FISH and FISHERIES

FISH and FISHERIES, 2008, **9**, 378–395

Functional genomics with microarrays in fish biology and fisheries

Frederick William Goetz[1] *& Simon MacKenzie*[2]

[1]Great Lakes WATER Institute, University of Wisconsin – Milwaukee, 600 E. Greenfield Ave., Milwaukee, WI 53204, USA; [2]Departament de Biología Cel.lular, Fisiologia i Immunologia, Facultat de Biociencies, Unitat de Fisiologia Animal, Universitat Autonoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Abstract

Over the past 10 years, DNA microarrays have been used for the analysis of gene expression during various physiological, developmental or cellular processes in fish. However, in the last few years, investigators have begun to use microarrays on fish to address ecological, evolutionary and environmental questions including the variability of gene expression in natural populations, speciation, ecotypic diversity, environmental remediation and host–pathogen interactions. Given their commercial importance and the availability of several microarray platforms, most of the research in these areas has focused on salmonids. However, the same issues could be studied in a number of fish species of interest to world fisheries. The application of array technology to analyse gene expression in exploited species may require the development of new array platforms containing genes derived specifically from these fish. The gene discovery required for developing these platforms will certainly be facilitated by the new sequencing technologies that have recently been developed. Further, given the quantum leap that has occurred in sequencing, and the likely improvements in these technologies in the immediate future, it may be possible in certain situations to use sequencing in place of arrays to measure global changes in gene expression. Given the current technological development, sequencing has an advantage over arrays in that it can be used as a tool for gene discovery as well as for quantifying gene expression.

Keywords Functional genomics, gene expression, microarrays, transcriptomics

Correspondence:
Frederick William Goetz, Great Lakes WATER Institute, University of Wisconsin – Milwaukee, 600 E. Greenfield Ave., Milwaukee, WI 53204, USA
Tel.: +1 414-382-1742
Fax: +1 414-382-1705
E-mail: rick@uwm.edu

Received 13 May 2008
Accepted 14 Aug 2008

Introduction

Functional genomics is the analysis of genetic information and its biological function. Deoxyribonucleic acid (DNA) sequencing provides the raw material (i.e. genes) for functional genomics. However, transcriptomics, the analysis of messenger ribonucleic acid (mRNA; transcript) expression, is currently the primary tool used to study gene function. The transcriptome is a collection of mRNA within a cell, tissue or organism and, when analysed, represents the available transcripts at a given point in time. Transcriptomes are not static and are influenced by many factors. Therefore, transcript expression profiles represent phenotypes shaped by the genotype of an organism and its environment.

It has been roughly 30 years since the development of Northern blotting as a technique to detect and quantify the expression of mRNA transcripts (Alwine *et al.* 1977). In Northern blotting, RNA or mRNA is separated on an agarose gel and transferred to a membrane that is probed with a labelled cDNA (complementary DNA) matching the gene of interest. Since the development of Northern blots, there has been an upsurge in molecular technology that has driven new methods for the measurement of gene expression. Many of these techniques rely on the polymerase chain reaction (PCR) and one of the first used was reverse transcription PCR (RT–PCR), a process in which mRNA is reverse transcribed to cDNA and then used in a PCR with gene-specific primers. An estimate of relative transcript abundance between samples can be obtained by observing the intensity of the resulting DNA fragments (amplicon) following staining and separation on an agarose gel. Using RT–PCR, the number of genes that could be simultaneously assayed was somewhat limited, but, more importantly, the quantification of the expression was generally crude. Therefore, RT–PCR analysis was quickly followed by the development of real-time quantitative PCR (qPCR) (Higuchi *et al.* 1992). Quantitative PCR still requires reverse transcription of the mRNA, but during the subsequent PCR, the accumulation of PCR product is estimated in 'real-time' by a fluorescent label (e.g. SYBR Green). The rate of increase of PCR product provides an accurate estimate of original mRNA copy number. New qPCR platforms (e.g. TaqMan® Custom Array) allow hundreds of samples to be assayed at a time, and even standard machines run samples in 96-well plate format. While there is still considerable discussion about the statistical analysis of qPCR data (Yuan *et al.* 2006; Karlen *et al.* 2007), there is no doubt that this technology provided two major contributions to the analysis of gene expression by increasing throughput and improving the accuracy of quantification.

As the techniques for the quantification of gene expression were being developed, sequencing technology was also improving rapidly, resulting in an exponential increase in the number of sequences that could be generated, even for non-model organisms (Hauser and Seeb 2008). Thus, it became relatively easy to generate cDNA libraries (collections of cloned cDNAs) and to perform large-scale sequencing of these libraries, producing what are referred to as expressed sequence tags (ESTs). ESTs are subsequences of full transcripts and represent collections of potentially expressed genes. As a result, ESTs became the core materials for the generation of cDNA microarrays, the next platform developed for large-scale expression analysis.

DNA arrays are ordered microscopic arrays of cDNAs or DNA oligonucleotides (small DNA fragments) that are spotted at very high density on substrates such as glass slides or nylon membranes. The DNAs arrayed on a chip can represent the genes expressed within an organism, within a specific tissue or within a tissue at a specific stage. Regardless, since the DNA sequences on the microarray are derived ultimately from mRNA transcripts (e.g. ESTs), they represent a collection of potentially expressed genes. For analysis, cDNA, derived from RNA, is labelled with different fluorescent dyes, commonly the cyanine dyes, Cy3 alone or also with Cy5 depending on the experimental design. The labelled cDNAs are hybridized (incubated) onto the microarray and the relative expression of individual mRNAs in the sample is compared between treatments by measuring and comparing the average fluorescence present at each replicate spot; a specific DNA sequence of a transcript that is present in multiple copies on an array. If a given treatment regulates specific mRNA transcripts, then changes in relative fluorescence for spots corresponding to those transcripts will be detected at their respective positions on the array. Based on the identities of the transcripts on the microarray, the user can then determine what genes or functional classes of genes are being regulated under different experimental conditions, and to what extent.

Over the past few years, a number of microarrays have been developed for fishes (review: Douglas

2006). High-density cDNA or oligoarrays have been developed for developmental studies (e.g. zebrafish) or aquaculture (e.g. salmonids, catfish, seabream and halibut). Most of these chips have been primarily used for fundamental investigations on cell biology or physiology involving development, growth, stress, effects of environmental pollutants and reproduction. In addition, because of the commercial emphasis on Atlantic salmon and rainbow trout, the focus of applications has been on salmonids. Salmonid and zebrafish arrays can be hybridized with samples from other fish species (Rise *et al.* 2004b; Kassahn *et al.* 2007), although some information will probably be lost depending on sequence similarity between species and on the levels of transcripts being assayed. However, the ability to sequence large numbers of ESTs has provided an opportunity to develop species-specific chips for non-model fish as well.

There are other extremely sensitive and comprehensive methods for the assessment of global gene expression including serial analysis of gene expression (SAGE – Velculescu *et al.* 1995) and polony multiplex analysis of gene expression (PMAGE – Kim *et al.* 2007). These techniques do not use hybridization technology like microarrays, but utilize the power of sequencing extremely large numbers of potential transcripts within a sample and then quantifying the frequency that given genes are expressed. However, in SAGE and PMAGE the sequences generated are relatively small (14–21 bp). Thus, for annotation (identification of gene function), the sequences must be compared with homologous genomic sequences, which are known for only a few fish species and not those generally associated with fisheries. SAGE has been employed with zebrafish (Knoll-Gellida *et al.* 2006) and medaka (Ahsan *et al.* 2008), but to our knowledge with no other fish species. There is potential for this type of analysis on other species in the future but that will depend on advances in sequencing technology (Hauser & Seeb 2008).

The intent here is not to review the many reports on gene expression (directed or global) in fish biology. Rather the objective is to discuss the new ways in which microarrays are being used to study biological topics that are, or could be, of high relevance to fisheries biology and management. Many of the examples are from studies on salmonids, primarily as a result of the availability of microarray platforms for this group. However, in most cases, it is easy to envisage a similar use for studies on non-salmonid, marine species. As the examples will show, microarray approaches have now been employed to determine the degree of natural genetic variation within populations, phenotypic divergence and parallel evolution in ecotypes, the basis for differences in life-history strategies, the genomic effects of environmental changes on fish populations and the strategies that fish use in reacting to different pathogens. There is still considerable controversy on the ways in which microarray studies are carried out and analysed (Marshall 2004; Salit 2006; Tarca *et al.* 2006), but that is not a focus of the review and not considered when discussing the examples.

Variation in gene expression in wild fish populations: evidence for natural selection and adaptation

Perhaps one of the most interesting uses of microarrays in fish has been to study genomic variation within natural populations (Oleksiak *et al.* 2002, 2005; Larsen *et al.* 2007, 2008) and between sympatric ecotypes (Derome and Bernatchez 2006; Derome *et al.* 2006; St-Cyr *et al.* 2008). Working with a microarray composed of *Fundulus heteroclitus* (Fundulidae) cardiac cDNAs, Oleksiak *et al.* (2002) demonstrated that, while there are some differences in the expression of certain genes between *Fundulus* sampled from different populations (e.g. northern vs. southern), there was also large variation in gene expression between individuals within a given population. In experiments on *Fundulus* hearts provided with different energy substrates, there was very high individual variation in the metabolic use of the substrate, in mRNA expression and genes associated with substrate-specific metabolism (Oleksiak *et al.* 2005). Similar results have also been reported for salmonids (Giger *et al.* 2006, 2008). These experiments indicate that there is significant variation in gene expression among individuals that should be taken into account when utilizing global gene expression approaches.

In one of the few microarray studies on global gene expression in wild populations of a marine fish species, Larsen *et al.* (2007) examined the differences in liver transcriptomes in European flounder (*Platichthys flesus*, Pleuronectidae) using a species-specific cDNA array derived from liver ESTs (Williams *et al.* 2003, 2006). Transcriptomic differences were observed directly from the livers of wild fish sampled from the Baltic and North Sea, and

after reciprocal transplantation and long-term acclimation to the differing salinities in the Baltic (9 ± 0.5 ppt) and North (33 ± 0.5 ppt) Seas. Microsatellite analysis indicated that neutral genetic differentiation between the two populations was very low (F_{ST} = 0.006). Even so, microarray analysis indicated differential expression of transcripts between the populations including ependymin, apolipoprotein 1A, ferritin H1, beta-2 macroglobulin, MHC class II genes and transferrin. Some of these genes have also been observed to be differentially expressed between populations of Atlantic salmon (*Salmo salar*, Salmonidae) (Roberge *et al.* 2006). In addition, Larsen *et al.* (2007) observed differences in the expression of genes involved with iron/haem biosynthesis that they hypothesized might be related to the adaptation of the two flounder populations to environmental conditions with different oxygen affinity requirements. In the transplantation/acclimation experiments, they observed differential expression of genes such as the angiotensinogen precursor, prothrombin, growth hormone and insulin-like growth factor 1 (Larsen *et al.* 2007). Changes in the expression of growth hormone and angiotensinogen under different salinity have also been shown in rainbow trout and sea bream (Shepherd *et al.* 2005; Wong *et al.* 2006). Further, studies on brown trout (*Salmo trutta*, Salmonidae) have demonstrated differences between separate populations in the expression of osmoregulatory and stress genes in relation to salinity change; suggesting the presence of local adaptations in gene expression (Larsen *et al.* 2008). However, it is important to point out that while the transplantation/acclimation approach used in the flounder study (Larsen *et al.* 2007) would decrease the effects of non-genetic factors on gene expression, it does not necessarily remove all such factors. Maternal and/or the effects of prior life-history attributes could still influence gene expression in wild fish acclimated and tested in the laboratory.

Variation in phenotypic traits across the geographic range of a species is common. Strong environmental selection can produce 'ecotypes,' local morphological and physiological variants associated with specific ecological conditions. Ecotypes are well documented in several subfamilies of the Salmonidae, including chars (Adams and Maitland 2006) and whitefishes (Lu and Bernatchez 1999; Ostbye *et al.* 2005). While there have been several investigations on the genetic relationships among ecotypes (Lu and Bernatchez 1999; Wilson *et al.* 2004), the cellular and genomic basis for phenotypic differences has only recently been examined.

In a series of papers, Bernatchez *et al.* used microarrays to examine differences in global gene expression between sympatric whitefish (*Coregonus clupeaformis*, Coregonidae) ecotypes, and between ciscoes (*Coregonus artedi*, Coregonidae) and whitefish. Dwarf and normal whitefish ecotypes are found in small lakes in northern Maine and southeastern Quebec (Lu and Bernatchez 1999). These ecotypes utilize different niches, dwarfs occupying the pelagic zone and feeding on plankton, whereas normal whitefish forage on benthic prey. In association with these different niches, dwarf whitefish are more active swimmers with higher metabolism and have lower growth rates than normal whitefish.

A study by Derome *et al.* (2006) used the Genomics Research on Atlantic Salmon Project (GRASP) 3557 cDNA microarray (Rise *et al.* 2004b) to examine global gene expression differences in white muscle between normal and dwarf whitefish populations in two separate lakes. The GRASP chip contains cDNAs from several salmonids and can be hybridized across species (Rise *et al.* 2004b). The question addressed in this study was whether parallel phenotypic evolution in different lakes is associated with parallel genomic expression and, if so, what types of genes were important. The results demonstrated that there were parallel expression differences for 16 genes on the chip between the two whitefish ecotypes sampled from two separate lakes. Of these 16 genes, 10 were directionally upregulated in dwarf ecotypes of both lakes relative to the normal whitefish, while five were higher in normal whitefish than in dwarf ecotypes (Fig. 1). However, six of the 10 genes upregulated in dwarfs aligned to parvalbumin homologues (Fig. 1), intracellular calcium binding proteins that are involved in the regulation of muscle contraction. Reportedly, parvalbumins decrease the time of muscle relaxation between contractions, thereby allowing for an increased number of contractions/time. Other genes that were upregulated in dwarfs included gamma-crystallin, which Derome *et al.* (2006) suggested was actually lactate dehydrogenase based on the evolution and function of the beta/gamma-crystallin superfamily. Lactate dehydrogenase is correlated with glycolytic capacity, and higher levels have been reported in the more active cisco relative to normal whitefish. Thus, the genes upregulated in dwarf ecotypes

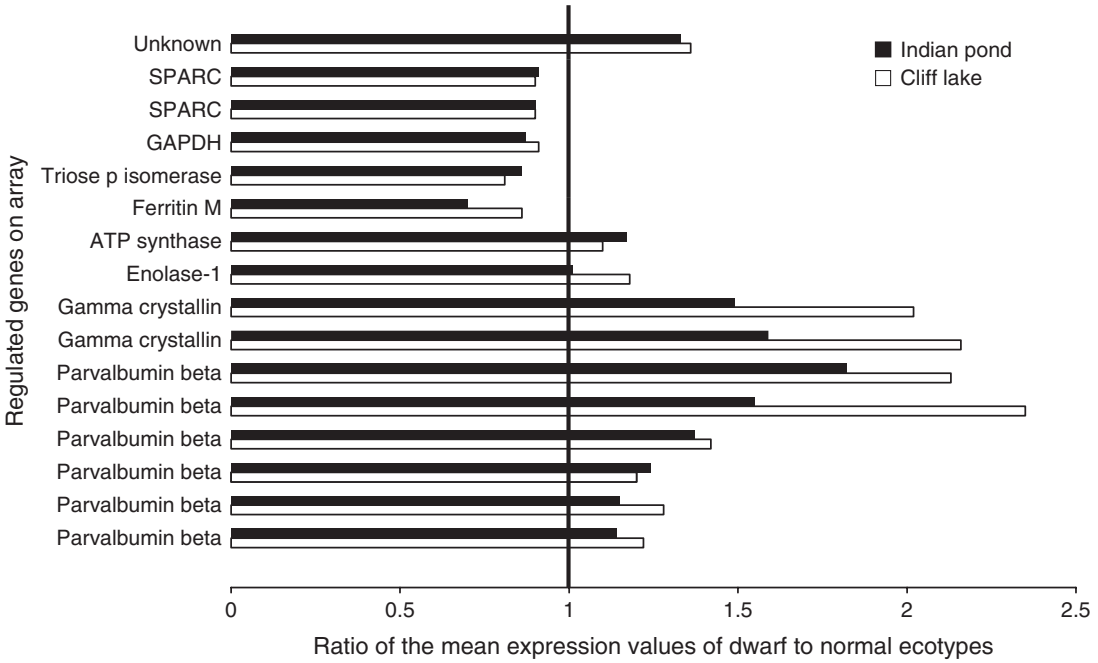

Figure 1 Parallel directional changes in white muscle genes between dwarf and normal whitefish (*Coregonus clupeaformis*, Coregonidae) ecotypes taken from two different populations (Indian Pond and Cliff Lake). Bars represent the ratio of the mean expression values of dwarf to normal ecotypes with bars >1 (dark line) implying upregulation in dwarfs, and <1 implying upregulation in normal ecotypes. Each gene corresponds to a different gene element on the microarray (GRASP 3557) even though they may have the same annotation (e.g. parvalbumin). SPARC, secreted protein, acidic, rich in cysteine; GAPDH, glyceraldehyde-3-phosphate dehydrogenase. Figure adapted from Table I – Derome *et al.* 2006.

would favour enhanced muscle contraction and energy production. The authors suggest that the genes exhibiting parallel directional changes in the ecotypes across the two lakes correlate with the phenotypic differences observed between these ecotypes, with dwarfs being more active swimmers with higher metabolic activity than normal whitefish. Their study provides evidence that parallel phenotypic evolution involves parallel genotypic evolution. Further, based on the small number of genes that they observed to be differentially regulated in parallel, the authors suggest that adaptive phenotypic divergence is accompanied by relatively few genetic changes. However, the limited gene coverage offered by the GRASP 3557 microarray should be taken into account when drawing these conclusions.

In a second study, Derome and Bernatchez (2006) used the GRASP 3557 cDNA microarray to compare gene expression in white muscle between white fish ecotypes and the cisco, another whitefish species that occupies the same ecological niche as the dwarf whitefish. However, dwarf whitefish do not co-occur with ciscoes, most likely as ciscoes outcompete the dwarf ecotype (Derome and Bernatchez 2006). Ciscoes and dwarfs are similar in that they are very active swimmers with high metabolic rates and correspondingly low growth rates compared with normal whitefish. In the cisco study, the hypothesis tested was whether the transcriptomic changes that were initially observed between dwarf and normal whitefish would also be observed in ciscoes given the trophic overlap between them and the dwarf whitefish. In comparison with normal whitefish, there were 18 genes that were differentially expressed in the same direction when comparing two cisco populations with normal whitefish populations. As with dwarfs, parvalbumin and gamma-crystallin were upregulated in ciscoes compared with normal whitefish. When ciscoes were compared with dwarfs, both pavalbumin and gamma-crystallin were upregulated in ciscoes suggesting that ciscoes may outcompete dwarfs because they are more active. Several structural genes, including those for actin and myosin proteins, were lower in ciscoes compared

with both dwarf and normal whitefish. The authors speculated that this pattern 'may reflect a trade-off between force and speed in muscle contraction as a reduced actomyosin filament concentration characterizes fastest twitch fibres, whereas high concentration of parvalbumin may be associated with increased rate of muscle contraction' (Rome 2006). Thus the cisco may favour speed over strength of contraction. Overall, when analysed on a functional group basis, the types of genes regulated in ciscoes relative to normal whitefish were similar to those regulated by the dwarf ecotype in relation to whitefish but more accentuated in the case of ciscoes (Fig. 2).

In a third paper from this group, the larger GRASP 16006 gene cDNA array (von Schalburg *et al.* 2005) was used to investigate global gene expression in the livers of dwarf and normal whitefish ecotypes (St-Cyr *et al.* 2008). Given the differences in growth, metabolism and overall activity, it was hypothesized that genes related to 'survival through enhanced activity' (e.g. genes related to energy and lipid metabolism) would be significantly elevated in dwarf vs. normal whitefish ecotypes, and that the opposite relationship would be observed with genes potentially associated with 'growth' (e.g. genes related to protein synthesis, cell growth and cell cycle). Given the larger number of genes represented on the 16006 GRASP array and the more complex tissue, more genes were differentially regulated between dwarf and normal whitefish than in the previous studies on muscle. When analysed as functional groups, the differential expression of genes generally supported the hypothesis that there appeared to be a 'trade-off' in the life-history strategy of these two ecotypes, with dwarfs having enhanced metabolism, activity and energy for foraging and predatory avoidance but slower growth and reduced fecundity. Dwarf whitefish had a higher percentage of upregulated genes in general (70.6%), and most of these genes were related to energy metabolism, lipid metabolism, iron metabolism and detoxification, genes that were considered to be involved with enhanced metabolic activity. However, the difference between dwarf and normal whitefish in expression of genes that could potentially be involved with growth and fecundity was less clear.

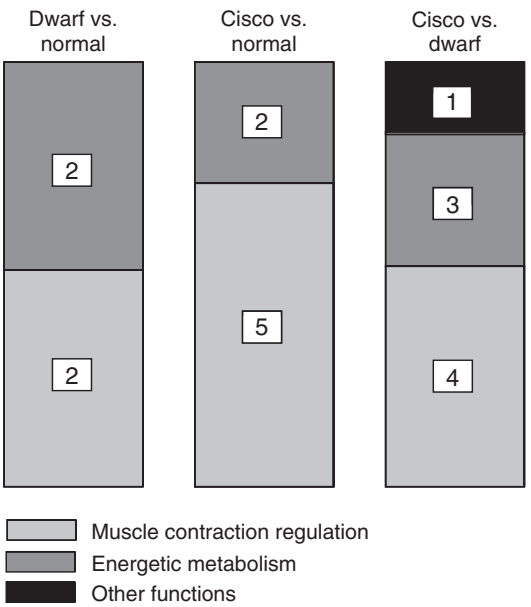

Figure 2 Partitioning of the main functional groups associated with upregulated white muscle genes in the dwarf vs. normal whitefish (*Coregonus clupeaformis*, Coregonidae), cisco (*Coregonus artedi*, Coregonidae) vs. normal whitefish and cisco vs. dwarf whitefish array (GRASP 3557) comparisons. Numbers in bars indicate the number of differentially expressed white muscle genes observed. Figure modified from Figure 4 – Derome and Bernatchez 2006.

For mammals, whole genome transcript profiling is possible (e.g. Illumina Sentrix® Human Expression BeadChips, Illumina, Inc., San Diego, CA, USA). The capacity to profile the entire transcriptome of an organism with a sequenced genome has led to the ability to determine the heritability of transcript expression and the characterization of what are called expression quantitative trait loci (eQTLs; Goring *et al.* 2007). eQTLs are obtained by combining QTL mapping (Naish and Hard 2008) with gene expression profiling, and allow for the identification of loci associated with the expression of a transcript. This approach is a powerful way to identify candidate genes that are involved in complex phenotypes. Recently, eQTL analysis was conducted on the dwarf and normal whitefish species pairs discussed above (Derome *et al.* 2008). There were 253 genes differentially expressed in white muscle between these two coregonid ecotypes, and 33 of them were associated with 53 eQTL. These differentially expressed genes belonged primarily to three functional groups, including (i) muscle contraction, energetics and repair, (ii) energetics and metabolism and (iii) protein synthesis. Of these, 12 genes were expressed in the same direction between dwarf and normal whitefish raised either in controlled envi-

ronmental conditions or between wild fish analysed for the original study by Derome *et al.* (2006). Besides the mapping of transcripts that were expressed differentially between ecotypes, major findings from these analyses were the larger number of eQTL in female datasets than in male datasets of equal size, and several major eQTL 'hotspots' (aggregations of eQTLs) in females.

A study was recently conducted to investigate possible differences in gene expression between brown trout exhibiting migratory and 'sedentary' (non-migratory) behaviour (Giger *et al.* 2008). Liver transcriptomes were surveyed for 80 juvenile brown trout sampled from three migratory and three sedentary populations. These were analysed on an individual basis using an array containing 1098 gene probes. Thus, this study could take into account the variation between individuals that was highlighted earlier as a possible issue for microarray studies. A migratory Atlantic salmon population was also compared to the brown trout populations. From the analysis of all pairwise population comparisons between and within different populations, there was a subset of five genes that consistently showed a difference in expression between life-history phenotypes (i.e. migratory vs. sedentary) but not between populations within a phenotype. They included two genes with unknown functions, and transaldolase 1, constitutive heat shock protein HSC70-1 and endozepine. An additional study (Amstutz *et al.* 2006) using qPCR has demonstrated a temporal correlation of transaldolase with changes in body condition factor in migratory and sedentary brown trout, validating the results observed with microarrays.

When pooled samples of sedentary and migratory populations were analysed, 227 genes showed differences between phenotypes. Further sorting of these genes using linear discriminant analysis pared this number to 12 genes, whose expression could correctly discriminate migratory vs. sedentary individuals by 98.8%. When novel (not analysed previously) individuals of both phenotypes were analysed using these 12 genes, 17 of 18 individuals were correctly classified as being migratory or sedentary. When all of the regulated genes were analysed according to their gene ontology (GO) annotation (classification based on function; http://www.geneontology.org/), genes within several related GO functions were observed that were consistently regulated between phenotypes (Fig. 3). Based on GO annotation, a number of these functions appeared to be focused on structural and binding aspects of blood, and therefore might be adaptations to oxygen concentrations in the environment. This is interesting as Larsen *et al.* (2007) also found that a number of genes related to iron/haem biosynthesis were different between flounder populations and speculated that this might also indicate adaptation to environments with different oxygen characteristics.

Finally, a recent study (Roberge *et al.* 2007a) used the GRASP 16 006 gene array to investigate the differences in transcriptomes between two Atlantic salmon subpopulations that reproduce in upstream and downstream areas of the same river

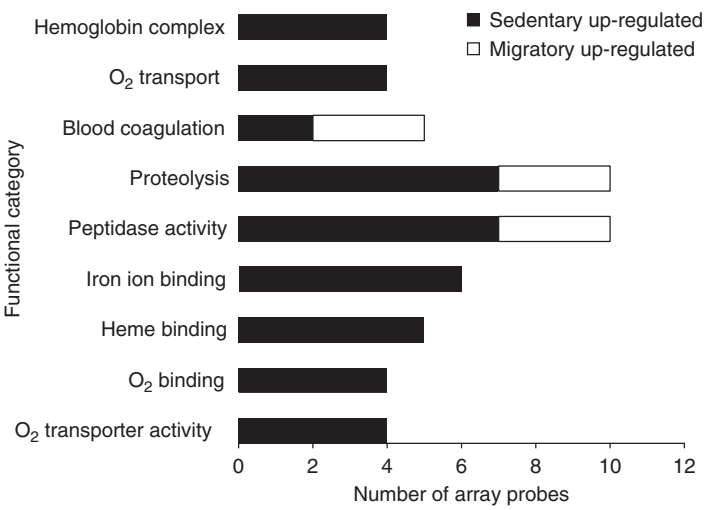

Figure 3 Gene ontology (GO) groups that showed significant over representation in the list of genes that were significantly differentially regulated between sedentary and migratory brown trout (*Salmo trutta*, Salmonidae) life-history phenotypes. 'Functional Category' is a description of the GO group and bars represent the number of gene elements on the array within a given GO category that were either upregulated in the livers of sedentary (black bars) or migratory (white bars) phenotypes. Figure adapted from Table 5 – Giger *et al.* 2008.

and, based on microsatellite analysis, exhibit some genetic divergence. The unique aspect of this investigation was the use of microarrays to analyse samples from half-sib families of crosses between these subpopulations and to determine the heritability of genes whose transcripts were detected on the microarrays. The authors applied a 'transcriptome scan' in which they calculated the distribution of quantitative genetic differences between the two populations (Q_{ST}) for genes that exhibited significant heritability. The hypothesis was that genes falling into the upper tail of this distribution would potentially be under directional selection and diverging between populations. The divergence of gene expression between the progeny of the two subpopulations was also analysed without considering heritability. The transcriptome scan identified 16 cDNAs on the microarray that fell within the upper 1.5% of the distribution and were, therefore, considered candidates for directional selection. Ten of these genes were not annotated and the remaining six represented various functional classes including oxidant defence, RNA processing/transcription, immunity (e.g. bacteriolysis, granulocyte markers), and tetrahydrobiopterin recycling and biosynthesis. Interestingly, the analysis of genes differentially expressed between the two populations in the absence of heritability revealed 12 genes that, while overlapping in some functional classes with the transcriptome scan (e.g. immunity), held only one gene in common, nucleolar RNA helicase II. Several possible reasons were presented to explain why this might have occurred (Roberge *et al.* 2008). This paper is of particular significance in that the authors used microarrays to determine genes that would potentially be under directional selection by analysing differences in gene expression as heritable phenotypic traits.

The studies described above demonstrate that gene expression profiles can successfully characterize life-history phenotypes and therefore can be used as predictors. What is unknown at this point is whether the transcriptional profiles are the consequence or the cause of these phenotypes.

Variation in gene expression in impacted fish populations: evidence for anthropogenic influences

Microarrays have been used recently to assess global changes in transcriptomes of wild fish populations resulting from human impacts including the escape of farmed fish and the release of pollutants to the environment. There has been considerable concern about the effects on wild populations of genetically engineered and/or artificially selected fish in the aquaculture industry. Fish that may escape from net pens are considered to be a significant risk to wild populations of the same species (McGinnity *et al.* 2003). This is a significant problem for the Atlantic salmon (*S. salar*) industry where there may be large numbers of escaped fish. However, the impact of these fish on wild populations is unknown as the overall difference at the genomic level of artificially selected salmon is not known. The aquacultured stocks have been selected for fast growth, age at sexual maturity and pathogen resistance, yet what that means from a genomic perspective is unclear. Therefore, a recent study used the GRASP 3557 gene cDNA array to investigate global differences in transcriptomes between farmed and wild Atlantic salmon in Canada and Norway (Roberge *et al.* 2006). Progeny from the breeding programmes in each country were compared to their population of origin (River Namsen, Norway, and St John River, Canada). While there were several genes differentially expressed between farmed and wild salmon, seven genes were differentially expressed in parallel (i.e. regulated in the same direction) between farmed and wild salmon in both Canada and Norway. These genes included several unknowns (no annotation), Ran protein, transducer of ERBB-2, ferritin H2, ATP synthase gamma chain, ATP synthase lipid-binding protein and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Farmed salmon have been selected for faster growth and also for increased fat. Some of the genes that were regulated in parallel between farmed and wild salmon might be implicated in fat metabolism (e.g. lipid binding protein) as well as metabolism in general (e.g. GAPDH). There appeared to be differences between Canadian and Norwegian salmon in the distribution of differentially regulated genes among functional groups, although only one of these differences was statistically significant. Genes involved in energy metabolism, cellular growth and division, metal ion transport and sequestration appeared in larger proportion in Canadian comparisons, while genes for chaperones, blood transport, signal transduction and lysosomal enzymes were in greater proportion in the Norwegian comparisons. These differences could be a result of different selection strategies but could also reflect differences in the population of

origin. As the authors indicate, the significance of their study is the demonstration of differences in transcription between farmed and wild Atlantic salmon after only five to seven generations of domestication, which could have impacts if escaped farmed salmon breed with wild fish.

More recently, Roberge *et al.* (2008) examined gene transcriptional differences between wild salmon and a backcross of farmed × wild hybrids with wild salmon. Even after correcting for differences in experimental design, the number of differentially expressed genes between wild salmon and the backcross was greater than in comparison between wild and farmed salmon (Roberge *et al.* 2006). A majority of the genes that were differentially regulated between wild and farmed salmon in the earlier study showed either non-significant or smaller transcriptional differences, suggesting an additive genetic control of gene expression. However, a number of genes that were differentially regulated between wild and farmed stocks showed higher transcriptional differences between wild and hybrid fish, indicating the occurrence of non-additive effects as well. Further, the authors suggested that the larger number and magnitude of differentially expressed genes observed between wild and backcross fish when compared to pure wild and farmed salmon indicated that interbreeding of wild and farmed salmon could modify the genetic control of gene transcription in natural populations (Roberge *et al.* 2008).

There have been many papers using microarrays to observe the effects of pollutants on fish (Douglas 2006); however, recently there have been several unique studies using microarrays to look at pollutant effects on wild fish populations. Fisher and Oleksiak (2007) used a directed 'metabolic' array containing genes generated from *Fundulus* heart and liver cDNA libraries to study the comparative transcriptomics of wild *Fundulus* from several superfund sites (highly polluted properties designated under the US. Comprehensive Environmental Response, Compensation, and Liability Act as to be remediated) along the Atlantic coast, including New Bedford Harbor (Massachusetts), Newark Bay (New Jersey) and Elizabeth River (Virginia). Array comparisons were made between each superfund site and two nearby reference sites as well as between all three superfund sites. As these investigators have previously reported (Oleksiak *et al.* 2002, 2005), there was considerable variation in gene expression among individuals within a given superfund or reference population. When comparing the fish at a particular superfund site with its two reference sites, the highest (28 genes) number of differentially expressed genes were observed in the New Bedford Harbor site and the lowest (five genes) were observed for comparisons in the Elizabeth River site. Of these, there were several genes in common between the different superfund sites (relative to the reference site), but only two that were common to all three sites: NADH-ubiquinone oxidoreductase AGGG subunit precursor (NDUB2) and thioredoxin. However, only NDUB2 was upregulated in all three superfund sites relative to the reference site. Thioredoxin was more highly expressed in one superfund site but less in the other two sites. NADH–ubiquinone oxidoreductase AGGG subunit precursor is involved in oxidative phosphorylation and the authors indicate that this gene would be a good candidate for 'resistance to pollution stress.' Thioredoxin is a general protein disulphide oxidoreductase. When looking at all genes that were differentially regulated (higher or lower) relative to the reference site, several had clear physiological implications including cytochrome P450 1B1 (CYP1B1) and cytochrome P450 2N2 (CYP2N2) that are involved in the biotransformation of pollutants. Based on the overall data, the authors suggest that 'multiple *Fundulus* populations have independently evolved common mechanisms of adaptive resistance to complex suites of pollutants.'

In another paper on *Fundulus*, Roling *et al.* (2007) used a targeted cDNA array of *Fundulus* genes involved in the response to exposure to chromium, arsenic and polyaromatic hydrocarbons, to assess the effectiveness of remediation at a superfund site in Charleston, South Carolina. The array was used to assay liver transcriptomes of *Fundulus* collected at the superfund site at the beginning of remediation (2000) and 3 (2003) and 5 (2005) years later. Besides transcriptomes, body burdens of chromium were also assessed. There were several differentially expressed genes between *Fundulus* livers from the superfund site relative to a nearby reference site, but the overall data indicated that the number of differentially regulated genes was correlated with chromium body burdens and not with the amount of chromium in the sediments (Fig. 4). The relationship for this particular site was complex since remediation resulted in the continual decrease in chromium in the sediments but not necessarily in body burden (Fig. 4), possibly because

of the release of chromium to the water by dredging during the 5-year study. The other interesting observation was that very few genes differentially regulated in wild *Fundulus* from the superfund site were detected in prior studies on *Fundulus* exposed to chromium in the laboratory (Roling *et al.* 2006). The authors suggested that the difference observed between wild and laboratory analyses might be a result of exposure conditions, in particular the length of exposure or the exposure to pollutant mixtures. In this regard, experiments using isogenic rainbow trout (to reduce inter-individual variation) exposed to six model toxicants including ethinylestradiol, trenbolone, 2,2,4,4-tetrabromodiphenyl ether, benzo[α]pyrene, Diquat and chromium VI, demonstrated that there were unique gene expression signatures for each pollutant on the GRASP 16006 cDNA array (Hook *et al.* 2006). Further, the regulation of some genes could be in opposing directions depending on the pollutant effect and mode of action. Additional studies by Hook *et al.* (2008) exposing isogenic trout to ethinylestradiol, 2,2,4,4-tetrabromodiphenyl ether and chromium VI individually or in a mixture, demonstrated that individual pollutant microarray signatures could definitely be observed in fish exposed to a mixture of the three chemicals. However, there were also unique gene expression changes when the three pollutants were mixed together that did not appear in any of the three individual exposures. Thus, a mixture of pollutants, as would be expected in the wild, might elicit a complex gene expression profile not resembling that caused by single pollutants.

Using transcriptomics to understand and combat disease

Pathogen emergence and the spread of infectious disease in natural and cultured fish populations are considered serious threats in marine and freshwater ecosystems and also in aquaculture. In fish, the

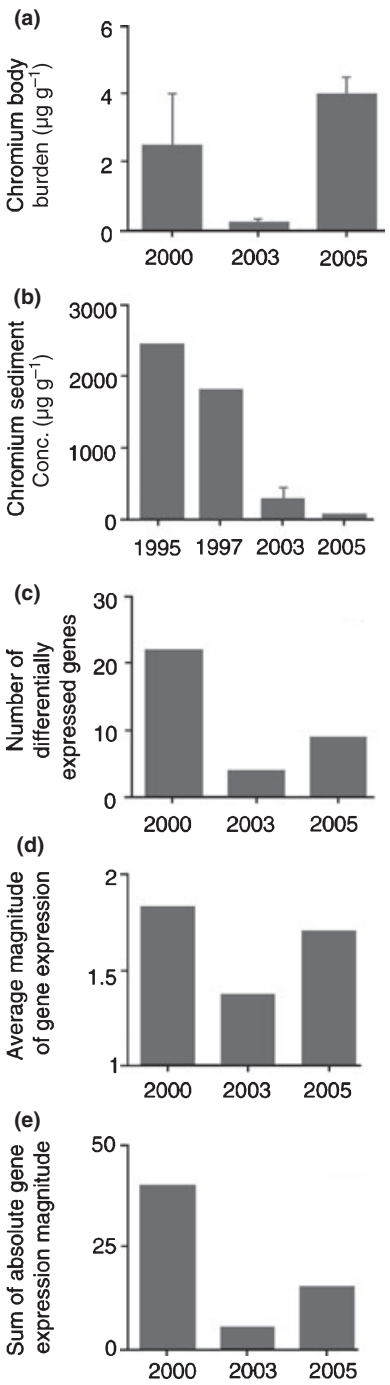

Figure 4 Comparison of chromium sediment concentrations, chromium body burden and number of differentially expressed genes in *Fundulus*. (a) Chromium body burden in fish captured at Shipyard Creek, South Carolina, USA, from 2000 to 2005. (b) Sediment chromium concentrations measured before and during remediation and measurements taken from years 1995 to 2005. (c) The number of differentially expressed genes as determined using targeted *Fundulus* arrays from fish captured from 2000–2005 at Shipyard Creek (contaminated/remediated site) compared to Winyah Bay National Estuarine Research Reserve (NERR – control site). (d) The average absolute magnitude of expression of those genes that are differentially expressed when compared to the NERR site. (e) The absolute magnitude of expression from each gene that was differentially expressed in Shipyard Creek captured fish (figure from Roling *et al.* 2007).

advent and availability of microarray technology has provided tools for researchers to describe viral, bacterial and parasitic disease processes in a number of host species with an emphasis upon aquacultured species (review: Douglas 2006). Host response and resistance to disease involves a complex interplay between the two arms of the immune system: the innate and adaptive immune responses. This response is orchestrated by pathogen recognition receptors, which are present in all animal species and detect specific pathogen groups via pathogen-associated molecular patterns; specific small molecular motifs conserved across pathogen groups (e.g. flagellin – the major component of bacterial flagellum). Through this recognition system, distinct transcription programmes are activated such as the antiviral response, leading to an immediate activation of specific innate host-defence modules (a heterogeneous group of leucocytes acting in a complementary fashion). This is followed by specific adaptive defence modules that act to eliminate a particular class of pathogens (Medzhitov 2007). A specific immune response involves more transcriptional changes than most other physiological responses (Abbas *et al.* 2005; Table 1). Although in its infancy in fish research, microarrays have begun to identify distinct gene expression profiles and specific cassettes of responsive genes whose regulatory patterns are conserved across different fish species in response to specific groups of pathogens such as the Rhabdoviridae.

Table 1 Performance of salmonid fish microarray platform (SFM 1.0) under diverse experimental conditions.

Experimental condition	Treatment (tissue type)	DEG	Relative DEG (%)
Developmental disturbance	M74 (embryos)	507	36
Ecotoxicology	Pyrene (liver)	141	8
	CCL4 (liver)	128	9
	Resin acids (liver)	238	13
	Cadmium (liver)	309	17
Bacterial pathogenicity (*in vitro* macrophage)	LPS	70	5
	LPS (ovary)	3	0.2
Viral pathogenicity (head kidney)	IHNV	426	30

DEG indicates the number of differentially expressed genes ($P < 0.01$) and relative DEG the percentage of expressed genes related to testing clones in each platform. See the references for details: M74 (Vuori *et al.* 2006); pyrene, CCL4, resin acids (Krasnov *et al.* 2005, 2007); cadmium (Merilainen *et al.* 2007); macrophage (MacKenzie *et al.* 2006a); ovary (MacKenzie *et al.* 2006b); IHNV (Mackenzie *et al.* 2008). M74, salmonid yolk-sac fry mortality; CCL4, hepatotoxic compound carbon tetrachloride; LPS, bacterial lipopolysaccharide; IHNV, infectious haemorrhagic necrosis virus.

A number of papers, mainly addressing disease processes in salmonids and in the Japanese flounder (*Paralichthys olivaceus*, Pleuronectidae), have addressed viral diseases including infectious salmon anaemia (Jorgensen *et al.* 2008) and rhabdovirus infections including infectious haematopoietic necrosis (IHN; Purcell *et al.* 2006; Mackenzie *et al.* 2008), viral haemorrhagic septicaemia (VHS; Byon *et al.* 2005, 2006) and hirame habdovirus infection (HIRRV; Yasuike *et al.* 2007). For bacterial diseases in salmonids, both *Aeromonas salmonicida* (Aeromonadaceae) infections (Tsoi *et al.* 2003; Ewart *et al.* 2005; Martin *et al.* 2006) and Salmon Rickettsial Disease (Rise *et al.* 2004a) have been investigated. In other fish species, infection with *Edwardsiella ictaluri* (Enterobacteriaceae), the causative agent of enteric septicaemia in the catfish (Peatman *et al.* 2007, 2008) and Japanese flounder (Matsuyama *et al.* 2007a), have been addressed. Differential gene expression profiles reported in fish for other pathogens include the fungal disease, saprolegniasis (Roberge *et al.* 2007b), and amoebic gill disease in the Atlantic salmon (Morrison *et al.* 2006), the protozoan *Myxobolus cerebralis* (Myxobolidae) infection in the rainbow trout (Baerwald *et al.* 2008), and infection by the monogenean parasite *Neoheterobothrium hirame* (Diclidophoridae) in the Japanese flounder (Matsuyama *et al.* 2007b). Significant progress has also been made in studies addressing pathogen genomes, which is highlighted by the publication of the first full genome of a fish pathogen, *Flavobacterium psychrophilum* (Flavobacteriaceae) the causative agent of rainbow trout fry syndrome and bacterial cold-water disease (Duchaud *et al.* 2007).

In a commercial setting, the contribution of transcriptomics to disease control management can take two significant forms: (i) the design, development and evaluation of vaccines, and (ii) the development of biomarkers to support selective breeding programmes for disease resistance. Due to the relatively large body of published research using transcriptomics to study host–pathogen interactions in fish, in the present review we have chosen to highlight selected examples of published research addressing rhabdoviral infection and recent progress towards prognostic markers for disease resistance in salmonids.

Development and evaluation of vaccines and characterization of host–vaccine responses

A good example of the potential of microarray analysis to help understand pathogenesis in fishes has been the elucidation of innate and adaptive immune responses to rhabdoviral infection. DNA vaccines against rhabdoviral infections (IHN, VHS, HIRR) have proven to be highly successful in fish (Lorenzen and LaPatra 2005). Intramuscularly administered DNA containing the viral G-glycoprotein gene confers protective immunity in all species so far vaccinated and tested (Boudinot *et al.* 1998; LaPatra *et al.* 2001; Byon *et al.* 2005; Seo *et al.* 2006; Yasuike *et al.* 2007). A DNA vaccine against IHN is currently under commercial license in Canada (Novartis Animal Health Canada, Inc.). Kim *et al.* (2000) used three DNA vaccine preparations of serologically unrelated rhabdoviruses, each encoding the G gene of either IHN virus (IHNV), snakehead rhabdovirus or spring viraemia of carp virus, and were able to show protective immunity in rainbow trout subjected to a lethal dose of IHNV. This suggests that rhadoviral G proteins elicit a non-specific antiviral immune response.

In a series of papers, Aoki *et al*. used a cDNA microarray enriched with 213 immune-related genes, to study the immune response and the efficacy of DNA vaccines containing the viral G-protein genes of VHS virus (VHSV) and HIRRV in the Japanese flounder (Byon *et al.* 2005; Yasuike *et al.* 2007). As expected, all DNA vaccines containing the viral G-glycoprotein gene conferred specific protection to fishes challenged 1 month after vaccination. The protection was hypothesized to occur via the type 1 interferon (IFN) system due to the number of IFN-related genes upregulated in both studies including IFN-stimulated gene 15 kDa (ISG15), IFN-stimulated gene 56 kDa (ISG56) and the Mx protein. In the previously mentioned studies on VHSV and HIRRV in the Japanese flounder, the majority of differentially upregulated genes were identified between 3 and 7 days post-vaccination. Interestingly, Mx, an antiviral protein that is commonly used as a marker for antiviral activity in animal species, was consistently upregulated across vaccinations (Yasuike *et al.* 2007). A similar observation, in which IRF-3, Mx, Vig-1 and Vig-8 were upregulated, was reported by Purcell *et al.* (2006) using the 16 006 feature GRASP cDNA microarray to expression profile rainbow trout at the site of DNA vaccination against IHNV 7 days post-vaccination. These results suggest that the host-expressed viral glycoprotein induces a systemic antiviral state indicative of the non-specific type 1 IFN innate immune response. However, the mechanism(s) for the development of a specific cytotoxic T- or B-lymphocyte-mediated humoural response in fish vaccinated with the plasmid DNA-IHNV G gene, which confers protective immunity over periods up to 25 months post-vaccination, is yet to be clearly identified and remains an area of uncertainty in fish immunology (Kurath *et al.* 2006).

In two recent studies in which virulent forms of IHNV and infectious salmon anemia virus (ISAV) were analysed using a targeted 1.8K salmonid array (Mackenzie *et al.* 2008), evidence from specific gene expression profiles suggests that the adaptive immune response is indeed activated at early stages in the response to viral infection. Mackenzie *et al.* (2008) compared IHNV, attenuated IHNV and LPS challenges in rainbow trout over a short time frame of 1 and 3 days post-challenge. Three days after infection, a significant change in the transcriptional programme of the head kidney revealed an immunological shift orientated towards the activation of adaptive immunity. This shift was IHNV-dependent as determined by differences between the attenuated and virulent IHNV-specific expression profiles. The rapid systemic spread of IHNV inhibited TNFα, MHC class I, and several macrophage and cell cycle/differentiation markers and favoured an MHC class II, immunoglobulin and MMP/TBX4 enhanced immune response. Using the same salmonid microarray, Jorgensen *et al.* (2008) reported a significant increase in adaptive immunity markers after ISAV infection in Atlantic salmon and, importantly, a progressive increase in IgZ mRNA in parallel with a decrease in IgM expression that peaked >30 days post-infection. A coordinated increase in a suite of genes related to B-lymphocyte differentiation and maturation and activation of T-lymphocyte-mediated immunity, including CD4, TGFβ, CD8a and IFNγ, was also reported. Therefore, from the results of microarray studies, evidence is accumulating for the coordinated regulation of the two arms of the piscine immune system in response to viral infection and illustrates the utility of this transcriptomic approach.

An interesting microarray study published by Dang *et al.* (2007) utilized a viral DNA microarray containing 92 putative transcripts for the red seabream iridovirus (RSIV) to characterize transcriptional profiles of RSIV in the spleen and kidney

of infected red seabream (*Pagrus major*, Sparidae). Of particular note is that the iridoviral disease has been described in a large number of marine species (Kawakami and Nakajima 2002), and in a recent study, 10 different ornamental fish species tested positive for infectious spleen and kidney necrosis virus disease (Jeong *et al.* 2008). Furthermore, a number of fish were asymptomatic carriers that were considered to constitute a significant risk for the transmission and outbreak of this viral disease. Using the viral microarray, Dang *et al.* (2007) were able to identify the spleen as being the most susceptible organ for viral gene expression and replication in which the classic temporal expression pattern for viral genes was observed including the coordinated expression of immediate-early, early and late genes throughout the 10-day infective period. Highly expressed viral genes may provide material for the development of DNA vaccines against this disease. Interestingly, the spleen has been identified as a target organ for both the Taiwan and Singapore Grouper viruses (TGIV, SGIV), both of which belong to the same viral group as RSIV (Qin *et al.* 2002; Chao *et al.* 2004).

Prognostic biomarkers for disease resistance

As the identification of virulence genes in pathogens is of key importance for vaccine development, the identification of prognostic biomarkers for disease resistance is a major aim for aquaculture. Disease resistance is normally measured by challenging with a pathogen of interest and then assessing cumulative mortalities. Surviving fish, or non-challenged siblings from the same family, are then considered 'resistant.' This process is costly and there is a need for prognostic measurements of resistance ideally based upon molecular determinants of resistance that can be measured by non-lethal methodologies. In a first example of this approach, Rise *et al.* (2004a) used the GRASP 3557 cDNA array to identify biomarkers of *in vitro* macrophage and *in vivo* head kidney (major haematopoietic tissue in fish) responses to *Piscirickettsia salmonis* (Piscirickettsiaceae) infection. This approach yielded a number of regulated genes common to both challenges, and the researchers proposed that 19 highly regulated transcripts could be used as potential biomarkers to evaluate the efficacy of vaccines against *P. salmonis*. C-type lectin 2-1, a gene whose product is involved in endocytosis and the C/EBP-driven inflammatory response (Matsumoto *et al.* 1999), was identified in the Rise *et al.*'s (2004a) study, and has been identified in almost all reports in which bacterial preparations have been used to challenge fish *in vivo* (Rise *et al.* 2004a; Ewart *et al.* 2005; MacKenzie *et al.* 2006a; Mackenzie *et al.* 2008; Martin *et al.* 2006). The utility of using C-type lectin and other potential biomarkers for disease resistance will require further effort to develop if individual variation in disease resistance is taken into account. This includes both genetic and epigenetic factors that will influence the immune performance of an individual in a given environment.

The ability to compare microarray experiments across different platforms still remains a difficult and relatively unaddressed issue in studies on fish. Using the salmonid cDNA 1.8K microarray discussed earlier, Mackenzie *et al.* (2008) compared the response of the rainbow trout head kidney to *in vivo* virulent IHNV, attenuated IHNV and bacterial lipopolysaccharide (LPS) challenge. From selected differentially expressed genes, the differential response was evaluated using a database containing over 155 experiments performed with the same array platform to ascertain those genes co-expressed in each or all experimental conditions. A total of 49 regulated genes specific to both LPS and viral stimuli were identified that were often directionally opposed. A further analysis restricted to viral-specific genes gave a total of 28 genes that were highly specific to viral responses when compared across multiple experiments. Using the same microarray platform, Jorgensen *et al.* (2008) extensively analysed a highly virulent ISAV infection in Atlantic salmon to identify differences between early and late mortalities and further characterize molecular determinants of resistance. Using linear discriminant analysis based upon qPCR, they were able to identify a minimum set of genes including 5-lipoxygenase activating protein, cytochrome P450 2K4, galectin-9 and annexin A1. These genes were chosen from an unbiased (expression profiles only, no functional inference) microarray dataset derived from individual liver samples and could be used to predict whether an individual fish fell into an early or late mortality group.

In summary, fish immunomics has provided a substantial number of datasets and potential tools that are mainly geared towards the study of disease processes in species of commercial aquaculture interest. The major aim has been to functionally identify the intensity of responses to specific

pathogens and their associated molecular components and to identify transcripts in a whole organism or specific tissue context that contribute to these responses. Microarrays have been used to investigate the basis for disease resistance in aquacultured strains of rainbow trout (e.g. whirling disease – Baerwald *et al.* 2008), but could be used for investigating resistance in wild populations of fish as well. However, the complex biology of the immune response, in which different spatial–temporal expression profiles will occur in multiple cell types at distinct body locations, make complete mapping of a response difficult and expensive. Thus, the identification through global transcriptomics of host- and pathogen-specific expression profiles and candidate gene markers will make an important contribution to vaccine design as well as the identification of asymptomatic pathogen carriers.

Future of global expression analysis

There have been many investigations that have used microarrays with fish, but the studies highlighted here were chosen based on their unique experimental designs or for the range of biological problems addressed. We also believe that the examples cover the use of transcriptomic analysis to address questions that could be applied generally to species relevant to the marine fisheries.

While there has been significant progress in the use of microarrays for analysing gene expression in fish, there are some significant limitations, some of which are a result of the array platforms currently available. In many cases, existing platforms may be missing a significant number of genes that could be important for the interpretation of array experiments. This issue was highlighted earlier in relation to the number of genes altered during specific immune responses (Table 1) but may apply to other physiological processes that we are not aware of.

Owing to the nature of cDNA hybridization and the inherent long-length of cDNAs spotted onto slides, significant cross-hybridization will be obtained from mRNAs that share a high level of homology. Thus, cDNA arrays are unable to discern between closely related mRNAs, and the resultant signal may represent a family of mRNAs hybridizing to the same target. Therefore, subtle variations in the differential expression of related genes may not be detected and would require a complex oligoarray design that is currently only available for model species. In the end, however, the biological question that is being addressed should be the primary issue and the approach that is used needs to take that into account together with the technological resources available to the investigator. As cDNA microarrays are capable of describing regulated genes, pathways or networks, this level of descriptive complexity is normally sufficient to identify regulatory patterns of interest that provide an initial road map for future studies. Further, cross-hybridization on cDNA arrays may not be a problem when addressing comparative gene expression studies across non-model fish species. Heterologous hybridization to cDNA microarrays using samples from closely related species has been reported in salmonids (von Schalburg *et al.* 2005) and in cichlids (Renn *et al.* 2004), and even between distantly related species (Kassahn *et al.* 2007), showing that cDNA-based transcriptomic profiling is useful as a method to identify both common and divergent traits in fish.

It is common practice to validate or confirm the results obtained from microarrays using qPCR on a subset of regulated genes. However, in many of the examples used in this review, particularly those investigating differential gene expression in ecotypes, alternative life-history strategies or fish under anthropogenic influences, there was no qPCR validation (exception: Roberge *et al.* 2008). This is an important issue given the strong conclusions that were frequently developed around specific genes or gene patterns observed to be regulated on microarrays. Arrays will be used more frequently in the future for the types of ecological and evolutionary investigations discussed in this review, and some validation of the results should be incorporated and expected by reviewers of these papers in the future.

Given the advances in sequencing technology and the reduction in sequencing costs, a reasonable question is whether microarrays will be the tool that is used in the future for analysing gene expression on a global scale (Salisbury 2008). Gene arrays currently have the advantage of being less expensive for large-scale experiments in which, for example, multiple treatments and sampling periods are being analysed. There have been some direct comparisons made for expression analyses between microarrays and sequencing (Chen *et al.* 2007), but those comparisons have not used some of the newest sequencing technologies.

New sequencing platforms such as the Roche Genome Sequencer FLX (454 Life Sciences, Branford, CT, USA), Illumina Solexa (Illumina Inc., San Diego, CA, USA) and ABI SOLiD (Applied Biosystems,

Foster City, CA, USA) have the advantage of providing hundreds of thousand to millions of sequences per run. With that number of sequences, comparative transcriptomics can be performed by analysing gene frequency as recently described for PMAGE (Kim *et al.* 2007). A drawback of some of the new sequence technologies is the short sequence lengths they produce, complicating gene annotation except for organisms with sequenced genomes or large EST databases. In addition, assembly of short reads to contiguous sequences (contigs) for annotation would be difficult. However, some sequencers such as the Roche Genome Sequencer FLX have longer read lengths (200–400 bp) and these have a greater potential for direct annotation and/or assembly to produce larger contigs for annotation. The advantage of using sequencers over arrays to compare transcriptomic profiles is that there are no preconceived ideas of which genes are important. Arrays are only as good as the genes they contain, while sequencers could theoretically provide the entire transcriptomic profile (Hauser and Seeb 2008). If this argument sounds familiar, it should. The same arguments were used 5–10 years ago to justify the use of subtractive cloning, differential display PCR and suppression subtractive hybridization vs. candidate gene approaches. It was argued that subtractive or differential cloning approaches would isolate all the genes important in a given comparison, not just the ones expected to be important based on preconception. The advantage in using sequencing over arrays would be that transcript quantification and gene discovery could be carried out simultaneously. Of course, if sequencing is used in this way, un-normalized libraries or cDNAs must be used. For some tissues (e.g. muscle), this will be a problem given their lack of transcript complexity. In the end, it will become a matter of the number of sequences that need to be generated to obtain a statistical difference between samples that are being compared. Currently there are few published examples other than SAGE to determine this number, and none on non-model organisms.

Microarrays certainly have an important role in that they can efficiently and cost-effectively quantify the difference in expression of many genes simultaneously and this can be done on a large number of samples. At this point, comparing transcriptomes of many samples by sequencing is not feasible. However, until arrays are developed for non-model species that contain every transcript that can theoretically be produced, it would make sense to use sequencing as a first approximation for global gene expression analysis. This approach would be discovery driven and could then be followed by the development of a focused array based on the initial sequencing analysis. The focused array would then be used for the quantification of large numbers of samples. In essence, microarrays would then become to sequencing what qPCR is currently to microarrays, further increasing our ability to resolve differences in gene expression and to quantify those differences across many genes simultaneously.

References

Abbas, A.R., Baldwin, D., Ma, Y. *et al.* (2005) Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* **6**, 319–331.

Adams, C.E. and Maitland, P.S. (2006) Arctic charr in Britain and Ireland – 15 species or one? *Ecology of Freshwater Fish* **16**, 20–28.

Ahsan, B., Kobayashi, D., Yamada, T. *et al.* (2008) UTGB/medaka: genomic resource database for medaka biology. *Nucleic Acids Research* **36**, D747–D752.

Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5350–5354.

Amstutz, U., Giger, T., Champigneulle, A., Day, P.J.R. and Largiader, C.R. (2006) Distinct temporal patterns of tansaldolase 1 gene expression in future migratory and sedentary brown trout (*Salmo trutta*). *Aquaculture* **260**, 326–336.

Baerwald, M.R., Welsh, A.B., Hedrick, R.P. and May, B. (2008) Discovery of genes implicated in whirling disease infection and resistance in rainbow trout using genome-wide expression profiling. *BMC Genomics* **9**, 37.

Boudinot, P., Blanco, M., de Kinkelin, P. and Benmansour, A. (1998) Combined DNA immunization with the glycoprotein gene of viral hemorrhagic septicemia virus and infectious hematopoietic necrosis virus induces double-specific protective immunity and nonspecific response in rainbow trout. *Virology* **249**, 297–306.

Byon, J.Y., Ohira, T., Hirono, I. and Aoki, T. (2005) Use of a cDNA microarray to study immunity against viral hemorrhagic septicemia (VHS) in Japanese flounder (*Paralichthys olivaceus*) following DNA vaccination. *Fish Shellfish Immunology* **18**, 135–147.

Byon, J.Y., Ohira, T., Hirono, I. and Aoki, T. (2006) Comparative immune responses in Japanese flounder, *Paralichthys olivaceus* after vaccination with viral hemorrhagic septicemia virus (VHSV) recombinant glycoprotein and DNA vaccine using a microarray analysis. *Vaccine* **24**, 921–930.

Chao, C.B., Chen, C.Y., Lai, Y.Y., Lin, C.S. and Huang, H.T. (2004) Histological, ultrastructural, and in situ hybridization study on enlarged cells in grouper *Epinephelus* hybrids infected by grouper iridovirus in Taiwan (TGIV). *Diseases of Aquatic Organisms* **58**, 127–142.

Chen, J., Agrawal, V., Rattray, M. *et al.* (2007) A comparison of microarray and MPSS technology platforms for expression analysis of *Arabidopsis*. *BMC Genomics* **8**, 414.

Dang, T.L., Yasuike, M., Hirono, I., Kondo, H. and Aoki, T. (2007) Transcriptional profile of red seabream iridovirus in a fish model as revealed by viral DNA microarrays. *Virus Genes* **35**, 449–461.

Derome, N. and Bernatchez, L. (2006) The transcriptomics of ecological convergence between 2 limnetic coregonine fishes (Salmonidae). *Molecular Biology and Evolution* **23**, 2370–2378.

Derome, N., Duchesne, P. and Bernatchez, L. (2006) Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. *Molecular Ecology* **15**, 1239–1249.

Derome, N., Bougas, B., Rogers, S.M. *et al.* (2008) Pervasive sex-linked effects on transcription regulation as revealed by expression quantitative trait loci mapping in lake whitefish species pairs (*Coregonus* sp., Salmonidae). *Genetics* **179**, 1903–1917.

Douglas, S.E. (2006) Microarray studies of gene expression in fish. *OMICS* **10**, 474–489.

Duchaud, E., Boussaha, M., Loux, V. *et al.* (2007) Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum*. *Nature Biotechnology* **25**, 763–769.

Ewart, K.V., Belanger, J.C., Williams, J. *et al.* (2005) Identification of genes differentially expressed in Atlantic salmon (*Salmo salar*) in response to infection by *Aeromonas salmonicida* using cDNA microarray technology. *Developmental Comparative Immunology* **29**, 333–347.

Fisher, M.A. and Oleksiak, M.F. (2007) Convergence and divergence in gene expression among natural populations exposed to pollution. *BMC Genomics* **8**, 108.

Giger, T., Excoffier, L., Day, P.J. *et al.* (2006) Life history shapes gene expression in salmonids. *Current Biology* **16**, R281–R282.

Giger, T., Excoffier, L., Amstutz, U. *et al.* (2008) Population transcriptomics of life-history variation in the genus *Salmo*. *Molecular Ecology* **17**, 3095–3108.

Goring, H.H., Curran, J.E., Johnson, M.P. *et al.* (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nature Genetics* **39**, 1208–1216.

Hauser L. and Seeb, J.E. (2008) Advances in molecular technology and their impact on fisheries genetics. *Fish and Fisheries* **9**, 473–486.

Higuchi, R., Dollinger, G., Walsh, P.S. and Griffith, R. (1992) Simultaneous amplification and detection of specific DNA sequences. *BioTechnology* **10**, 413–417.

Hook, S.E., Skillman, A.D., Small, J.A. and Schultz, I.R. (2006) Gene expression patterns in rainbow trout, *Oncorhynchus mykiss*, exposed to a suite of model toxicants. *Aquatic Toxicology* **77**, 372–385.

Hook, S.E., Skillman, A.D., Gopalan, B., Small, J.A. and Schultz, I.R. (2008) Gene expression profiles in rainbow trout, *Onchorynchus mykiss*, exposed to a simple chemical mixture. *Toxicological Science* **102**, 42–60.

Jeong, J.B., Kim, H.Y., Jun, L.J. *et al.* (2008) Outbreaks and risks of infectious spleen and kidney necrosis virus disease in freshwater ornamental fishes. *Diseases of Aquatic Organisms* **78**, 209–215.

Jorgensen, S.M., Afanasyev, S. and Krasnov, A. (2008) Gene expression analyses in Atlantic salmon challenged with infectious salmon anemia virus reveal differences between individuals with early, intermediate and late mortality. *BMC Genomics* **9**, 179.

Karlen, Y., McNair, A., Perseguers, S., Mazza, C. and Mermod, N. (2007) Statistical significance of quantitative PCR. *BMC Bioinformatics* **8**, 131.

Kassahn, K.S., Caley, M.J., Ward, A.C., Connolly, A.R., Stone, G. and Crozier, R.H. (2007) Heterologous microarray experiments used to identify the early gene response to heat stress in a coral reef fish. *Molecular Ecology* **16**, 1749–1763.

Kawakami, H. and Nakajima, K. (2002) Cultured Fish Species Affected by Red Sea Bream Iridoviral Disease from 1996 to 2000. *Fish Pathology* **37**, 45–47.

Kim, C.H., Johnson, M.C., Drennan, J.D., Simon, B.E., Thomann, E. and Leong, J.A. (2000) DNA vaccines encoding viral glycoproteins induce nonspecific immunity and Mx protein synthesis in fish. *Journal of Virology* **74**, 7048–7054.

Kim, J.B., Porreca, G.J., Song, L. *et al.* (2007) Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. *Science* **316**, 1481–1484.

Knoll-Gellida, A., Andre, M., Gattegno, T., Forgue, J., Admon, A. and Babin, P.J. (2006) Molecular phenotype of zebrafish ovarian follicle by serial analysis of gene expression and proteomic profiling, and comparison with the transcriptomes of other animals. *BMC Genomics* **7**, 46.

Krasnov, A., Koskinen, H., Rexroad, C., Afanasyev, S., Molsa, H. and Oikari, A. (2005) Transcriptome responses to carbon tetrachloride and pyrene in the kidney and liver of juvenile rainbow trout (*Oncorhynchus mykiss*). *Aquatic Toxicology* **74**, 70–81.

Krasnov, A., Afanasyev, S. and Oikari, A. (2007) Hepatic responses of gene expression in juvenile brown trout (*Salmo trutta lacustris*) exposed to three model contaminants applied singly and in combination. *Environmental Toxicology and Chemistry* **26**, 100–109.

Kurath, G., Garver, K.A., Corbeil, S., Elliott, D.G., Anderson, E.D. and LaPatra, S.E. (2006) Protective immunity and lack of histopathological damage two years after DNA vaccination against infectious hematopoietic necrosis virus in trout. *Vaccine* **24**, 345–354.

LaPatra, S.E., Corbeil, S., Jones, G.R. *et al.* (2001) Protection of rainbow trout against infectious hematopoietic necrosis virus four days after specific or semi-specific DNA vaccination. *Vaccine* **19**, 4011–4019.

Larsen, P.F., Nielsen, E.E., Williams, T.D. *et al.* (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Molecular Ecology* **16**, 4674–4683.

Larsen, P.F., Nielsen, E.E., Koed, A., Thomsen, D.S., Olsvik, P.A. and Loeschcke, V. (2008) Interpopulation differences in expression of candidate genes for salinity tolerance in winter migrating anadromous brown trout (*Salmo trutta* L.). *BMC Genetics* **9**, 12.

Lorenzen, N. and LaPatra, S.E. (2005) DNA vaccines for aquacultured fish. *Revue Scientifique et Technique – Office International des Epizooties* **24**, 201–213.

Lu, G. and Bernatchez, L. (1999) Correlated trophic specialization and genetic divergence in sympatric lake whitefish (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution* **53**, 1491–1505.

MacKenzie, S., Iliev, D., Liarte, C. *et al.* (2006a) Transcriptional analysis of LPS-stimulated activation of trout (*Oncorhynchus mykiss*) monocyte/macrophage cells in primary culture treated with cortisol. *Molecular Immunology* **43**, 1340–1348.

MacKenzie, S., Montserrat, N., Mas, M. *et al.* (2006b) Bacterial lipopolysaccharide induces apoptosis in the trout ovary. *Reproductive Biology and Endocrinology* **4**, 46.

Mackenzie, S., Balasch, J.C., Novoa, B. *et al.* (2008) Comparative analysis of the acute response of the trout, *O. mykiss*, head kidney to *in vivo* challenge with virulent and attenuated infectious hematopoietic necrosis virus and LPS-induced inflammation. *BMC Genomics* **9**, 141.

Marshall, E. (2004) Getting the noise out of gene arrays. *Science* **306**, 630–631.

Martin, S.A., Blaney, S.C., Houlihan, D.F. and Secombes, C.J. (2006) Transcriptome response following administration of a live bacterial vaccine in Atlantic salmon (*Salmo salar*). *Molecular Immunology* **43**, 1900–1911.

Matsumoto, M., Tanaka, T., Kaisho, T. *et al.* (1999) A novel LPS-inducible C-type lectin is a transcriptional target of NF-IL6 in macrophages. *Journal of Immunology* **163**, 5039–5048.

Matsuyama, T., Fujiwara, A., Nakayasu, C. *et al.* (2007a) Gene expression of leucocytes in vaccinated Japanese flounder (*Paralichthys olivaceus*) during the course of experimental infection with *Edwardsiella tarda*. *Fish and Shellfish Immunology* **22**, 598–607.

Matsuyama, T., Fujiwara, A., Nakayasu, C. *et al.* (2007b) Microarray analyses of gene expression in Japanese flounder *Paralichthys olivaceus* leucocytes during monogenean parasite *Neoheterobothrium hirame* infection. *Diseases of Aquatic Organisms* **75**, 79–83.

McGinnity, P., Prodohl, P., Ferguson, A. *et al.* (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 of London Series B – Biological Sciences* **270**, 2443–2450.

Medzhitov, R. (2007) Recognition of microorganisms and activation of the immune response. *Nature* **449**, 819–826.

Merilainen, P.S., Krasnov, A. and Oikari, A. (2007) Time- and concentration-dependent metabolic and genomic responses to exposure to resin acids in brown trout (*Salmo trutta m. lacustris*). *Environmental Toxicology and Chemistry* **26**, 1827–1835.

Morrison, R.N., Cooper, G.A., Koop, B.F. *et al.* (2006) Transcriptome profiling the gills of amoebic gill disease (AGD)-affected Atlantic salmon (*Salmo salar* L.): a role for tumor suppressor p53 in AGD pathogenesis? *Physiological Genomics* **26**, 15–34.

Naish, K.A. and Hard, J.J. (2008) Bridging the gap between the genotype and the phenotype: linking genetic variation, selection and adaptation in fishes *Fish and Fisheries* **9**, 396–422.

Oleksiak, M.F., Churchill, G.A. and Crawford, D.L. (2002) Variation in gene expression within and among natural populations. *Nature Genet* **32**, 261–266.

Oleksiak, M.F., Roach, J.L. and Crawford, D.L. (2005) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nature Genet* **37**, 67–72.

Ostbye, K., Naesje, T.F., Bernatchez, L., Sandlund, O.T. and Hindar, K. (2005) Morphological divergence and origin of sympatric populations of European whitefish (*Coregonus lavaretus* L.) in Lake Femund, Norway. *Journal of Evolutionary Biology* **18**, 683–702.

Peatman, E., Baoprasertkul, P., Terhune, J. *et al.* (2007) Expression analysis of the acute phase response in channel catfish (*Ictalurus punctatus*) after infection with a Gram-negative bacterium. *Developmental and Comparative Immunology* **31**, 1183–1196.

Peatman, E., Terhune, J., Baoprasertkul, P. *et al.* (2008) Microarray analysis of gene expression in the blue catfish liver reveals early activation of the MHC class I pathway after infection with *Edwardsiella ictaluri*. *Molecular Immunology* **45**, 553–566.

Purcell, M.K., Nichols, K.M., Winton, J.R. *et al.* (2006) Comprehensive gene expression profiling following DNA vaccination of rainbow trout against infectious hematopoietic necrosis virus. *Molecular Immunology* **43**, 2089–2106.

Qin, Q.W., Shi, C., Gin, K.Y. and Lam, T.J. (2002) Antigenic characterization of a marine fish iridovirus from grouper, *Epinephelus* spp. *Journal of Virological Methods* **106**, 89–96.

Renn, S.C., Aubin-Horth, N. and Hofmann, H.A. (2004) Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics* **5**, 42.

Rise, M.L., Jones, S.R., Brown, G.D., von Schalburg, K.R., Davidson, W.S. and Koop, B.F. (2004a) Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiological Genomics* **20**, 21–35.

Rise, M.L., von Schalburg, K.R., Brown, G.D. *et al.* (2004b) Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Research* **14**, 478–490.

Roberge, C., Einum, S., Guderley, H. and Bernatchez, L. (2006) Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. *Molecular Ecology* **15**, 9–20.

Roberge, C., Guderley, H. and Bernatchez, L. (2007a) Genomewide identification of genes under directional selection: gene transcription Q(ST) scan in diverging Atlantic salmon subpopulations. *Genetics* **177**, 1011–1022.

Roberge, C., Paez, D.J., Rossignol, O., Guderley, H., Dodson, J. and Bernatchez, L. (2007b) Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Molecular Immunology* **44**, 1374–1383.

Roberge, C., Normandeau, E., Einum, S., Guderley, H. and Bernatchez, L. (2008) Genetic consequences of interbreeding between farmed and wild Atlantic salmon: insights from the transcriptome. *Molecular Ecology* **17**, 314–324.

Roling, J.A., Bain, L.J., Gardea-Torresdey, J., Bader, J. and Baldwin, W.S. (2006) Hexavalent chromium reduces larval growth and alters gene expression in mummichog (*Fundulus heteroclitus*). *Environmental Toxicology and Chemistry* **25**, 2725–2733.

Roling, J.A., Bain, L.J., Gardea-Torresdey, J., Key, P.B. and Baldwin, W.S. (2007) Using mummichog (*Fundulus heteroclitus*) arrays to monitor the effectiveness of remediation at a superfund site in Charleston, South Carolina, U.S.A. *Environmental Toxicology and Chemistry* **26**, 1205–1213.

Rome, L.C. (2006) Design and function of superfast muscles: new insights into the physiology of skeletal muscle. *Annual Reviews of Physiology* **68**, 193–221.

Salisbury, M.W. (2008) *Sequencers take on Microarrays*. Available at: http://WWW.genome-technology.com, February, GenomeWeb LLC, New York, pp. 38–42.

Salit, M. (2006) Standards in gene expression microarray experiments. *Methods in Enzymology* **411**, 63–78.

von Schalburg, K.R., Rise, M.L., Cooper, G.A. *et al.* (2005) Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics* **6**, 126.

Seo, J.Y., Kim, K.H., Kim, S.G. *et al.* (2006) Protection of flounder against hirame rhabdovirus (HIRRV) with a DNA vaccine containing the glycoprotein gene. *Vaccine* **24**, 1009–1015.

Shepherd, B.S., Drennon, K., Johnson, J. *et al.* (2005) Salinity acclimation affects the somatotropic axis in rainbow trout. *American Journal of Physiology – Regulatory Integrative and Comparative Physiology* **288**, R1385–R1395.

St-Cyr, J., Derome, N. and Bernatchez, L. (2008) The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Molecular Ecology* **17**, 1850–1870.

Tarca, A.L., Romero, R. and Draghici, S. (2006) Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology* **195**, 373–388.

Tsoi, S.C., Cale, J.M., Bird, I.M., Ewart, V., Brown, L.L. and Douglas, S. (2003) Use of human cDNA microarrays for identification of differentially expressed genes in Atlantic salmon liver during *Aeromonas salmonicida* infection. *Marine Biotechnology (New York)* **5**, 545–554.

Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science* **270**, 484–487.

Vuori, K.A., Koskinen, H., Krasnov, A. *et al.* (2006) Developmental disturbances in early life stage mortality (M74) of Baltic salmon fry as studied by changes in gene expression. *BMC Genomics* **7**, 56.

Williams, T.D., Gensberg, K., Minchin, S.D. and Chipman, J.K. (2003) A DNA expression array to detect toxic stress response in European flounder (*Platichthys flesus*). *Aquatic Toxicology* **65**, 141–157.

Williams, T.D., Diab, A.M., George, S.G. *et al.* (2006) Development of the GENIPOL European flounder (*Platichthys flesus*) microarray and determination of temporal transcriptional responses to cadmium at low dose. *Environmental Science and Technology* **40**, 6479–6488.

Wilson, A.J., Gislason, D., Skulason, S. *et al.* (2004) Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Molecular Ecology* **13**, 1129–1142.

Wong, M.K., Takei, Y. and Woo, N.Y. (2006) Differential status of the renin-angiotensin system of silver sea bream (*Sparus sarba*) in different salinities. *General and Comparative Endocrinology* **149**, 81–89.

Yasuike, M., Kondo, H., Hirono, I. and Aoki, T. (2007) Difference in Japanese flounder, *Paralichthys olivaceus* gene expression profile following hirame rhabdovirus (HIRRV) G and N protein DNA vaccination. *Fish and Shellfish Immunology* **23**, 531–541.

Yuan, J.S., Reed, A., Chen, F. and Stewart, C.N. Jr (2006) Statistical analysis of real-time PCR data. *BMC Bioinformatics* **7**, 85.

