the workshop on the future of genomic research in aquaculture held in Bergen (September 2007).
A2PN members have received regularly the porject News Letter aiming to keep them informed about the activities of Aquagenome.

c- Dissemination actions related to the Aquagenome project.

During the first year of the project, several dissemination actions have been carried out. They all rely with communication/exchanges with aquaculture industry and stakeholder, with whom our connexion have to be improved. Moreover, the aquaculture sector is probably poorly informed and motivated by possible application of genomics tools and it is important for the AQUAGENOME objectives that this analysis will be developped with a large participation of aquaculture stakeholders.
Various actions have been developped which include production of a flyer presenting Aquagenome objectives (see bellow), participation to workshop where many aquaculture stakeholder were present, publication of a paper on Aquagenome project in a pfrofessional journal related to aquauclture (Fish Farming International). Through these communication actions, we hope to be able to mobilize these people on Aquagenome objectives.

Planned/ actual dates	Type	Type of audience	Partner responsible /involved
22-23 November Athens, Greece.	PROFETPOLICY workshop **What future for Mediterranean Marine Aquaculture?**	Aquaculture stakehoders	HCMR
14-17 August Tronheim, Norway	AQUA NOR 2007 The international Aquaculture Venue	Aquaculture Stakeholders	UGOT
13-14 December, Warsaw, Pologne	**PROFET POLICY workshop on 'Governance in Continental Freshwater Aquaculture' 13-14 December 2007 - Warsaw**	Aquaculture Stakeholders	VNIRO
November	**Publication in Fish Farming International (see text bellow)**	Aquaculture Stakeholder	GFP

Publication in Fish farming International (Nov. 2008)

AQUAGENOME – Genomics in fish and shellfish: From research to aquaculture

Project financed by the EU for 2 years – completion date 31st December 2008

Partners:

INRA – France
University of Stirling – UK
Genesis Faraday – UK
Hellenic Centre for Marine Research – Greece
SYSSAF – France
Goteburg University – Sweden
Algarve University – Portugal
Institute of Marine Research – Norway
Instituto Investigaciones Marinas – Spain
Russian Federal Institute of Fisheries & Oceanography – Russia

Objectives:

Co-ordinate ongoing and future national and international research projects in genomics of aquaculture species, and support the diffusion of genomic approaches within research laboratories
Enhance transfer of information and knowledge to the aquaculture industry

Aims:

Inventory existing genomic resources and identify and evaluate new bioinformatics resources, genomic information and tools that are vital for the development of European aquaculture
Identify specific research domains in which genomic approaches should be developed in order to support the European aquaculture industry
Support exchanges of scientists and genomic resources between members of the Aquagenome Associated partner Network by offering mobility grants and resource exchange grant (see information on the project website).
Deterimine the required actions to apply genomic tools and knowledge in an effective and sustainable manner in aquaculture, by interacting with producers and stakeholders

Aquagenome Associated Partner Network (A2PN):

To extend impact of the project to a wide network of stakeholders.
Members of the A2PN will have access to resources and results of the project, will be able to access the exchange grants offered by the project and will be invited to participate in relevant tasks and workshops.
Relatively easy to join the A2PN and so interested organisations should get in touch.

Sign Up to receive more information:

To receive newsletters and other information about open calls for applications for exchanges of scientists and resources
Contact carol.didcock@genesis-faraday.org

Site-Visit Workshops:

Two workshops will be held – one in France focussing on southern aquaculture species and one in Scotland focussing on northern aquaculture species.
Target audience – farm managers, breeding programme managers, health and welfare managers, industrial stakeholders.
Examine current possibilities for use of genetics and genomics in aquaculture and discuss future areas of development.
Site visits to see aquaculture genetics and genomics in practice.
Case studies of how genetic and genomic approaches have been effectively integrated into terrestrial livestock breeding programs to examine possibilities for aquaculture.
Also examine some of the ethical issues involved with this.

Website:

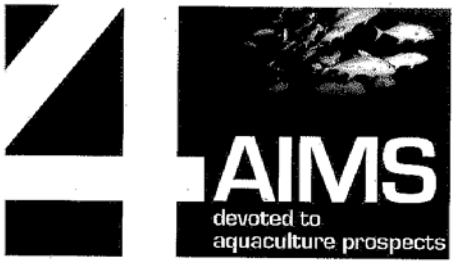
The Aquagenome website contains a wide range of information, including previous newsletters, information about the project and further information of the site-visit workshops.

Aquagenome Flyer distributed at Aqua Nor 2007 and PROFET workshop.

Analyze the accessibility, the quality and also the sustainability of these resources and what would be the bottleneck for such access by the community

To inventory existing genomic resources & identify and evaluate new bioinformatic resources, genomic information & tools that are vital for the development of European aquaculture

1



2

To identify specific research domains in which genomic approaches should be developed in order to support the European aquaculture industry.

The research domains will include:
Studies of immune system and pathogenic agents as well as vaccine development and side-effects of vaccines
Population genetics, genotyping and marker assisted selection
Chronic stress and genomics of environmental adaptations in farmed species
Sex differentiation, puberty and broodstock management
Early development and feeding as well as metabolism toxicogenomics

3

To support Exchanges of scientists and genomic resources between members of the AQUAGENOME Associated Partner Network created.

Exchange and dissemination of information through organization of thematic meetings.
Development of training activities
_ distribution of training grants by the Steering Committee of the project.
Development of a program for benchmarking exercises
_ dissemination of good practices and initiation of work
Fostering of the exchange of information and resources

4

To determine the required actions to apply genomic tools & knowledge in an effective & sustainable manner in aquaculture by interacting with producers & stakeholders.

To establish a network of contacts
_ regular knowledge transfer
To improve awareness of 'Best Practice' and 'Best Available Technology'
To develop a consensus Technology Road-Map
To prioritize short, medium and long term research
To provide practical advice to interested breeders through the organization of face-to-face

GENOMICS IN FISH AND SHELLFISH
FROM RESEARCH TO AQUACULTURE

Context and achievement

In recent years, a relatively large number of projects investigating the biology of aquaculture species through genomic approaches have been funded both at European levels and national levels. These have generated an extensive core of genomic tools and expertise that is spread among European research laboratories.

However, this scientific resource is not coordinated within research and transfer to the European aquaculture industry is still in its infancy.

In this context, the objectives of AQUAGENOME are (i) to coordinate the ongoing and future national and international research projects in the field of genomics in European fish and shellfish aquaculture and support diffusion of genomic approaches within research laboratories (ii) to enhance the transfer of information and knowledge towards the aquaculture industry.

These achievements will require the use of a number of instruments, e.g. working groups to gather experts in specific domains, organization of workshops where aquaculture industry will meet research and construct a roadmap for genomic transfer to industry, creation of a website to share information and provide links, development of a database and software to broaden access to genomic resources and the distribution of grants to support exchanges.

More informations

www.aquaculture-europe.org/

AQUAGENOME

www.aquaculture-europe.org/

10 Partners

Institut National de la Recherche Agronomique France

University of Stirling United Kingdom

Genesis Faraday United Kingdom

Hellenic Centre for Marine Research Greece

Syndicat des Sélectionneurs Avicoles et Aquacoles Français France

Zoological Institute of Göteborg University Sweden

Algarve University Portugal

Institute of Marine Research Norway

Instituto Investigaciones Marinas Spain

Russian Federal Institute of Fisheries and Oceanography Russia

AQUA

GENOMICS IN FISH AND SHELLFISH
FROM RESEARCH TO AQUACULTURE

© photo Christophe MAITRE/INRA & Emilie BONNET

SIXTH FRAMEWORK
PROGRAMME
PRIORITY 8.1
Scientific Support to Policies

During the second year of the project, various dissemination actions have been developped in order to increase dissemination of knowledge related to Aquagenome project. They all rely with communication/exchanges with aquaculture industry and stakeholder, with whom our connexion have to be improved. Moreover, the aquaculture sector is probably poorly informed and motivated by possible application of genomics tools and it is important for the AQUAGENOME objectives that this analysis will be developped with a large participation of aquaculture stakeholders.
These actions included:

- Presentation of Aquagenome project (context, objectives, main achievments) and also conclusions raised by the White Paper have been performed during the two workshops organized by the project for the aquaculture stakeholders.
- Presentation of the Aquagenome project at the Agriculture and Biotehcnology Conference (Cork, Ireland, 24-27 August 2009)
- Presentation of the Aquagenome project at the Aquabreeding/Reprofish meeting “Future prospects for aquauculture breeding in Europe”, last 1-3 October 2008.

Initiated during the Aquabreeding/Reprofish meeting in October 2008 and persued during the last Aquagenome project meeting in Paris (december 2008), discussions on the future of the Aquagenome network were developped by Aquagenome partners. Several arguments support continuation of the activity of the Aquagenome network, including updating of the databases on genomic resources in Europe presented on the Aquagenome website and continuation of the transfer of genomic knowledge towards aquaculture stakeholders. For this last objective, it appeared relevant to merge our questioning with Aquabreeding and Reprofish discussions on the same issue. One of the major question was the transfer of our project knowledge towards the European Aquaculture Technology & Innovation Plateform (EATIP) which appeared as the best place for discussing and exchanging with Aquaculture stakeholders. We finally decided to participate actively in close links with Aquabreeding and Reprofish projects to the different working groups established by EATIP. We anticipate that this strategy will be the most efficient strategy for disseminating Aquagenome knowledge and conclusion towards fish and shellfish aquaculture stakeholders.

Part 7: Summary of the main achievements and recommendation.

Main achievements.

The initial objectives of the aquagenome project, as indicated at the start of the project, **can be summarize as follow**:

- To inventory existing genomic resources and identify and evaluate new bioinformatic resources, genomic information and tools that are vital for the development of European aquaculture.
- To identify specific research domains in which genomic approaches should be developed in order to support the European aquaculture industry.
- To support exchanges of scientists and genomic resources between members of the AQUAGENOME Associated Partner Network created.
- To determine the required actions to apply genomic tools and knowledge in an effective and sustainable manner in aquaculture, by interacting with producers and stakeholders.

At the end of the project, two years later, we can consider that the following achievements have been reached:

1) We have set up an inventory of the existing resources in genomic of fish and shellfish aquaculture species. This inventory is accessible on project website (www.aquaculture-europe.org) but without relevant updating, this information will quickly be obsolete.
2) In relation with the quality of the gene annotation which a major bioinformatic problems faced by these fish and shellfish genomic resources, we have investigated 2 possible strategies (phyllgenomic analysis and high-throughput new sequencing methods) which should bring significant imporvement in the quality of the annotation.
3) In close link with the main european researchers involved in fish and shellfish genomic, we have produced a “White Paper on Genomics in European Aquaculture Research” which identifies and recommands future research needs in order to support European aquaculture.
4) In order to favorite developpment of genomics in aquaculture, we have carried out a benchmarking exercise on the use of different microarray tehcnologies and comparaison with new sequencing technologies.
5) Exchanges of young scientists and of genompic resources between research groups were supported by the distribution of grants (total 15 mobility grants and 9 resources exchange grants).
6) Improvement of awareness of aquaculture stakeholders on ‘best practises’ and ‘best available technology’ have been carried out through 2 specific workshop organized for aquaculture producers.
7) To engage the industry in consideration of the potential of genomics in aquaculture, we have develop a consensus Technology Road-Map for the sector and prioritized research and knowlege transfer needs.

Altogether, these achievements cover the initial objectives of the project and this work leads to recommendations in 3 domains:

- Research to be develop in the future
- Actions to be develop to support transfer of fish and shellfish genomic knwoledge towards aquaculture.

Recommendations.

a- In the White Paper on genomics in European Aquaculture Research, several recommandations have been proposed and include:

a. **strategic recommendations** on the importance of genomic tools and approaches for basic knwoledge on biology of fish and shellfish, the importance of such knwoldege for development of aquaculture, the importance of the follwing-up of the EU-funded project and of the cross-talk bewteen projects which should be systematically encouraged and supported.
b. **Specific recommendations** on the application of genomics in aquaculture research in a large variety of domains including health, reproduction, breeding, nutrition and growth, aquaculture and environment.
c. **Recommendation for the development and maintenance of genomic resources** which put sequencing of new fish and shellfish whole genome as a first objective. Gene annotation and resource mainetance have been also highlighted as important domains to invest.

b- In the report of the roadmapping exercise aiming to prioritize research and knowledge transfer needs, the following points have been recommanded:
From a list of prioritized desired traits, a list of prioritized research statements have been proposed. Associated with these research priorities, knowledge transfer statement have been also ranked;
Finally, this analysis leads to (i) the offer of 38 **'quick win'** opportunities for collboration (ii) give attention to prioritized **'nexte steps'** identified from the raod mapping exercise (iii) develop **the outputs** into, published Aquageneome raodmapping exercisen (iv) propose an integrated, strategic and phased portfolio of research and knowledge transfer (v) engage relevant stakeholders (vi) tackle the **promotors and the blockers** of the project (vii) improve **networking and coordination** between relevant organization and projects.

c- Dissemination of Aquagenome knowledge.
One of the major question was the transfer of our project knowledge towards aquaculture sector through the European Aquaculture Technology & Innovation Plateform (EATIP) which appeared as the best place for discussing and exchanging with Aquaculture stakeholders. We finally decided to participate actively in close links with Aquabreeding and Reprofish projects to the different working groups established by EATIP. We anticipate that this strategy will be the most efficient strategy for disseminating Aquagenome knowledge and conclusion towards fish and shellfish aquaculture stakeholders.