ELSEVIER

Aquatic Toxicology 67 (2004) 143–154

AQUATIC TOXICOLOGY

www.elsevier.com/locate/aquatox

Ecotoxicogenomics: the challenge of integrating genomics into aquatic and terrestrial ecotoxicology

Jason R. Snape[a,*], Steve J. Maund[b], Daniel B. Pickford[a], Thomas H. Hutchinson[a]

[a] *AstraZeneca Global Safety Health and Environment, Brixham Environmental Laboratory, Freshwater Quarry, Brixham, Devon TQ5 8BA, UK*

[b] *Syngenta Crop Protection AG, 4002 Basel, Switzerland*

Received 20 October 2002; accepted 30 November 2003

Abstract

Rapid progress in the field of genomics (the study of how an individual's entire genetic make-up, the genome, translates into biological functions) is beginning to provide tools that may assist our understanding of how chemicals can impact on human and ecosystem health. In many ways, if scientific and regulatory efforts in the 20th century have sought to establish which chemicals cause damage to ecosystems, then the challenge in ecotoxicology for the 21st century is to understand the mechanisms of toxicity to different wildlife species. In the human context, '*toxicogenomics*' is the study of expression of genes important in adaptive responses to toxic exposures and a reflection of the toxic processes per se. Given the parallel implications for ecological (environmental) risk assessment, we propose the term '*ecotoxicogenomics*' to describe the integration of genomics (transcriptomics, proteomics and metabolomics) into ecotoxicology. Ecotoxicogenomics is defined as the study of gene and protein expression in non-target organisms that is important in responses to environmental toxicant exposures. The potential of ecotoxicogenomic tools in ecological risk assessment seems great. Many of the standardized methods used to assess potential impact of chemicals on aquatic organisms rely on measuring whole-organism responses (e.g. mortality, growth, reproduction) of generally sensitive indicator species at maintained concentrations, and deriving 'endpoints' based on these phenomena (e.g. median lethal concentrations, no observed effect concentrations, etc.). Whilst such phenomenological approaches are useful for identifying chemicals of potential concern they provide little understanding of the mechanism of chemical toxicity. Without this understanding, it will be difficult to address some of the key challenges that currently face aquatic ecotoxicology, e.g. predicting toxicant responses across the very broad diversity of the phylogenetic groups present in aquatic ecosystems; estimating how changes at one ecological level or organisation will affect other levels (e.g. predicting population-level effects); predicting the influence of time-varying exposure on toxicant responses. Ecotoxicogenomic tools may provide us with a better mechanistic understanding of aquatic ecotoxicology. For ecotoxicogenomics to fulfil its potential, collaborative efforts are necessary through the parallel use of model microorganisms (e.g. *Saccharomyces cerevisiae*) together with aquatic (e.g. *Danio rerio*, *Daphnia magna*, *Lemna minor* and *Xenopus tropicalis*) and terrestrial (e.g. *Arabidopsis thailiana*, *Caenorhabdites elegans* and *Eisenia foetida*) plants, animals and microorganisms.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Genomics; Microarray; Microbiology; Ecotoxicology; Risk assessment

* Corresponding author. Tel.: +44-1803-884273; fax: +44-1803-882974.
E-mail address: jason.snape@brixham.astrazeneca.com (J.R. Snape).

0166-445X/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.aquatox.2003.11.011

1. Introduction

Winkler first used the term genomics in 1920 to describe the complete set of chromosomes and their associated genes (McKusick, 1997). Today, genomics is a broadly used term that encompasses numerous scientific disciplines and technologies. These disciplines include: genome sequencing; assigning function to identified genes; determining genome architecture; studying gene expression at the transcriptome level; studying protein expression at the proteome level; and investigating metabolite flux (metabolomics). Due to the magnitude and complexity of '-omic' data, these disciplines are underpinned by information technology support through bioinformatics. *Toxicogenomics* is the subdiscipline combining the fields of genomics and (mammalian) toxicology (Nuwaysir et al., 1999). It has also been described as the study of genes and their products important in adaptive responses to chemical-derived exposures (after Iannaconne, 2001; see also Rockett and Dix, 1999; Lovett, 2000; Pennie et al., 2000; ECETOC, 2001). Stimulated by the sensational advances in the Human Genome Mapping Project (http://www.hgmp.mrc.ac.uk/) and on-going sequencing programmes in numerous other species (Rockett and Dix, 2000), the toxicogenomic approach presents important opportunities to improve understanding of the molecular mechanisms underlying toxic responses to environmental contaminants (Bradley and Theodorakis, 2002; Moore, 2001).

Aside from mammals, organisms that are now the focus of genomic sequencing efforts include populations of microbes, plants, insects, nematodes, amphibians and fish. It is difficult to provide an exact number for organisms that have had their genomes sequenced as sequence data is spread between several discrete databases. This difficulty in providing an exact number of genome sequences is compounded by a number of the sequence databases including individual chromosomes and numerous plasmids within their total genome counts. Our investigations have identified that approximately 84 prokaryotic and 9 eukaryotic genomes have been sequenced.

Whilst the number of whole genomes that have been sequenced is small, sequence data does exist for a number of phyla, ranging from a single reported

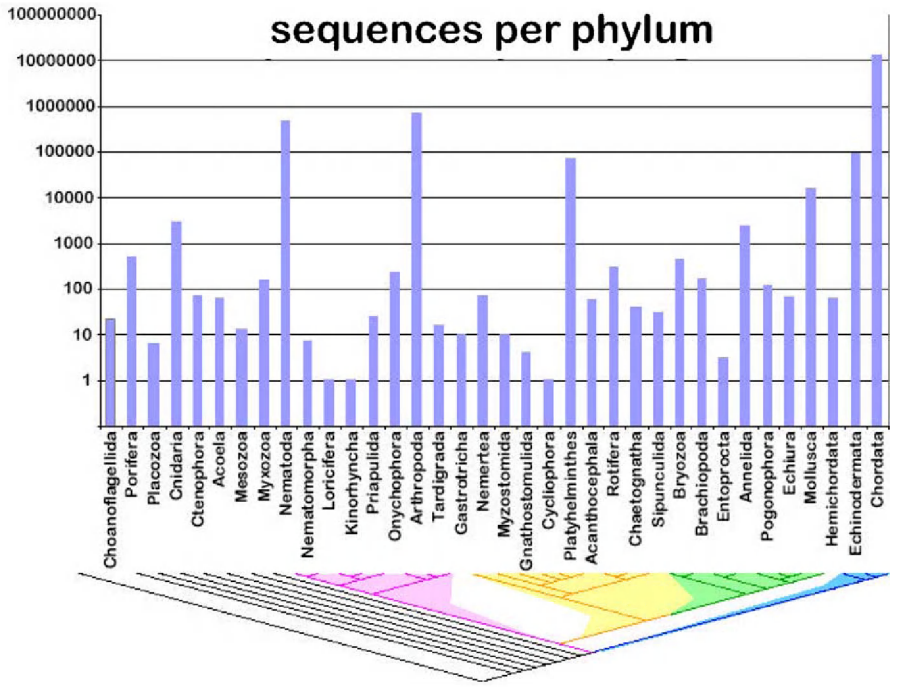

Fig. 1. The total number of DNA sequences published per phylum (correct as of the end of September 2002).

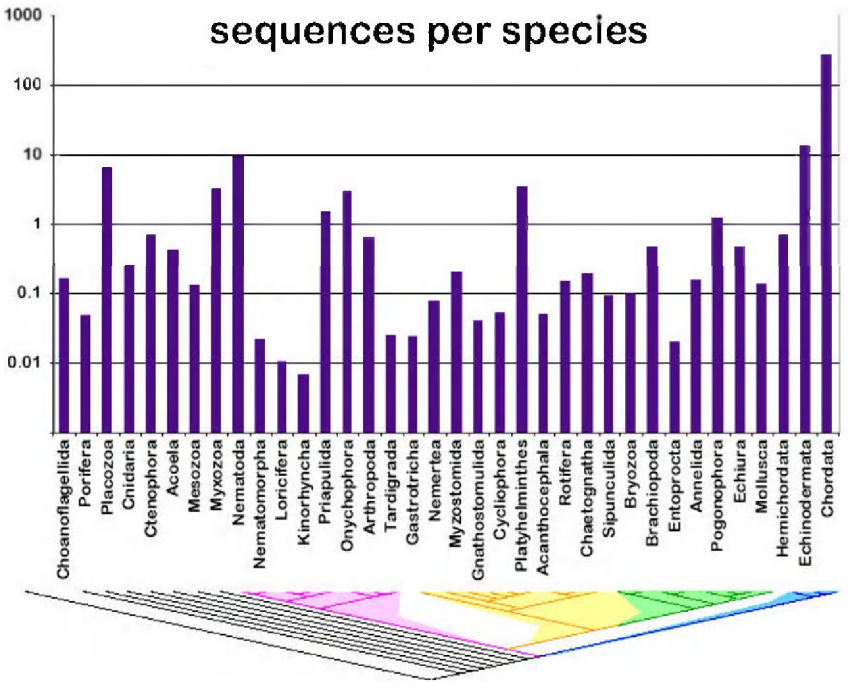

Fig. 2. The total number of published DNA sequences per species per phylum indicating that most phyla are under represented with less than one published sequence per species being available.

sequence for Cycliophora to over 10,000,000 sequences for Chordates (Fig. 1). However, if you examine the number of sequences per species within each phylum a quite different picture emerges with the majority of phyla having less that a single sequence per species (Fig. 2). Clearly, there is a current lack of DNA sequence resource for the ecotoxicologist to utilise. However, numerous national and international sequencing efforts and environmental genomics programmes are in place that will provide additional sequence resource for the ecotoxicologist to improve the certainty with which cross-species genomic comparisons can be drawn and allow the impact of chemicals to be assessed mechanistically. One such sequence programme is the "Tree of Life" initiative announced by the National Science Foundation (NSF) that has prioritised the sequencing of a further 63 genomes covering a broad range of phyla (Table 1).

Recently, Chapman (2001) has introduced the term *ecogenomics* to describe the application of genomics to ecology. The proposed application of genomics to organisms outside the laboratory aims to provide insight into their physiological status and to translate this into an understanding of their responses to each other and to the environment. Chapman (2001) emphasises the immunological aspects of this approach to life and environmental stressors ("to live is to be stressed—but the only alternative is worse"). Given the evidence of immunotoxicity in aquatic organisms exposed to heavy metals and organic chemicals (Zelikoff, 1993; Hutchinson et al., 1999, 2002), this approach also has implications for environmental pollutants and emerging diseases in wildlife populations (Harvell, 1999).

Against this background, we propose the term '*ecotoxicogenomics*' to describe the integration of genomic-based science into ecotoxicology. Given the need to balance the growing requirement for ecotoxicity data with animal welfare, there is a need to maximise the knowledge output from limited testing with lower vertebrates in order to help reduce the future level of routine ecotoxicity testing. Ecotoxicogenomic tools are likely to provide a vital role in this context. Indeed, elements of this type of approach have been described by Moore (2001), identifying

Table 1
The NSF "Tree of Life" list of organisms prioritised for genome sequencing

Scientific name	Common name
Astyanax mexicanus	Blind Cavefish
Metriaclima zebra	Lake Malawi Zebra Cichlid
Chrysemys picta	Painted Turtle
Sphenodon punctatus	Tuatara
Amphisbaena alba	South American Amphisbaenian
Alligator mississippiensis	American Alligator
Dromaius novaehollandiae	Emu
Petromyzon marinus	Lamprey
Brachiostoma floridae	Lancelet, Amphioxus
Manduca sexta	Tobacco Hornworm
Heliconius erato	Heliconid Butterfly
Heliothis virescens	Noctuid Tobacco Budworm
Oncopeltus faciatus	Milkweed Bug
Nasonia vitripennis	Jewel Wasp
Tribolium castaneum	Red Flour Beetle
Thermobia domestica	Firebrat
Schistoicerca americana	Grasshopper
Daphnia pulex	Water Flea
Parhyale hawaiensis	Amphipod
Tigriopus californicus	Copepod
Spisula solidissima	Atlantic Surfclam
Ilyanassa obsoleta	Eastern Mud Snail
Platynereis dumerilii	Marine Polychaete
Helobdella robusta	Freshwater Leech
Themiste langiformis	Sipunculid
Cerebratulus lacteus	Milky Ribbon Worm
Neochildea fusca	Platyhelminthes
Mnemiopsis leidyi	Ctenophora
Callyspongia diffusa	Marine Sponge
Volvox carteri	Colonial alga
Caulerpa taxifolia	Ulvophyceae
Mesostigma viride	Prasinophyceae
Coleochaete orbicularis	Coleochaetales
Chara aspera	Spoonwort
Marchantia polymorpha	Liverwort
Anthoceros sp.	Hornwort
Lycopodium lucidulum	Shining Club Moss
Angiopteris erecta	Tree King Fern
Ceratopteris richardii	Fern
Marsilea quadrifolia	Water Clover
Amborella trichopoda	Oldest living angiosperm
Nuphar adventa	Waterlily
Acorus gramineus	Sweet Flag
Lirodendron tulipifera	Yellow or Tulip Poplar
Mimulus guttatus	Monkeyflower
Ananas comosus	Pineapple
Gossypium herbaceum	A-genome Cotton
Gossypium longicalyx	F-genome Cotton
Gossypoides kirkii	Cotton
Opuntia cochellinifera	Prickly-Pear Cactus
Oryza rufipogon	Rice
Oryza glaberrima	Rice
Oryza punctata	Rice
Oryza officinalis	Rice
Oryza minuta	Rice
Oryza australiensis	Rice
Oryza latifolia	Rice
Oryza schlechteri	Rice
Oryza ridleyi	Rice
Oryza brachyantha	Rice
Oryza granulata	Rice

the need to understand mechanisms of toxicity (including genomic and proteomic aspects), develop predictive simulation models and QSARs of toxic effects; link molecular and cellular biomarkers with higher level population and ecosystem responses; and use this knowledge to anticipate potential ecological risk assessment issues for new chemicals and emerging technologies. Moreover, there is a recognised need for ecotoxicology to move toward a more holistic approach (Chapman, 2002), a proposal that is in harmony with the power of genomics as a tool for understanding toxicant impacts in a diversity of organisms. The identification of endpoints or responses of broader ecological relevance, from such a holistic approach, will enable a simplistic and focused screening approach to be adopted that utilises custom arrays, reporter assays and relevant biomarkers.

2. Ecotoxicogenomics: the ecological relevance of genomic responses

In principle, environmental contaminants may induce genomic responses in an organism. In biomedicine, almost without exception, gene expression is altered in toxicity, as either a direct or indirect result of toxicant exposure (Nuwaysir et al., 1999). Depending upon the severity and duration of the contaminant exposure, genomic measures may be short-term toxicological responses leading to impacts on 'fitness' (survival and reproduction), or the 'genotoxic disease syndrome' defined by Kurelec (1993). A number of studies have demonstrated genotype-dependent impacts in animals exposed to toxicants (Oakshott, 1976; Hawkins et al., 1989; Schat and Ten Bookum, 1992). Furthermore, using traditional biochemical genetic

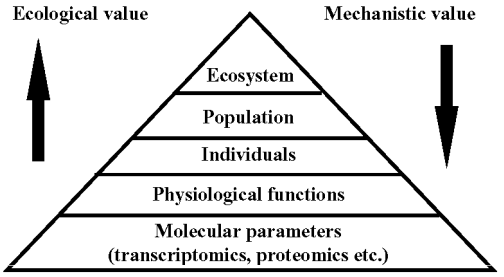

Fig. 3. Conceptual framework for ecotoxicogenomics.

techniques, Shugart and Theodorakis (1996) reported an inverse correlation between DNA strand breakage (a nonspecific genotoxic response) and the fecundity of mosquitofish (*Gambusia affinis*) inhabiting ponds heavily contaminated with radionuclides.

As the field of molecular biology is in post-genomic era (with the sequencing of *Haemophilus influenzae* in 1995), the availability of additional DNA sequence data coupled with technological advances in genomic science will facilitate the direct interrogation of both gene and protein expression in organisms exposed to a diversity of environmental and chemical stimuli. Provided that genome sequencing is extended to include an increased number of species with ecological relevance (including animals, plants and microorganisms), there will be increased opportunity to address the phenotypic and genotypic basis of fitness using a 'bottom-up' approach (molecular to population; Fig. 3). For example, in the UK, the Natural Environment Research Council (NERC) Environmental Genomics Programme recognises this opportunity and will apply genomics to advance fundamental issues in ecology and evolutionary biology, providing the context to future questions of chemical impacts on environmental and human health. Objectives of the NERC programme include: identification of gene loci affecting ecological performance; functional characterisation of traits affecting ecological performance; assessing individual, population, community and ecosystem responses to the environment; determining the extent and significance of genetic variation among traits affecting ecological performance (NERC, 2001).

In terms of the ecological significance of molecular variation, microarrays and other technologies enable variation among populations to be quantified rapidly at multiple gene loci. In the NERC programme, specific strategic objectives include: determining the consequence of variation in the levels of gene expression on ecological performance; defining the extent of spatial and temporal variation at regulatory versus structural loci; identifying numbers of loci and their genomic distribution; and determining community structure and investigating the impact of environmental change. Also, for a targeted range of environments and contaminated field sites, environmental genomics will be used to address key questions, including: which genes are turned on, and what do they do; is there variation in gene expression in response to environmental change, and is this variation adaptive; what are the ecosystem-, community-, and population-level consequences of the molecular transformations performed by these genes?

One of the major challenges facing biologists is the understanding of how phenotype maps on to genotype. Many opportunities exist for exploiting whole or partial genome data to address aspects of this problem. There is a need to understand both the responses of genomes to genetic and environmental stimuli, and conversely the impact the environment plays on the phenotype (Thompson, 1991; Depledge, 1994; Shugart and Theodorakis, 1996). The relationship between DNA sequence, form and function is poorly understood. Genome sequencing, in combination with the tools of functional genomics, offers opportunities to develop a better understanding of phenotypic evolution, to explain (at a mechanistic level) the evolution of novel traits, to determine rules of adaptive evolution and to understand the contribution of regulatory pathways to adaptive evolution. For example, while use of the *Caenorhabdites elegans* and *Drosophila melanogaster* model has facilitated significant advances in understanding the developmental and reproductive consequences of mutation in laboratory invertebrates, there is comparatively little known of the genotype–phenotype relationship in other invertebrate taxa. Traditional techniques for examining specific gene expression provide glimpses of the insights that may be gained in studying gene expression and phenotypic effects in invertebrates exposed to contaminants. Recently, for example, Jessen-Ellner et al. (2002) described the expression of p53, p97 and p120 (a new member of the p53 gene family) in surf clams, with PCB-induced suppression of the serotonergic nervous system in embryonic

animals associated with modified p120 gene expression. Other invertebrate models such as the nematode *C. elegans* offer potentially powerful tools to address the genomic basis for developmental and reproductive effects of environmental chemicals (Hood et al., 2000).

3. Technological advances

There have been various methods developed to investigate gene expression at a genome wide level (Patanjali et al., 1991; Arnheim and Erlich, 1992; Schena et al., 1995; Velculescu et al., 1995; Rockett and Dix, 2000; Clark et al., 2002). Parallel to these advances in gene expression technologies there have been advances in methodologies to investigate protein and metabolite profiles at the proteome (Gevaert and Vandekerckhove, 2000; Mann et al., 2001) and metabolome level (Nicholson et al., 1999; Thomas, 2001; Watkins and German, 2002; Buchholz et al., 2002). To date, few of these strategies have been applied to study the response of wildlife population to chemical stressors in complex natural environments, such as sediments and soil. An important objective will be the refinement of existing, and the development of new, technologies capable of measuring gene, protein and metabolite profiles for wildlife populations subject to multiple stressors in complex natural environments. Approaches that are currently being used to monitor gene, protein and metabolite expression are briefly discussed.

3.1. DNA array-based technologies

Traditional hybridisation-based assays developed over the past 30 years (Grunstein and Hogness, 1975; Benton and Davis, 1977) use flexible membranes, such as nitrocellulose and nylon, radioactivity, and autoradiography to look at the expression of one or two genes in isolation (Schena et al., 1998). By contrast, modern high-density DNA array-based techniques use solid surfaces such as glass with fluorescent labelling and detection (Schena et al., 1995; Rockett and Dix, 2000). These array-based assays are becoming widely available for a number of model and non-model organisms and tissues (Table 2) and have an advantage of being data rich (it is possible to analyse many thousands of genes simultaneously). Using this approach it is possible to identify transcripts that are markedly up- or down-regulation following experimentation. However, this approach like the traditional methods of gene expression analysis does rely on the measurement of signal intensity from nucleic acid hybridisation. The efficiency of detection is reliant on the efficiency of labelling and hybridisation of an individual clone. As a result the data generated is at best only semi-quantitative. It can also prove difficult to detect low-copy number transcripts (Bartlett, 2001).

The major strength of array-based technologies is the sheer number of transcripts that can be analysed simultaneously in a single experiment and the rapidity with which information can be gathered (Bartlett, 2001). The technology also benefits from the ability

Table 2
Web addresses for major genome mapping projects (after Rockett and Dix, 2000)

Species	Sequencing project web address
Arabidopsis	http://www.nsf.gov/bio/pubs/arab
Barley	http://www.css.orst.edu/barley/nabgmp/nabgmp.htm
Chicken	http://www.ri.bbsrc.ac.uk/chickmap/ChickMapHomePage.htm
Cow	http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl
Dog	http://mendel.Berkeley.edu/dog.html
Fruitfly	http://www.fruitfly.org
Human	http://www.hgmp.mrc.ac.uk http://www.er.doe.gov/production/ober/hugtop.html http://www.nhgri.nih.gov/HGP
Maize	http://www.math.iastate.edu/danwell/maize/homepage.html
Mouse	http://www.informatix.jax.org
Nematode	http://www.wormbase.org
Pig	http://www.ri.bbsrc.ac.uk/pigmap/pigmap.htm
Pufferfish	http://jgi.doe.gov/tempweb/programs/fugu.htm
Rat	http://ratmap.gen.gu.se
Rice	http://www.staff.or.jp
Zebrafish	http://zfish.uoregon.edu/zfinfo/zfmap.html
Microarray data repositories	
GEO	http://www.ncbi.nlm.nih.gov/geo/
EBI	http://www.ebi.ac.uk/arrayexpress
Organisations	http://www.ciit.org/toxicogenomics http://www.niehs.nih.gov/nct/concept http://www.nih.gov/science/models http://www.sanger.org http://genomicglossaries.com/content/omes.asp http://welcometrust

Table 3
Organisations for further information

Region	Organisation web address
Asia-Pacific	
Europe	http://www.sanger.org
	http://genomicglossaries.com/content/omes.asp
	http://www.wellcome.ac.uk
North America	http://www.ciit.org/toxicogenomics
	http://www.niehs.nih.gov/nct/concept
	http://www.nih.gov/science/models

to generate custom arrays against specific tissues, families or sub-sets of genes. One major drawback of array-based techniques is its reliance on a priori knowledge of the transcripts likely to be present in a sample such that gene probes can be designed in advance (Clark et al., 2002). Other techniques for studying gene expression, that are not reliant on a priori knowledge of DNA sequence data, include serial analysis of gene expression (SAGE; Velculescu et al., 1995) and differential display (Stein and Liang, 2002; Liang, 2002).

The lack of robust quality control procedures and capture of adequate metadata has caused problems with the analysis and reproducibility of array-based transcriptomics investigations (Night, 2001). As a consequence the international Microarray Gene Expression Data (MGED) group has written an open letter to scientific journals proposing standards for publication (Nature, 2002). This was designed to clarify the Minimal Information About a Microarray Experiment (MIAME) guidelines (Nature, 2001). As a result *Nature* and the Nature family of journals require that articles received on or after 1 December 2002 that contain microarray experiments must be compliant with the MIAME standard and that the data integral to the paper's conclusions should be submitted to the ArrayExpress or GEO databases (Table 3; Nature, 2002). It is anticipated that other journals will also require researchers to adhere to these quality standards.

3.2. Proteome analysis

Proteomics describes the study of the total complement of proteins in a cell or tissue type, expressed under specific conditions. The first of the two key requirements for studying the proteome is the need for high-resolution separation and display of the proteins in the tissue in a form, which is amenable to subsequent analysis and comparison. This first stage does not require that the identity of the proteins be known. The second requirement is the need to identify those proteins that are either expressed under a given condition or not expressed relative to the control specimen. The most suitable separation technique to study protein expression and post-translational modification is 2-D gel electrophoresis, in which proteins are separated first by charge, and then by size. However, for peptides or protein fragments, 2-D gel electrophoresis is not always effective, and high-resolution chromatography may be more appropriate. For characterisation of the proteome, only mass spectrometry has the sensitivity, selectivity and throughput to identify each protein. Individual proteins separated by 2-D gel electrophoresis are excised, and digested in situ using an endopeptidase such as trypsin. The resultant set of peptides are then analysed as a mixture using matrix assisted, laser desorption ionisation, time-of-flight (MALDI-TOF) mass spectrometry. This gives a peptide mass fingerprint (PMF) with which comparisons can be made to fragment databases to yield the identity of the protein spot (Gevaert and Vandekerckhove, 2000; Mann et al., 2001). Pratt et al. (2002) have recently coupled stable isotope labelled amino acids (in this case decadeuterated leucine) with PMF to aid protein identification in vivo. As with DNA-array-based technologies, a drawback to the application of proteomics to study the impact of environmental contaminants on wildlife populations is the absence of relevant DNA sequence information. Other limitation include: limited existing PMF databases that result in few positive matches being made; and the inherent difficulties associated with integrating transcriptomic and proteomic observations due to differences in protein decay rates.

3.3. Metabolomics

Metabolomics (metabonomics in a pharmacological context) describes the quantitative measurement of the dynamic metabolic response of a living system to a toxic or physiological challenge (Nicholson et al., 1999; Bundy et al., 2002). Metabolomics shares two distinct advantages with proteomics in terms of the elucidating gene function (Raamsdonk et al., 2001). That is, the total complement of proteins or

metabolites changes according to the physiological, developmental or pathological state of a cell, tissue, organ or organism, and unlike transcript (mRNA) analysis proteins and metabolites are functional entities within the cell (Raamsdonk et al., 2001). A third advantage that is distinct to metabolomics is that there are far fewer metabolites than genes or gene products to be studied (Raamsdonk et al., 2001).

The majority of metabolomic studies use NMR-based technologies to complement the information provided by measuring the transcriptomic and proteomics responses to contaminant exposure (Nicholson et al., 1983, 1985, 1999; Bales et al., 1984; Nicholson and Wilson, 1989). These H NMR-based are usually coupled to pattern recognition, expert systems and related bioinformatic tools to interpret and classify complex NMR-generated datasets (Nicholson et al., 1999). Nicholson et al. (1999) have reviewed the use of metabolomics for toxicological classification in vivo and Bundy et al. (2002) has used a metabonomic strategy to identify new endogenous biomarkers resulting from the exposure of the earthworm *Eisenia veneta* (Rosa) to 4-fluoroaniline, 3,5-difluoroaniline and 2-methylalinine exposure.

4. Microbial genomics

Molecular biology entered the post-genomic era in 1995 with the publication of the first complete genome from a free-living organism, *H. influenzae* (Fleischman, 1995). Since then 84 microbial genomes have been sequenced (http://wit.integratedgenomics.com/GOLD; http://www.tigr.org/tdb/mdb/mdbcomplete.html) and approximately 200 more are currently in progress.

Whilst most of the early sequencing efforts have focused on pathogenic microorganisms, an increased number of ecologically important microorganisms, involved in wastewater treatment and biogeochemical cycles, have been or are being sequenced. Importantly, from the perspective of reducing toxic exposure within Wastewater Treatment Plants (WTPs) recent sequencing efforts also include some of the standard species *(Escherichia coli, Pseudomonas putida, Nitrosomonas* sp.) used in microbial toxicity assessment to help protect the function of WTPs and the quality of the receiving environment. In this context, standard bacterial toxicity tests focus on a small group of microbes, selected to derive the Predicted No-Effect Concentration for a chemical or effluent against indigenous representatives of sewage bacteria ($PNEC_{microorganisms}$), for the protection of WTPs (TGD, 2002). The availability of complete genome sequence and arrays for a number of environmentally important microbes will enable traditional toxicity measurements such as respiration inhibition, growth inhibition or reduced ammonia oxidation to be coupled with specific changes in gene and protein expression. Such an approach may provide an insight into how the cells are being inhibited and the extent to which other microbes may also be susceptible.

Coupled to microbial genome sequence efforts has been a vast investment in microbial biodiversity programmes exploiting other molecular biological techniques. These programme have served to highlight the importance of some key microbial groups and their associated ecological role (especially mineral cycling—C, N, S). However, the key microbes that underpin the biogeochemical cycles, ecological condition, and productivity are not included within current ecological risk assessments. Two such examples are highlighted below. Species of *Prochlorococci* are a group of primary producers that can be responsible for up to 50% of total ecosystem photosynthesis and are responsible for 30% of global carbon dioxide fixation (West et al., 2001; Ting et al., 2002; Rocap et al., 2003). The biological fixation of dinitrogen gas to ammonia is restricted to solely microbial populations, namely groups of free-living aerobic microorganisms (including *Cyanobacteria, Thiobacillus, Methylomonas*), free-living anaerobic microorganisms (including *Desulfovibrio, Methanosarcina, Heliobacterium*) and symbiotic bacteria (including *Rhizobium*; Madigan et al., 1997). Species and strains from both of these primary producers and nitrogen-fixing organisms have full genome sequence data available (http://wit.integratedgenomics.com/GOLD; http://www.tigr.org/tdb/mdb/mdbcomplete.html). It is important that ecotoxicology is extended to include these important groups of bacteria as they underpin ecological quality and sustainability. It is equally important that the role of microbes as surrogates for higher organisms is determined through investigating microbial toxicology and ecotoxicology in parallel at the '-omic' level. Greer et al. (2001) have recently

reviewed the wide range of opportunities genomic science offers to environmental microbiologist and a number of the issues and benefits highlighted equally apply to the ecotoxicologist.

5. Genomics and non-target organisms

A wide range of non-target organisms (including plants, invertebrates and fish) are routinely including in regulatory testing schemes to provide data for use in ecological risk assessment (Walker et al., 2001). For aquatic environments, plant species used include freshwater and marine microalgae, together with macrophytes such as *Lemna minor*. For terrestrial environments, a range of species may be employed for ecotoxicity testing, including radish (*Raphanus sativus*) and onion (*Allium cepa*), together with other monocotyledons and dicotyledons. While there is considerable knowledge on the genetic basis of metal tolerance in plants (see Depledge, 1994), there is potential to develop a mechanistic understanding of the comparative phytotoxicity of organic chemicals through the use of genomics. One option would be to include the species with fully described genomes for phytotoxicity testing and research (Table 3).

In terms of invertebrates often used in ecotoxicology research, the most popular species (e.g. *Chironomus riparius*, *Daphnia magna* and *Mytilus edulis*) have as yet not been sequenced as genomic models. For fish, the current situation is more encouraging since the announcement of plans to fully sequence the zebrafish genome in the near future (Table 3). Perhaps one of the greatest incentives for using genomic technologies is those aspects of ecotoxicology where long-term tests are required, for example, in order to optimally design reproductive and trans-generational effects of chemicals in fish or other animals (Pennie et al., 2000). Alternatively, gene or protein expression signatures could be developed for a limited range of aquatic or terrestrial organisms and such information used to identify the mechanism(s) of action of a single chemical or complex environmental mixtures (see schematic approach in Fig. 4). A successful example of this approach has recently been described by Bradley et al. (2002) who used protein expression signatures in

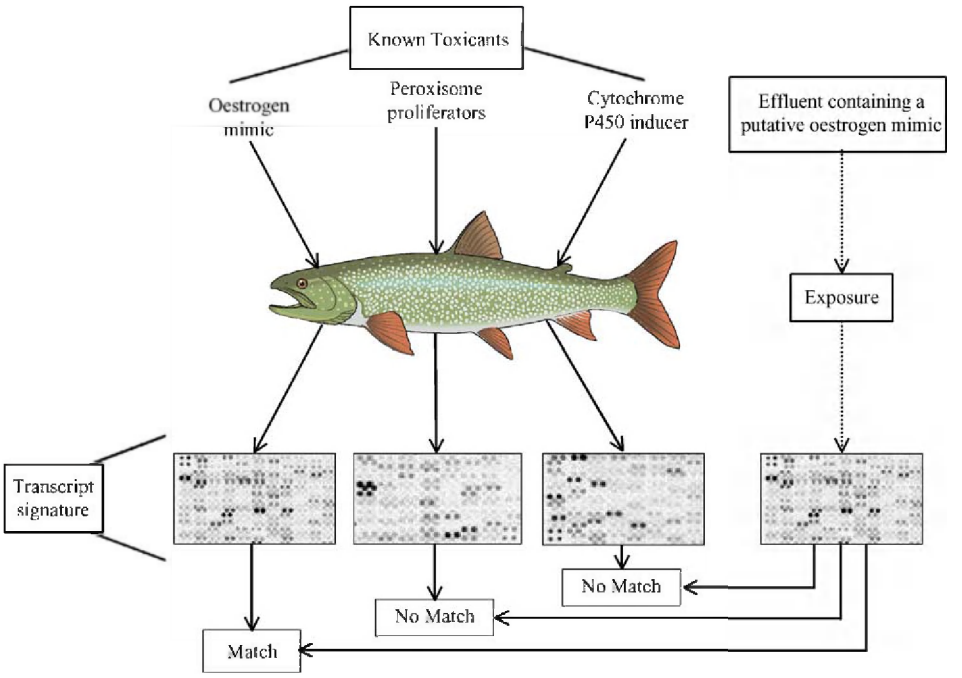

Fig. 4. A simplistic schematic illustration of the application of gene expression signatures to identify the potential mechanism of action of an unknown environmental contaminant (adapted from Nuwaysir et al., 1999).

rainbow trout to investigate endocrine disrupting chemicals in sewage effluent. There is even an outside chance to reduce the ever-growing batteries of ecotoxicity tests by using ecotoxicogenomic information to extrapolate between wildlife taxa through a comparative genomic approach (e.g. freshwater cladocerans versus terrestrial annelids).

6. Conclusions

The rapid development of genomic science will give rise to numerous challenges and scientific opportunities to develop improved knowledge to understand the potential risks of chemicals to human and ecosystem health at a mechanistic level. The application of transcriptomic, proteomics and metabolomic technologies allows the expression profile of hundreds to many thousands of genes, gene products and metabolites to be generated simultaneously, providing for the first time a broad impression of how organisms or cells respond to a given stimulus. The power of such an approach will be maximised further through linking changes at the "-omic" level with traditional ecotoxicological end points and life history characteristics. The integration genomic approach will also provide "leads" or genes of interest that can be validated through detailed functional studies. However, the technology is not fool proof as highlighted by recent concerns over quality control problems in commercially available DNA microarray technology (Night, 2001). The requirement for researchers to ensure their data is MIAME compliant prior to publication may start to improve the quality of data in the public domain.

It is clear from the accelerating list of published literature on genomics (and post-genomics) that this will offer significant opportunities across all the life sciences, from medicine to ecology. Ecotoxicology is inevitably being drawn into this arena and the future will no doubt witness intensive debate over the direction and wise application of genomic information in the ecological risk assessment of chemicals. Essentially, all the cautions raised for mammalian *toxicogenomics* apply equally to *ecotoxicogenomics* (Nuwaysir et al., 1999; Pennie et al., 2000; ECETOC, 2001). In ecotoxicology, such concerns are probably best addressed by research that incorporates genomics, proteomics and metabolomics into well designed in vivo studies, at the system biology level, using environmentally relevant exposures and a range of time points which measure established endpoints of population relevance (e.g. survival, development and reproduction). The sensational increase in genomic data for an increasing number of non-mammalian species suggests that such experiments become more feasible by the day and ecotoxicologists should not be afraid to seize the day.

Acknowledgements

The authors acknowledge Dr. Mark Blaxter, University of Edinburgh, for permission to use the sequence-based analyses versus phyla, and species, within this manuscript. J.R.S. acknowledges the continued support of the Natural Environment Research Council (NERC) in his role as Science Coordinator to their Environmental Genomics and Post-Genomics and Proteomics Science Programmes.

References

Arnheim, N., Erlich, H., 1992. Polymerase chain reaction strategy. Annu. Rev. Biochem. 61, 131–156.

Bales, J.R., Higham, D.P., Howe, I., Nicholson, J.K., Sadler, P.J., 1984. Use of high resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine. Clin. Chem. 20, 426–432.

Bartlett, J., 2001. Technology evaluation: SAGE, Genzyme molecular oncology. Curr. Opin. Mol. Ther. 3 (1), 85–96.

Benton, W.D., Davis, R.W., 1977. Screening lgt recombinant clones by hybridization to single plaques in situ. Science 196, 180–183.

Bradley, B., Theodorakis, C., 2002. The post-genomic era and ecotoxicology. Ecotoxicology 11, 7–9.

Bradley, B., Shrader, E.A., Kimmel, D.G., Meiller, J.C., 2002. Protein expression signatures: an application of proteomics. Mar. Env. Res., in press.

Buchholz, A., Hurlebaus, J., Wandrey, C., Takors, R., 2002. Metabolomics: quantification of intracellular metabolite dynamics. Biomol. Eng. 19 (1), 5–15.

Bundy, J.G., Lenz, E.M., Bailey, N.J., Gavaghan, C.L., Svendsen, C., Spurgeon, D., Hankard, P.K., Osborn, D., Weeks, J.M., Trauger, S.A., Speir, P., Sanders, I., Lindon, J.C., Nicholson, J.K., Tang, H., 2002. Metabonomic assessment of toxicity of 4-fluoroaniline, 3-5-difluoraniline and 2-fluoro-4-methylaniline to the earthworm *Eisenia veneta* (Rosa): identification of new exogenous biomarkers. Environ. Toxicol. Chem. 21, 1966–1972.

Chapman, R.W., 2001. EcoGenomics—a consilience for comparative immunology. Dev. Comp. Immunol. 25, 549–551.

Chapman, P.M., 2002. Integrating toxicology and ecology: putting the eco into ecotoxicology. Mar. Poll. Bull. 44, 7–15.

Clark, T., Lee, S., Ridgway Scott, L., wang, S.M., 2002. Computational analysis of gene identification with SAGE. J. Comput. Biol. 9 (3), 513–526.

Depledge, M.H., 1994. Genotypic toxicity: implications for individuals and populations. Environ. Health Perspect. 102 (Suppl. 12), 101–104.

ECETOC, 2001. White Paper on Genomics, Transcript Profiling, Proteomics and Metabonomics (GTPM)—An Introduction. ECETOC Document No. 42, European Centre for the Ecotoxicology & Toxicology of Chemicals, Brussels, Belgium, 27 p.

Fleischman, R.D., 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269, 496–512, plus 39 coauthors.

Gevaert, K., Vandekerckhove, J., 2000. Protein identification methods in proteomics. Electophoresis 21, 1145–1154.

Greer, C.W., Whyte, L.G., Lawrence, R.R., Masson, L., Brousseau, R., 2001. Genomics technologies for environmental science. Env. Sci. Tech. 35 (17), 360A–366A.

Grunstein, M., Hogness, D.S., 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. 72 (10), 3961–3965.

Harvell, C.D., 1999. Emerging marine diseases—climate links and anthropogenic factors. Science 285, 1505–1510, plus 12 coauthors.

Hawkins, A.J.S., Rusia, J., Bayne, B.L., Day, A.J., 1989. The metabolic/physiological basis of genotype-dependent mortality during copper exposure in *Mytilus edulis*. Mar. Env. Res. 28, 139–144.

Hood, T.E., Calabrese, E.J., Zuckermann, B.M., 2000. Detection of an estrogen receptor in two nematode species and inhibition of binding and development by environmental chemicals. Ecotox. Environ. Safety 47, 74–81.

Hutchinson, T.H., Field, M.D.R., Manning, M.J., 1999. Evaluation of immune function in juvenile turbot (*Scopthalmus maximus* L.) exposed to sediments contaminated with polychlorinated biphenyls. Fish. Shellfish Immunol. 9, 457–472.

Hutchinson, T.H., Field, M.R.D., Manning, M.J., 2002. Evaluation of non-specific immune functions in dab, *Limanda limanda* (L). following short-term exposure to sediments contaminated with polyaromatic hydrocarbons and/or polychlorinated biphenyls. Mar. Env. Res., in press.

Iannaconne, P.M., 2001. Toxicogenomics—the call of the wild chip. Environ. Health Perspect. 109, A8–A11.

Jessen-Ellner, K., Kreiling, J.A., Begley, G.S., Steele, M.E., Walker, C.W., Stephens, R.E., Reinisch, C.L., 2002. A new invertebrate member of the p53 gene family is developmentally expressed and responds to polychlorinated biphenyls. Environ. Health Perspect. 110, 377–385.

Kurelec, B., 1993. The genotoxic disease syndrome. Mar. Env. Res. 35, 341–348.

Liang, P., 2002. A decade of differential display. Biotechniques 33 (2), 338–344.

Lovett, R.L., 2000. Toxicologists brace for genomics revolution. Science 289, 53–57.

Madigan, M.T., Martinko, J.M., Parker, J., 1997. Biology of Microorganisms. Prentice Hall, London, UK.

Mann, M., Hendrickson, R.C., Pandey, A., 2001. Analysis of proteins and proteomes by mass spectroscopy. Ann. Rev. Biochem. 70, 437–473.

McKusick, V.A., 1997. Genomics: structural and functional studies of gemones. Genomics 45 (2), 244–249.

Moore, M.N., 2001. Biocomplexity: the post-genome challenge in ecotoxicology. Aquat. Toxicol. 59, 1–15.

Nature, G., 2001. Minimal Information About a Microarray Experiment (MIAME). Nat. Genet. 29, 365–371.

Nature, 2002. Microarray standards at last. Nature 419, 323.

NERC, 2001. See the NERC Environmental Genomics programme webpage at http://www.nerc.ac.uk/funding/thematics/envgen/index.htm.

Nicholson, J.K., Buckingham, M.J., Sadler, P.J., 1983. High resolution proton NMR studies of vertebrate blood and plasma. Biochem. J. 221, 605–615.

Nicholson, J.K., Higham, D., Timbrell, J.A., Sadler, P.J., 1985. Quantitative ^{1}H-NMR urinalysis studies on the biochemical effects of cadmium exposure in the rat. Mol. Pharmacol. 36, 125–132.

Nicholson, J.K., Wilson, I.D., 1989. High resolution proton NMR spectroscopy of biological fluids. Progress NMR Spectrosc. 21, 449–501.

Nicholson, J.K., Lindon, J.C., Holmes, E., 1999. Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29, 1181–1189.

Night, J., 2001. When the chips are down. Nature 410, 860–861.

Nuwaysir, E.F., Bittner, M., Trent, J., Barrett, J.C., Afshari, C.A., 1999. Microarrays and toxicology: the advent of toxicogenomics. Mol. Carcinogenesis 24, 153–159.

Oakshott, J.G., 1976. Selection at the Adh locus in *Drosophila melanogaster* imposed by environmental ethanol. Genet. Res. 26, 265–274.

Patanjali, S.R., Parimoo S., Weissman, S.M., 1991. Construction of a uniform-abundance (normalized) cDNA library. Proc. Natl. Acad. Sci., 88 (5), 1943–1947.

Pennie, W.D., Tugwood, J.D., Oliver, G.J., Kimber, I., 2000. The principles and practice of toxicogenomics: applications and opportunities. Toxicol. Sci. 54, 277–283.

Pratt, J.M., Robertson, D.H.L., Gaskell, S.J., Riba-Garcia, I., Hubbard, S.J., Sidhu, K., Oliver, S.G., Butler, P., Hayes, A., Petty, J., Beynon, R.J., 2002. Stable isotope labelling in vivo as an aid to protein identification in peptide mass fingerprinting. Proteomics 2, 157–163.

Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M., Berden, J.A., Brindle, K.M., Kell, D.B., Rowland, J.J., Westerhoff, H.V., van Dam, K., Oliver, S.G., 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat. Biotechnol. 19, 45–50.

Rocap, G., Distel, D.L., Waterbury, J.B., Chisholm, S.W., 2003. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed

spacer sequences. Appl. Environ. Microbiol. 68 (3), 1180–1191.

Rockett, J.C., Dix, D.J., 1999. Application of DNA arrays to toxicology. Environ. Health Perspect. 107, 681–685.

Rockett, J.C., Dix, D.J., 2000. DNA arrays: technology, options and toxicological applications. Xenobiotica 30, 155–177.

Schat, H., Ten Bookum, W.M., 1992. Genetic control of copper tolerance in *Silene vulgaris*. Heredity 68, 219–229.

Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E., Davis, R.W., 1998. Microarrays: biotechnology's discovery platform for functional genomics. Trends Biotechnol. 16 (7), 301–306.

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.

Shugart, L.R., Theodorakis, C., 1996. Genetic ecotoxicology: the genotypic diversity approach. Comp. Biochem. Physiol. 113C, 273–276.

Stein, J., Liang, P., 2002. Differential display technology: a general guide. Cell Mol. Life Sci. 59 (8), 1235–1240.

TGD 2002. See the European Chemicals Bureaus webpage at http://ecb.jrc.it/cgi-bin/reframer.pl?A=ECB&B=/Technical-Guidance-Document/

Ting, C.S., Rocap, G., King, J., Chisholm, S.W., 2002. Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. Trends Microbiol. 10 (3), 134–142.

Thomas, G.H., 2001. Metabolomics breaks the silence. Trends Microbiol., 9 (4), 158.

Thompson, J.D., 1991. Phenotypic plasticity as a component of evolutionary change. Trends Evol. Ecol. 6, 246–249.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. Science 270 (5235), 484–487.

Walker, C.H., Hopkin, S.P., Sibly, R.M., Peakall, D.B., 2001. Principles of Ecotoxicology. Taylor & Francis, 309 p.

Watkins, S.M., German, J.B., 2002. Towards the implementation of metabolomic assessments of human health and nutrition. Curr. Opin. Biotechnol. 13 (5), 512–516.

West, N.J., Schonhuber, W.A., Fuller, N.J., Amann, R.I., Rippka, R., Post, A.F., Scanlan, D.J., 2001. Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridization using 16S rRNA-targeted oligonucleotides. Microbiology 147 (7), 1731–1744.

Zelikoff, J.T., 1993. Metal pollution induced immunomodulation in fish. Ann. Rev. Fish Dis. 3, 305–326.