

# Hepatic transcriptomic profiles of European flounder (*Platichthys flesus*) from field sites and computational approaches to predict site from stress gene responses following exposure to model toxicants

Falciani, Francesco; Diab, AM; Sabine, V; Williams, Timothy; Ortega, Fernando; George, SG; Chipman, James

DOI:  
[10.1016/j.aquatox.2008.07.020](https://doi.org/10.1016/j.aquatox.2008.07.020)

License:  
None: All rights reserved

*Document Version*  
Peer reviewed version

*Citation for published version (Harvard):*  
Falciani, F, Diab, AM, Sabine, V, Williams, T, Ortega, F, George, SG & Chipman, J 2008, 'Hepatic transcriptomic profiles of European flounder (*Platichthys flesus*) from field sites and computational approaches to predict site from stress gene responses following exposure to model toxicants', *Aquatic Toxicology*, vol. 90, no. 2, pp. 92-101. <https://doi.org/10.1016/j.aquatox.2008.07.020>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.



Contents lists available at ScienceDirect

## Aquatic Toxicology

journal homepage: [www.elsevier.com/locate/aquatox](http://www.elsevier.com/locate/aquatox)



# Hepatic transcriptomic profiles of European flounder (*Platichthys flesus*) from field sites and computational approaches to predict site from stress gene responses following exposure to model toxicants

F. Falciani<sup>a,\*</sup>, A.M. Diab<sup>b</sup>, V. Sabine<sup>b</sup>, T.D. Williams<sup>a</sup>, F. Ortega<sup>a</sup>, S.G. George<sup>b,1</sup>, J.K. Chipman<sup>a,1</sup>

<sup>a</sup> School of Biosciences, The University of Birmingham, Birmingham B15 2TT, UK

<sup>b</sup> Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland, UK

### ARTICLE INFO

#### Article history:

Received 3 June 2008

Received in revised form 24 July 2008

Accepted 31 July 2008

#### Keywords:

Ecotoxicogenomics

Microarray

Flounder

Environmental diagnostics

Pollution

GALGO

Modeling

### ABSTRACT

Genomic technologies offer opportunities to gain a more global assessment of the health status of an organism through an understanding of the functional pathways that are responding to pollutant exposure. We have developed a 13,000 clone cDNA toxicogenomics microarray for *Platichthys flesus*, the European flounder (EU-GENIPOL Project). We aimed to distinguish the origins of flounder taken from six sampling sites of different pollution status in Northern Europe according to their hepatic gene expression profile using bioinformatic approaches. To determine which gene expression differences may relate to pollutant impact, we have completed complementary laboratory exposures of flounder to selected toxicants and determined the associated gene expression profiles. Using multivariate variable selection coupled with a statistical modelling procedure (GALGO) we can predict geographical site but the accuracy is limited to specific sites. The search space for a combination of genes that effectively predicts class membership is very large, however, by combining the signatures derived from acute laboratory exposure to individual chemicals to limit the search space, a very accurate model for classification of all the different environmental sites was achieved. The final model utilised the expression profiles of 16 clones and validation with a qPCR array comprising these genes correctly assigned the site of origin for fish obtained from three of the sites in an independent sampling. These data would imply that the gene expression fingerprints obtained with these arrays are primarily attributable to variations in chemical pollutant responses at the different sites, indicating their potential utility in environmental impact assessment.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

River estuaries often contain complex mixtures of xenobiotics derived from industrial and domestic effluents in high concentrations. Thus, organisms living in them are exposed to these compounds which may be detrimental to their health. The monitoring of both chemical content and biological responses in sentinel species play important roles in assessing the impact of such anthropogenic inputs. To achieve this, biomarkers have been employed as indicative or predictive of exposure to, or damage by, toxicants.

Such biomarkers include molecular, biochemical, histological, and morphological parameters (van der Oost et al., 2003). The measurement of expression of single biomarker genes cannot provide an integrated assessment of the effects of different chemical classes, however, with the advent of the “omic” technologies (Schena et al., 1995), interest has turned to the possibility of characterizing in more depth the health status of individual animals through a more complete analysis. Toxicogenomic studies in genetic model organisms have already shown the potential of this approach (McMillian et al., 2005) and there is great potential to apply such methods to environmentally relevant species (ecotoxicogenomics) (Neumann and Galvez, 2002; Snell et al., 2003; Snape et al., 2004; Denslow et al., 2007). Through a detailed characterisation of gene expression it is now possible to establish perturbation of biological pathways in a tissue of an organism and then to relate such pathways to toxic, stress or adaptive responses. Thus, microarrays have been used to determine expression profiles that may be diagnostic of exposure to certain classes of chemicals (Bartosiewicz et al., 2001; Waring et al., 2002; Hamadeh et al., 2002) and can allow class-prediction, where samples are classified according to a pre-existing database.

**Abbreviations:** AWI, Alfred Wegener Institute; CEFAS, Centre for Environment, Fisheries and Aquaculture Science; GENIPOL, genomic tools for biomonitoring of pollutant coastal impact; I.C.E.S., International Council for Exploration of the Seas; MIAME, minimum information about a microarray experiment; OSPAR, convention for the protection of the marine environment of the North-East Atlantic; PPAR, peroxisome proliferator activated receptor; SAM, statistical analysis of microarray.

\* Corresponding author. Tel.: +44 121 4143037; fax: +44 121 4145925.

E-mail address: [Falciani@bham.ac.uk](mailto:Falciani@bham.ac.uk) (F. Falciani).

<sup>1</sup> Joint senior authors.

Current methodology and strategies in biomonitoring (e.g. use of histopathology) is likely to be efficient in the detection of a wide range of toxicities. However, the value of, and need for, a complementary suite of molecular biomarkers is in the provision of information on the mechanisms of toxicity and the additional detection of subtle changes that are adaptive or compensatory to exposure to the toxicant and may be predictive of potential toxicity. These markers, which may be obtained through the non-biased microarray approach, can thus act as sensitive, early warning sensors.

Our studies have focussed on the use of a flatfish, European flounder (*Platichthys flesus*) as a sentinel species. *P. flesus* is a teleost flatfish common in estuarine environments throughout north-west Europe, many of which are significantly impacted by toxicants including endocrine disrupters, heavy metals, polycyclic aromatic hydrocarbons (PAHs) and dioxins. These are usually present as complex mixtures closely associated with the sediment and flounder are exposed through feeding on associated invertebrates. It is a key sentinel species for biological effects monitoring in initiatives such as the OSPAR Joint Assessment and Monitoring Programme (OSPAR, 1997).

One of the complications of using microarrays for the simultaneous measurement of expression of thousands of genes in tissues from animals taken from the environment is that the complex pattern of gene expression profiles will be determined by many factors such as genetic make-up, age, diet, salinity, temperature and reproductive state as well as being influenced by pollutant inputs. Moreover, inter-individual variability in gene expression profiles within organisms in the same environment can be substantial (e.g. Pritchard et al., 2001; George et al., 2004). Gene expression profiles will also vary depending on the doses of toxicants, by the length of exposure and the interval between exposure and sampling, as will many other toxicological parameters. In order to be able to interpret the expression profiles of fish in the environment, we have used a strategy within the present study to compare the gene expression of fish from different environmental sites with respect to a number of conventional biomarkers and also in relation to the “stress gene responses” observed in flounder exposed acutely to a range of different toxicants under controlled laboratory conditions (Williams et al., 2008 co-submission). Using a supervised classification procedure in combination with a multivariate gene selection procedure we have assessed the ability of subsets of genes to allow the prediction of the site at which fish were sourced. This approach has helped to establish stress genes that contribute to the differences in hepatic gene expression observed in flounder obtained from different sites. Integration of laboratory and field studies in this way should help to discriminate between pollutant mediated gene expression changes and those associated with other factors.

## 2. Materials and methods

### 2.1. Environmental fish samples

Adult gonadally quiescent male *P. flesus* of length 13–22 cm were caught at six North Sea sampling sites, our reference site, the Alde estuary ( $n = 29$ ; latitude  $52^{\circ}95'N$ , longitude  $1^{\circ}33'E$  Suffolk, UK), two sites on the Tyne estuary (Tyne and Wear, UK), Team ( $n = 8$ ; latitude  $54^{\circ}57'N$ , longitude  $1^{\circ}27'W$ ) and Howdon ( $n = 10$ ; latitude  $54^{\circ}57'N$ , longitude  $1^{\circ}38'W$ ), two sites on the Elbe estuary, Cuxhaven ( $n = 4$ ; latitude  $53^{\circ}53'N$ , longitude  $08^{\circ}15'-19'E$ ) and Brunsbuttel ( $n = 21$ ; latitude  $53^{\circ}52'N$ , longitude  $09^{\circ}09'-10'E$ ) and one site close to the island of Helgoland ( $n = 5$ ; latitude  $54^{\circ}06'N$ , longitude  $07^{\circ}15'-08^{\circ}00'E$ ) (Schleswig-Holstein, Germany), during statutory monitoring programs carried out in October 2002 and 2004, and in April 2005 and 2007 by CEFAS, Burnham laboratory and AWI, Bremerhaven. Fish were caught using beam trawls and held in tanks of

flowing sea water onboard ship and were sacrificed either onboard ship or immediately on return to shore. Livers were immediately removed and samples flash frozen in liquid nitrogen for RNA extraction. Each individual sample of liver was processed and arrayed individually as described previously (Williams et al., 2006).

### 2.2. Flounder laboratory exposures to which gene expression profiles were compared

Juvenile immature *P. flesus* of 50–120 g body wt. were previously exposed separately to six chemical substances (Sigma–Aldrich, Gillingham, UK) a pro-oxidant heavy metal salt, cadmium chloride ( $50 \mu\text{g}/\text{kg}$  in saline), a polycyclic aromatic hydrocarbon, 3-methylcholanthrene ( $25 \text{mg}/\text{kg}$  in olive oil), a technical PCB mixture Aroclor 1254 ( $50 \text{mg}/\text{kg}$  in olive oil), an organic pro-oxidant, *tert*-butyl-hydroperoxide ( $5 \text{mg}/\text{kg}$  in saline), a chlorinated hydrocarbon pesticide, lindane ( $25 \text{mg}/\text{kg}$  in olive oil), a peroxisomal proliferative fatty acid analog, perfluoro-octanoic acid ( $100 \text{mg}/\text{kg}$  in olive oil), or their carriers, olive oil ( $1 \text{ml}/\text{kg}$ ) or saline (0.9%). The European flounder cDNA microarray was developed as part of the GENIPOL consortium and we have previously reported its construction and use in determining responses of flounder to disease, toxicant and hormone treatments (Williams et al., 2006, 2007; Diab et al., 2008). Details of the exposures and gene expression responses are given in detail in Williams et al. (2007 and co-submission) and Diab et al. (2008). Briefly, hepatic gene expression changes were determined 1, 2, 4, 8 and 16 days post-injection in comparison with time-matched carrier controls. Through gene ontology analyses, the key biological processes disturbed by the individual treatments were characterised.

### 2.3. cDNA microarray analysis

Hepatic gene expression profiles were determined in environmentally sampled fish using the GENIPOL cDNA microarray which was fully described by Diab et al. (2008) following established protocols described earlier (loc. cit). Briefly, flounder hepatic total RNA was treated to remove contaminating genomic DNA with DNA-free (Ambion, Huntingdon, UK), reverse transcribed to cDNA with SuperscriptII (Invitrogen) using oligo-dT primers (Alta Biosciences, Birmingham, UK) then labelled with either Cy5-dCTP or Cy3-dCTP (Amersham, Amersham, UK) using Klenow polymerase (Invitrogen, Paisley, UK) before purification (Qiagen, Crawley, UK). All arrays utilised a synthetic reference DNA, labelled by a similar procedure. Each array consisted of labelled cDNA from one individual fish hybridized versus the differentially labelled synthetic reference. Hybridisations were carried out for 18 h, before stringent washing and scanning (Axon 4000B) (Molecular Devices, Wokingham, UK). Data were captured using Genepix software (Molecular Devices), and each slide was checked in detail, with spots showing poor morphology or arrays showing gross experimental artefacts being discarded. The data used in analyses consisted of local background-subtracted median intensities. MIAME-compliant data was input to Array-Express using maxdLoad2 (Hancock et al., 2005), accession number E-MAXD-40.

### 2.4. Microarray data analysis

#### 2.4.1. Data normalisation

Each labelled cDNA sample was hybridized to one microarray slide (Williams et al., 2006; Diab et al., 2008). Quantile normalisation was applied to the raw intensities ( $\log_2$  values) across the microarrays using the affy package (<http://www.bioconductor.org>) under the programming environment R (<http://cran.t-project.org>). The mean of the log intensities was then calculated for each gene

on the array. The expression of each gene was determined from average Cy5 intensities of at least two spots; where coefficients of variation were greater than 30% data were removed.

2.4.2. Gene expression and gene ontology

Individual samples from each site were considered as biological replicates. Differentially expressed genes were selected for each site by comparison with the Alde samples using a *t*-test with *P* < 0.05 and Benjamini and Hochberg multiple testing correction (Benjamini and Hochberg, 1995), resulting in a false discovery rate (FDR) of 0.05. These statistically significantly changing genes were filtered to remove those altered in expression in comparison with Alde samples by less than two-fold. Gene ontology (GO) analysis was carried out using Blast2GO (Conesa et al., 2005). The lists of differentially expressed genes were compared with all reliably detected genes using the Fisher’s Exact test within the Gossip package, which output the GO terms statistically significantly (FDR < 0.1) over-represented or under-represented in the test list.

2.4.3. Cluster analysis

Hierarchical clustering analysis (HCAs) was performed in R (<http://cran.t-project.org>). The degree of similarity of each pair of patterns was computed using Euclidean distance on standardized gene expression profiles.

2.4.4. Classification methods

To identify molecular signatures predictive of the geographical sampling site of the fish we have used two complementary approaches.

The first uses a univariate variable selection strategy where each gene is tested individually for its ability to discriminate between the sample groups. In this procedure genes were selected using an *F*-test. A predictive model was built using the smallest subset of most differentially expressed genes which gave the highest classification accuracy with a *K*-nearest-neighbor (KNN) classifier. In using this procedure we used the implementation available in the software application “Prophet” (Herrero et al., 2003).

A multivariate variable selection method was also used in combination with a KNN classifier for prediction of the sampling site based on the selection of gene sets. In this procedure combination of

genes are tested for their ability to discriminate between different sample groups. The GALGO package (Trevino and Falciani, 2006), running under the R language was employed. In short, GALGO incorporates a Genetic Algorithm procedure for selecting models with a fitness value above a given threshold. The supervised classification method was *k*-nearest neighbours algorithm (KNN) with *k* = 3. The classification efficiency (error estimation) was evaluated using a leave-one-out cross validation procedure. The models developed by GALGO were chosen to have a size of five “genes” (Trevino and Falciani, 2006). One thousand independent searches were performed and the best “chromosomes” of each run were selected. The genes represented in the model population were sorted accordingly to their frequency. A single representative model was built by applying a forward selection procedure to the sorted list using the criteria that maximized the average fitness.

2.5. Validation by quantitative PCR (qPCR) array

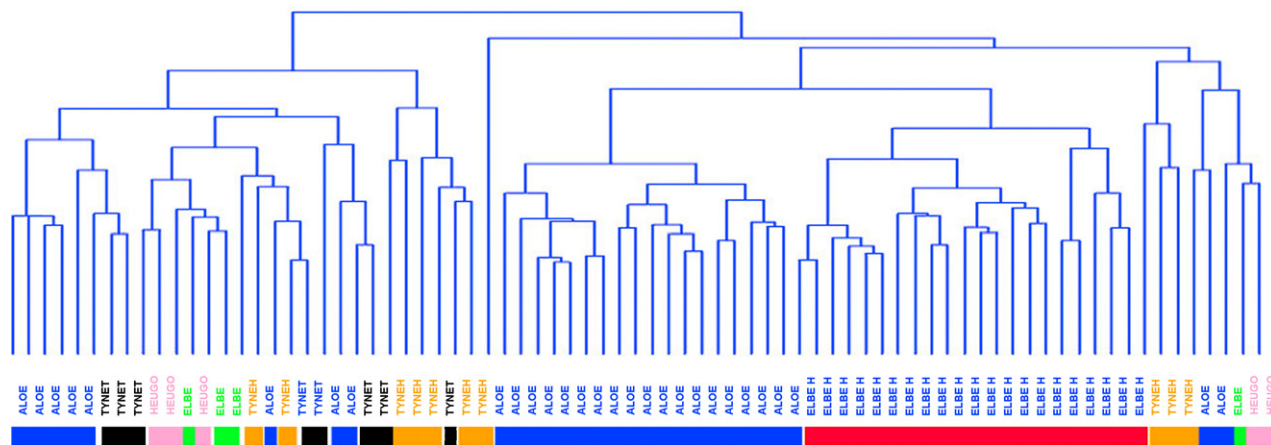
The validation of this microarray by real-time quantitative PCR was previously shown in experiments carried out with Cd and estradiol (Williams et al., 2006, 2007). Microarray and qPCR data for vitellogenin B, glutathione reductase, Cu/Zn superoxide dismutase and metallothionein correlated at *r*<sup>2</sup> = 0.95, 0.98, 0.79 and 0.89, respectively. In order to validate the results of the present experiment further qPCR was carried out using additional independently sampled and prepared flounder RNA.

Total RNA was extracted from liver samples obtained from fish collected from each of the Alde, Brunsbuttel and Tyne (Howdon) sites (*n* = 5 per site) obtained at a different sampling, using Tri reagent (Sigma) as above. Samples of 1 µg total RNA were reverse transcribed for 60 min at 42 °C using a *Reverse-IT* MAX™ first strand synthesis kit (Abgene, Epsom, UK) with 0.5 µM of a 3:1 mixture of random hexamers and anchored oligo-dT in a total volume of 20 µl. The reaction was terminated by heating to 75 °C for 10 min, diluted 20-fold with distilled water and then stored at –20 °C. PCR primers for the 15 “site diagnostic” and 3 “normalisation” genes (Table 1) were synthesized by MWG Biotech (London, UK), adjusted to 100 pmoles/µl with water and stocks stored at –20 °C. Quality control was assessed by product analysis of PCR reactions using agarose gel electrophoresis to verify formation of single products of

**Table 1**  
Genes and their primers used for qPCR array analysis

Putative identity	Accession no.	Forward primer	Reverse Primer
<b>Diagnostic clones</b>			
40S ribosomal protein S14	DV566842	ttttggcgtctgtcacatc	gacctggagtcttggctcgtg
40S ribosomal protein S16	DV566540	tcagctgagacgaaagatgc	atgtccaccaccttcaactc
Aldehyde dehydrogenase 9A1	DV568325	gggagaagattgcaaaagctg	gagcaggagcagacttccac
cDNA clone OLf01.07c	DV565610	caaaactgcagctgctcatc	ccatgagcgtgatcataaa
Choriogenin	DV570128	ctcccagaagtccagtgaa	gtggcagggcattgagttac
Copper/zinc superoxide dismutase	DV566819	tactgaaaggagccggagaa	ctgctccagcagctcacattc
Cytochrome P450 1A	DV569379	gaacaaatatcacttatgacacca	catccctggaactcttttgg
EST PFIL007F01	EC377593	tgagtttttgtatggaagaaaga	tgccatggttaacagccatta
EST PFSHT4C7	AJ543380	tcaccataacaggaagatgtg	agccaactagagccaccaga
EST22	DV565543	ttgtctggattcgaattgga	tactctgtcggaggcagat
F-actin capping protein β subunit	DV567493	gaggtgcaggagaagtccag	tgtccttggcttcccaaag
Glutamate carboxypeptidase	DV565571	cgacatggatctgatctga	ggagctcgcttaacaatgct
EST Ldt-0086	DV568781	agggcaaatggcctttagtt	acaggcgtgtggagaaag
LEAP-2	DV566771	ctccgaccacagctcagact	ccaggtttggatcacgctc
Metallothionein	DV566510	cctggagaagaaccagacca	tacagtgcatcaggctctctc
up059 transmembrane protein	DV566792	tgctcctcagctctctgtctg	tcggtcaatgggtgattgaa
<b>“Normalisation” genes</b>			
Beta actin	AF135499	gaccaactgggatgacatgg	gcgtacagggacagcacagc
Elongation factor alpha	EC378399	tgtccactctgctaaggctg	cttgaggcgttctgtctcct
Alpha tubulin'	AJ291985	cacagcctcactctgcttttg	agatgacaggggcataggtg

Genes identified as “diagnostic” by statistical modelling are shown with addition of those genes used for “normalisation” of data. GenBank accession numbers and primer sequences are shown.



**Fig. 1.** Unsupervised cluster analysis of transcriptomic responses in male flounder from different environments. Hierarchical dendrogram using Euclidean distances. Sampling locations are shown by coloured blocks: Alde – blue; Tyne, Team – black; Tyne Howden – orange; Helgoland – pink; Cuxhaven – green; Elbe, Brunsbützel – red.

expected molecular size and examination of melting point analysis data during real time PCR to verify single peak dissociation profiles.

For qPCR array each forward and reverse primer pair were combined and diluted to yield enough mix for 100 reactions at a final primer concentration of 267 nM. 96-well plates were set up with three sub arrays containing 2 µl of each primer mix per well (15 wells with “diagnostic” genes, 4 wells with “normalisation” genes and 4 additional wells containing actin primers as “linearity” estimates), whilst 2 µl of water was added to 2 additional wells for no template and no transcriptase controls. For each subarray, 5 µl of the diluted cDNA reaction mixture was pipetted into the “diagnostic” and “normalization” wells, 5 µl of four decadally diluted cDNA samples were pipetted into the “actin linearity” wells and 5 µl water into the control wells. 10 µl SYBR Green PCR mix (ABgene Inc.) and 3 µl water were then added to each well and the plate subjected to real-time qPCR on a Quantica cyler (Techne Instruments, Stone, UK) using universal cycling conditions, activation at 95 °C for 15 min, cycling at 95 °C 15 s, 58 °C for 15 s and 72 °C for 30 s for 45 cycles and then melting at 75–95 °C. The cycle number ( $C_t$ ) signals for each well were recorded and used for data analysis.  $C_t$  values for all samples across multiple plates were normalized to beta-actin, centered by median of gene, normalized to ±1, and ordered by hierarchical clustering (Unweighted pair-group average, 1 – Pearson  $r$ ) using the analysis package Statistica (StatSoft, Bedford, UK).

### 3. Results

#### 3.1. Gene expression changes

Comparison of gene expression with Alde samples resulted in 481 genes identifiable with known proteins that were statistically significantly differentially expressed ( $P < 0.05$   $t$ -test) and

with a greater than two-fold change in expression from Alde at one or more sites (Supplementary Table 1). Gene ontology analyses showed 167 statistically significantly altered GO terms (Supplementary Table 2).

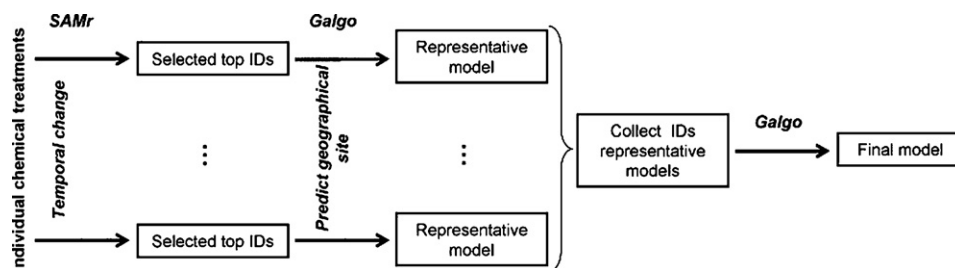
#### 3.2. Clustering

Unsupervised hierarchical clustering analysis resulted in imperfect aggregation based on the sites from which the samples of fish were collected. The hierarchical dendrogram revealed that no sample set perfectly segregates accordingly to its site apart from the highly impacted Brunsbützel samples (Fig. 1).

#### 3.3. Class prediction

Initially, a univariate variable selection methodology to develop biomarkers predictive of the geographical site was used. This methodology ranks each individual clone based on its ability to discriminate among the geographical sites and then selects a final model that uses the smallest possible subset of clones that is most predictive. Initial analysis revealed that the classification accuracy was relatively high. However, this methodology does not take into account the fact that genes work in the context of integrated networks and pathways. To address this issue and identify genes that, in combination, are predictive of the geographical site, we used the statistical modelling (GALGO) system that implements a genetic algorithm as a multivariate variable selection method (Trevino and Falciani, 2006). The overall strategy is shown in Fig. 2.

First, the full microarray dataset was used to identify a subset of genes predictive of geographical site. In Table 2, the results of the multivariate analysis of the transcription dataset using GALGO are



**Fig. 2.** Flow diagram illustrating the procedure implemented to select the clones that were used to build the final statistical representative model based on the different chemical treatments.

**Table 2**  
(A) Classification accuracy and (B) the list of clones that constitutes the statistical model, with putative identities

Samples	Classification sensitivity						(A)
	Alde	Cuxhaven	Brunsbüttel	Helgoland	Tyne H.	Tyne T.	
Alde	<b>0.979</b>	0	0	0	0	0	0
Cuxhaven	0.020	<b>0.809</b>	0	0.100	0	0	0
Brunsbüttel	0	0	<b>1.000</b>	0	0	0	0
Helgoland	0	0.191	0	<b>0.900</b>	0	0	0
Tyne H.	0	0	0	0	<b>0.700</b>	0.133	0
Tyne T.	0.001	0	0	0	0.300	<b>0.867</b>	0

Clone	Accession	Putative identity	Clone	Accession	Putative identity
(B)					
PfCR1B3	AM041917	40S Ribosomal S7	PfIL226G11	DV566685	Tryptophan 2,3-dioxygenase
PfIL003A07	DV565345	Hemoglobin beta A-chain	PfIL227G09	EC378161	No similarity
PfIL003A08	DV565346	Glutamate carboxypeptidase	PfIL229G01	DV566792	EST up059
PfIL006B05	EC377575	Similar to Heat shock cognate 71kDa protein	PfIL237G09	DV567131	Alpha-2 microglobulin
PfIL006F05	DV565411	40S Ribosomal S11	PfIL241D03	DV567285	EST 22
PfIL006G01	DV565416	Cytochrome P450 1A (CYP1A)	PfIL249C07	DV567616	EST no168c05
PfIL006G04	DV565417	Small inducible cytokine	PfIL251A07	DV567701	Alpha-tubulin (6)
PfIL006H12	DV565427	60S Ribosomal L37a	PfIL253D06	DV567816	No similarity
PfIL009G06	DV565552	Hemoglobin alpha chain	PfIL254E09	DV567868	S100 Calcium-binding protein A14
PfIL009G07	DV565553	Unnamed protein product CAG12952	PfIL258H07	DV568070	No similarity
PfIL011B03	DV565568	Nascent polypeptide-associated complex alpha	PfIL260G07	DV568170	Stomatin-like 2
PfIL011H12	DV565599	Complement factor B/C2-B	PfIL260G08	DV568171	Hemoglobin alpha A chain
PfIL050F03	DV565721	EST irn13ks	PfIL265H03	DV568392	40S Ribosomal S15a
PfIL201B01	DV565784	Retinol-binding protein II, cellular	PfIL267B03	DV568444	60S Ribosomal L23
PfIL202C05	DV565828	High choriolytic enzyme 1	PfIL267F02	DV568466	60S Ribosomal L36
PfIL203F07	DV565919	Beta-2 microglobulin	PfIL276C02	EC378905	Chemotaxin
PfIL207C08	EC377807	MHC Class IIb antigen	PfIL282H11	na	EST Sp13f
PfIL213D01	DV566259	40S Ribosomal S29	PfIL289D06	EC379111	EST liver-5-G06_57_14
PfIL216G02	DV566352	40S Ribosomal S28	PfIL300G04	DV569695	Aldehyde reductase-like 6
PfIL216G09	DV566356	Unnamed protein product CAG10196	PfIL311F08	DV570202	Alkylated DNA repair protein AlkB
PfIL221A02	EC378046	Toxin-1	PfIL314F08	DV570229	Aprataxin homolog
PfIL226B02	EC378132	Elastase I precursor	PfIL320F08	EC379649	Unnamed protein product CAG12708

The horizontal axis represents the individual samples grouped according to the geographical site classes whereas the vertical axis represents the predicted sites. The model was built using all microarray data. Classification accuracy of for each site prediction are marked in bold along the diagonal of the table.

presented. The prediction accuracy (Table 2A) is high (larger than 80%) for all sites except Tyne Howdon for which it is 70% as a result of the uncertainty in discriminating the geographically close Tyne Team site.

These results demonstrated that it was possible to identify gene signatures that are highly predictive of geographical site. We postulated that these complex signatures may be, in part, the result of a combined cellular response to the individual chemicals that are present in the environment; however, this model does not include

many of the known classical pollutant responsive biomarker genes. This may be artefactual or a reflection of the fact that our procedure is somehow identifying a set of alternative, and perhaps better, predictors. In order to test this hypothesis another model was developed which was specifically designed to represent the response of a set of chosen “classical” biomarker genes (listed in Table 3B). The classification accuracy of the representative model obtained using GALGO is shown in Table 3A. Whilst the model is accurate in predicting some of the sites such as Alde, Elbe Brunsbüttel and to some

**Table 3**  
(A) Classification accuracy using only the clones that represent biomarker genes and (B) the list of clones that constitutes the statistical model, with putative identities

Samples	Classification sensitivity						
	Alde	Cuxhaven	Brunsbüttel	Helgoland	Tyne H.	Tyne T.	Unclassified
(A)							
Alde	<b>0.850</b>	0.003	0.005	0.004	0.031	0.024	0.084
Cuxhaven	0.200	<b>0.038</b>	0.106	0.067	0.053	0.041	0.495
Brunsbüttel	0.001	0.002	<b>0.936</b>	0.004	0.001	0.006	0.050
Helgoland	0.169	0.064	0.338	<b>0.022</b>	0.035	0.044	0.327
Tyne H.	0.138	0.016	0.046	0.011	<b>0.369</b>	0.174	0.246
Tyne T.	0.035	0.004	0.035	0.009	0.103	<b>0.539</b>	0.275

Clone	Accession	Putative identity	Clone	Accession	Putative identity
(B)					
PfIL258E10	DV568057	CYP1A Cytochrome P450 1A	PfIL264C12	DV568325	Aldehyde dehydrogenase 9A1
PfGST-o1	AJ310428	Glutathione S-transferase GSTA1	PfIL222C07	DV566510	Metallothionein
PfIL263F01	DV568296	Copper/zinc superoxide dismutase	PfIL289A06	DV569240	Manganese superoxide dismutase
PfIL242D06	DV567331	Choriogenin L	PfIL257D11	DV567998	Vitellogenin
PfIL262F06	DV568249	UDP-glucuronosyltransferase 3	PfIL265A03	DV568358	Catalase

The horizontal axis represents the individual samples grouped according to the geographical site classes whereas the vertical axis represents the predicted sites. Classification accuracy of for each site prediction are marked in bold along the diagonal of the table.

**Table 4**  
Measures of the overall prediction per geographical site

Site-treatment	Alde	Cuxhaven	Brunsbüttel	Helgoland	Tyne Howdon	Tyne Team
Cadmium	0.848	0.75	1.0	0.6	0.671	0.441
tBHP	0.903	0.5	1.0	0.493	0.378	0.828
3-MC	0.96	0.75	1.0	1.0	0.471	0.527
PFOA	0.905	0.75	0.992	0.855	0.711	0.851
Lindane	0.848	0.558	0.99	0.684	0.72	0.31
Aroclor	0.805	0.0	0.935	0.0	0.346	0.616

Each row shows the sensitivity values of the representative models that were built using only the most statistically significantly changing clones after the administration of the six specific chemical treatments.

extent the Tyne Team site, it shows a rather poor accuracy for all the other sites.

We then hypothesized that a wider selection of stress-responsive genes should improve predictivity. In order to test this hypothesis, the predictive power of hepatic gene signatures identified from laboratory exposure of flounder to the model xenobiotics, cadmium, *tert*-butyl hydroperoxide, perfluorooctanoic acid, 3-methylcholanthrene, Aroclor and lindane (Williams et al., 2008 co-submission) was assessed. Models were developed from the 200 most statistically significantly differentially expressed clones for each chemical treatment (Supplementary Table 3). Table 4 shows the classification accuracy of the models developed using these gene lists. It can be seen that some gene signatures such as those found following cadmium treatment, were predictive of some geographical locations with an average weighted sensitivity of the model of 84%, but its overall accuracy compared to the use of the entire gene pool analyzed above, was decreased. Similar findings were obtained with the other gene signatures and the average weighted sensitivity of the representative models were in the range from 49% to 86%.

**Table 5**  
(A) Overall classification accuracy using the combined list of clones that were selected from statistical modeling of the independent chemical experiments and (B) the list of clones that constitutes the statistical model, with putative identities

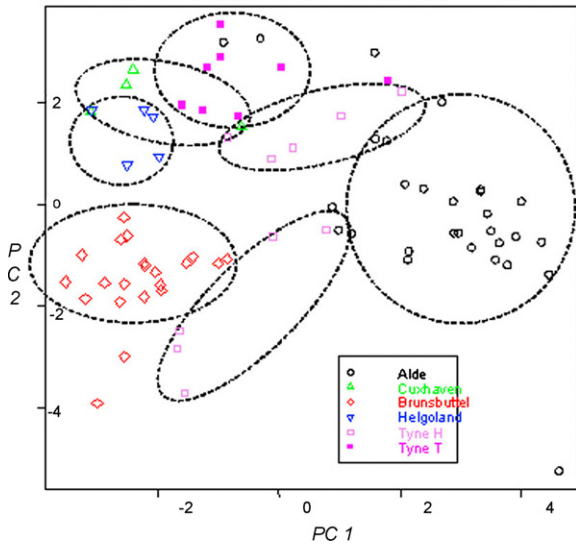
Samples	Classification sensitivity						Unclassified
	Alde	Cuxhaven	Brunsbüttel	Helgoland	Tyne Howden.	Tyne Team.	
(A)							
Alde	<b>0.934</b>	0.018	0	0.001	0	0.029	0.022
Cuxhaven	0.068	<b>0.821</b>	0	0.182	0	0	0
Brunsbüttel	0	0	<b>1.000</b>	0	0	0	0
Helgoland	0	0.082	0	<b>1.000</b>	0	0	0
Tyne H.	0	0	0.004	0	<b>0.788</b>	0.113	0.001
Tyne T.	0	0	0	0	0.159	<b>0.905</b>	0
Clone	Accession		Putative identity				
(B)							
PFIL231A08	DV566842		40S ribosomal protein S14				
PFIL223B02	DV566540		40S ribosomal protein S16				
PFIL264C12	DV568325		Aldehyde dehydrogenase 9 A1				
PFIL315D10	DV570246		Aldehyde reductase-like				
PFIL012C05	DV565610		EST OLf01.07c				
PFIL312G04	DV570128		Choriogenin				
PFIL230D11	DV566819		Copper/zinc superoxide dismutase				
PFIL292A02	DV569379		CYP1A Cytochrome P450 1A				
PFIL007F01	EC377593		EST PFIL007F01				
PFSHT4C7	AJ543380		EST PFSHT4C7				
PFIL009F01	DV565543		EST22				
PFIL246D07	DV567493		F-actin capping protein beta subunit				
PFIL011B11	DV565571		Glutamate carboxypeptidase (Darmin)				
PFIL276B07	DV568781		EST Ldt-0086				
PFIL229B08	DV566771		Liver-expressed antimicrobial peptide 2 precursor (LEAP-2)				
PFIL222C07	DV566510		Metallothionein				
PFIL229G01	DV566792		up059 similar to transmembrane protein				

Each column corresponds to one of the six different environmental conditions. Classification accuracy of for each site prediction are marked in bold along the diagonal of the table.

On the basis that a larger pool of “stress-related” genes are more likely to be able to have a predictive capacity, all the clones that were selected in the representative models developed from the gene signatures of each chemical treatment (Table 4) and the set of selected ‘biomarker’ genes (Table 3) were merged to develop a representative model based on this list. We found that this statistical model greatly improved our prediction of the geographical sites compared to the use of individual stressors (Table 5A) and that the average weighted sensitivity of the model improved to 92.5%. To ensure that this solution was not discovered by random chance we tested the probability of finding a model with given accuracy based on selecting a random list of 100 clones from the full environmental sites dataset. The distribution of average weighted sensitivity of these models is in the range 68–78%. Principal components analysis was then used to illustrate the overall differences in response between sampling sites, based on the expression of genes incorporated into the final representative model (Fig. 3). The test-site fish are clearly separated from those sampled at our reference site, the Alde. The fish from Helgoland and Cuxhaven are well separated from other samples but show some overlap with each other. There is an overlap between some individuals from Tyne Howdon, and Brunsbüttel with Tyne Team fish, whilst the remainder of Tyne Howdon and Brunsbüttel fish fall into well defined areas.

#### 3.4. Verification of predictive gene set by qPCR array

PCR primers were designed to each of the 17 diagnostic clones used in the final genetic model, serial dilutions of input cDNA confirmed near 100% yield for all primer sets with one exception where more than one product was obtained and this was omitted from the final qPCR array assembly. No-template and no-transcriptase controls were used as contamination checks. Serial dilutions of



**Fig. 3.** Principal components analysis of environmentally sampled flounder using the expression of clones shown in Table 5.

input cDNA were utilised with actin primers to confirm linearity of response under the conditions used and the amount of input cDNA to use was optimised by using the beta-actin gene primers so that a  $C_t$  value of around 20 was obtained. RNA extracted from five fish from the reference Alde and two polluted sites, Tyne (Howdon) and Elbe (Brunsbuttel) were obtained from an independent sampling and analyzed using the PCR array.  $C_t$  values for actin differed by less than  $\pm 1$ , therefore in this study only beta-actin expression was used for normalisation. Cluster analysis (Fig. 4) showed that all samples were segregated by site of origin.

**4. Discussion**

In the present study, we have investigated by cDNA microarray the gene expression profiles of 77 hepatic samples of *P. flesus* collected from six different environmental sites around north-west European estuaries. Heavy metal, petroleum hydrocarbon, carcinogenic polyaromatic hydrocarbon and polychlorinated biphenyl concentrations in the sediments of the highly industrialized sites

**Table 6**

Concentrations of major pollutant classes in sediments from UK and German estuaries where flounder were sampled

Analyte	Alde	Tyne (Howdon)	Elbe (Brunsbuttel)
Cd (mg/kg dry wt)	<0.02	1.7	0.8
Cu (mg/kg dry wt)	10	48.8	27.8
Hg (mg/kg dry wt)	0.01	0.24	nd
Pb (mg/kg dry wt)	25	248	80
Zn (mg/kg dry wt)	63	536	241
PAH's <sup>a</sup> ( $\mu$ g/kg dry wt)	23	3830	2720
PCB's <sup>a,b</sup> ( $\mu$ g/kg dry wt)	0.06	4.48	4.85

Chemistry data courtesy of CEFAS, UK; Deutsches Ozeanographisches Datenzentrum, Germany; ICES, Denmark.

<sup>a</sup> PAH's sum of carcinogenic PAHs: benzo[a]anthracene, benzo(a)pyrene, benzo[g,h,i]perylene, chrysene, indeno[1,2,3-cd] pyrene.

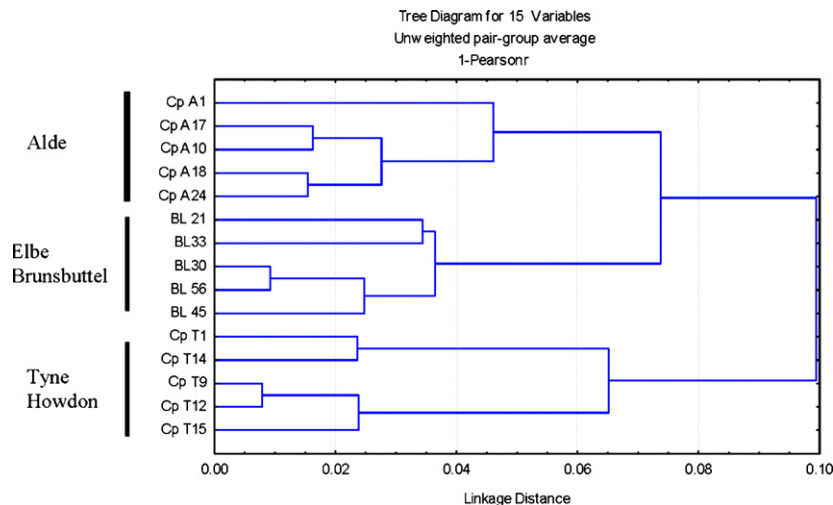
<sup>b</sup> PCB's sum of CB28, CB52, CB101, CB118, CB138, CB153, CB180.

in the Tyne and Elbe estuaries are generally one to two orders of magnitude greater than in the Alde estuary (Table 6) and residue analysis has shown detectable levels of these pollutants in tissues of fish from these sites (ICES data, not shown), thus it is to be expected that fish from these sites will display transcriptomic responses to these pollutant classes. The sites at Cuxhaven and Helgoland were of intermediate contamination (data not shown).

**4.1. Gene expression and gene ontology**

The Alde site has low anthropogenic impact (Lyons et al., 2004) and is regarded as a non-polluted reference site due to consistently low plasma vitellogenin and hepatic ethoxyresorufin-*o*-deethylase (EROD) activity in flounder from this site (George et al., 2004; Kirby et al., 2004a,b). In comparison with the Alde fish, the expression of many genes was altered in livers of fish from one or more of the other sampling sites. Of these changes a majority showed repression in comparison with the Alde, however, there was an upregulation of many genes coding for xenobiotic metabolism and detoxification proteins.

The functionally related groups of genes showing clearest overall induction in the highly impacted Brunsbuttel and Tyne Howdon sites were those encoding the phases I and II metabolic enzymes. Of these the most prominent was the well-known biomarker cytochrome P450 1A (CYP1A), which was induced at all sites in comparison with the Alde and induced most highly (greater than



**Fig. 4.** Hierarchical dendrogram of cluster analysis of gene expression of clones shown in Table 5 determined by qPCR array with an independent sample set to Figure 3. Sampling locations are A, Alde; B, Brunsbuttel and T, Tyne Howdon.



seven-fold) at the Brunsbittel site. CYP1A is inducible by aromatic hydrocarbon receptor (AhR) ligands such as polycyclic aromatic hydrocarbons (PAH) and halogenated aromatic compounds such as PCB's and its mRNA and enzyme activity (EROD) have previously been shown to be induced at pollutant-impacted sites including the Tyne (George et al., 2004; Kirby et al., 2004b; Williams et al., 2003). Hepatic neoplasia have been found to be more frequent in flounder sampled from the Tyne in comparison with the Alde (Stentiford et al., 2003) and in the Elbe estuary in comparison with Helgoland (Koehler, 2004) and this is believed to be due to PAH impact. Aldehyde dehydrogenase (ADH) activity has been reported to be induced in flounder treated with polycyclic aromatic hydrocarbons (Koehler et al., 1998) and ADH mRNA was found to be induced in Tyne fish (Williams et al., 2003). In the present study, a transcript with homology to aldehyde dehydrogenase 9A1 (originally associated with gamma-amino butyraldehyde dehydrogenase activity, Kikonyogo and Pietruszko, 1996) was induced in fish from all sites compared with the Alde. Cytochrome P450 8B1, a bile acid synthetic enzyme (Eloranta and Kullak-Ublick, 2005), is upregulated at all Elbe sites but not in Tyne samples. The induction of such genes at the Helgoland site, which was previously relatively uncontaminated, may in part be due to deposition of contaminants by the severe floods of 2002 which carried much estuarine sediment from the Elbe into the German Bight (Einsporn et al., 2005). At the Tyne Team site there is induction of glutathione-S-transferases which detoxify lipid peroxides produced by oxidative stress, and this response is further supported by the induction of other oxidative stress-responsive genes, catalase and glutathione peroxidase 1. Only one oxidative stress responsive transcript was induced at all sites, peroxiredoxin 6, with selenium-binding protein 1 displaying induction at the Cuxhaven and the Tyne sites. Additionally, cytochrome P450 2 and 3 family transcripts were induced at the Tyne Team site, with the GO term mono-oxygenase activity being over-represented.

Surprisingly, a number of chaperones and heat-shock proteins, which we previously found to be very highly induced by acute heavy metal and organic xenobiotic exposures in the laboratory (Williams et al. (2008) co-submission), were broadly repressed in fish from the polluted sites. Such a lack of induction may reflect the differences between acute laboratory exposures and long-term environmental exposure where a degree of adaptation has occurred.

There was little evidence of a co-ordinated induction of protein synthetic genes such as those encoding ribosomal proteins, again in contrast to laboratory exposures. Indeed certain translation initiation factors were repressed at our test sites, as previously found in Tyne fish (Williams et al., 2003). There were, however, a number of inductions of amino-acid metabolic genes, with additional support from the over-representation of related GO terms in Tyne Howdon and Brunsbittel samples. The lipid transporters cholesteryl ester transfer protein and fatty acid-binding protein were induced at all Elbe sites, whereas other lipid transporters such as apolipoprotein E and vigilin were repressed. This disruption of lipid transport may reflect that flounder from this area have shown hepatic steatosis as reported in a histopathology study by Koehler (2004). The alteration of expression in these gene classes and those of carbohydrate and lipid metabolism, as well as those of energy generation pathways, may be due to pollution effects or to differences in nutritional status between the groups of fish. Vigilin has been shown to stabilise vitellogenin mRNA in *Xenopus* (Cunningham et al., 2000). The suppression of vigilin gene expression noted at Brunsbittel, Cuxhaven and Helgoland may therefore have reduced any apparent induction of vitellogenin transcript, however, it is not currently known if this mechanism is conserved in flounder.

No clear conclusions could be made as to whether gene expression responses favoured apoptosis or cell survival since

mixed responses were observed. Expression of cytochrome C, programmed cell death 2 and 6 were repressed at most sites compared with the Alde. At the Brunsbittel site there was induction of the diablo gene (Verhagen and Vaux, 2002) and associated transcripts, at the Tyne sites the expression of survival motor neuron domain containing 1 (Talbot et al., 1998) and programmed cell death 4 were induced, whilst that of survivin, the anti-apoptotic gene was suppressed. These responses would indicate that the fish have been affected by a mixture of pro- and anti-apoptotic stimuli, including toxicants.

Fish from all sites showed an unexpected repression of transcription of DNA repair enzymes such as alkylated DNA repair protein AlkB and aprataxin, a repair enzyme of single strand breaks (Takahashi et al., 2007) when compared with the Alde fish. It may be possible that this repression contributes to the increased prevalence of DNA adducts reported in flounder from the Tyne (Lyons et al., 1999). Cell cycle genes apart from cyclin G and geminin, a DNA replication inhibitor, are generally repressed at all sites. Such cell cycle transcriptional disruption may also contribute to the pro- and anti-apoptotic gene expression responses observed.

Haemoglobin transcripts were induced and associated GO terms were over-represented in Tyne Howdon and Brunsbittel fish which may be consistent with tissue damage allowing erythrocyte infiltration into the liver at these two sites, a condition confirmed by histopathology analysis (results not shown). Significant induction of hepatic metallothionein gene expression was only observed in the Cuxhaven fish, which is consistent with the elevated protein levels found in a caging study with cod (*Gadus morhua*) in this area (Chesman et al., 2007). qPCR analysis of metallothionein mRNA and analysis of Cd/Zn metallothionein protein of these and other flounder from the Alde and Tyne did not show significantly different levels between sites (George et al., 2004).

Inflammatory and immune system genes displayed mixed responses. Whilst expression of the antimicrobial proteins hepcidin, lysozyme and pleurocidin were elevated at most sites, a variety of other transcripts were repressed in the Helgoland and Cuxhaven samples compared with the Tyne and Brunsbittel samples. Disruption of immune function has previously been shown in flounder from the Elbe estuary (Broeg et al., 2005). Tyne Team and Howdon samples showed induction of complement component C3 and Tyne Howdon induction of nuclear factor kappa B, with over-representation of the GO term inflammatory response.

Although, historically, high plasma vitellogenin levels have been reported in flounder from the Tyne (Allen et al., 1999), a more recent study reported a decline towards levels found at the Alde which was attributed to the installation of secondary treatment facilities in 2000 (Kirby et al., 2004a). In the present study where sampling was carried out in 2002, significant induction of vitellogenin mRNA in comparison with Alde fish was apparent only at the Tyne Howdon site, and physiological anchoring by measurement of plasma vitellogenin levels confirmed these findings (George et al., 2004), however, in common with results obtained in fish from other polluted sites there was a high inter-individual variability (George et al., 2004 and unpublished). Although there is previous evidence of induced vitellogenesis in bream (*Abramis brama*) from the Elbe, this was relatively low in flounder from the Elbe estuary (Hecker et al., 2002) and plasma vitellogenin analysis of three of the fish subjected to microarray analysis in the present study also showed no induction (unpublished results).

The numbers of gene ontology terms whose representation was statistically significantly changed in comparison with the Alde reference site may be useful as a rough indicator of impact. These were, Tyne Howdon 106, Brunsbittel 91, Tyne Team 73, Cuxhaven 46 and Helgoland 35. Comparing Brunsbittel and Tyne Howdon

fish, from the GO classes affected it appeared that the additional changes at the Tyne Howdon site were related to inflammation and haem proteins, whilst both sites showed alterations in classes indicative of contamination with organic compounds such as PAHs and PCBs compared with the other sites.

#### 4.2. Class prediction

Using a hierarchical clustering method, we inferred that samples of each geographical site do not display a unique expression pattern. Although the cluster analysis demonstrated that the molecular profile of fish liver cells could be associated with the geographical site in some cases, it suffered from several limitations. First, it did not identify genes which are associated strongly with any specific site and secondly it did not provide any indication of how useful a gene (or a set of genes) would be as biomarkers either for discriminating between sampling sites or as an indicator of exposure to different forms of stress.

In this study, we merged the results of acute stress responsive genes derived from controlled laboratory exposure of fish to individual toxicants with the gene expression profiles that were generated by long-term exposure in different environmental conditions. To integrate both types of analysis we built a series of representative statistical models using a multivariate analysis approach. Neither the use of any of the representative models derived from single chemical treatments or the use of a battery of conventional biomarker genes were capable at efficiently summarizing the complexity measured in the environmental samples. However, by merging the clones selected from each representative stressor model we acquired a representative model that showed a comprehensive prediction of the site of origin of fish based on the gene expression generated in each of the environmental conditions. We emphasize that this type of analysis does not negate the use of specific biomarkers of stressors, the conventional biomarkers employed were efficient at site prediction of Brunsbittel fish and CYP1A, for example, gave marked differences between the clean Alde site and the more polluted sites of Tyne and Brunsbittel. However, the statistical model has been successful in the use of a battery of stress-related genes for site prediction. This implies that one of the distinguishing features of the expression profiles between the different sites is the influence of stress genes. Since the models developed using a combination of all signatures developed from the individual chemical exposures are even more accurate in predicting geographical location it is suggested that it is a combination of different stressors that contribute to the molecular profile of liver tissue chronically exposed in the environment.

The genes that were identified by modeling as efficiently predicting sampling sites included those encoding ribosomal proteins, haemoglobin and aldehyde dehydrogenase discussed above. Other predictive transcripts are indicative of an oxidative stress response in polluted environments, including an aldo-keto reductase and copper/zinc superoxide dismutase. The glutamate carbonylpeptidase-like clone, which is associated with folate and homocysteine levels and is similar to anserinase (Yamada et al., 2005) also contributed to prediction, warranting further study on its function and regulation. The well documented biomarkers CYP1A, choriogenin and metallothionein were also strong predictors. Induction of the innate immune response gene, liver expressed antimicrobial peptide (LEAP-2) may indicate increased bacterial exposure of pollutant-impacted fish. The unidentified expressed sequence tags (ESTs) present opportunities for novel biomarker gene discovery, indeed, one was isolated by suppressive, subtractive hybridisation between flatfish from reference and polluted estuaries (Straub et al., 2004).

## 5. Conclusions

Using a multivariate variable selection coupled with a statistical modelling procedure (GALGO) we demonstrated that gene expression signatures in livers of flounder can predict their geographical site of origin but that the accuracy is limited to specific sites. The search space for a combination of genes that effectively predict sample origin is very large and since the selected sites contained varying levels of complex mixtures of pollutants we limited the search space by incorporating prior knowledge and data from laboratory exposures to prototypical pollutants. This produced a very accurate model utilising the expression profiles of only 17 genes for classification of all the different environmental sites. This model was further verified by use of a qPCR array to determine the expression of 16 clones which predicted with 100% accuracy the site of origin of an independent sample of 15 fish from 3 of the sites. This indicates that expression changes derived from short-term exposures can, in some cases, be used to infer longer term exposures accumulated over months or years of exposure. These gene expression changes combine recent and longer term responses to the environment. We aim, in future experiments, to define the bases of individual and population susceptibility and adaptation to environmental pollutants in fish, using a combination of 'omic technologies and genetic analyses.

## Acknowledgements

This work was funded by EU 'GENIPOL' grant EKV-2001-0057 and NERC UK grant NEH/B/S/22000/00255, UK NERC Post-genomic and Proteomic programme grant NE/C507661/1, Birmingham functional genomics facilities were funded by BBSRC grant 6/JIF 13209 and bioinformatics by MRC infrastructure grant G.4500017. We thank R.E. Godfrey, A. Jones, and L. Klovzra for technical assistance, and A. Koehler, K. Winzer, K. Dempsey, W. Reynolds and B. Lyons for help in obtaining flounder samples during statutory sampling programs.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.aquatox.2008.07.020.

## References

- Allen, Y., Matthiessen, P., Scott, A.P., Haworth, S., Feist, S., Thain, J.E., 1999. The extent of oestrogenic contamination in the UK estuarine and marine environments—further surveys of flounder. *Sci. Total Environ.* 233 (1–3), 5–20.
- Bartosiewicz, M., Penn, S., Buckpitt, A., 2001. Applications of gene arrays in environmental toxicology: fingerprints of gene regulation associated with cadmium chloride, benzo(a)pyrene, and trichloroethylene. *Environ. Health Perspect.* 109 (1), 71–74.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* 57 (1), 289–300.
- Broeg, K., Westernhagen, H.V., Zander, S., Korting, W., Koehler, A., 2005. The "bio-effect assessment index" (BAI). A concept for the quantification of effects of marine pollution by an integrated biomarker approach. *Mar. Pollut. Bull.* 50 (5), 495–503.
- Chesman, B.S., O'Hara, S., Burt, G.R., Langston, W.J., 2007. Hepatic metallothionein and total oxyradical scavenging capacity in Atlantic cod *Gadus morhua* caged in open sea contamination gradients. *Aquat. Toxicol.* 84 (3), 310–320.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21 (18), 3674–3676.
- Cunningham, K.S., Dodson, R., Nagel, M.A., Shapiro, D.J., Schoenburg, D.R., 2000. Vigilin binding selectively inhibits cleavage of the vitellogenin mRNA 3'-untranslated region by the mRNA endonuclease polysomal ribonuclease 1. *PNAS* 97 (23), 12498–12502.
- Denslow, N.D., Garcia-Reyero, N., Barber, D.S., 2007. Fish 'n' chips: the use of microarrays for aquatic toxicology. *Mol. Biosyst.* 3 (3), 172–177.
- Diab, A.M., Williams, T.D., Sabine, V.S., Chipman, J.K., George, S.G., 2008. The GENIPOL European flounder *Platichthys flesus* L. toxicogenomics microarray: application

- for investigation of the response to furunculosis vaccination. *J. Fish Biol.* 72 (9), 2154–2169.
- Einsporn, S., Broeg, K., Koehler, A., 2005. The Elbe flood 2002—toxic effects of transported contaminants in flatfish and mussels of the Wadden Sea. *Mar. Pollut. Bull.* 50 (4), 423–429.
- Eloranta, J.J., Kullak-Ublick, G.A., 2005. Coordinate transcriptional regulation of bile acid homeostasis and drug metabolism. *Arch. Biochem. Biophys.* 433 (2), 397–412.
- George, S., Gubbins, M., MacIntosh, A., Reynolds, W., Sabine, V., Scott, A., Thain, J., 2004. A comparison of pollutant biomarker responses with transcriptional responses in European flounders (*Platichthys flesus*) subjected to estuarine pollution. *Mar. Environ. Res.* 58 (2–5), 571–575.
- Hamadeh, H.K., Trouba, K.J., Amin, R.P., Afshari, C.A., Germolec, D., 2002. Coordination of altered DNA repair and damage pathways in arsenite-exposed keratinocytes. *Toxicol. Sci.* 69 (2), 306–316.
- Hancock, D., Wilson, M., Velarde, G., Morrison, N., Hayes, A., Hulme, H., et al., 2005. maxLoad2 and maxBrowse: standards-compliant tools for microarray experimental annotation, data management and dissemination. *BMC Bioinform.* 6, 264.
- Hecker, M., Tyler, C.R., Hoffmann, M., Maddix, S., Karbe, L., 2002. Plasma biomarkers in fish provide evidence for endocrine modulation in the Elbe River, Germany. *Environ. Sci. Technol.* 36 (11), 2311–2321.
- Herrero, J., Al-Shahrour, F., Diaz-Urriarte, R., Mateos, A., Vaquerizas, J.M., Santoyo, J., et al., 2003. GEPAS: a web-based resource for microarray gene expression data analysis. *Nucleic Acids Res.* 31 (13), 3461–3467.
- Kikonyogo, A., Pietruszko, R., 1996. Aldehyde dehydrogenase from adult human brain that dehydrogenates gamma-aminobutyraldehyde: purification, characterization, cloning and distribution. *Biochem. J.* 316, 317–324.
- Kirby, M.F., Allen, Y.T., Dyer, R.A., Feist, S.W., Katsiadaki, I., Matthiessen, P., et al., 2004a. Surveys of plasma vitellogenin and intersex in male flounder (*Platichthys flesus*) as measures of endocrine disruption by estrogenic contamination in United Kingdom estuaries: temporal trends, 1996 to 2001. *Environ. Toxicol. Chem.* 23 (3), 748–758.
- Kirby, M.F., Neall, P., Bateman, T.A., Thain, J.E., 2004b. Hepatic ethoxyresorufin O-deethylase (EROD) activity in flounder (*Platichthys flesus*) from contaminant impacted estuaries of the United Kingdom: continued monitoring 1999–2001. *Mar. Pollut. Bull.* 49 (1–2), 71–78.
- Koehler, A., 2004. The gender-specific risk to liver toxicity and cancer of flounder (*Platichthys flesus* (L.)) at the German Wadden Sea coast. *Aquat. Toxicol.* 70 (4), 257–276.
- Koehler, A., Broeg, K., Bahns, S., 1998. Localisation of a tumor-associated phenotype of Benz-aldehyde dehydrogenase in liver carcinogenesis of flounder by quantitative histochemistry. *Mar. Environ. Res.* 46, 185–189.
- Lyons, B.P., Stewart, C., Kirby, M.F., 1999. The detection of biomarkers of genotoxin exposure in the European flounder (*Platichthys flesus*) collected from the River Tyne Estuary. *Mutat. Res. Genet. Toxicol. Environ. Mutag.* 446, 111–119.
- Lyons, B.P., Stentiford, G.D., Green, M., Bignell, J., Bateman, K., Feist, S.W., et al., 2004. DNA adduct analysis and histopathological biomarkers in European flounder (*Platichthys flesus*) sampled from UK estuaries. *Mutat. Res.* 552 (1–2), 177–186.
- McMillian, M., Nie, A., Parker, J.B., Leone, A., Kemmerer, M., Bryant, S., et al., 2005. Drug-induced oxidative stress in rat liver from a toxicogenomics perspective. *Toxicol. Appl. Pharmacol.* 207 (2 Suppl.), 171–178.
- Neumann, N.F., Galvez, F., 2002. DNA microarrays and toxicogenomics: applications for ecotoxicology? *Biotechnol. Adv.* 20 (5–6), 391–419.
- OSPAR, April 1997. OSPAR, Oslo and Paris conventions for the prevention of marine pollution, Copenhagen. Summary record and JAMP technical annexes. OSPAR, London.
- Pritchard, C.C., Hsu, L., Delrow, J., Nelson, P.S., 2001. Project normal: defining normal variance in mouse gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 98 (23), 13266–13271.
- Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270 (5235), 467–470.
- Snape, J.R., Maund, S.J., Pickford, D.B., Hutchinson, T.H., 2004. Ecotoxicogenomics: the challenge of integrating genomics into aquatic and terrestrial ecotoxicology. *Aquat. Toxicol.* 67 (2), 143–154.
- Snell, T.W., Brogdon, S.E., Morgan, M.B., 2003. Gene expression profiling in ecotoxicology. *Ecotoxicology* 12 (6), 475–483.
- Stentiford, G.D., Longshaw, M., Lyons, B.P., Jones, G., Green, M., Feist, S.W., 2003. Histopathological biomarkers in estuarine fish species for the assessment of biological effects of contaminants. *Mar. Environ. Res.* 55 (2), 137–159.
- Straub, P.F., Higham, M.L., Tanguy, A., Landau, B.J., Phoel, W.C., Hales Jr., L.S., et al., 2004. Suppression subtractive hybridization cDNA libraries to identify differentially expressed genes from contrasting fish habitats. *Mar. Biotechnol. (N.Y.)* 6 (4), 386–399.
- Takahashi, T., Tada, M., Igarashi, S., Koyama, A., Date, H., Yokoseki, A., et al., 2007. Aprataxin, causative gene product for EAOH/AOA1, repairs DNA single-strand breaks with damaged 3'-phosphate and 3'-phosphoglycolate ends. *Nucleic Acids Res.* 35 (11), 3797–3809.
- Talbot, K., Miguel-Aliaga, I., Mohaghegh, P., Ponting, C.P., Davies, K.E., 1998. Characterization of a gene encoding survival motor neuron (SMN)-related protein, a constituent of the spliceosome complex. *Hum. Mol. Genet.* 7 (13), 2149–2156.
- Trevino, V., Falciani, F., 2006. GALGO: an R package for multivariate variable selection using genetic algorithms. *Bioinformatics* 22 (9), 1154–1156.
- van der Oost, R., Beyer, J., Vermeulen, N.P.E., 2003. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. *Environ. Toxicol. Pharmacol.* 13, 57–149.
- Verhagen, A.M., Vaux, D.L., 2002. Cell death regulation by the mammalian IAP antagonist Diablo/Smac. *Apoptosis* 7 (2), 163–166.
- Waring, J.F., Gum, R., Morfitt, D., Jolly, R.A., Ciurlionis, R., Heindel, M., et al., 2002. Identifying toxic mechanisms using DNA microarrays: evidence that an experimental inhibitor of cell adhesion molecule expression signals through the aryl hydrocarbon nuclear receptor. *Toxicology* 181–182, 537–550.
- Williams, T.D., Gensberg, K., Minchin, S.D., Chipman, J.K., 2003. A DNA expression array to detect toxic stress response in European flounder (*Platichthys flesus*). *Aquat. Toxicol.* 65 (2), 141–157.
- Williams, T.D., Diab, A.M., George, S.G., Godfrey, R.E., Sabine, V., Conesa, A., et al., 2006. Development of the GENIPOL European flounder (*Platichthys flesus*) microarray and determination of temporal transcriptional responses to cadmium at low dose. *Environ. Sci. Technol.* 40 (20), 6479–6488.
- Williams, T.D., Diab, A.M., George, S.G., Sabine, V., Chipman, J.K., 2007. Gene expression responses of European flounder (*Platichthys flesus*) to 17-beta estradiol. *Toxicol. Lett.* 168 (3), 236–248.
- Williams, T.D., Diab, A., Ortega, F., Sabine, V.S., Godfrey, R.E., Falciani, F., Chipman, J.K., George, S.G. et al., 2008. Transcriptomic responses of European flounder (*Platichthys flesus*) to model toxicant. *Aquat. Toxicol.*, in press, doi:10.1016/j.aquatox.2008.07.019.
- Yamada, S., Tanaka, Y., Ando, S., 2005. Purification and sequence identification of anserinase. *Febs J.* 272, 6001–6013.