Assessment of tools for identifying the genetic origin of fish and monitoring their occurrence in the wild

*D. Blohm[1], F. Bonhomme[2], G. Carvalho[3], D. Crosetti[4], T. Cross[5], G. Dahle[6], D. Danancher[7], R. H. Devlin[8], E. Garcia-Vazquez[7], K. Glover[6], B. Guinand[2], G. Hulata[9], K. Jørstad[6], K. Kohlmann[10], S. Lapègue[11], P. McGinnity[12], G. Marteinsdottir[13], P. Moran[14], C. Primmer[15], P. A. Prodöhl[16], M. L. Rise[17], C. Saavedra[18], Ø. Skaala[6], T. Svåsand[6], A. Triantafyllidis[19], E. Verspoor[20]

[1] University of Bremen, Germany
[2] UM2-IFREMER-CNRS, Montpellier, France
[3] University of Wales, Bangor, UK
[4] ICRAM, Rome, Italy
[5] University College Cork, Ireland
[6] Institute of Marine Research, Bergen, Norway
[7] University of Oviedo, Spain
[8] Fisheries and Ocean Canada, Vancouver, Canada
[9] Institute of Animal Science, Bet Dagan, Israel
[10] Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
[11] IFREMER, La Tremblade, France
[12] Marine Institute, Furnace, Ireland
[13] University of Iceland, Iceland,
[14] University of Vigo, Spain
[15] University of Turku, Finland
[16] Queen's University, Belfast, Northern Ireland, UK
[17] Memorial University of Newfoundland, St John's, Canada
[18] Instituto de Acuicultura de Torre la Sal-CSIC, Castellón, Spain
[19] University of Thessaloniki, Greece
[20] FRS Marine Laboratory, Scotland, UK

* All authors have contributed equally to this work

Introduction

The potential genetic effects of aquaculture activities have aroused a great deal of concern, and the perceived risks are often associated with interbreeding with natural populations and the adverse effects of ecosystem interactions (1). The EU-funded project Genimpact (http://genimpact.imr.no) reviews specific aspects of potential risks and concerns on interbreeding and aquaculture-ecosystem interactions. In workshop 2, emphasis was given on the current knowledge and state of art of the tools available for the study of monitoring escapees of the species under study i.e. Atlantic salmon, Atlantic cod, European sea bass, gilthead sea bream, turbot, common carp, Atlantic halibut, scallops, mussels, oysters (Pacific oyster and European flat oyster) and European lobster. Additionally emphasis was given on the future research objectives for better and improved monitoring methods.

Current knowledge on non genetic tools

Identification of escapees is assured if all farmed fish are tagged, however tagging may not always be a realistic option due to high cost or biological restrictions such as the size of the cultured individuals. The most commonly used non-genetic methods for discrimination among farmed and wild fish include: i) *External identification tags:* this has been widely practised since the dawn of aquaculture. The major problem has been the difficulty of tagging small fish; however recently, new methods using fluorescent implants have been used successfully; ii) *Morphology/Morphometry:* since farmed fish are often characterized by body changes or defects that can be used for visual detection by professionals or laymen; accuracy and consistency are however, in many cases uncertain; iii) *Scale and otolith pattern recognition:* growth patterns in scales and otoliths can be used for identification of farmed fish; and iv) *Biochemical and physiological markers:* such as carotenoids and fatty acids due to different diets of wild and reared fish, or analysis of trace elements in otoliths and bone structures; however, so far, most work has been performed to identify Atlantic salmon escapees and relatively little on the other species.

Current knowledge on genetic tools

In aquaculture, molecular markers are increasingly employed for purposes of monitoring genetic variation in domestic stocks and for identifying domestic individuals in the wild (2, 3). Molecular markers are inherent to the individual, thus they can't be lost. If polymerase chain reaction (PCR) is used to resolve variation, tissue sampling does not require animals to be killed. Their main advantage with respect to physical markers is that they are inheritable, enabling identification of the offspring of aquaculture individuals released or escaped into wild populations. Thus genetic markers allow the estimation of the true impact of aquaculture escapes in wild populations through fitness studies. The main genetic markers that have been used are allozymes, mitochondrial DNA and microsatellite DNA. Future markers include coding DNA variation based on single nucleotide polymorphisms (SNPs) and / or using DNA microarrays.

The first genetic tool to be used involved the analysis of inherited variation in proteins (enzymatic proteins) within and among wild and farmed fish and shellfish populations. The different allelic forms, referred to as allozymes, result from DNA base variation in protein coding genes which causes amino acid changes and changes in either the protein's electrical charge or its molecular shape. Some of the major problems of this method include i) that it generally requires destructive sampling of individuals to obtain required tissue samples, ii) that allozymes lose their activity very fast and therefore the maximum storage time for allozymes is much shorter than DNA samples, and iii) the variation resolved has proven of limited use for monitoring farm escapes, or for studying the genetic effects of cultured fish on wild populations. Of the species considered, the method has been most widely used in relation to the Atlantic salmon (4), though extended studies have also been done on carp. In general allozyme markers are now seldom employed and have been largely superseded by DNA based analyses.

Mitochondrial markers were the first DNA markers to be extensively used. Mitochondrial genomes, in animal species, consist of single circular pieces of double-strain DNA As the mitochondrial genome mostly contains coding DNA, its main source of variation is SNPs. The mutation rate is higher in mitochondrial than in nuclear coding DNA regions resulting in higher levels of variation and population differentiation. Significantly different levels of nucleotide diversity are detected among different mitochondrial genes and DNA regions with different levels of variation can be chosen as markers of different biological units: highly conserved regions, for developing species-specific or race-specific markers; less conserved regions, for markers of stocks, and so on. Additionally, in higher animals,mitochondria are typically inherited via females, therefore DNA cannot detect introgression via males, but can be used for exploring sex-biased gene flow. The only exceptions are mussels and other bivalves in which a second mitochondrial genome is paternally inherited (the M genome).

Microsatellites have, in the last decade, developed into the most popular genetic markers (2). Microsatellites or simple sequence repeats (SSRs) are tandemly repeated motifs of 1-6 DNA bases, which are abundantly distributed within genomes and usually characterized by a high degree of polymorphism in the number of repeats. With the advent of PCR technology, microsatellite loci can be easily amplified using specific primers which bind to the region flanking the variable microsatellite. Recently, with the availability of high-throughput capillary sequencers or mass spectrometry, the characterisation of variant types, once a bottleneck, has become

easy and rapid. By analysing a panel of multiple microsatellite loci, a unique combined SSR genotype profile can be produced for each individual tested and studies show that such genotype profiles are highly discriminating, with randomly chosen individuals having a low probability of having matching genotypes. In the field of fisheries and aquaculture, microsatellites are useful for the characterization of breeding populations and stocks, for paternity and relatedness analysis of natural populations, hatchery broodstock selection, constructing dense linkage maps, and mapping economically important quantitative traits. With this type of information and the development of powerful analytical methods/statistical programmes in recent years, the focus has increasingly shifted from defining populations to discriminating individuals (5); it is now often possible to assign or exclude individuals originating from a claimed population. This methodology has applications to the identification of the genetic origin of specific individuals, of immigrants into populations, the occurrence of hybridization or admixture, the assessment of introgression of hatchery individuals into wild populations, and evaluation of the success of stock enhancement programmes.

More recently efforts have focused on direct analysis of coding DNA. This encompasses the small fraction (normally 5% or less) of the genome of any higher organism which is transcribed and used to produce messenger RNA, much of which is further translated to produce proteins. Variation in coding DNA is generally assessed by means of DNA sequencing and most commonly involves SNPs; these are individual point mutations in genomic DNA at which different sequence alternative (alleles) exist in a population. Studies of coding DNA to date have tended to focus on a few particular genes, possibly because they are believed to be functionally important and related to particular biological traits. Examples include the MHC, or growth hormone genes groups. However, overall, as yet relatively few gene sequences have been characterized for the aquatic species considered.

The characterization of variation in these regions, inlcuding the analysis of expressed sequence regions or tags (ESTs) in a number of commercially important aquatic species offers an important source of coding DNA sequences in the future. SNPs can be used for many purposes because they are very common (their frequency is estimated to be 1 each 1000 bp). These can be identified by producing and comparing nucleotide sequences of a given region for several individuals. Once the SNPs have been identified they can be typed easily by a number of methods from "easy" ones like restriction enzyme digestion, to "sophisticated" methods like DNA chips, real time PCR machines using TAQ-man probes, or molecular beacons. Their digital nature means that in the future SNPs will probably be the method of choice for monitoring animal and plant species to help define the interactions between natural and cultured populations.

Increasing genomic resources for many species, such as cDNA libraries, EST databases, and DNA microarrays are having a profound impact on research in areas such as agriculture and medicine. Since DNA microarrays can provide expression information for thousands of genes simultaneously, they are now the principal tools for conducting functional genomics research. The fabrication of cDNA microarrays involves gene/clone selection, PCR amplification of transcript sequences using universal primers, purification of PCR products, and robotic printing of PCR products (cDNAs) onto glass slides (6). cDNA or oligonucleotide microarray platforms have been built for several aquaculture species, including Atlantic salmon and common carp. Since microarray-based experiments have identified fish genes potentially contributing to fitness-relevant traits, such as precocious ovary development and

rapid growth, they may be useful in future research aimed at evaluating the impact of escaped aquaculture fish on wild fish populations.

In conclusion, identification of aquaculture escapees and their offspring in the wild may require different markers depending on species and situations. Some factors are decisive regarding how to develop / apply a genetic marker for a given situation; these include the genetic distance between domestic stocks and wild populations, the number of generations after accidental escapes or deliberate introductions occur, the genome structure of a species (with more or less multi-allelic loci). Practical feasibility (cost, equipment, technical difficulty, expertise required) is another aspect which needs to be taken into account, as well as the number of samples and sample sizes (one single episode of escapes? routine surveys?) and finally statistical analysis. When absolute markers exist for differentiating aquaculture escapees and native individuals, statistical analysis does not pose particular problems because immigrants and their offspring can be directly identified. However absolute markers (stock-specific) are rarely found within species, particularly for those with high dispersal capacity, as is the case of many marine animals. Fortunately, many statistical methods, implemented in readily available computer programs, are available which make it possible to detect immigrants in such cases with a high level of probability (7).

Main conclusions- Future research priorities

A number of conclusions can be reached based on the available tools for each of the studied species, (Table 1) as well as the various discussions that followed. These are:

- Several different types of genetic markers are currently available for monitoring cultured and wild fish, each with its own advantages and disadvantages and different genetical features (dominant/codominant, nuclear/mitochondrial DNA, maternal/ biparental mode of inheritance, number of loci detected per assay, number of alleles detected per locus) and different types of analytical procedures are required in each case with different technical and expertise requirements.
- The main markers developed in the studied species and which are currently used involve allozymes, PCR RFLPs (nuclear and mtDNA), RAPDs, AFLPs and microsatellites. The choice of marker in a given context depends on a number of factors including the specific question to be addressed (e.g. population structure, levels of genetic diversity, measuring differential fitness between farmed and wild fish, identification of farmed and wild stock) and the species under investigation (i.e. there is great variance on the amount of information available, in addition to technical and logistical considerations). Therefore, molecular markers should be assessed on a case by case basis. Some generalizations can be made, such as that microsatellites have more power to detect subtle differences, but other, older markers such as allozymes may still be of use in certain situations (they have proven to be valuable genetic tags in cod and in brown trout).
- In many situations, microsatellites will be the marker of choice in attempts to detect cultured individuals in the wild. Success in doing so will depend on having baseline data which can be used to assign the origin of individuals being tested, especially where makers diagnostic for farmed individuals are lacking as is the case in most situations. In such cases even the use of numerous, highly polymorphic loci cannot guarantee success, particularly where differentiation between cultured and wild fish is low. The possibility to detect escapees can be best assessed if there is a detailed knowledge of population structure of wild as

well as farmed populations for the markers considered.

- Information on population structuring is limited or lacking for most of the species of interest and research in this area should be a major priority in areas of aquaculture and situations where stocking is undertaken. Research should include extended spatial as well as temporal monitoring of populations, to ascertain if spatial structuring found is stable across space and time; a single set of samples from one time period may or may not reveal the true situation.
- Despite numerous works on various species with various molecular markers, existing research, data analysis and comparative studies on each species cannot easily be exploited due to lack of standardization among studies and marker characterization (sampling design, appropriate use of controls, replicate screening within and between labs). Intercalibration of results of different laboratories is still minimal. Databases of produced genotypes and of the genetic material of control individuals are still needed in most cases in order for European Science to take advantage of the previous efforts that have been spent on the genetic analyses of species of aquaculture interest.
- Communication with aquaculture stakeholders is also essential in order to be able to record and monitor information on the origin and number of broodstock for every species.
- A particular concern as regards data base integration is how to minimize genotyping errors (which have often been proven to significantly affect results) as well as how to maximize the information obtained from different statistical analyses. More work is needed to assess the statistical properties of the different theoretical models used and more research should be done with real data sets.
- Their distinct advantages mean that in the future SNPs and microarrays are likely to see increasing use. Much work is still needed for identification of SNPs and construction of microarrays. Genomic programmes are in progress or in the process of being initiated in most aquaculture species of interest, but studies of association between specific phenotypes (domestication) and possible linked QTL are still missing. This could help to identify markers for farmed fish. Priority should be given to identifying the genes involved in domestication, i.e. the changes in the genetic architecture of wild populations when brought under farming practices. This should allow the identification of the true functional differences of wild to farmed individuals and should therefore facilitate the identification of farmed escapees based on the genes that matter and not only with supposedly neutral markers. Where genomic resources are still not available for a species of interest, searching for functional polymorphisms and differentiation between wild and farmed individuals should focus on finding informative EST linked loci, which show differentiation between cultured and wild populations.
- More immediate alternatives are to consider the development of new genetically tagged farmed strains where possible, especially for ranching studies. In cases where extended breeding programmes have already started, like salmon, it will be difficult, but the possibility of selecting for a few molecular markers, to provide diagnostic variation, as a required part of selective breeding programmes should be examined for other species.
- The use of non-genetic tags and other strategies e.g. triploids, sterile fish to reduce impact of escapees should also be considered when reviewing monitoring technology options.

Species	Allozymes	Mitochondrial DNA	Micros	AFLPs	SNPs	Genomics	Other
Common carp *Cyprinus carpio*	19 variable loci of 60	D-loop, Cytb, ND3/4, ND5/6	>100	One study	PCR RFLPs	In progress	RAPD, SSCP
Atlantic salmon *Salmo salar*	33 variable loci of 110	D-loop, Cytb, ND1, ND5/6 >20 RFLPs >40 SNPS	~1700	Yes	Yes	Yes, well underway	Blood proteins
European sea bass *Dicentrarchus labrax*	>25	D-loop, Cytb	>250	> 200	In progress	In progress	Linkage map
Gilthead sea bream *Sparus aurata*	21-27	Dloop	127	147	Yes	In progress	
Turbot *Scophthalmus maximus*	>25	Dloop, Cytb, rRNAs, tRNAs	>35	No	No	In progress	
Atlantic cod *Gadus morhua*	>30	Whole	>80	No	>90 ; more in progress	>20 000 ESTs ; more in progress	
Atlantic halibut *Hippoglossus hippoglossus*	>40	No	>25	No	No	In progress	
European flat oyster *Ostrea edulis* *Crassostrea gigas*	22 24	Cytb, rRNAs Cytb, 16S rRNA	24 >120	296 > 300	No >50	No In progress	Genetic map Genetic maps
European lobster *Hommarus gammarus*	>40	Whole	>50	No	No	No	
Mussels *Mytilus edulis* *M. galloprovincialis* *M. trossulus*	15	Whole	7	No	Yes (PCR-FFLP)	In progress	
Scallop *Pecten maximus*	13	Cytb, COI, ND1, rRNAs, tRNAs	9	No	In progress	In progress	

Table 1. *Availability of different molecular markers in the Genimpact species*

References

(1) McGinnity P., Prödohl P. A., Ferguson A., Hynes R., Maoileidigh N. O., Baker N., Cotter D., O'Hea B., Cooke D., Rogan G., Taggart J. and Cross T. (2003). Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. Proceedings of the Royal Society, London. B., 270: 2443-2450.

(2) Liu Z.J. and Cordes J.F. (2004). DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238: 1-37.

(3) Bravington M.V. and Ward R.D. (2004). Microsatellite markers: evaluating their potential for estimating the proportion of hatchery-reared offspring in a stock enhancement programme. Molecular Ecology, 13: 1287-1297.

(4) Verspoor E., Beardmore J.A., Consuegra S., García de Leániz C., Hindar K., Jordan W.C., Koljonen M.-L., Mahkrov A.A, Paaver T., Sánchez, J. A., Skaala Ø., Titov S. and Cross T. F. (2005). Population Structure in the Atlantic

Salmon: Insights From 40 Years of Research into Genetic Protein Variation. Journal of Fish Biology, 67 (Suppl. A): 3-54.

(5) Hansen M.M., Kenchington E. and Nielsen E.E. (2001). Assigning individual fish to populations using microsatellite DNA markers. Fish and Fisheries, 2: 93–112.

(6) Rise M.L., von Schalburg K.R., Brown G.D., Mawer M.A., Devlin R.H., Kuipers N., Busby M., Beetz-Sargent M., Alberto R., Ross A. , Gibbs A., Hunt P., Shukin R., Zeznik J.A., Nelson C., Simon R.M., Jones S.R.M., Smailus D.E., Jones S.J.M., Schein J.E., Marra M.A, Butterfield Y.S.N., Stott J.M., Ng S.H.S., William S. Davidson W.S. and Koop B.F. (2004). Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. Genome Research, 14: 478-490.

(7) Excoffier L. and Heckel G. (2006). Computer programs for population genetic data analysis: a survival guide. Nature Reviews Genetics, 7: 745-758.